
Whole-exome sequencing of cases with familial cardiomyopathy

Timothy Spracklen

SPRTIM002

Thesis presented for the degree of

Doctor of Philosophy

in the Department of Medicine

Faculty of Health Sciences

University of Cape Town

December 2020

Supervisors:

Professor Ntobeko Ntusi

Associate Professor Gasnat Shaboodien

Dedication

To Granny

And in loving memory

Acknowledgements

In memory of Professor Bongani Mayosi, without whom this work would not have been possible.

- I am very grateful to my supervisor, Professor Ntobeko Ntusi, for his guidance and support, even in the most difficult of circumstances (not to mention a pandemic!). Thank you for your insights and encouragement – I valued our meetings, and always walked away from your office feeling both enthused and edified.

- Thank you to Professor Karen Sliwa for hosting me and fostering such a stimulating working environment at the Hatter Institute for Cardiovascular Research in Africa. I am tremendously appreciative of you and Professor Sandrine Lecour for your support throughout this project. To all my lab- and office-mates at the Hatter Institute, thanks so much for your friendship and company over the years. Thank you to Associate Professor Gasnat Shaboodien for accepting me into the Cardiovascular Genetics research group.

- To Professor Liesl Zühlke, thank you for being the first person to read this thesis! Your kindness and generosity of spirit is inspiring. From her research unit, thank you to Nicky Saacks for your friendship and support – it was a joy to work with you.

- Thank you to the clinical team at Groote Schuur Hospital, especially Dr Sarah Kraus for her cardiomyopathy expertise, enthusiasm and infectious passion for making a difference to the lives of the patients and their families. Thank you also for making this project

possible, through IMHOTEP and all the behind-the-scenes support you gave me when things got tough – my gratitude to you cannot be overstated.

- I thank Professor Bernard Keavney and Dr Paul Kasher for hosting and mentoring me at the University of Manchester, and for sharing your expertise so graciously. I learnt so much from you and your research groups – this project simply would not have been possible without your valuable contributions. Thank you also to everyone else who made the DGEMBE exchange possible (NRF grant number: 98563), particularly Associate Professor Mark Engel.

- I also thank Drs Peixin Zhu and Guoqiang Zhang for giving me the incredible opportunity to work and learn at Novartis Institutes for BioMedical Research. Thank you to everyone who made the Global Health Fellowship possible, including Dr Ross Tracey, Sherri Schwaninger, and Dr Jonathan Spector.

- The financial assistance of the National Research Foundation of South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

- Finally, an enormous “thank you” to the rest of the Spracklen Family for your unwavering support and belief in me, for which I am eternally grateful. Thank you for keeping me sane (no small feat!) throughout this whole doing-a-PhD process, and for tolerating me for all these years.

Declaration of own work

I hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

Signature: _____

Date: _____

Plagiarism declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the Chicago author-date convention for citation and referencing. Each contribution to, and quotation in, this dissertation/thesis from the work(s) of other people has been attributed, and has been cited and referenced. Any section taken from an internet source has been referenced to that source.
3. This dissertation/thesis is my own work, and is in my own words (except where I have attributed it to others).
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

Signature: _____

Date: _____

Table of contents

List of abbreviations

1000G	One Thousand Genomes Project
ACM	Arrhythmogenic cardiomyopathy
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
AFR	African (1000 Genomes Project super population group)
AMR	Admixed American (1000 Genomes Project super population group)
ANOVA	Analysis of variance
AP	Alkaline phosphatase
AR	Autosomal recessive
ARVC	Arrhythmogenic right ventricular cardiomyopathy
ATP	Adenine triphosphate
AVF	Augmented vector foot
AVL	Augmented vector left
Brady	Bradycardia
bp	Base pairs
CADD	Combined Annotation Dependent Depletion
CAF	Central Analytical Facilities
Cas9	Caspase 9
cDNA	Complementary DNA
CMR	Cardiac magnetic resonance imaging
CPGR	Centre for Proteomic and Genomic Research
CRISPR	Clustered regularly-interspaced short palindromic repeats
CTP	Cytosine triphosphate
CVD	Cardiovascular disease
dbSNP	The Single Nucleotide Polymorphism Database
DCM	Dilated cardiomyopathy
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
dpf	Days post fertilisation

EAS	East Asian (1000 Genomes Project super population group)
ECG	Electrocardiography
ECM	Extracellular matrix
E. coli	Escherichia coli
EF	Ejection fraction
ESC	European Society of Cardiology
EUR	European (1000 Genomes Project super population group)
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
ExoI	Exonuclease I
GL-3	Globotriaosylceramide
gnomAD	Genome Aggregation Database
GSH	Groote Schuur Hospital
GTex	Genotype-Tissue Expression
GTP	Guanine triphosphate
HCl	Hydrogen chloride
HCM	Hypertrophic cardiomyopathy
HGNC	Human Genome Organisation Gene Nomenclature Committee
HGVS	Human Genome Variation Society
HICRA	Hatter Institute for Cardiovascular Research in Africa
Hr	Hour
HREC	Human Research Ethics Committee
HRM	High resolution melt
ICC	Inherited cardiac condition
ICTS	Information and Communication Technology Services
ID	Identifier
IMHOTEP	African Cardiomyopathy and Myocarditis Registry Programme
iPSC	Induced pluripotent stem cell
IQR	Interquartile range
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LAS	Low amplitude signal duration
LB	Lysogeny broth

LBBB	Left bundle branch block
LMIC	Low- and middle-income country
LVNC	Left ventricular noncompaction
MAF	Minor allele frequency
M-CAP	Mendelian Clinically Applicable Pathogenicity
$MgCl_2$	Magnesium chloride
MGI	Mouse Genome Informatics
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
n	Number
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
NCD	Non-communicable disease
NEB	New England BioLabs
NGS	Next-generation sequencing
NIBR	Novartis Institutes for BioMedical Research
No.	Number
NYHA	New York Heart Association
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
PolyPhen	Polymorphism Phenotyping
PPCM	Peripartum cardiomyopathy
RBBB	Right bundle branch block
RCH	Red Cross War Memorial Children's Hospital
RCM	Restrictive cardiomyopathy
RE	Restriction enzyme
RefSeq	Reference Sequence
RNA	Ribonucleic acid
RV	Right ventricular
SAS	South-East Asian (1000 Genomes Project super population group)
SCD	Sudden cardiac death

sgRNA	Single guide RNA
SIFT	Sorting Intolerant from Tolerant
SND	Sinus node dysfunction
S.O.C.	Super Optimal broth with Catabolite repression
T_a	Annealing temperature
TALEN	Transcription activator-like effector nuclease
TM	Trademark
U	Enzyme unit
UCT	University of Cape Town
UTP	Uracil triphosphate
VEP	Variant Effect Predictor
VT	Ventricular tachycardia
VUS	Variant of uncertain significance
WHO	World Health Organisation

List of gene identifiers

The following is a list of the Human Genome Organisation Gene Nomenclature Committee (HGNC) approved gene names for all genes mentioned in the main text of this submission (not including supporting appendices)

ABCB4	ATP binding cassette subfamily B member 4	HGNC:45
ACTC1	Actin alpha cardiac muscle 1	HGNC:143
ADAL	Adenosine deaminase like	HGNC:31853
ADRB2	Adrenoceptor beta 2	HGNC:286
AGTR1	Angiotensin II Receptor Type 1	HGNC:336
AKAP9	A-kinase anchoring protein 9	HGNC:379
ALPK3	Alpha kinase 3	HGNC:17574
ASNA1	Guided entry of tail-anchored proteins factor 3, ATPase	HGNC:752
ATG2A	Autophagy related 2A	HGNC:29028
CCSER1	Coiled-coil serine rich protein 1	HGNC:29349
CDH2	Cadherin 2	HGNC:1759
CTNNA3	Catenin alpha 3	HGNC:2511
DES	Desmin	HGNC:2770
DHTKD1	Dehydrogenase E1 and transketolase domain containing 1	HGNC:23537
DMD	Dystrophin	HGNC:2928
DPP6	Dipeptidyl peptidase like 6	HGNC:3010
DSC2	Desmocollin 2	HGNC:3036
DSG2	Desmoglein 2	HGNC:3049
DSP	Desmoplakin	HGNC:3052
ENGASE	Endo-beta-N-acetylglucosaminidase	HGNC:24622
FAM104B	Family with sequence similarity 104 member B	HGNC:25085
FBXO32	F-box protein 32	HGNC:16731
FLNC	Filamin C	HGNC:3756
FPGS	Folylpolyglutamate synthase	HGNC:3824
GATA4	GATA binding protein 4	HGNC:4173
GATAD1	GATA zinc finger domain containing 1	HGNC:29941
GBE1	1,4-alpha-glucan branching enzyme 1	HGNC:4180
GLA	Galactosidase alpha	HGNC:4296
GOLGA5	Golgin A5	HGNC:4428
HCN4	Hyperpolarisation activated cyclic nucleotide gated potassium channel 4	HGNC:16882
HEXB	Hexosaminidase subunit beta	HGNC:4879
ITGB5	Integrin subunit beta 5	HGNC:6160
ITPR2	Inositol 1,4,5-trisphosphate receptor type 2	HGNC:6181
JUP	Junction plakoglobin	HGNC:6207
KCNH2	Potassium voltage-gated channel subfamily H member 2	HGNC:6251
KCNJ12	Potassium voltage-gated channel subfamily J member 12	HGNC:6258
KCNK10	Potassium two pore domain channel subfamily K member 10	HGNC:6273
KCNN3	Potassium calcium-activated channel subfamily N member 3	HGNC:6292
KIAA1671	KIAA1671	HGNC:29345
KIF20A	Kinesin family member 20A	HGNC:9787
LAMA2	Laminin subunit alpha 2	HGNC:6482
LAMA3	Laminin subunit alpha 3	HGNC:6483
LFNG	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	HGNC:6560
MTO1	Mitochondrial tRNA translation optimisation 1	HGNC:19261
MYBPC3	Myosin binding protein C, cardiac	HGNC:7551
MYBPH	Myosin binding protein H	HGNC:7552
MYBPHL	Myosin binding protein H like	HGNC:30434

MYH6	Myosin heavy chain 6	HGNC:7576
MYH7	Myosin heavy chain 7	HGNC:7577
MYOM1	Myomesin 1	HGNC:7613
NDUFB1	NADH:ubiquinone oxidoreductase subunit B1	HGNC:7695
NNT	Nicotinamide nucleotide transhydrogenase	HGNC:7863
NRAP	Nebulin related anchoring protein	HGNC:7988
NRP1	Neuropilin 1	HGNC:8004
OBSCN	Obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF	HGNC:15719
PDLIM5	PDZ and LIM domain 5	HGNC:17468
PKP2	Plakophilin 2	HGNC:9024
PLEKHM2	Pleckstrin homology and RUN domain containing M2	HGNC:29131
PLN	Phospholamban	HGNC:9080
POLG	DNA polymerase gamma, catalytic subunit	HGNC:9179
PPCS	Phosphopantothenoylcysteine synthetase	HGNC:25686
PTEN	Phosphatase and tensin homolog	HGNC:9588
PXDNL	Peroxidasin like	HGNC:26359
RAD54L2	RAD54 like 2	HGNC:29123
RBM10	RNA binding motif protein 10	HGNC:9896
RYR2	Ryanodine receptor 2	HGNC:10484
SH3BGR	SH3 domain binding glutamate rich protein	HGNC:10822
SMYD1	SET and MYND domain containing 1	HGNC:20986
SOD2	Superoxide dismutase 2	HGNC:11180
SORBS2	Sorbin and SH3 domain containing 2	HGNC:24098
SRPRA	SRP receptor subunit alpha	HGNC:11307
SYMPK	Symplekin	HGNC:22935
SYNE2	Spectrin repeat containing nuclear envelope protein 2	HGNC:17084
SYNM	Synemin	HGNC:24466
TAF1A	TATA-box binding protein associated factor, RNA polymerase I subunit A	HGNC:11532
TGFB3	Transforming growth factor beta 3	HGNC:11769
TJP1	Tight junction protein 1	HGNC:11827
TMEM43	Transmembrane protein 34	HGNC:28472
TMEM87B	Transmembrane protein 87B	HGNC:25913
TNNI3	Troponin I3, cardiac type	HGNC:11947
TNNT2	Troponin T2, cardiac type	HGNC:11949
TP63	Tumour protein p63	HGNC:15979
TRMT13	tRNA methyltransferase 13 homologue	HGNC:25502
TRPM4	Transient receptor potential cation channel subfamily M member 4	HGNC:17993
TRPV1	Transient receptor potential cation channel subfamily V member 1	
TTN	Titin	HGNC:12403
UBA1	Ubiquitin like modifier activating enzyme 1	HGNC:12469
USP36	Ubiquitin specific peptidase 36	HGNC:20062
XIRP2	Xin actin binding repeat containing 2	HGNC:14303
ZNF284	Zinc finger protein 284	HGNC:13078
ZSCAN10	Zinc finger and SCAN domain containing 10	HGNC:12997

List of amino acid abbreviations and classifications

The classifications below are adapted from Zhang J. 2000. J Mol Evol*,* ***50****(1):57. When assessing missense variation, amino acid replacements within a category are conservative, and between categories are considered radical, as adapted from Dagan T et al. 2002.* Mol Evol Biol*,* ***19****(7):1023.*

Amino acid	Abbreviation	Charge	Volume and polarity
Alanine	A (Ala)	Neutral	Neutral, small
Arginine	R (Arg)	Positive	Polar, relatively large
Asparagine	N (Asn)	Neutral	Polar, relatively small
Aspartic acid	D (Asp)	Negative	Polar, relatively small
Cysteine	C (Cys)	Neutral	Special
Glutamic acid	E (Glu)	Negative	Polar, relatively small
Glutamine	Q (Gln)	Neutral	Polar, relatively small
Glycine	G (Gly)	Neutral	Neutral, small
Histidine	H (His)	Positive	Polar, relatively large
Isoleucine	I (Ile)	Neutral	Nonpolar, relatively small
Leucine	L (Leu)	Neutral	Nonpolar, relatively small
Lysine	K (Lys)	Positive	Polar, relatively large
Methionine	M (Met)	Neutral	Nonpolar, relatively small
Phenylalanine	F (Phe)	Neutral	Nonpolar, relatively large
Proline	P (Pro)	Neutral	Neutral, small
Serine	S (Ser)	Neutral	Neutral, small
Threonine	T (Thr)	Neutral	Neutral, small
Tryptophan	W (Trp)	Neutral	Nonpolar, relatively large
Tyrosine	Y (Tyr)	Neutral	Nonpolar, relatively large
Valine	V (Val)	Neutral	Nonpolar, relatively small

List of figures

List of tables

List of publications emanating from this thesis

Spracklen TF, Kasher PR, Kraus S, Botha TL, Page DJ, Kamuli S, Booi Z, Chin A, Laing N, Keavney BD, Ntusi NAB, Shaboodien G. Identification of a POLG variant in a family with arrhythmogenic cardiomyopathy and left ventricular fibrosis. *Circ Genom Precis Med* 2021; **14**(1): e003138.

Shaboodien G, Spracklen TF, Kamuli S, Ndibangwi P, Van Niekerk C, Ntusi NAB. Genetics of inherited cardiomyopathies in Africa. *Cardiovasc Diagn Ther* 2020; **10**(2): 262-278.

Spracklen TF, Keavney BD, Laing N, Ntusi NAB, Shaboodien G. Next-generation sequencing in the identification of genetic causes of cardiomyopathy. Submitted to *Cardiovasc Res*, November 2020.

Spracklen TF, Chakafana G, Schwartz PJ, Kotta MC, Shaboodien G, Ntusi NAB, Sliwa K. The genetics of peripartum cardiomyopathy: current knowledge, future directions and clinical implications. *Genes (Basel)* 2021; **12**(1): 103.

Abstract

Introduction: Cardiomyopathies are disorders of the myocardium that can lead to heart failure, arrhythmias and sudden death. Heritable forms include dilated, hypertrophic and arrhythmogenic cardiomyopathy (DCM, HCM and ACM respectively). As heterogeneous disorders, over 50 genes have been implicated in these cardiomyopathies to date. However, the yield of genetic testing ranges from less than 40% in idiopathic DCM to over 50% in ACM and HCM, indicating that many causal genes are yet to be identified. This is particularly true in African populations, where the genetics of cardiomyopathy is underexplored. In a review of the role of next-generation sequencing in gene discovery, over 20 new cardiomyopathy genes were found to have been identified through exome sequencing of cardiomyopathy patients. The literature review also highlighted the need for functional validation of newly identified disease genes. Therefore, the aims of this investigation were to utilise exome sequencing to identify disease-causing mutations in South African families with heritable cardiomyopathy, and to establish methods of variant validation through functional modelling in zebrafish.

Methods: Five probands and 34 relatives were included in this investigation. The probands and their relatives were clinically examined and diagnosed with DCM, HCM or ACM at Groote Schuur Hospital, Cape Town. Exome sequencing was performed on each of the five probands as well as at least one other family member. Variants of interest were identified by filtering the exome sequencing data by allele frequency, variant quality, variant consequence, predicted deleteriousness, and the potential inheritance patterns as determined by family history analysis. Variants occurring in known cardiomyopathy genes were prioritised, but genes outside the cardiac panel were considered based on literature mining, expression in the heart, and results of prior animal models. Candidate variants were validated by Sanger sequencing and assessed using international criteria for pathogenicity. The candidate ACM gene *POLG* was investigated in zebrafish larvae using two genetic manipulations. Firstly, zebrafish *polg* was disrupted using CRISPR/Cas9 in single-cell embryos and, at three days post-fertilisation, the phenotypic effects were compared to uninjected control larvae, as well as larvae in which other known cardiomyopathy genes were disrupted. Secondly, human *POLG* cDNA was cloned, and the c.2492A>G variant introduced using site-directed mutagenesis; this construct was used to generate variant *POLG* mRNA that was injected into zebrafish embryos. Larvae were phenotypically examined at four days post-fertilisation and compared to three control groups (unmutated *POLG*-injected, water-injected, and uninjected embryos).

Results: In three families, genotype-phenotype correlations were identified that have not yet been reported in South Africa, although this genetic overlap between cardiomyopathies has been described elsewhere. *Family 1:* the mutation *MYH7* c.4394C>T (p.S1465L) was identified in three siblings with DCM. Although *MHY7* is typically associated with HCM, mutations in this region have been reported in DCM patients in other populations. *Family 2:* the mutation *GLA* c.774_775del (p.R259Rfs*5) was found in a mother and her son, both of whom had been diagnosed with HCM. The finding of a pathogenic truncating *GLA* mutation in this family resulted in the genetic re-diagnosis of those individuals with Fabry disease, an HCM phenocopy. *Family 3:* in this large DCM family consisting of three affected brothers and their nephew, no pathogenic variants were identified, but two variants of uncertain significance (VUSs) were found in the genes *DSC2* and *PKP2*. Both variants fulfilled some criteria for pathogenicity, but have not been associated with DCM in South African patients before. In Families 4 and 5, no mutations in known cardiomyopathy-causing genes were identified. *Family 4:* exome sequencing revealed the variant *POLG* c.2492A>G (p.Y831C) in this ACM family, with a clinical phenotype consisting of arrhythmia and left ventricular fibrosis. This was a VUS, but *in vivo* modelling using CRISPR/Cas9 in zebrafish larvae demonstrated that disruption of the gene may impair cardiac development, while expression of the c.2492A>G variant in zebrafish larvae resulted in a significant reduction in heart rate, ventricle size and cardiac output. These results indicate that *POLG* variation may underly the arrythmia observed in the family, while prior mouse models reported that *POLG* mutations can induce cardiac fibrosis. *Family 5:* rare, compound heterozygous missense mutations in *ITGB5* were identified as the candidate causative variants in this small family with severe paediatric DCM, possibly affecting adhesion of cardiomyocytes to the extracellular membrane.

Conclusion: In total, pathogenic or likely pathogenic mutations were identified in two out of five families studied, while three VUSs with moderate or strong pathogenic potential were identified in two other families. The potential role of *POLG* in human cardiomyopathy and arrhythmic phenotypes is a finding that should be explored further, as should the putative role of *ITGB5* in paediatric cardiomyopathy. This study indicates how exome sequencing, combined with *in vivo* functional analysis, can identify variants that are likely to contribute to disease in human patients. These techniques may prove useful in bridging the gap in cardiomyopathy knowledge in Africa.

Thesis overview

The work presented in this thesis addresses the genetics of familial cardiomyopathy. Although a genetic aetiology is often suspected in cardiomyopathy, with over 50 cardiomyopathy genes described to date, many causal genes have not yet been identified. In particular, the genetics of cardiomyopathy in Africa remains largely unexplored.

In **Chapter 1**, heritable cardiomyopathies are introduced, and the rationale and aims of the study are presented. Chapter 1 also contains a literature review titled “Next-generation sequencing in cardiomyopathy” in which modern sequencing techniques are explained and the new cardiomyopathy genes that have been identified using these techniques are discussed. This chapter also gives an overview of functional models of cardiomyopathy.

Chapter 2 describes the methods used to investigate the genetics of cardiomyopathy in selected patients from a South African cohort. Included is the study design as well as an overview of the molecular process from DNA extraction to variant identification and functional validation using zebrafish genetic manipulations.

Subsequently, **Chapters 3 and 4** present the results of these genetic investigations. In Chapter 3, families are described in which mutations in known cardiomyopathy genes were found as the causes of dilated and hypertrophic cardiomyopathy, giving rise to genotype-phenotype correlations which have not been characterised in South African patients to date. Chapter 4 focusses on potential new cardiomyopathy genes, *POLG* and *ITGB5*, identified in two families with arrhythmogenic and paediatric cardiomyopathy, respectively, in which the genetic analysis did not identify disease-causing mutations in established cardiomyopathy genes.

Chapter 5 explores the zebrafish model as a functional model for validating genetic findings in African cardiomyopathy patients, using two different genetic techniques and one of the genes from Chapter 4 (*POLG*) as a candidate.

Chapter 6 summarises the principal findings of the study and considers its practical implications as well as directions for future research in the field.

Chapter 1 Introduction

1.1 Cardiovascular disease and cardiomyopathy

Cardiovascular disease (CVD) is the leading cause of mortality in the world as evidenced by the latest release from the Global Burden of Disease study (Global Burden of Disease Collaborators 2018, Lopez and Adair 2019). Non-communicable diseases (NCDs) such as those of the cardiovascular system are typically chronic and progressive in nature; consequently, they are becoming an increasing healthcare burden in low- and middle-income countries (LMICs) due to social and economic factors, as well as the high prevalence of communicable diseases in these populations (Mayosi et al. 2009). Indeed, according to the World Health Organisation (WHO) up to 80% of NCD-related deaths, and 90% of those before the age of 60 years, are reported to occur in the developing world (WHO 2018). It has been estimated that over 75% of deaths due to CVD, similarly, are in LMICs, where the sufficient detection and treatment of CVD may not be possible (WHO 2017). Although many instances of CVD can be predisposed by lifestyle choices, with risk factors including physical inactivity, obesity and tobacco use (Perk et al. 2012, Yusuf et al. 2004), some forms of CVD such as cardiomyopathy cannot be influenced in this way, and may be heritable (Lahrouchi, Behr, and Bezzina 2016, McNally and Puckelwartz 2015).

Cardiomyopathies are disorders of the myocardium, the muscle layer of the heart. When this layer is compromised structurally and/or functionally, the heart's ability to contract and pump blood may be affected. An impairment in myocardial function may progress to cardiac dysfunction and heart failure; furthermore, patients with cardiomyopathy are at risk of arrhythmias and sudden cardiac death (SCD) (Towbin 2014). Although few studies of the incidence of the disease in African populations have been conducted (Figure 1.1), it is recognised as a prevalent healthcare burden with considerably greater incidences of, for example, dilated cardiomyopathy (DCM) and peripartum cardiomyopathy (PPCM), when compared to European and American populations (Sliwa, Damasceno, and Mayosi 2005). Out of 1,006 patients presenting with acute heart failure in 12 countries in Africa, cardiomyopathy was identified as the cause of heart failure in 26.5% of cases (Damasceno et al. 2012).

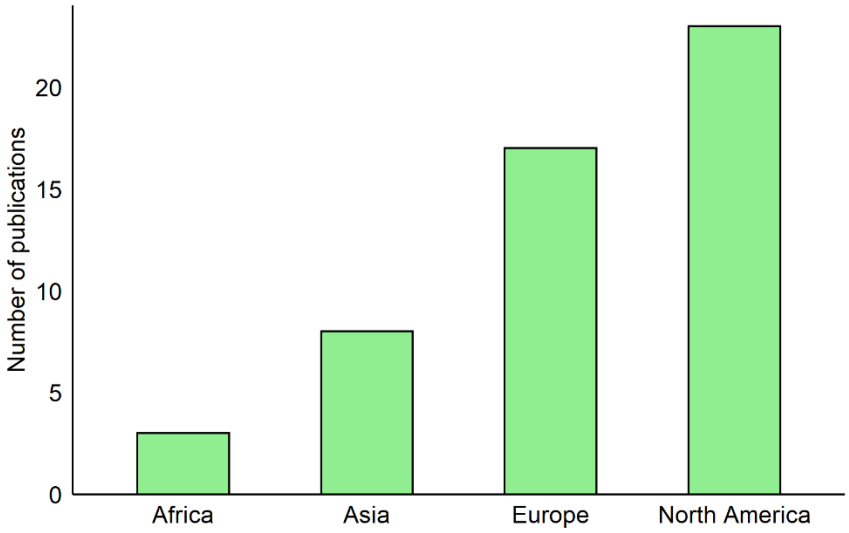


Figure 1.1: Comparison of cardiomyopathy studies over the last decade by region. Shown is the number of publications that were found to report on the incidence or prevalence of cardiomyopathy in Africa, Asia, Europe and North America, between 2011 and 2021. The numbers were determined by a search of PubMed using the Medical Subject Heading terms “incidence” OR “prevalence” AND “cardiomyopathy”, followed by manual refinement of the 2,140 results by excluding papers that were unrelated or did not report on the frequency of cardiomyopathy. Publications from Australia, Middle East and South America were not included in this analysis.

There are many types of cardiomyopathy, some of which are part of a systemic disorder while others affect only the heart. These conditions can be divided into several morphological phenotypes, although the basis for these classifications is not always clear (Elliott et al. 2008). Nevertheless, a group of heritable cardiomyopathies is evident; these are caused by mutations in genes expressed in myocardial tissues, tend to present in younger individuals and are usually more progressive in nature than acquired forms of the disease (McNally, Golbus, and Puckelwartz 2013). Cardiomyopathies with documented genetic influences include DCM, as well as arrhythmogenic, hypertrophic and restrictive cardiomyopathy (ACM, HCM and RCM, respectively) and left ventricular noncompaction (LVNC), where up to 50% of cases may have a familial origin (Baig et al. 1998, Corrado, Basso, and Thiene 2009). The work in this thesis is primarily concerning ACM, DCM and HCM (Figure 1.2); these and other forms of heritable cardiomyopathy will be described in more detail below (Section 1.2).

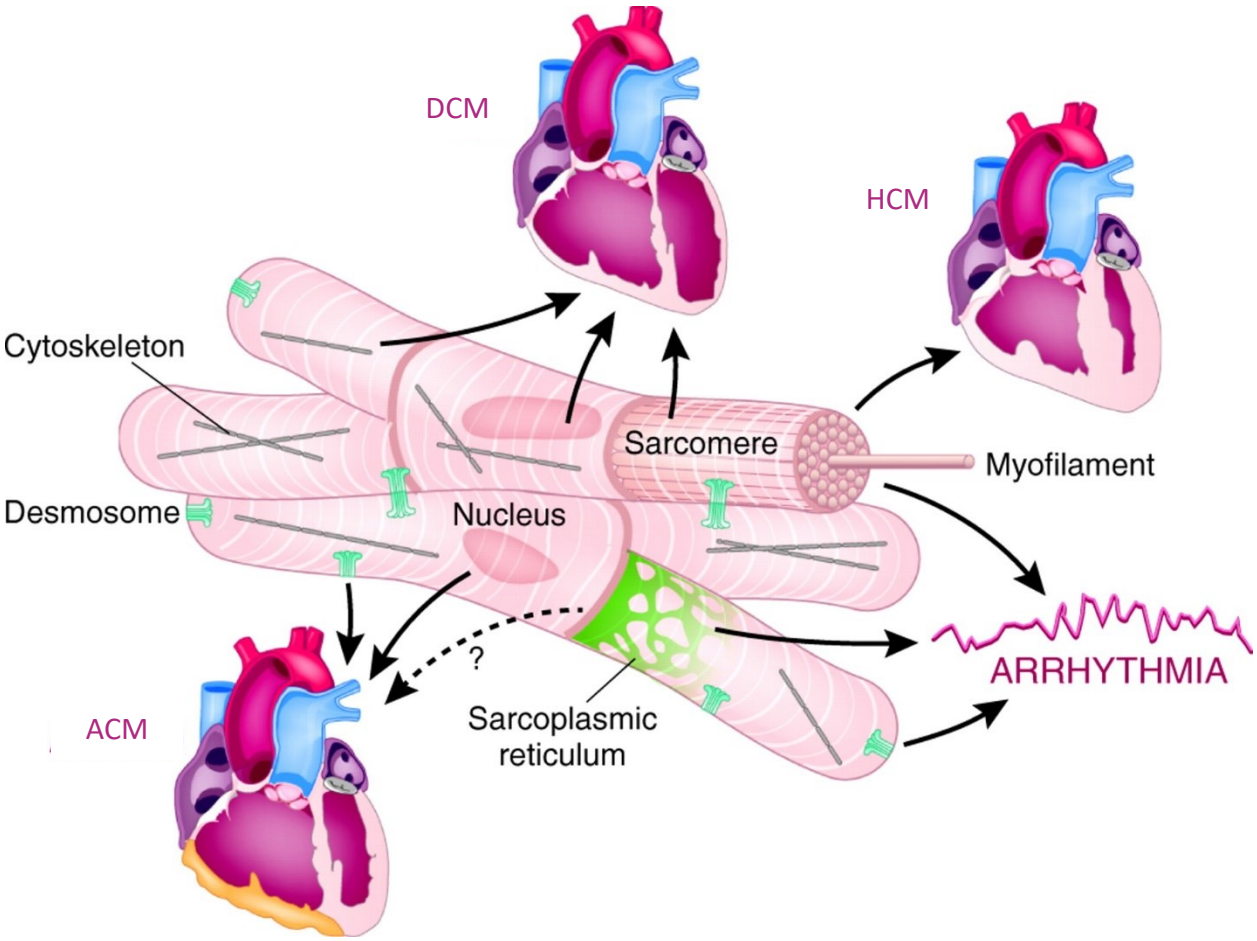


Figure 1.2: Illustration of DCM, ACM and HCM phenotypes. Also indicated are the myocardial cell components that inherited mutations typically affect. HCM is traditionally caused by mutations in genes encoding the sarcomere, the contractile unit of the heart. ACM is usually caused by mutations in genes encoding the desmosome, which is involved in cell adhesion. DCM is the most heterogeneous form of cardiomyopathy and may be caused by mutations in cytoskeletal, sarcomeric, nuclear or transcription co-activator genes. Adapted from McCauley MD and Wehrens XH. 2009. *Dis Model Mech*, **2**(11-12):565. *ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy*

1.2 Clinical genetic description of the heritable cardiomyopathies

1.2.1 Clinical genetic description of dilated cardiomyopathy

The most common form of cardiomyopathy, DCM, is also the most genetically heterogeneous. The exact prevalence of DCM is unclear but may be as high as 1/250 (Hershberger, Hedges, and Morales 2013). DCM is also the most common form of cardiomyopathy in Africa where it is frequently reported to be the leading or second most common cause of heart failure, after

hypertension (Adebayo et al. 2009, Damasceno et al. 2012, Falase and Ogah 2012, Makubi et al. 2014, Sliwa et al. 2008).

According to the European Society of Cardiology (ESC) guidelines, DCM is defined as the presence of left ventricular dilation and left ventricular systolic dysfunction. Characterised by enlargement of the left ventricle and systolic dysfunction (Elliott et al. 2008), the left ventricular dilatation with impaired contraction typical of DCM may result from many insults, both genetic and environmental. Familial DCM, if defined as the presence of two or more family members manifesting DCM, constitutes around half of all DCM cases. Familial DCM has been attributed to mutations in genes encoding various components of the cardiac myocyte. These include cytoskeletal, sarcomeric and nuclear envelope proteins, as well as transcription factors and ion channels (McNally and Mestroni 2017). Candidate gene screening has identified over 50 genes associated with DCM, although few of these have been shown to be common causes of the disease. The most frequently mutated gene in DCM is the giant sarcomeric protein-coding gene *TTN*, truncating mutations in which can cause DCM through impaired force generation or transmission in the myocardium (McNally and Mestroni 2017). However, *TTN* and other genes only account for approximately 40% of patients with familial DCM (Sturm and Hershberger 2013, Sweet, Taylor, and Mestroni 2015), indicating that there are yet more causative genes to be discovered.

1.2.2 Clinical genetic description of hypertrophic cardiomyopathy

Defined as cardiac hypertrophy in the absence of loading conditions, HCM is typically characterised by hypertrophy most severely involving the interventricular septum, the presence of a dynamic left ventricular outflow tract obstruction, and a propensity to SCD due to ventricular arrhythmias (Elliott et al. 2006, 2008). The increased stiffness of the ventricular walls can also lead to diastolic dysfunction, and some patients enter a “burned out” phase of left ventricular dilatation and systolic dysfunction mimicking primary DCM. It is a fairly common disease, with a reported prevalence of up to 1/500 (Maron et al. 1995, Zou et al. 2004). The prevalence of the disease in Africa is unclear, but between 0.07% and 5.75% of African patients referred for echocardiography were diagnosed with HCM (Abegaz 1990, Jingi et al. 2013, Maro, Janabi, and Kaushik 2006, Ogah et al. 2008, Raphael et al. 2018). In contrast to DCM, the genetics of HCM are less diverse, and up to 50% of HCM cases carry mutations in known cardiomyopathy genes

(Gruner et al. 2013, Jensen et al. 2013). The genes that are primarily implicated in HCM are those involved in structure and function of the cardiac sarcomere (Marian and Braunwald 2017).

1.2.3 Clinical genetic description of arrhythmogenic cardiomyopathy

ACM is a form of cardiomyopathy in which the ventricular myocardium is progressively infiltrated by fibrofatty tissue (Marcus et al. 2010). ACM is an inclusive term that refers to an arrhythmogenic disease of the myocardium characterised by ventricular arrhythmias and underlying structural abnormalities, not caused by loading conditions such as valvular heart disease, hypertensive heart disease, or ischaemia (Towbin et al. 2019). ACM includes, but is not limited to, ARVC and may overlap with other cardiomyopathy phenotypes, in particular DCM (Towbin et al. 2019). Many have advocated for the use of the term ACM, rather than ARVC, due to the frequent involvement of structural and functional abnormalities of the left ventricle (Corrado, Basso, and Judge 2017), a recommendation that is adhered to in this thesis. Often exacerbated by exercise, ACM is an important cause of arrhythmia and SCD amongst the young and athletic. The prevalence of ACM ranges from 1/5000 to 1/2000 in European populations (Pilichou et al. 2016), but the prevalence of ACM in Africa is largely unknown (Watkins et al. 2009). It has been reported that between 0.25% and 0.43% of African patients referred for echocardiography are subsequently diagnosed with ACM (James et al. 2012, Raphael et al. 2018).

Most of the implicated genes in ACM encode components of the cardiac desmosome: mutations in *DSC2*, *DSG2*, *DSP*, *JUP* and *PKP2* have been associated with impaired electrical and mechanical stability of the myocardial tissue, leading to ACM phenotypes (Haugaa et al. 2016, Ohno 2016). However, other cardiomyopathy genes such as *CTNNA3*, *DES*, *RYR2*, *TGFB3*, *TMEM43* and *TTN* have been associated with ACM as well (Ohno 2016). The genetic cause of the disease has not been identified in up to 50% of cases (Quarta et al. 2011).

1.2.4 Clinical genetic description of left ventricular noncompaction

LVNC is a form of cardiomyopathy in which the left ventricular walls develop prominent trabeculations (spongy collections of muscle fibres caused by failure of compaction during cardiac development) which can lead to heart failure, arrhythmias and SCD (Captur and Nihoyannopoulos 2010). The prevalence of the disease is largely unknown and may be due to indistinct diagnostic criteria. In a prospective study of 700 patients referred for cardiac magnetic resonance in New

York, the LVNC diagnosis rate ranged from 3% to 39% when using different imaging criteria (Ivanov et al. 2017). Although the incidence of the disease in Africa is unclear, approximately 7% of 780 sub-Saharan Africans referred to a cardiomyopathy clinic were diagnosed with LVNC (Peters et al. 2012).

Similar to DCM, LVNC has a heterogeneous genetic basis, with candidate mutations identified in cytoskeletal, ion channel, sarcomeric and nuclear membrane genes (Dong et al. 2017). The pathogenesis of LVNC and the process of noncompaction are unclear at this stage, and a greater understanding of the genetics of the condition may help clarify how it occurs.

1.2.5 Clinical genetic description of restrictive cardiomyopathy

RCM is a disorder characterised by stiffening of the myocardium, while chamber size and systolic function remain normal. Impaired filling of the ventricles can lead to arrhythmias and symptoms of heart failure (Elliott et al. 2008). It is thought to be the least common form of cardiomyopathy, although the exact prevalence is unclear (Muchtar, Blauwet, and Gertz 2017). Limited data on the incidence in Africa is available; however, when investigating 6,275 Ethiopian and 3,908 Malawian CVD patients, RCM was reported in 0.11% and 0.08% of cases, respectively (Jingi et al. 2013, Yadeta et al. 2017). RCM can be caused by mutations in the genes *ACTC1*, *MYH7*, *TNNT2* and *TNNI3* (Muchtar, Blauwet, and Gertz 2017).

1.3 The genetics of familial cardiomyopathy

In excess of 50 genes have been associated with cardiomyopathy to date (McNally, Golbus, and Puckelwartz 2013). Although the different cardiomyopathies are clinically distinct, there is a degree of genetic overlap between them, as mutations within some genes have been reported to induce more than one type of cardiomyopathy (Kalyva et al. 2014). The mutations which cause cardiomyopathies usually occur within genes involved in cardiac muscle contraction, cytoskeletal organisation and cell adhesion (Bao et al. 2013, Zhao et al. 2016); however, the known cardiomyopathy-related genes cannot explain all familial cases of the disease, with as many as 40-60% of patients lacking a conclusive genetic diagnosis (Gruner et al. 2013, Quarta et al. 2011, Sturm and Hershberger 2013, Sweet, Taylor, and Mestroni 2015). This is particularly true in African populations, where relatively few genetic studies have been conducted to date and the genetics of cardiomyopathy is largely unknown (Shaboodien et al. 2020).

In previous work conducted in South African populations, a founder mutation within the gene *PKP2* was described in several Afrikaner families with ACM (Watkins et al. 2009), while similar founder mutations have been described in *MYBPC3* and *MYH7* in South African HCM families (Moolman-Smook et al. 1999). Additional disease-causing variants were described in South African cardiomyopathy patients in the known cardiomyopathy genes *DSP* (Fish 2010), *MYBPC3* (Moolman-Smook et al. 1998, 1999, Ntusi et al. 2016), *MYH7* (Moolman, Brink, and Corfield 1993, 1995, Ntusi et al. 2016, Posen et al. 1995), *PLN* (Fish 2016, Fish et al. 2016), *TNNI3* (Mouton et al. 2015) and *TNNT2* (Moolman et al. 1997). The novel ACM gene *CDH2* was first identified in a large South African family (Fish 2016, Mayosi et al. 2017). Outside South Africa, limited investigations have been performed on African cohorts, in which mutations were reported in known cardiomyopathy genes in Egyptian and Tunisian HCM patients (Jaafar et al. 2016, Kassem et al. 2013), a large Moroccan family with DCM (Adadi et al. 2018), and a single patient from East Africa (Choung et al. 2017).

Familial cases of cardiomyopathy are typically observed to have an autosomal dominant (AD) pattern of inheritance, in which a single copy of the mutated gene is sufficient to cause disease (Towbin 2014). However, autosomal recessive (AR), X-linked, mitochondrial and syndromic forms of cardiomyopathy have also been described (Csányi et al. 2016, Theis et al. 2011, Towbin 2014). Identifying the causative mutations in families is of importance, both to family members who may be at risk of developing the disease, and also in the elucidation of disease pathways in the heart. The study of inherited cardiomyopathy, however, can be complicated by the genetic heterogeneity of the disease, as well as factors such as incomplete penetrance, phenotypic diversity and the presence of genetic modifiers.

1.4 Review: Next-generation sequencing in cardiomyopathy

Due to the large number of known cardiomyopathy genes, as well as the need to consider previously unreported genes for their role in cardiomyopathy, techniques such as candidate gene screens are not suitable for uncovering the genetic basis of the disease. While several cardiomyopathy-causing variants have been identified using traditional Sanger-based DNA sequencing of candidate genes, next-generation sequencing (NGS) is likely the technique that will enable the elucidation of additional cardiomyopathy genes (McNally and Puckelwartz 2015). NGS refers to massively parallel non-Sanger-based high-throughput DNA sequencing

technologies that include, amongst others, exome sequencing, whole genome sequencing and targeted sequencing of known or candidate disease genes (Figure 1.3).

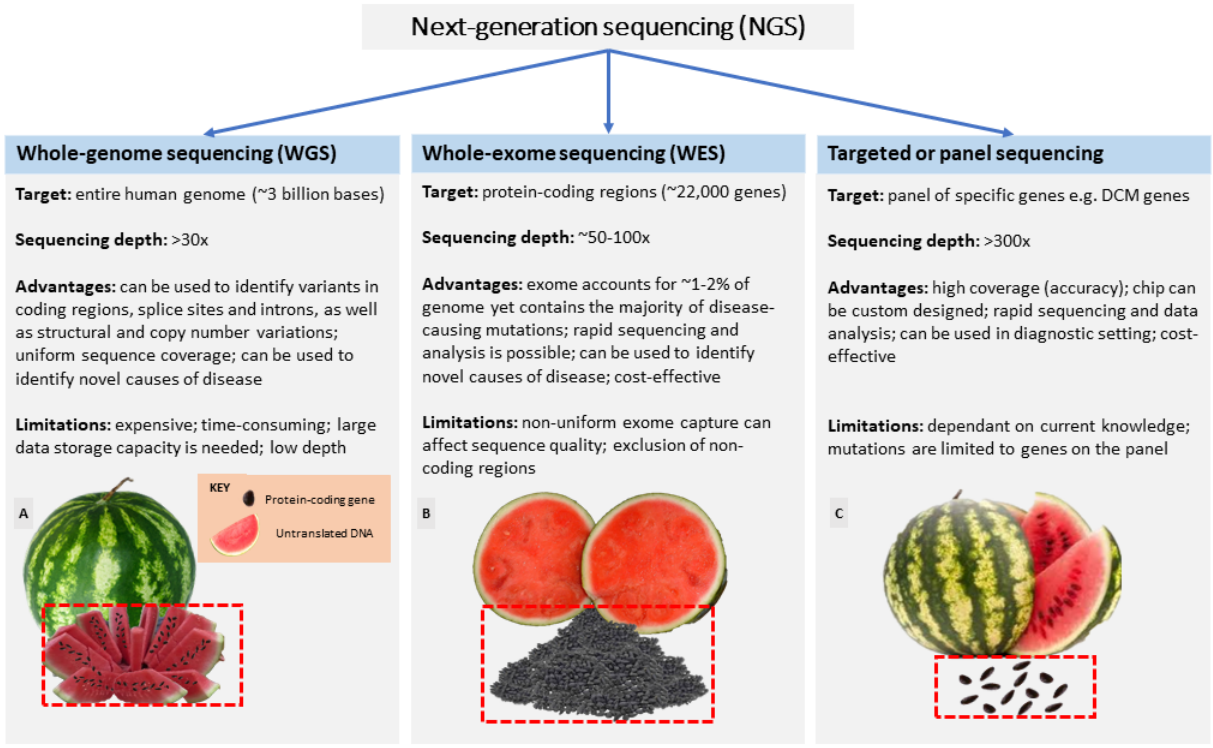


Figure 1.3: An overview of the next-generation sequencing techniques, their advantages and limitations. *NGS, next-generation sequencing*

NGS has allowed the identification of novel variants and genes for cardiomyopathy, although the detection of variants of uncertain clinical significance is a diagnostic challenge. One of the early benefits of NGS has, perhaps unexpectedly, accrued from exome and genome sequencing of large populations who were not selected for the presence of genetic cardiomyopathies, but generally in the context of exome sequencing studies of complex traits. The aggregation of these sequences in readily queriable formats in the Exome Aggregation Consortium (ExAC) and later Genome Aggregation Database (gnomAD) databases has provided new scientific insights and directly impacted clinical variant interpretation in cardiomyopathy (Karczewski et al. 2019, Lek et al. 2016). These datasets have clearly shown that the standard of proof for a variant in a new gene to be considered causal in the literature has been too low in some instances. An important

paper considering some published findings in the light of variant frequencies in the ExAC database illustrated that interrogation of a sufficiently large NGS database of disease-free controls cast doubt on the pathogenicity of some previously published variants, or in some cases even whole genes, due to their high population frequency (Tayal et al. 2017, Walsh et al. 2017). The ExAC and/or gnomAD databases have been incorporated into a semi-automated decision-making tool for cardiomyopathies and other inherited cardiac conditions (ICCs) that integrates case-specific data with expertly curated computational annotations, providing decision support for clinically detected variants according to the internationally accepted American College of Medical Genetics and Genomics (ACMG) guidelines (Whiffin et al. 2018). The impact of large databases such as gnomAD on variant interpretation in non-European populations is limited hitherto by the small number of non-Europeans that have had exome or genome sequencing; much work remains necessary to catalogue the extent of genetic variation in non-European populations in equal detail.

Although several putative cardiomyopathy genes have been discovered through NGS experiments, the exact impact of these genes on the disease phenotype is yet to determined. Understanding these may improve our knowledge of how primary genetic mutations, possibly in concert with modifier variants, could alter the phenotypic expression of a disease, leading to a better understanding of genotype-phenotype correlations. This review (Section 1.4) investigates how NGS technologies have advanced the field of genetic cardiomyopathy by enabling the discovery of new disease-causing genes, and considers the likely future roles for panel testing, exome sequencing and genome sequencing in cardiomyopathy.

1.4.1 An overview of next-generation sequencing

The sequencing of genetic information (such as genes) is a fundamental first step in the study of hereditary diseases. The traditional method of sequencing, referred to as Sanger sequencing, involves directly sequencing short fragments of DNA using dye-labelled terminator nucleotides. Sanger sequencing is often prohibitively expensive and time consuming in large-scale sequencing projects. In the last decade, the emergence of second-generation, or NGS, technologies has enabled the rapid and simultaneous sequencing of large tracts of DNA. Although several different NGS platforms are available and are reviewed more thoroughly elsewhere (Heather and Chain 2016, Reuter, Spacek, and Snyder 2015), they all involve the clonal amplification and parallel reading of spatially separated target regions of DNA (Morey et al. 2013). Third-generation

sequencing technologies such as Oxford Nanopore are being introduced, in which much longer fragments of DNA can be sequenced (Magi et al. 2018); these technologies, however, are beyond the scope of this review as they are not yet in routine use.

Briefly, NGS involves the fragmentation of DNA (usually by mechanical means) into smaller pieces, followed by targeted enrichment to capture specific genomic regions (i.e. target genes or the exome; this step is omitted in genome sequencing). Clonal amplification is then necessary to increase the signal generated in the sequencing stage (Morey et al. 2013). Sequencing involves the parallel, cyclic reading of each target DNA fragment. The most commonly used chemistries for sequencing include Illumina sequencing-by-synthesis, Ion Torrent semiconductor sequencing and pyrosequencing (Heather and Chain 2016, Reuter, Spacek, and Snyder 2015). With each incorporation of a nucleotide, the technologies produce a signal such as light, fluorescence or hydrogen ions – the detection of these signals by a camera or a semiconductor, and the interpretation of these signals by a computer, allows the nucleotide sequence of each fragment to be determined.

The sequencing process produces short sequence reads for each DNA fragment, and the first challenge in data processing is to determine the order in which these fragments occur in the human genome (Morey et al. 2013). Another important step in bioinformatic processing is the removal of low-quality sequence reads or alignments, and the identification of sequence variants. These variants are typically defined as any sequence deviation from the human reference genome. It is these deviations that are investigated for potential pathogenic mutations in human patients, and the quandary that many researchers face is differentiating benign polymorphisms from true disease-causing mutations.

1.4.2 Exome sequencing in cardiomyopathy

Exome sequencing, also known as whole exome sequencing, is an NGS technique in which only the protein-coding region of the genome (accounting for 1-2% of the genome) is sequenced. By focussing on coding regions, variants within introns and possible regulatory regions may be omitted; however, approximately 85% of all disease-causing variants have been reported to occur in coding regions or splice sites (Choi et al. 2009). It should be noted that as the role of intronic and regulatory variants in disease continues to be elucidated, this 2009 statistic may eventually prove to be an overestimate. An advantage of exome sequencing is a reduction in data generation

(compared to sequencing the entire genome), while enriching the data for variants likely to cause disease (Bamshad et al. 2011, Choi et al. 2009). Additionally, the technique is unbiased and allows for the investigation of genes that have never been associated with cardiomyopathy before (Rabbani, Tekin, and Mahdieh 2014). Limitations of exome sequencing include the heterogeneity in capture efficiency across the exome, and the difficulty of assessing genes not previously reported to cause cardiomyopathy (Rehm et al. 2013). Despite this, exome sequencing remains the method of choice in cardiomyopathy research due to the relative cost-effectiveness of the technique and has been used to identify new disease genes for ACM, DCM, HCM, RCM and LVNC. These will be discussed below.

1.4.2.1 Exome sequencing in dilated cardiomyopathy

To date, 12 new DCM genes have been identified through exome sequencing, compared to one identified through targeted sequencing and one through whole genome sequencing (Table 1.1). Despite familial DCM being inherited in a typically dominant pattern, all but one of the 12 new genes were described in families with AR disease, and of these, eight were reported in paediatric cardiomyopathies. Exome sequencing has so far had limited applicability in novel gene discovery in AD or adult-onset DCM.

Mutations in *ALPK3* have been identified in five cardiomyopathy families, four of which were consanguineous (Table 1.1). Originally reported by Almomani et al. (2016) in two DCM families of Dutch and Moroccan ancestry, further genetic screening identified additional *ALPK3* mutations in a DCM family of Turkish descent as well as a Pakistani HCM family (Almomani et al. 2016, Phelan et al. 2016). In total, five patients were discovered to carry homozygous truncating mutations in *ALPK3*; three of these patients died within five days of birth, but two survived to the age of 11 years, at which stage their disease showed a severe HCM phenotype (Almomani et al. 2016). In two of the families, *ALPK3* truncating mutations were associated with AR DCM that was later reclassified as HCM (Çağlayan et al. 2017, Phelan et al. 2016).

Table 1.1: New DCM disease genes identified through NGS

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
ADRB2	c.742delG (p.V248Wfs*19)	0	Digenic (inherited with *TTN* mutation)	AD DCM	Not specified	**Targeted sequencing**	Waldmüller et al. (2015)
ALPK3	c.2018delC (p.N675Sfs*30) c.3792G>A (p.W1264*)	0.000021 0	Compound heterozygous	AR DCM to HCM	Consanguineous Pakistani family	*SNP array*	Phelan et al. (2016)
	c.3781C>T (p.R1261*)	0.000034	Homozygous	AR DCM	Consanguineous Moroccan family	**Exome sequencing**	Almomani et al. (2016)
	c.3792G>A (p.W1264*)	0	Homozygous	AR DCM	Consanguineous Turkish family	*Candidate gene screening*	Almomani et al. (2016)
	c.4736-1G>A (p.V1579Gfs*30)	0.000015	Homozygous	AR DCM	Consanguineous Dutch family	**Exome sequencing**	Almomani et al. (2016)
	c.2018delC (p.N675Sfs*30)	0.000021	Homozygous	AR DCM to HCM	Not specified; non-consanguineous family	*Exome sequencing*	Çağlayan et al. (2017)
ASNA1	c.913C>T (p.Q350*) c.488T>C (p.V163A)	0 0	Compound heterozygous	AR DCM	Not specified; non-consanguineous family	**Exome sequencing**	Verhagen et al. (2019)
FBXO32	c.727G>C (p.G243R)	0	Homozygous	AR DCM	Consanguineous Saudi family	**Exome sequencing**	Al-Yacoub et al. (2016)
GATAD1	c.304T>C (p.S102P)	0	Homozygous	AR DCM	Consanguineous Norwegian family	**Exome sequencing**	Theis et al. (2011)
KCNJ12	c.1000_1002delGAG (p.E334del)	0	Heterozygous	AD DCM	Chinese family	**Exome sequencing**	Yuan et al. (2017)
MYBPHL	c.763C>T (p.R255*)	0.001403	Heterozygous	AD DCM	Not specified; non-consanguineous family	**Genome sequencing**	Barefield et al. (2017)
NRAP	c.4504C>T (p.R1502*)	0.000343	Homozygous	AR DCM	Not specified; non-consanguineous family	**Exome sequencing**	Truszkowska et al. (2017)
PLEKHM2	c.2156delAG (p.K645A*12)	0	Homozygous	AR DCM	Consanguineous Bedouin family	**Exome sequencing**	Muhammad et al. (2015)
PPCS	c.320_334del (p.107_111del) c.538G>C (p.A180P)	0.000100 0	Compound heterozygous	AR DCM	Non-consanguineous European family	**Exome sequencing**	Iuso et al. (2018)
	c.698A>T (p.E233V)	0	Homozygous	AR DCM	Consanguineous Arabic family	**Exome sequencing**	Iuso et al. (2018)
SOD2	c.542G>T (p.G181V)	0.000004	Homozygous	AR DCM	Consanguineous Dutch family	**Exome sequencing**	Almomani et al. (2020)
SYNM	c.1612T>C (p.W538R)	0.000049	Heterozygous	AD DCM	Non-consanguineous Chinese family	**Exome sequencing**	Zhang et al. (2019)
TAF1A	c.251T>C (p.L84S) c.1021G>A (p.G341R)	0.000033 0.000029	Compound heterozygous	AR DCM	Not specified; non-consanguineous family	**Exome sequencing**	Long et al. (2017)
XIRP2	c.5460delT (p.F1820Lfs*2) c.6968dupA (p.N2323Kfs*15)	0 0	Digenic (inherited with *TNNT2* mutation)	Sporadic DCM	Not specified; non-consanguineous family	**Exome sequencing**	Long et al. (2015)

***** Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; MAF, minor allele frequency; SNP, single nucleotide polymorphism

PPCS was implicated as a novel DCM gene when rare mutations in the gene were detected using exome sequencing in two families of Arabic and European ancestry with early-onset AR DCM (Iuso et al. 2018). The age of onset in mutation carriers ranged from two weeks to three years, and three of the five affected individuals' cardiomyopathy was fatal. Yeast cells transformed with the mutations were unable to grow efficiently, while patient fibroblasts had reduced levels of Coenzyme A which could be rescued by increasing expression of unmutated *PPCS*. Reducing expression of the gene in fruit flies led to a significant impairment of heart function (Iuso et al. 2018).

In contrast to *ALPK3* and *PPCS*, mutations in the genes *ASNA1*, *FBXO32*, *GATAD1*, *KCNJ12*, *NRAP*, *PLEKHM2*, *SOD2*, *SYNM* and *TAF1A* were reported in single DCM families (Almomani et al. 2020, Al-Yacoub et al. 2016, Long et al. 2015, Long et al. 2017, Theis et al. 2011, Truszkowska et al. 2017, Verhagen et al. 2019, Yuan et al. 2017, Zhang et al. 2019). These genes have not been associated with cardiomyopathy since their original report.

XIRP2 was not reported as a primary disease-causing gene, but rather as a potential modifier of the disease. The reason for this was that a patient, who had severe, early-onset DCM, carried *TNNT2* mutations that had previously been associated with a milder form of DCM (Li et al. 2001, Long et al. 2015): compound heterozygous truncating mutations in *XIRP2* were identified in this patient using exome sequencing and were suggested to contribute to the disease phenotype (Long et al. 2015). Such genetic modifiers were not considered by other authors but may explain phenotypic variability, where the effect of mild mutations can be exacerbated by the presence of additional genetic variation. Several rare deleterious *XIRP2* gene variants were identified by sequencing the gene in Chinese individuals with sudden unexplained nocturnal death or Brugada syndrome, indicating that *XIRP2* may contribute to arrhythmic phenotypes (Huang et al. 2018). An exome sequencing investigation of 142 DCM patients indicated that loss-of-function mutations in *PDLIM5* may act as modifiers which increase the susceptibility of the heart to the development of cardiomyopathy (Verdonschot et al. 2019).

1.4.2.2 Exome sequencing in hypertrophic cardiomyopathy

Exome sequencing has enabled the identification of an additional three genes, *FLNC*, *MTO1* and *SMDY1*, in familial HCM (Table 1.2; Appendix A).

Table 1.2: New HCM disease genes identified through NGS

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
FLNC	c.4824G>A (p.A1539T)	0	Heterozygous	AD HCM	Spanish family	**Exome sequencing**	Valdés-Mas et al. (2014)
	Various†	Various†	Various†	ACM, DCM, HCM, RCM, LVNC	Various†	*Various*†	Brodehl et al. (2016); Brun et al. (2020); Gómez et al. (2017); Hall et al. (2020) Jaafar et al. (2016); Liu et al. (2019); Ortiz-Genga et al. (2016); Reinstein et al. (2016); Rojnueangnit et al. (2019); Tucker et al. (2017); Valdés-Mas et al. (2014)
MTO1	c.1282G>A (p.A428T) c.1858dup (p.R620Kfs*8)	0.000051 0	Compound heterozygous	AR HCM	Italian family	**Exome sequencing**	Ghezzi et al. (2012)
	Various†	Various†	Various†	HCM	Various†	*Various*†	Baruffini et al. (2013); Ghezzi et al. (2012); O'Byrne et al. (2018); Taylor et al. (2014)
SMYD1	c.675delA (p.K225Nfs*8)	0	Not specified	LVNC	Chinese patient	*Exome sequencing*	Liu et al. (2019)
	c.814T>C (p.F272L)	0.000004	*De novo*	HCM	Chinese patient	**Exome sequencing**	Fan et al. (2019)

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

† See Appendix A for a full list of reported mutations

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Two siblings with severe infantile recessive HCM, both of whom died within 40 days of birth, were found to carry compound heterozygous mutations in *MTO1* upon exome sequencing (Ghezzi et al. 2012). Mutations in the gene have subsequently been reported in several sporadic and familial cases of HCM (Baruffini et al. 2013, Ghezzi et al. 2012, O'Byrne et al. 2018, Taylor et al. 2014). Fibroblasts from mutation-carrying patients demonstrated a reduced mitochondrial respiration rate which could be rescued by transfection with wildtype *MTO1* (Ghezzi et al. 2012). Subsequent investigation of *Mto1* null mice indicated slight ventricular dilation, arrhythmia, fibrosis and degeneration of myofibres in the absence of *Mto1* (Becker et al. 2014).

The muscle filamin gene *FLNC* was first associated with myofibrillar myopathy before mutations in the gene were reported by Valdés-Mas et al. (2014) in patients with sporadic and familial forms of HCM. It has since emerged as an important cause of cardiomyopathy, with numerous rare truncating mutations found in additional HCM patients as well as individuals with ACM, DCM, RCM and LVNC (Begay et al. 2016, Brodehl et al. 2016, Brun et al. 2020, Gómez et al. 2017, Hall et al. 2020, Jaafar et al. 2016, Janin et al. 2017, Liu et al. 2019, Nozari et al. 2018, Ortiz-Genga et al. 2016, Reinstein et al. 2016, Rojnueangnit et al. 2019, Tucker et al. 2017). FLNC is a sarcomeric, actin-binding protein that has also been shown to interact with muscular intercalated discs and play a role in cellular signalling. Expression of the human *FLNC* mutations in rat cardiomyocytes led to abnormal FLNC protein aggregation and disrupted protein structure (Valdés-Mas et al. 2014), while *FLNC* mutations introduced into human stem cell-derived cardiomyocytes reduced their contractility (Tucker et al. 2017). Transcriptomic analysis of left ventricular tissue indicated a potential dysregulation of cell adhesion, extracellular matrix organisation and cardiovascular development amongst *FLNC* mutation carriers compared to age-matched controls (Hall et al. 2019).

Using exome sequencing, a Chinese patient with severe HCM and atrial enlargement was found to carry a *de novo* missense mutation in the gene *SMYD1* (Fan et al. 2019). Due to its known role in heart development (Nagandla et al. 2016), as well as a prior mouse model demonstrating cardiac hypertrophy in the absence of the gene (Franklin et al. 2016), *SMYD1* was previously considered a candidate gene for HCM. The missense mutation was predicted to affect the nuclear localisation signal domain of the protein, and an impaired localisation to the nucleus was demonstrated in human cardiomyocytes in cell culture transfected with the mutant gene (Fan et al. 2019). This is the first association of the gene with human disease phenotypes (Fan et al. 2019), although two mutations in the gene have been reported since: a truncating mutation in

LVNC (Liu et al. 2019), and an unspecified homozygous mutation in severe paediatric DCM (Coyan et al. 2019).

1.4.2.3 Exome sequencing in left ventricular noncompaction

Through exome sequencing, the genes *HCN4*, *NNT* and *PTEN* have been associated with LVNC (Table 1.3) (Bainbridge et al. 2015, Milano et al. 2014, Tang et al. 2018).

In a large family with AD LVNC and sinus bradycardia, the only variant that segregated with disease was a mutation in *HCN4* (Milano et al. 2014). *HCN4* encodes a pacemaking ion channel protein which is expressed in myocardial cells (primarily in the sino-atrial node), where it is suggested to play a role in heart rate modulation (D'Souza et al. 2017). Although mutations in the gene had been linked to bradycardia and other arrhythmias before (Baruscotti et al. 2010), this was the first time the gene had been associated with structural heart defects. Subsequently, mutations in *HCN4* have been described in seven additional European families with similar clinical phenotypes (sinus bradycardia and LVNC) (Bainbridge et al. 2015, Millat et al. 2015, Schweizer et al. 2014). Knockout of *HCN4* in mice led to the development of sinus arrhythmia, while overexpression of the gene attenuated bradycardia (Kozasa et al. 2018).

NNT was identified as a potential novel gene for familial and sporadic LVNC by exome sequencing (Bainbridge et al. 2015). Silencing *NNT* orthologues in zebrafish caused cardiac oedema, bradycardia and significant contractile dysfunction, supporting the role of the gene in LVNC pathogenesis (Bainbridge et al. 2015). However, the gene has not been identified as a cause of cardiomyopathy in other studies as yet. Exome sequencing was also used to reveal the digenic inheritance of *GATA4* and *PTEN* mutations in a single LVNC patient (Tang et al. 2018). While *GATA4* mutations have been associated with DCM phenotypes (Li et al. 2013), this was the first association of *PTEN* with cardiomyopathy in humans. *GATA4* and *PTEN* have both been implicated in signalling pathways in cardiomyocyte differentiation (Ang et al. 2016, Aoyagi and Matsui 2011), and Tang et al. (2018) suggest that dysregulation of these pathways may underlie the development of LVNC in this patient.

Table 1.3: New LVNC disease genes identified through NGS

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
HCN4	c.1241C>G (p.A414G)	0	Heterozygous	LVNC with sinus brady	Not specified; non-consanguineous family	*Candidate gene screen*	Milano et al. (2014)
	c.1441T>C (p.Y481H)	0	Heterozygous	LVNC with sinus brady	Not specified; non-consanguineous family	*Candidate gene screen*	Milano et al. (2014)
	c.1444G>C (p.G482R)	0	Heterozygous	LVNC with sinus brady	Consanguineous Dutch family	**Exome sequencing**	Milano et al. (2014)
	c.1444G>C (p.G482R)	0	Heterozygous	LVNC with SND	German family	*Targeted sequencing*	Schweizer et al. (2014)
	c.1444G>C (p.G482R)	0	Heterozygous	LVNC with sinus brady	French family	*Targeted sequencing*	Millat et al. (2015)
	Not specified (p.695*)	0	Heterozygous	LVNC with sinus brady	German family	*Candidate gene screen*	Schweizer et al. (2014)
	c.2648C>G (p.P883R)	0.007727	Heterozygous	LVNC with sinus brady	Not specified; single proband	*Candidate gene screen*	Schweizer et al. (2014)
NNT	c.638_639insT (p.T216Yfs*14)	0.000013	Not specified	LVNC	Not specified; non-consanguineous family	**Exome sequencing**	Bainbridge et al. (2015)
	c.829G>T (p.D277Y)	0	Not specified	LVNC	Not specified; single proband	*Candidate gene screen*	Bainbridge et al. (2015)
PTEN	c.517C>T (p.R173C)	0	Digenic (inherited with *GATA4* mutation)	LVNC	Not specified; non-consanguineous family	**Exome sequencing**	Tang et al. (2018)
TRPM4	c.858G>A (p.T286T)	0.000008	Heterozygous	LVNC with RBBB	Not specified; non-consanguineous family	**Targeted sequencing**	Saito et al. (2018)

***** Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

brady, bradycardia; gnomAD, Genome Aggregation Database; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RBBB, right bundle branch block; SND, sinus node dysfunction

1.4.2.4 Exome sequencing in arrhythmogenic cardiomyopathy

Two new ACM genes, *CDH2* and *TJP1*, have been identified through exome sequencing to date (Table 1.4).

CDH2 was first associated with ACM when Mayosi et al. (2017) reported a missense mutation in the gene in a multi-generation South African family with ACM, as well as an additional *de novo* mutation in a sporadic ACM patient. The latter mutation was found to segregate with disease in a European family (Turkowski et al. 2017). Similarly, four *TJP1* missense mutations were identified using exome sequencing and candidate gene screens in four European ACM families and patients (De Bortoli et al. 2018). *CDH2* and *TJP1* both encode protein components of gap junctions and area composita, and are known to interact with each other as well as cardiac actin and other known ACM genes (Roberts 2018, Vermij, Abriel, and van Veen 2017), suggesting that the pathogenesis of adhesion-related cardiomyopathy may extend beyond the cardiac desmosome.

Table 1.4: New ACM disease genes identified through NGS

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
CDH2	c.686A>C (p.N229P)	0	Heterozygous	AD ACM	South African family	**Exome sequencing**	Mayosi et al. (2017)
	c.1219G>A (p.D407N)	0	Heterozygous	Sporadic ACM	South African patient	*Candidate gene screen*	Mayosi et al. (2017)
	c.1219G>A (p.D407N)	0	Heterozygous	AD ACM	European family	*Exome sequencing*	Turkowski et al. (2017)
TJP1	c.793C>T (p.R265W)	0.000011	Heterozygous	ACM	Italian patient	*Targeted sequencing*	De Bortoli et al. (2018)
	c. 986C>T (p.S329L)	0.000059	Heterozygous	ACM	Dutch/German patient	*Targeted sequencing*	De Bortoli et al. (2018)
	c.1079A>T (p.D360V)	0	Heterozygous	ACM	Dutch/German patient	*Targeted sequencing*	De Bortoli et al. (2018)
	c.2006A>G (p.Y669C)	0	Heterozygous	ACM	Not specified; non-consanguineous family	**Exome sequencing**	De Bortoli et al. (2018)
TP63	c.796C>T (p. R266*)	0	Heterozygous	ACM	Not specified; non-consanguineous family	**Targeted sequencing**	Poloni et al. (2019)

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; gnomAD, Genome Aggregation Database; MAF, minor allele frequency

1.4.2.5 Exome sequencing in restrictive cardiomyopathy

Through exome sequencing, two novel genes for RCM have been identified, *KIF20A* and *TMEM87B*.

TMEM87B was identified as a potential RCM gene by Yu et al. (2016) when they described a paediatric patient presenting with several symptoms suggestive of syndromic disease in addition to RCM. This patient was found to have a deletion at chromosome 2q13, and many of the phenotypic features observed in the patient were consistent with 2q13 microdeletion syndrome. However, the cardiovascular phenotype was more severe than expected, including atrial septal defect, biatrial enlargement and RCM. Exome sequencing revealed a potentially deleterious mutation, c.1366A>G (p.N456D), in *TMEM87B* (Yu et al. 2016). The missense variant p.N456D was hemizygous in the patient, as *TMEM87B* maps to the microdeletion region of chromosome 2. Although little is known about the function of *TMEM87B*, and mutations in the gene have not been identified in cardiomyopathy since, prior experiments silencing expression of the gene in zebrafish led to the development of heart defects (Russell et al. 2014).

Compound heterozygous *KIF20A* variants were reported in two siblings with a severe, congenital, right ventricular form of RCM (Louw et al. 2018). *KIF20A* encodes a member of the kinesin family of mitotic proteins, and patient fibroblasts demonstrated reduced microtubule motility during cell division. Although it has not been implicated in cardiomyopathy since, *KIF20A* was demonstrated by the authors to play a role in zebrafish heart development and function, with silencing expression of the gene leading to thicker ventricular walls and signs of heart failure (Louw et al. 2018).

1.4.3 Whole genome sequencing in cardiomyopathy

Whole genome sequencing is a technique which has great potential for unbiased screening of all genes for their role in cardiomyopathy, as it involves the capture and sequencing of the entire genome. This allows for the generation of vast amounts of genetic information, including variation in intronic and intergenic regions. This means that promotor region variation or variation in other regulatory regions can be assessed; these areas are usually omitted when using exome or targeted sequencing approaches. However, the challenges involved in the interpretation of non-coding variants remain considerable; an even higher proportion of variants of uncertain

significance (VUSs) in noncoding regions may confidently be expected than have been encountered in panel testing and exome sequencing thus far. Because of the elimination of a capture step in whole genome sequencing, coverage of exons tends to be more uniform; in other disease settings, genome sequencing has proven more sensitive than exome sequencing to detect rare germline exonic SNVs for this reason (Meynert et al. 2014, Turner et al. 2016). It is important to note that neither exome nor genome sequencing entirely cover their genomic targets; the extent and reasons for reduced coverage have been explored in the literature (Sanghvi et al. 2018).

A comprehensive search of the PubMed database for all eligible studies encompassing certain criteria (heritable or sporadic non-syndromic forms of cardiomyopathy were investigated, NGS was used to study the patients, and a clear genetic cause of disease was reported) produced very few whole genome sequencing cardiomyopathy studies, published between 2011 and March 2021, presumably due to the current challenges with data storage, data analysis and sequence coverage. Although many variants may occur in introns and between genes, analysing their potential pathogenic effects remains difficult. A singe novel cardiomyopathy gene, *MYBPHL*, was identified through genome sequencing when a coding, truncating variant in the gene was discovered in a family with AD DCM and conduction abnormalities (Barefield et al. 2017). Although the function of the gene is unknown, and mutations in *MYBPHL* have not been reported in other cardiomyopathy cases to date, the authors demonstrated its high expression in human and mouse atria and interactions with the cardiac myofilament (Barefield et al. 2017). Deletion of the gene in mice led to a DCM phenotype and susceptibility to arrhythmias. In other genome sequencing studies, only mutations in known cardiomyopathy genes were reported (Guo et al. 2017, Esslinger et al. 2017).

1.4.4 Targeted sequencing in cardiomyopathy

The application of NGS to the sequencing of a finite number of genes is known as targeted sequencing. Gene panels typically include the major disease-associated genes and, although predesigned panels are commercially available, custom panels are often used. In some countries such as the UK, a process of national standardisation of genetic testing is underway to eliminate variations within healthcare systems regarding how genetic testing is offered to patients, and how it may form part of patient management. Targeted panel sequencing is now used for diagnostic purposes in these countries, as established cardiomyopathy genes are screened for reported and

new disease-causing mutations. Diagnostic genetic panel testing for ICCs is not currently available in South Africa, or any other African country to our knowledge.

The advantage of NGS panels is that high sequence coverage can be achieved. Panels are also easily customisable and can vary considerably, ranging from few genes to hundreds of genes (Glotov et al. 2015, Waldmüller et al. 2015). A single ICC panel involving many genes can be run straightforwardly on a DNA sample, and only the genes consistent with the patient's clinical presentation interrogated, facilitating standardisation. However, targeted sequencing is often limited to the coding regions and exon/intron boundaries, so intronic and promoter region variants may not be detected. Another disadvantage is that, by focussing on a gene panel, targeted sequencing is largely unsuitable for novel gene discovery unless candidate genes are incorporated into the design of the panel. Panel size is also a matter of some contention, and there is no clear consensus on whether larger panels incorporating disease genes from various heritable cardiomyopathies are more useful than cardiomyopathy-specific gene panels. Several reports have indicated that increasing the number of genes on a panel does not correlate with an increase in genetic diagnoses, but merely increases the number of VUSs and unclear genetic findings (Alfares et al. 2015, Burns et al. 2017, Ouellette et al. 2018, Pugh et al. 2014, Thomson et al. 2019, Walsh et al. 2017).

Although NGS panels are usually limited to known cardiomyopathy genes, targeted sequencing has been used to identify novel genotype-phenotype relationships, illustrating the genetic overlap which is characteristic of the cardiomyopathies. Gene panels have been used to describe novel associations of ACM, HCM and channelopathy gene mutations in DCM (Haas et al. 2015, Mahdieh et al. 2018), ACM gene mutations in LVNC and DCM (Klauke et al. 2017, Ramond et al. 2017), HCM gene mutations in LVNC and RCM (Brodehl et al. 2017, Schaefer et al. 2014, Wu et al. 2015) and channelopathy or muscular dystrophy genes (*AKAP9*, *DMD* and *OBSCN*) with cardiomyopathy (Forleo et al. 2017). In these studies, the variants were rare or absent in the ExAC and/or gnomAD populations, and many of them met established criteria for pathogenicity (Richards et al. 2015), supporting a causative role of these genes in other forms of cardiomyopathy.

Although exome or genome sequencing may be more suitable for research and discovery, NGS panels can be designed to include candidate genes and this way, suspected cardiomyopathy genes can be sequenced. This approach was used by Waldmüller et al. (2015) to identify a

truncating mutation in *ADRB2* as a candidate cause of DCM (Table 1.1). Although biologically a plausible candidate gene, in the absence of further evidence or additional family members, this variant was classed as a VUS. Similarly, a truncating mutation in *TP63* was reported in an ACM patient (Poloni et al. 2019) (Table 1.4) and a *TRPM4* mutation was identified in a case of LVNC and cardiac conduction disease, although it is probable that other genetic factors may be contributing to the LVNC phenotype in this case (Saito et al. 2018) (Table 1.3).

In a Web-based discrete choice experiment, Buchanan et al. (2019) sought to identify the key attributes of panel testing, exome and genome sequencing that drove test selection in UK practitioners who order genetic tests for ICCs. Their results indicated that as of early 2019, practitioners had a strong preference for panel tests over exome or genome sequencing. The key drivers of choice, perhaps unsurprisingly, were identification of more causative variants, a yield of fewer VUSs, and lower cost. Respondents were willing to pay an additional 117 GBP for each 1% of additional diagnostic yield from a genetic test; we interpret this figure as high and somewhat at odds with the strong preference expressed for panel testing (Buchanan et al. 2019).

1.4.5 Challenges and opportunities in genomic research of cardiomyopathy

NGS has enabled global progress in the field of heritable cardiomyopathy chiefly through the availability of panel testing that can screen known cardiomyopathy genes with a high degree of accuracy. The data presented here also indicates that NGS has proved of some value through the elucidation of new disease genes, though this has primarily been in paediatric or syndrome-associated cardiomyopathy. Moreover, NGS has not yet found traction in LMICs such as those in Africa, where there are large gaps in knowledge of the genetics of cardiomyopathy. Aside from two genes identified in Chinese patients, and a single gene reported in South African patients, gene discovery has been largely limited to high-income countries (Figure 1.4).

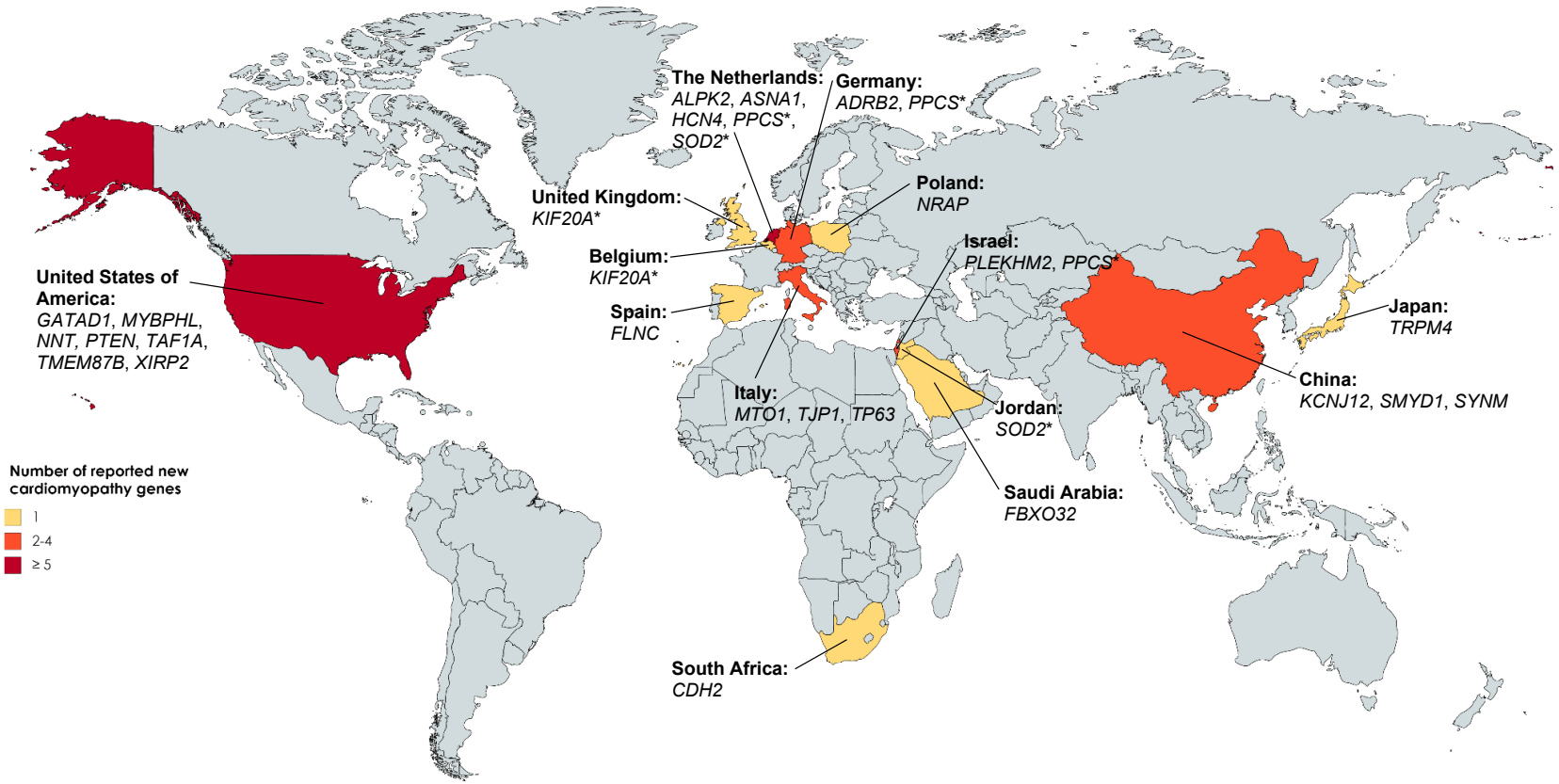


Figure 1.4: Global distribution of new disease genes identified by next-generation sequencing. Countries of origin were identified by the first author's affiliation. Where more than one affiliation was present or the first authorship was joint (gene symbols marked by an asterisk), all countries were included. Countries shaded in light orange indicate one cardiomyopathy gene was identified by a primary author from that country, countries shaded in dark orange indicate two to four such authors were identified. Countries shaded in red indicate more than five new cardiomyopathy genes were identified by a primary author from that country. Created with MapChart.net ©

The investigation of variants at the genome- or exome-level should become mainstream when NGS is used for non-diagnostic purposes (research and discovery). As the cost of sequencing decreases and NGS becomes more accessible, it will be applied more routinely in the study and genetic diagnosis of cardiomyopathy. Although progress has been made, there are still challenges in NGS data analysis that must be overcome. The continued use of exome and genome sequencing in research will require an improvement in our ability to analyse genes outside of the known cardiomyopathy genes, and better analysis of noncoding and intronic variation for their contribution to heritable disease. It has been suggested, for example, that deep-intronic variation in known cardiomyopathy genes may contribute to the development of HCM by affecting transcription and splicing factor binding sites (Mendes de Almeida et al. 2017). An intronic variant in *MYBPC3* has recently been shown to account for as much as 1% of HCM (Harper et al. 2020). Furthermore, even when the disease-causing variant is identified in a family, substantial phenotypic variability is often observed and unexplained (Hershberger and Siegfried 2011). This may be due to genetic modifiers, but understanding this variability may lie within transcriptomics and epigenomics (Alimadadi et al. 2020, Haywood et al. 2020, Ito et al. 2017), other NGS technologies that are rarely used in the study of cardiomyopathy.

Currently, exome and targeted sequencing appear to be the methods of choice when investigating the heritable cardiomyopathies, possibly due to the costs of whole genome sequencing combined with challenges in data analysis. Targeted sequencing has limited scope for novel gene discovery, and although several new genotype-phenotype correlations have been described, the applicability of panel sequencing is most beneficial in a diagnostic setting. The identification of known disease-causing mutations through cardiomyopathy gene panels requires little interpretation; caution is advisable, however, when analysing VUSs that arise from panel sequencing. The majority of reported novel cardiomyopathy genes were identified through exome sequencing. As a great proportion (> 85%) of pathogenic mutations occurs in the coding regions of the genome (Choi et al. 2009), the suitability of exome sequencing for novel gene discovery is as expected.

There is little consensus as to what qualifies a genetic variant as disease-causing, particularly in the cardiomyopathies, and it is likely that there are notable interlaboratory differences in variant interpretation, especially regarding VUSs. However, the use of sequencing data in medical genetics is contingent on the ability to distinguish pathogenic from benign variants. ACMG guidelines place the greatest pathogenic potential, regardless of disease type, upon truncating nonsense, frameshift and splice region mutations, particularly when these variants occur in genes

where loss-of-function mutations are known to cause disease (Richards et al. 2015). When these mutations occur early in the protein-coding sequence, or disrupt functionally important regions, the disease-causing potential is evident. Many cardiomyopathy-causing mutations are missense, and the challenge often arises in the distinction between low-impact VUSs and pathogenic missense variants. There are numerous *in silico* prediction algorithms which can be used to quantify the potential impact of a variant based on factors such as the protein sequence and the evolutionary conservation of the affected residue (Adzhubei et al. 2010, Kumar, Henikoff, and Ng 2009, Schwarz et al. 2014). However, these tools may over-call variants as pathogenic and are not reliable when assessing variants of milder impact (Choi et al. 2012, Richards et al. 2015). Indeed, there are variants that would be classified as benign using bioinformatic prediction tools that have been shown, using *in vitro* functional work, to have a pathogenic effect (Hedley et al. 2013). Moreover, these computational tools were not designed to predict pathogenicity in a diagnostic setting, and some may be biased due to the training and validation datasets used (Grimm et al. 2015). The outputs of these prediction tools should therefore be interpreted cautiously and are not definitive proof of pathogenicity. These difficulties in data analysis are only magnified when investigating variants at the genome- or exome-level. Due to the vast amount of data these techniques generate, exome and genome sequencing may produce many variants which are clinically ambiguous based on existing knowledge. There are currently no guidelines for reporting of VUSs, and some laboratories may choose to report VUSs in candidate genes to their patients (Vears, Sénécal, and Borry 2017). However, it is advisable to retain data about VUSs until knowledge improves, with the aim to definitively classify these variants as pathogenic or benign (Matthijs et al. 2016).

A total of 26 potential new cardiomyopathy genes were reported in the literature between 2011 and 2019, 23 of which were from exome sequencing experiments. Although most of the genes were identified in DCM patients, arguably the greatest impact has come from the genes reported in ACM, LVNC and HCM. *FLNC*, *HCN4* and *MTO1* have been demonstrated as mutation hotspots for numerous forms of heritable cardiomyopathy (Baruffini et al. 2013, Janin et al. 2017, Millat et al. 2015, Nozari et al. 2018, O'Byrne et al. 2018, Ortiz-Genga et al. 2016, Schweizer et al. 2014, Taylor et al. 2014, Tucker et al. 2017). *FLNC* and *HCN4* now often appear on cardiomyopathy panels for diagnostic screening (Dewar et al. 2017, Oliveira et al. 2015), demonstrating the potential power of exome studies and the ultimate goal in cardiomyopathy research. The identification of *CDH2* and *TJP1* as novel ACM genes suggests the pathogenic site of adhesion-

related ACMs to extend beyond the cardiac desmosome, encompassing the entirety of the intercalated disc (De Bortoli et al. 2018, Mayosi et al. 2017, Turkowski et al. 2017).

In contrast, 15 novel DCM genes were identified in family studies but only mutations in *ALPK3* and *PPCS* have been reported in additional cardiomyopathy families. The variants in *ADRB2*, *ASNA1*, *FBXO32*, *GATAD1*, *KCNJ12*, *MYBPHL*, *NRAP*, *PLEKHM2*, *SOD2*, *SYNM*, *TAF1A* and *XIRP2* may therefore qualify as VUSs even though most of them are considered rare enough to cause disease (Table 1.1). The absence of additional families does not necessarily contradict a variant's disease-causing potential, and clarity may be provided through appropriate functional assays. For example, significant impairments in autophagy were reported in patient tissues carrying *FBXO32* and *PLEKHM2* mutations, suggesting a mechanism whereby these mutations may cause disease (Al-Yacoub et al. 2016, Muhammad et al. 2015). Knockouts of *taf1a* and *asna1* in zebrafish were able to recapitulate the DCM phenotype, causing heart failure and death within 6 to 11 days (Long et al. 2017, Verhagen et al. 2019), whereas similar experiments with *gatad1* in zebrafish were inconclusive with one out of six mutant fish exhibiting an enlarged heart (Yang et al. 2016).

Identifying the genetic cause of disease using NGS is challenging due to the numerous variables that are involved and inherent limitations in the techniques. It is unclear how many NGS studies have failed to characterise disease-causing mutations in human patients to date. Nevertheless, the validation of findings is an important aspect of NGS study design. As demonstrated above, approaches to this include the identification of other patients with similar phenotypes who also carry mutations in the putative cardiomyopathy gene, or the investigation of these genes using *in vitro* or *in vivo* models to determine their role in pathogenesis.

1.5 Modelling cardiomyopathy in cell culture

One approach to the functional investigation of candidate genes is to characterise the effects of gene loss or mutation in human cells in culture. Induced pluripotent stem cells (iPSCs) are stem cells that can be derived from adult fibroblasts through expression of just four transcription factors (Takahashi and Yamanaka 2006). They may therefore be used to model human disease by differentiating the stem cells to form the tissue of interest.

When investigating heritable cardiomyopathy, it is noteworthy that iPSCs can be differentiated into cardiomyocytes (Lian et al. 2013). One strategy is to compare iPSC lines derived from control individuals and patients carrying known disease-causing mutations; however, inter-individual genetic variation may influence phenotypic expression or otherwise reduce confidence that observed phenotypic differences are due to the mutation alone (Reinhardt et al. 2013, Soldner and Jaenisch 2012). An alternative approach which overcomes these limitations is the generation of isogenic iPSC lines which are propagated and subsequently genetically engineered to carry the mutations of interest, for example using the clustered regularly-interspaced short palindromic repeats (CRISPR)/Cas9 system (Soldner and Jaenisch 2012). This technique will be useful in determining whether a candidate gene is necessary for normal cardiomyocyte function and will indicate the potential of mutations within the gene to cause cardiomyocyte dysfunction. However, as the heart is a complex organ comprised of multiple cell types including cardiomyocytes, the effect of specific mutations on the structure and functioning of the entire heart cannot be elucidated through cell culture alone. In order to investigate the organ-level effects of genetic mutations, *in vivo* animal models may be preferable.

1.6 Animal models of cardiomyopathy

Disease models are an important tool in research, allowing novel insights into the mechanisms of human disease. In the cardiomyopathies, animal models may be used to investigate the role of novel genes in heart development or function and consider whether proposed pathogenic mutations are likely to cause disease in humans. Although mouse models are widely used to study various human conditions, they are of limited use in cardiomyopathy research due to significant physiological and genetic differences between human and murine hearts. For instance, the mouse heart beats at a rate approximately ten times faster than that of human hearts, and has a higher ventricular expression of the *Myh6* gene, whereas in humans the slower *MYH7* isoform is predominantly expressed in the ventricles (Duncker et al. 2015). Furthermore, homozygous mutations of known cardiomyopathy genes are often embryonic lethal in mice, while mice heterozygous for these mutations are phenotypically normal (Gramlich et al. 2009), although heterozygosity is the typical disease state in human cardiomyopathy. The mouse heart may therefore not be considered sufficiently representative of the human myocardium in the modelling of cardiomyopathies (Duncker et al. 2015). Given these limitations of the mouse model, alternative model organisms such as the zebrafish (*Danio rerio*) have been described.

The zebrafish has been used as a successful model of numerous human disorders, varying from tuberculosis and neurological conditions (Tobin et al. 2010, Zdebik et al. 2013), to CVD. Although anatomically very different from humans, zebrafish have proved successful models of human disease because of the similarities between the species in terms of the genes expressed during development, and the regulatory networks which drive cell fate. The zebrafish genome project, which was initiated in 2001, demonstrated that approximately 70% of human genes have a zebrafish orthologue, and 82% of known disease-causing genes (according to the Online Mendelian Inheritance in Man database) have at least one zebrafish orthologue (Howe et al. 2013).

Characteristics of zebrafish which make it a favourable organism in the modelling of disease include their high prolificacy, rapid *ex utero* development and the fact that they can be maintained in large numbers at relatively low cost (Lieschke and Currie 2007). Zebrafish larvae are transparent for up to three weeks post fertilisation, which makes the high-resolution visualisation of inner organs possible, for example through the use of fluorescent markers or proteins (Pan et al. 2013). When modelling CVD, zebrafish may be preferred because the circulatory system is not required for oxygenation for up to seven days after fertilisation – during this time, their larvae are able to oxygenate through diffusion alone (Sehnert et al. 2002). Therefore, genetic defects which cause severe cardiac dysfunction can be analysed *in vivo*. While superficially different from mammalian hearts, the zebrafish heart displays various similarities at the developmental and cellular levels, and has been used to model congenital heart defects, arrhythmias and cardiomyopathies (Asnani and Peterson 2014).

Zebrafish are particularly amenable to genetic manipulation, and have been used to identify novel genes involved in cardiac development by forward genetic screens (Bendig et al. 2006, Knöll et al. 2007). In relation to cardiomyopathies, a potential use of the zebrafish model is to investigate the role of genetic mutations identified in familial cases of the disease, as described for *asna1*, *taf1a* and *gatad1* above (Section 1.4.5) (Long et al. 2017, Verhagen et al. 2019, Yang et al. 2016). Mechanisms of genetic manipulation which can be employed in the zebrafish include gene silencing through antisense morpholino oligonucleotide injection, gene overexpression by the injection of mRNAs, and genome editing using transcription activator-like effector nucleases (TALENs) or CRISPR/Cas9 (Bedell and Ekker 2015, Bill et al. 2009, Chang et al. 2013). With these techniques, the function of different genes in zebrafish cardiac development can be analysed. Zebrafish models of cardiomyopathy have been successfully used to validate the

pathogenicity of variants by demonstrating their ability to impair cardiac function (Bainbridge et al. 2015, Dhandapany et al. 2014, Norton et al. 2011, Shehata et al. 2015). It is this potential of the zebrafish model to recapitulate human disease phenotypes, within a short timeframe, that makes it an attractive *in vivo* system, particularly where possible pathogenic genetic variants are identified but which have yet to be functionally validated. Given the advantages of modelling CVD in zebrafish, pursuing such a model may be beneficial to understanding the genetics of cardiomyopathy in Africa, and would assist in the characterisation of novel disease-causing genes.

1.7 Rationale

Given that approximately 50% of all cardiomyopathy patients lack a genetic determinant of disease, the genes which have been identified to date cannot account for the total genetic contribution to heritable cardiomyopathy. The yield of genetic testing in idiopathic DCM is considerably lower than other forms of cardiomyopathy, although the role of genetics is likely underestimated in these cases due to the influences of low mutation penetrance, *de novo* mutations, small family sizes or mutations in presently uncharacterised disease genes (Mestroni et al. 2014, McNally and Mestroni 2017). It is therefore expected that additional genetic causes of cardiomyopathy exist, especially in DCM, a disease characterised by much genetic heterogeneity and a high proportion of idiopathic cases.

As discussed previously, exome sequencing is a technique that may allow the detection of these causes of cardiomyopathy, by focussing genetic analysis on variants that are likely to have a functional effect. Applying this technique in a family setting should reduce genetic ambiguity and increase the chances of mutation detection. The genetics of cardiomyopathy in Africa is largely unknown, and it is probable that further analysis of this genetically unique population may identify more genes of interest (Shaboodien et al. 2020). Research previously conducted in a South African family identified a novel ACM-causing gene using exome sequencing (Mayosi et al. 2017), and it is anticipated that similar analyses in the local population may help close the gap in our knowledge of cardiomyopathy genetics.

1.8 Aim and objectives

The aim of this study is to identify the disease-causing variants in South African cases of familial cardiomyopathy. This will be addressed through the following objectives:

1. Identify possible disease-causing variants in families affected by cardiomyopathy using exome sequencing.
2. Investigate candidate genes for their role in heart development or function *in vivo* using CRISPR/Cas9 knockout in zebrafish.
3. Investigate specific mutations for their potential to cause cardiomyopathy phenotypes *in vivo* using mRNA overexpression in zebrafish.

Chapter 2 Methods

2.1 Study design

The African Cardiomyopathy and Myocarditis Registry Programme (IMHOTEP) is a broad investigation into the aetiology and genetics of cardiomyopathy in Africa, for which ethical approval was obtained from the Human Research Ethics Committee (HREC), UCT (HREC reference no. 766/2014). The first primary objective of IMHOTEP is to define the clinical, electrocardiographic, autonomic, imaging, histological and genetic characteristics of cardiomyopathy and myocarditis in children and adults in Africa (Kraus 2019). HREC approval was also obtained for this sub-study of IMHOTEP (HREC reference no. 111/2019) (Appendix B). In this study, patients from IMHOTEP with clear familial forms of cardiomyopathy, and their relatives, were investigated using exome sequencing in an attempt to identify the genetic causes of their disease, in accordance with the first objective of IMHOTEP.

2.1.1 Study participants

In total, five probands and 34 relatives who were recruited between 1997 and 2017 were included in this investigation. Written informed consent was obtained from each participant, as well as a blood sample for molecular analysis (Appendix C). Assent was obtained when recruiting children over the age of 8 years, in addition to consent from their parents or legal guardians, which was obtained from minors of all age. Eligible patients were identified during routine screening at the cardiac clinics at Groote Schuur Hospital (GSH) and Red Cross War Memorial Children's Hospital (RCH), Cape Town, South Africa. Eligibility was determined by a primary diagnosis of cardiomyopathy (Section 2.1.2) and a positive family history for the disease (at least one additional family member with a similar diagnosis). Index patients were excluded if they were diagnosed with secondary causes of cardiomyopathy, systemic arterial hypertension (> 160/100 mmHg), coronary artery disease, pericardial diseases, congenital heart disease, pulmonary disease or valvular heart disease.

2.1.2 Diagnosis and screening

All patients and their relatives were screened for cardiomyopathy at the cardiac clinics at GSH or RCH. Each case was reviewed by a diagnostic panel of experts at GSH, consisting of specialists

in cardiomyopathy, electrophysiology, cardiac imaging and histopathology, to determine each patient's diagnosis as following ESC guidelines (Elliott et al. 2008). Study participants were assessed with chest X-ray, electrocardiography (ECG), echocardiography and basic blood investigations of haemoglobin levels, white blood cell count, as well as renal and liver function (Kraus 2019). When clinically indicated, patients and their family members were also screened using cardiac magnetic resonance imaging (CMR) and/or Holter monitoring.

DCM and HCM were diagnosed according to ESC guidelines, contingent on the absence of hypertension or valve disease that may cause the observed abnormalities (Elliott et al. 2008). The ESC defines ACM histologically as progressive replacement of right ventricular myocardium with adipose and fibrous tissue, and the presence of right ventricular dysfunction with or without left ventricular involvement (Elliott et al. 2008). The diagnosis of ACM was made on the basis of International Task Force criteria (Table 2.1), in which ACM diagnoses are classified using major and minor criteria as definite (two major criteria, one major and two minor criteria, or four minor criteria), borderline (one major and one minor criteria, or three minor criteria) or possible (one major or two minor criteria) (Marcus et al. 2010).

Demographic and clinical information was collected during patient enrolment into the IMHOTEP parent study. Demographic data that was used in this investigation was patient sex, self-reported ethnicity and family history. Detailed family histories were obtained by a genetic counsellor during the recruitment process and were used in this investigation in the determination of potential inheritance patterns of disease. Clinical information was used to correlate identified genetic mutations with disease status. This clinical data included disease status (classified as affected, unaffected, at-risk or unknown), diagnosis, age of disease onset, disease severity and outcomes.

Table 2.1: International Task Force criteria for ACM

Major criteria	Minor criteria
1. Global and/or regional dysfunction and structural alterations	
By 2D echocardiogram Regional RV akinesia, dyskinesia or aneurysm and one of the following (end diastole): • Parasternal long axis view ≥ 32 mm or ≥ 19 mm/m^2 • Parasternal short axis view ≥ 36 mm or ≥ 21 mm/m^2 • Fractional area change ≤ 33% *By MRI* Regional RV akinesia or dyskinesia or dyssynchronous RV contraction and one of the following: • RV end diastolic volume ≥ 110 ml/m^2 male or ≥ 100 ml/m^2 female • RV ejection fraction ≤ 40% *By RV angiography* Regional RV akinesia, dyskinesia or aneurysm	*By 2D echocardiogram* Regional RV akinesia or dyskinesia and one of the following (end diastole): • Parasternal long axis view 29-32 mm or 16-19 mm/m^2 • Parasternal short axis view 32-36 mm or 18-21 mm/m^2 • Fractional area change 33-40% *By MRI* Regional RV akinesia or dyskinesia or dyssynchronous RV contraction and one of the following: • RV end diastolic volume 100-110 ml/m^2 male or 90-100 ml/m^2 female • RV ejection fraction 40-45%
2. Tissue characterisation of wall	
Residual myocytes < 60% by morphometric analysis, (or < 50% if estimated), with fibrous replacement of the RV free wall myocardium in at least one sample, with or without fatty replacement of tissue on endomyocardial biopsy	Residual myocytes 60-75% by morphometric analysis, (or 50-65% if estimated), with fibrous replacement of the RV free wall myocardium in at least one sample, with or without fatty replacement of tissue on endomyocardial biopsy
3. Repolarisation abnormalities	
Inverted T waves in right precordial leads (V_1, V_2 and V_3) or beyond in individuals > 14 years of age (in the absence of complete RBBB QRS ≥ 120 msecs)	Inverted T waves in leads V_1 and V_2 in individuals > 14 years of age (in the absence of complete RBBB), or in V_4, V_5, or V_6. Inverted T waves in leads V_1, V_2, V_3 and V_4 in individuals > 14 years of age in the presence of complete RBBB
4. Depolarisation/conduction abnormalities	
Epsilon wave (reproducible low amplitude signals between end of QRS complex to onset of the T wave) in the right precordial leads (V_1 to V_3)	• Late potentials by signal averaged ECG in at least one of three parameters in the absence of a QRS duration of ≥110 msecs on the standard ECG • Filtered QRS duration: ≥ 114 ms • Duration of terminal QRS < 40 µV (LAS): ≥ 38 ms • Root mean square voltage of terminal 40 ms: ≥ 20 µV • Terminal activation duration of QRS ≥ 55 ms measured from the nadir of the S wave to the end of the QRS, including R', in V_1, V_2 or V_3, in the absence of complete RBBB
5. Arrhythmias	
Non-sustained or sustained VT of left bundle branch morphology with superior axis (negative or indeterminate QRS in II, III, AVF and positive in AVL)	• Non sustained or sustained VT of RV outflow configuration, LBBB morphology with inferior axis (positive QRS in II, III, AVF and negative in AVL) or of unknown axis • Greater than 500 ventricular extrasystoles/24 hours by Holter
6. Family history	
• ACM confirmed in a first-degree relative who meets current task force criteria • ACM confirmed pathologically at autopsy or surgery in a first-degree relative • Identification of a pathogenic mutation categorised as associated or probably associated with ACM in the patient under evaluation	• History of ACM in a first-degree relative in whom it is not possible or practical to determine if the family member meets current task force criteria • Premature sudden death (< 35 years) due to suspected ACM in a first-degree relative • ACM confirmed pathologically or by current Task Force Criteria in second-degree relative

Adapted from Marcus FI. 2010. *Circulation*, **121**(13):15-18. *µV, microvolt; 2D, two-dimensional; ACM, arrhythmogenic cardiomyopathy; AVF, augmented vector foot; AVL, augmented vector left; ECG, electrocardiography; LAS, low amplitude signal duration; LBBB, left bundle branch block; m, metre; ml, millilitre; mm, millimetre; ms, millisecond; MRI, magnetic resonance imaging; RBBB, right bundle branch block; RV, right ventricular; VT, ventricular tachycardia*

2.2 DNA extraction

Blood samples obtained from the participants during recruitment were used to isolate DNA for use in downstream molecular processes.

2.2.1 DNA extraction and storage

Genomic DNA was extracted from all probands and family members using either the Gentra® Puregene® blood kit (Qiagen, Hilden, Germany) or the PAXgene® Blood DNA kit (Qiagen), following the manufacturer's instructions (Appendix D). DNA samples were anonymised and archived at -80°C in the Cardiovascular Genetics laboratory, Hatter Institute for Cardiovascular Research in Africa (HICRA), UCT.

2.2.2 DNA quality control

DNA extracts were quantified using a NanoDrop® 2000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE USA) and diluted to working concentrations of 100 ng/µl. The integrity of the DNA was investigated using agarose gel electrophoresis on 1% (w/v) agarose gels (Appendix E). A 100 bp DNA ladder (New England BioLabs (NEB), Ipswich, MA USA) was included as a marker on each gel. Both the DNA samples and ladder were stained with 3X GelRed® nucleic acid stain (Biotium, Fremont, CA USA). DNA was visualised using a Uvitec Xplorer D55 (Uvitec Cambridge, Cambridge, UK) and the Xplorer 1D software version 15.08 (Uvitec Cambridge).

2.3 Exome sequencing

Exome sequencing was performed on the proband of each family, as well as at least one other clinically screened family member (Table 2.2). The sequencing reactions were performed at the Centre for Proteomic and Genomic Research (CPGR), Cape Town, or Central Analytical Facilities (CAF), Stellenbosch University (Table 2.2). At CPGR, exonic regions were captured using the Ion AmpliSeq™ Exome RDY kit (Thermo Fisher Scientific, Waltham, MA USA) and Ion Xpress™ Barcode Adapters 1-16 kit (Thermo Fisher Scientific), while the IonCode™ Barcode Adapters 1-384 kit (Thermo Fisher Scientific) was used at CAF. At both facilities, templates for sequencing were prepared using the Ion PI™ Hi-Q™ Chef kit (Thermo Fisher Scientific) or the Ion PI™ Chip

kit v3 (Thermo Fisher Scientific). The enriched exonic regions were sequenced using an Ion Proton™ high-throughput sequencer (Thermo Fisher Scientific).

Table 2.2: Samples used for exome sequencing

Family	Individuals	Disease status	IMHOTEP code
Family 1			
	III:4 (proband)	Affected	DCM 389.5ELM
	III:6	Affected	DCM 389.7GAR
Family 2			
	I:2 (proband)	Affected	HCM 49.1PAM
	II:2	Affected	HCM 50.1GAR
Family 3			
	III:8	Affected, deceased	DCM 3.5ANT
	III:9 (proband)	Affected, deceased	DCM 3.1CAR
	IV:1	Affected, deceased	DCM 3.11IRV
Family 4			
	III:2	Affected	ACM 142.2WIL
	IV:2 (proband)	Affected	ACM 142.5ROB
	IV:3	Affected	ACM 142.6JEA
	IV:4	Affected, deceased	ACM 142.4MIC
Family 5			
	I:1	Unaffected, possible carrier	DCM 435.3MUT
	I:2	Affected, deceased	DCM 435.1CLE
	II:3 (proband)	Unaffected, possible carrier	DCM 435.2DAV

ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; IMHOTEP, the African Cardiomyopathy and Myocarditis Registry Programme

2.4 Exome sequencing data analysis

Sequencing reads were aligned to the *H. sapiens* reference genome sequence (UCSC version hg19) using the Ion Reporter™ platform (Thermo Fisher Scientific) and the AmpliSeq™ Exome Hi-Q germline workflow version 5.10 (Thermo Fisher Scientific). Exome variant data was downloaded for annotation and filtering.

2.4.1 Variant annotation

Computations were performed using facilities provided by UCT's ICTS High Performance Computing team (hpc.uct.ac.za). Variant files were annotated using Ensembl Variant Effect Predictor (VEP) release 94 (McLaren et al. 2016), on the UCT High Performance Cluster (Appendix F). The annotation databases that were included in the pipeline were gene symbol, Ensembl transcript and protein identifiers, Human Genome Variation Society (HGVS)

nomenclature, gnomAD frequency data, dbSNP identifiers, as well as SIFT and PolyPhen predictions of pathogenicity.

2.4.2 Filtering and bioinformatic analysis of exonic variants

The VEP output from each family was subjected to standard filters: variants were filtered based on global gnomAD minor allele frequency (MAF) ≤ 0.01, variant consequence (including insertions and deletions, as well as missense, nonsense and splice donor or acceptor variants), allele ratio > 0.3, and Combined Annotation Dependent Depletion (CADD) score > 20. CADD scores were not included in the VEP output and were determined manually (https://cadd.gs.washington.edu/) for all variants that passed the first three filters (Rentzsch et al. 2019). All intronic, intergenic and other non-coding variants were excluded from the analysis. Variants were then filtered according to all potential inheritance patterns determined from the family history.

All variants in a panel of 250 known or candidate cardiomyopathy genes (Appendix G) were explored for possible pathogenic variants. The bioinformatics tools MutationTaster (http://www.mutationtaster.org/), SIFT (https://sift.bii.a-star.edu.sg/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and M-CAP (http://bejerano.stanford.edu/mcap/) were used to determine the predicted pathogenicity of each variant (Adzhubei et al. 2010, Jagadeesh et al. 2016, Schwarz et al. 2014, Sim et al. 2012).

The population databases 1000 Genomes (https://www.ensembl.org/index.html) (The 1000 Genomes Project Consortium 2015), Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS/), ExAC (http://exac.broadinstitute.org/) (Lek et al. 2016) and gnomAD (https://gnomad.broadinstitute.org/) (Karczewski et al. 2019), were used to determine the frequency of each variant in known population groups. Variants with allele frequencies > 0.01 in any database were excluded from further analysis.

The conservation of mutated amino acids was checked in chimpanzee (*P. troglodites*), macaque (*M. mulatta*), mouse (*M. musculus*), cat (*F. catus*), chicken (*G. gallus*), zebrafish (*D. rerio*) and frog (*X. xenopus*) by using the ClustalW multiple sequence alignment tool in BioEdit version 7.2.5 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) to compare the protein sequence of each gene to the *H. sapiens* reference sequence. Conservation was assessed according to amino acid classification as previously described (Dagan, Talmor, and Graur 2002, Zhang 2000).

2.4.3 Gene prioritisation

All variants occurring in the cardiomyopathy gene panel (Appendix G) were prioritised, particularly mutations occurring in genes reported to cause a similar phenotype. However, genes outside of the cardiomyopathy panel were also considered. This process was of particular importance when no mutations were found in any known cardiomyopathy genes. These genes were searched in the PubMed (https://www.ncbi.nlm.nih.gov/pubmed/), Genotype-Tissue Expression (GTex) (https://gtexportal.org/home/) and Mouse Genome Informatics (MGI) databases (http://www.informatics.jax.org/) to assess any links to the heart and its functioning.

PubMed was used to find any prior literature implicating the gene in the development of cardiomyopathy or normal heart function. GTex was used to determine the expression of the gene in the heart; any genes with no recorded expression in the heart (left ventricle) were excluded. MGI is a database of mouse models, and any genes that have been functionally investigated in mice and showed cardiovascular defects were prioritised. However, genes that have no mouse models recorded were not excluded. Genes of interest were identified based on their function, expression in the heart and the results of prior animal studies. The potential impact of the variants in these genes was assessed as for those occurring in the cardiomyopathy panel.

2.4.4 Candidate selection

Variants in both known and putative cardiomyopathy genes were assessed based on the results of the bioinformatics and gene prioritisation analyses (Sections 2.4.2 and 2.4.3). Truncating (nonsense and frameshift) and splice site mutations were prioritised, as were missense variants that were predicted deleterious by at least three of the pathogenicity prediction tools (Table 2.3). Priority was given to variants affecting genes with functional links to the heart and its function (as determined by PubMed and MGI database searches). On this basis, candidate variants were selected for validation and segregation analysis.

Table 2.3: Pathogenicity prediction tools used for the scoring of variants

Prediction software	Scoring system	Interpretation
CADD	Ranks variants using a scaled score that combines conservation, protein-level changes, functional genomic data and transcript information (Rentzsch et al. 2019)	≤ 20: benign > 20: deleterious
M-CAP	Combines previous prediction tools, such as SIFT and PolyPhen-2, with machine learning and sequence conservation data (Jagadeesh et al. 2016)	≤ 0.025: benign > 0.025: deleterious
MutationTaster	Uses a Bayes classifier to produce a binary output ('polymorphism' or 'disease causing') reported with a probability value instead of a score (Schwarz et al. 2014)	'Polymorphism': benign 'Disease causing': deleterious
PolyPhen-2	Scores missense variants based on sequence, structural and evolutionary (phylogenetic) information, as well as the protein function (Adzhubei et al. 2010)	≤ 0.450: benign > 0.450: deleterious
SIFT	Scores missense variants based on sequence conservation and the physical properties of the amino acids (Sim et al. 2012)	> 0.05: benign ≤ 0.05: deleterious

CADD, Combined Annotation Dependent Depletion; M-CAP, Mendelian Clinically Applicable Pathogenicity; PolyPhen, Polymorphism Phenotyping; SIFT, Sorting Intolerant From Tolerant

2.5 Validation and segregation analysis

All candidate variants were assessed using polymerase chain reaction (PCR) followed by high resolution melt (HRM) analysis of the PCR products. The presence of each variant in additional family members was also investigated using PCR-HRM. PCR products were purified and sequenced by Sanger sequencing to validate the variants.

2.5.1 PCR conditions and analysis

Target DNA regions from probands and family members were amplified using singleplex PCR-HRM. Each reaction contained 50 ng DNA, 1X GoTaq® Colourless Flexi buffer (Promega, Madison, WI USA), 3 mM $MgCl_2$ (Promega), 0.4X EvaGreen® dye in water (Biotium), 0.4 µM of each primer (Table 2.4) (Integrated DNA Technologies, Inc., Coralville, IA USA), 0.2 mM dNTPs (NEB) and 0.5 U GoTaq® DNA polymerase (Promega), in a final volume of 25 µl. The reactions were conducted in a Corbett Rotor-Gene 6000 (Qiagen). Standard thermal cycling conditions included an initial denaturation step of 95°C for 10 mins, followed by 50 cycles of denaturation at 95°C for 5 s, annealing at 55°C for 10 s and extension at 72°C for 10 s. A final HRM step was included in which the reaction temperature was increased from 72°C to 95°C in 0.1°C increments.

Table 2.4: Primers for validation and segregation analysis

Gene	Variant (HGVS code)	Direction	Primer sequence (5` - 3`)	Product size (bp)	Protocol deviation
ABCB4	c.523A>G	Sense Antisense	GATTGGGTGCTGGAGTTCTTG AGACATGGCTGCCAGATGATC	314	T_a 60°C
DSC2	c.2642T>A	Sense Antisense	GAATCCATTAGAGGACACACT GAACACACTCATCTCTTCATG	338	None
GLA	c.774_775del	Sense Antisense	ACTCAAGAGAAGGCTACAAGTGC TTGAACAAGGAGGGCTCAAGTT	284	None
HEXB	c.1250C>T	Sense Antisense	TCACATGGCACTAACTCTGAAGA AAGGATTACAGGGAAGCCAGA	251	None
ITGB5	c.146G>A	Sense Antisense	CCTGCACCCTTTAGACTCAGA AGCATTCCTTCCCTTCTTCCT	251	None
	c.1929G>C	Sense Antisense	CAAGAAGAAAGGCAGAGCAGA GGAAGGAGAAAAACCGTGACT	306	None
KCNK10	c.1052A>G	Sense Antisense	TTGTTGCTCTTCTGTCTCTTGGC TTGTTGATGCTCTCCTGGGATG	283	Type-It®
KCNH2	c.2805del	Sense Antisense	CCCATTCTCTGTTTCCCAC AGAAGCTGAAAATGTTGGACAC	421	Q5; T_a 56°C
KCNN3	c.202C>T	Sense Antisense	TGGATGAAGACCCCAAGTGC GAAAGCGGTGGGAGAGGAG	282	None
LAMA3	c.4643A>G	Sense Antisense	AAAATACCATCCTCAACCCCACC TGCCACTTCCTGCCACTTAGG	298	None
MYH7	c.4394C>T	Sense Antisense	AGGTCCTGAGACAGACCCTGG GCTTGAAGGTCTCCAGATGTTCCA	216	T_a 60°C
MYOM1	c.139A>G	Sense Antisense	CACCAGCACTATGATCTCAG ACAAAGCGAGAAGTAACTGG	348	1 mM $MgCl_2$; 60 cycles
NDUFB1	c.257G>C	Sense Antisense	CTATGGGATTTGTCATTGGATGTT TCCAAGGATTCAATAACAAACAAGA	244	None
POLG	c.2492A>G	Sense Antisense	ATGGTCTGCTGAGTGGTTGTAGG GCCCTCAGAGCCCAGTTTCTA	294	None
PXDNL	c.2218C>T	Sense Antisense	TATCTGGATGCACAGCTCGC ATGGGGAAACAAGGAGGGTC	418	Q5; T_a 64°C
SORBS2	c.322T>C	Sense Antisense	GAATGGAAACTCAGGAGGTGC TACTTACCGAGGGATGTGCC	132	None
TJP1	c.1412A>G	Sense Antisense	CTGAGAGTATCTGTGTCCTGCGT CCATTTTCTGACAGCATCTCACC	211	None
TRPV1	c.860C>T	Sense Antisense	GGCAGAGACAGAGGGAGTTT CCCCAGCATCAGAATCTCATTG	295	T_a 60°C; 2 mM $MgCl_2$
TTN	c.3899A>G	Sense Antisense	TCTGAAGCTGTTGAATCAGGATTTG CGCTTGCCATCTTTGTACCAA	244	None
	c.24083G>C	Sense Antisense	CCTTCCCCACATAGATCGGAT TGTATATGCCTGTGTCGGAGG	242	None

bp, base pairs; HGVS, Human Genome Variation Society; T_a, annealing temperature

PCR products were separated on a 2% (w/v) agarose gel in order to determine the specificity of the reaction. Agarose gel electrophoresis was performed and visualised as described previously (Section 2.2.2). In instances where non-specific amplification was observed, the PCR conditions were adjusted by optimising the annealing temperature (T_a) or concentration of $MgCl_2$ in the reaction, or the reactions were performed using the Type-It® HRM PCR kit (Qiagen) or the Q5® High-Fidelity PCR kit (NEB) (Table 2.4). To optimise T_a and $MgCl_2$ concentration, the amplification of target product was determined at T_a ranging from 50°C to 60°C and/or $MgCl_2$ concentrations

ranging from 1 mM to 5 mM. For amplification using the Type-It® HRM PCR kit, 1X HRM PCR Master Mix was added to 0.2 µM of each primer (Table 2.4) and 50 ng DNA, in a total reaction volume of 25 µl. Thermal cycling conditions were as standard (T_a 55°C) and were conducted in a Corbett Rotor-Gene 6000 (Qiagen). When amplifying using the Q5® High-Fidelity PCR kit, 1X Q5 High-Fidelity Master Mix was added to 2.5 µM of each primer (Table 2.4) and 50 ng template DNA, to a total volume of 25 µl. Reaction conditions included initial denaturation at 98°C for 30 s, followed by 30 cycles of denaturation at 98°C for 10 s, annealing for 15 s (at a temperature determined by gradient PCR) and extension at 72°C for 30 s. A final extension step at 72°C was conducted for 2 mins. These reactions were performed on an Eppendorf Mastercycler Pro S (Eppendorf, Hamburg, Germany). Upon completion, they were HRM analysed using a Corbett Rotor-Gene 6000 (Qiagen), in which the reaction temperature was increased from 72°C to 95°C in 0.1°C increments, in the presence of 0.4X EvaGreen® dye (Biotium).

The Corbett Rotor-Gene 6000 Application software version 1.7 was used to visualise the amplification of the target DNA regions and their melt profiles. Melt curves from the patient DNA were compared to those from their family members in order to determine the melt profiles of samples containing and lacking the variants of interest. The presence of the variants was confirmed by purification and Sanger sequencing of the PCR amplicons (Sections 2.5.2 and 2.5.3).

2.5.2 PCR product purification

PCR products were purified using alkaline phosphatase (AP) and exonuclease I (ExoI) digestion. Each reaction contained 1 U FastAP™ (Thermo Fisher Scientific) and 2 U ExoI (Thermo Fisher Scientific), with 5 µl PCR product in a final volume of 20 µl. Purification reactions were conducted using a Labnet Multigene™ Gradient thermal cycler (Labnet International, Inc., Edison, NJ USA) and consisted of a 37°C incubation period of 1 hr followed by a 75°C step for 15 mins.

2.5.3 Sanger sequencing reaction conditions

Sequencing was performed using the BigDye® Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific). Sequencing reactions included 3 µl purified PCR product, 1X sequencing buffer, 4 µl BigDye® Terminator 3.1 Ready Reaction Mix and 0.1 µM primer, to a total volume of 20 µl. The reactions were conducted in a Labnet Multigene™ Gradient thermal cycler (Labnet

International, Inc.), where thermal cycling conditions consisted of initial denaturation at 98°C for 5 mins, followed by 30 cycles of denaturation at 96°C for 30 s, annealing at 50°C for 30 s and extension at 60°C for 4 mins. Sequencing products were visualised by capillary electrophoresis using an ABI PRISM® 3130*xl* Genetic Analyser (Applied Biosystems, Waltham, MA USA) at the DNA Sequencing Unit, CAF, Stellenbosch University.

2.5.4 Analysis of sequencing products

Sequencing electropherograms were visualised and edited using FinchTV software version 1.4.0 (http://www.geospiza.com/Products/finchtv.shtml). This involved the removal of indistinct regions at the ends of the electropherograms and confirming each base call in the remaining sequence. The edited sequences were aligned to a reference sequence downloaded from the Ensembl database (https://www.ensembl.org/index.html) (Table 2.5) using the ClustalW multiple sequence alignment tool in the BioEdit Sequence Alignment Editor version 7.2.5 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The alignment was analysed at the position of interest to determine if the expected variant from the exome sequencing output was present.

Table 2.5: Gene and transcript database accession numbers

Gene	Ensembl gene ID	Ensembl transcript ID	NCBI RefSeq ID
ABCB4	ENSG00000005471	ENST00000265723	NM_000443
DSC2	ENSG00000134755	ENST00000280904	NM_024422
GLA	ENSG00000102393	ENST00000218516	NM_000169
HEXB	ENSG00000049860	ENST00000261416	NM_000521
ITGB5	ENSG00000082781	ENST00000296181	NM_002213
KCNK10	ENSG00000100433	ENST00000319231	NM_138317
KCNH2	ENSG00000055118	ENST00000262186	None
KCNN3	ENSG00000143603	ENST00000271915	NM_002249
LAMA3	ENSG00000053747	ENST00000313654	NM_198129
MYH7	ENSG00000092054	ENST00000355349	NM_000257
MYOM1	ENSG00000101605	ENST00000356443	None
NDUFB1	ENSG00000183648	ENST00000329559	NM_004545
NRP1	ENSG00000099250	ENST00000265371	None
POLG	ENSG00000140521	ENST00000268124	NM_001126131
PXDNL	ENSG00000147485	ENST00000356297	NM_144651
SORBS2	ENSG00000154556	ENST00000355634	None
TJP1	ENSG00000104067	ENST00000346128	NM_003257
TRPV1	ENSG00000196689	ENST00000571088	None
TTN	ENSG00000155657	ENST00000589042	NM_001267550

ID, identifier; NCBI, National Centre for Biotechnology Information; RefSeq, Reference Sequence

2.5.5 Assessment of variant pathogenicity

Variants were classified as 'pathogenic', 'likely pathogenic', 'benign', 'likely benign' or 'of uncertain significance', using the ACMG's Standards and Guidelines for the Interpretation of Sequence Variants (Richards et al. 2015). Following these guidelines, each variant was assessed for pathogenicity using population, computational, functional, segregation, *de novo* and other data.

2.6 Functional modelling of candidate variants in zebrafish embryos

Candidate genes and/or variants were modelled using genetic manipulations in zebrafish. These included CRISPR/Cas9 targeting of the zebrafish orthologues of known and candidate cardiomyopathy genes (*pkp2*, *myh7* and *polg*) and mRNA overexpression of the human *POLG* variant c.2492A>G. This work was performed in collaboration with Novartis Institutes for BioMedical Research (NIBR) Inc., Cambridge, MA USA, and the University of Manchester, UK.

2.6.1 CRISPR/Cas9 knockout of cardiomyopathy genes

In this model, the CRISPR/Cas9 system was used to target and disrupt the candidate ACM gene *POLG* (Chapter 4) as well as the known cardiomyopathy genes *MYH7* and *PKP2*. *PKP2* and *MYH7* are genes which have well-characterised associations with ACM and DCM/HCM phenotypes, respectively, and were used as positive controls.

2.6.1.1 Single guide RNA design and synthesis

Single guide RNAs (sgRNAs) were designed to target three genes in zebrafish: *myh7* (ENSDARG00000079564), *pkp2* (ENSDARG00000023026) and *polg* (ENSDARG00000060951). An in-house UCSC genome browser at NIBR (Cambridge, MA USA) was used to identify potential sgRNAs of 20-23 bp, with a terminal PAM sequence, a 5` GG and no predicted off-target effects (Table 2.6).

Table 2.6: Single guide RNA sequences for CRISPR/Cas9

Gene	Human orthologue	Similarity to human sequence	sgRNA coordinates (strand)	sgRNA sequence (5` - 3`)
myh7	*MYH7*	85.6%	8:16543250-16543273 (-)	GCTTTCACATACTCTTCATCTGG
pkp2	*PKP2*	36.8%	4:17057073-17057096 (-)	GTTGCTGTAGTCGGTTCAGCCGG
polg	*POLG*	67.6%	25:9339993-9340016 (+)	GGCGGATGGAGAGAGCCAAAAGG

sgRNA, single guide RNA

Oligonucleotides containing the selected sgRNA sequence (Table 2.6), flanked by a T7 promoter site (5`-GAAATTAATACGACTCACTATAGG-3`) and a 20 bp universal primer recognition sequence (5`-GTTTTAGAGCTAGAAATAGC-3`) were PCR amplified using the Q5® High-Fidelity PCR kit (NEB) to create double-stranded DNA. Each reaction contained 1X Q5® High-Fidelity master mix, 2 µM universal primer (5` - AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGAT AACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC - 3`) and 2 µM of the sgRNA-specific oligonucleotide (Integrated DNA Technologies, Inc.) to a total volume of 50 µl. Reactions were conducted in an Eppendorf Mastercycler Pro S (Eppendorf). The reaction conditions consisted of a 2 min 95°C hold, followed by 5 cycles of 95°C for 20s, 58°C for 20s and 72°C for 30s. A final step of 72°C for 2 mins was also included.

The resulting double-stranded DNA product was purified using the QIAquick PCR purification kit (Qiagen) following the manufacturer's instructions (Appendix H). The purified PCR products were analysed using a NanoDrop® 8000 (NanoDrop Technologies Inc.) and electrophoresis on a 1.2% E-gel® agarose gel (Thermo Fisher Scientific), to check the quantity and integrity of the DNA.

SgRNAs were generated by *in vitro* T7 transcription using the MEGAshortscript™ T7 Transcription kit (Thermo Fisher Scientific). Each reaction contained 1X T7 Reaction Buffer, 7.5 mM T7 ATP, CTP, GTP and UTP Solutions, 2 µl T7 Enzyme Mix and 1 µg template DNA, to a total volume of 20 µl. The reactions were incubated at 37°C overnight before treatment with 1 µl TURBO DNase I (Thermo Fisher Scientific) and incubation at 37°C for a further 15 mins. SgRNAs were then purified using the MEGAclear™ Transcription Clean-Up kit (Thermo Fisher Scientific) following the standard protocol (Appendix H). The sgRNAs were quantified using a NanoDrop® 8000 spectrophotometer (NanoDrop Technologies Inc.). Aliquots of sgRNA were stored in working solutions of up to 500 ng/µl at -80°C.

2.6.1.2 Zebrafish used for CRISPR/Cas9

The zebrafish line used for CRISPR/Cas9 knockout was previously generated at NIBR (Cambridge, MA USA). The fish were genetically engineered to express the monomeric red fluorescent protein mCherry in cardiomyocytes driven by the *cmlc2* promotor (*cmcl2*:mCherry), resulting in a fluorescent red heart transgenic line. Males and females from this line were housed and maintained at NIBR on a 14:10 hr light:dark cycle at 28°C.

2.6.1.3 Zebrafish breeding and embryo collection

The *cmcl2*:mCherry zebrafish were bred in paired matings to generate fertilised embryos. The afternoon before embryos were required, males and females were separated and placed in breeding tanks filled with system water. A single male and single female were transferred to each breeding tank, separated by a divider overnight. The following morning, the dividers were removed and the fish allowed to breed. Within 10-30 mins, embryos were collected and washed with E3 water containing methylene blue (0.002 g/l) (Appendix E).

2.6.1.4 Injection of CRISPR/Cas9

Injections were performed using a FemtoJet® 5247 (Eppendorf). Approximately 1 ng Cas9 protein (PNA Bio Inc., Thousand Oaks, CA USA) and 0.5 ng sgRNA were injected into zebrafish embryos at the single-cell stage of development (n = 125 per experimental group). Uninjected control embryos (n = 125) from the same clutch were allowed to develop normally. All embryos were incubated at 28°C in E3 water with methylene blue (Appendix E) for 3-4 days.

2.6.1.5 Imaging and phenotypic analysis

At three days post fertilisation (dpf), injected and control zebrafish larvae (n = 5 per group) were anaesthetised using tricaine (Appendix E; dose 0.0167%) and visualised using two-photon microscopy. A custom-built microscope at NIBR was used (Figure 2.1). Images were generated at approximately 60 frames per second at a wavelength of 720 nm, using a 16x objective and 2x optical zoom. The hearts were viewed in 10 planes at a total depth of 140 µm. Video files were analysed using DanioScope (Noldus, Wageningen, The Netherlands) to capture still images of each heart during ventricular diastole. Images of the ventricles were taken from when they were most relaxed. These images were then used to measure the ventricle area, length and thickness in each fish larva.

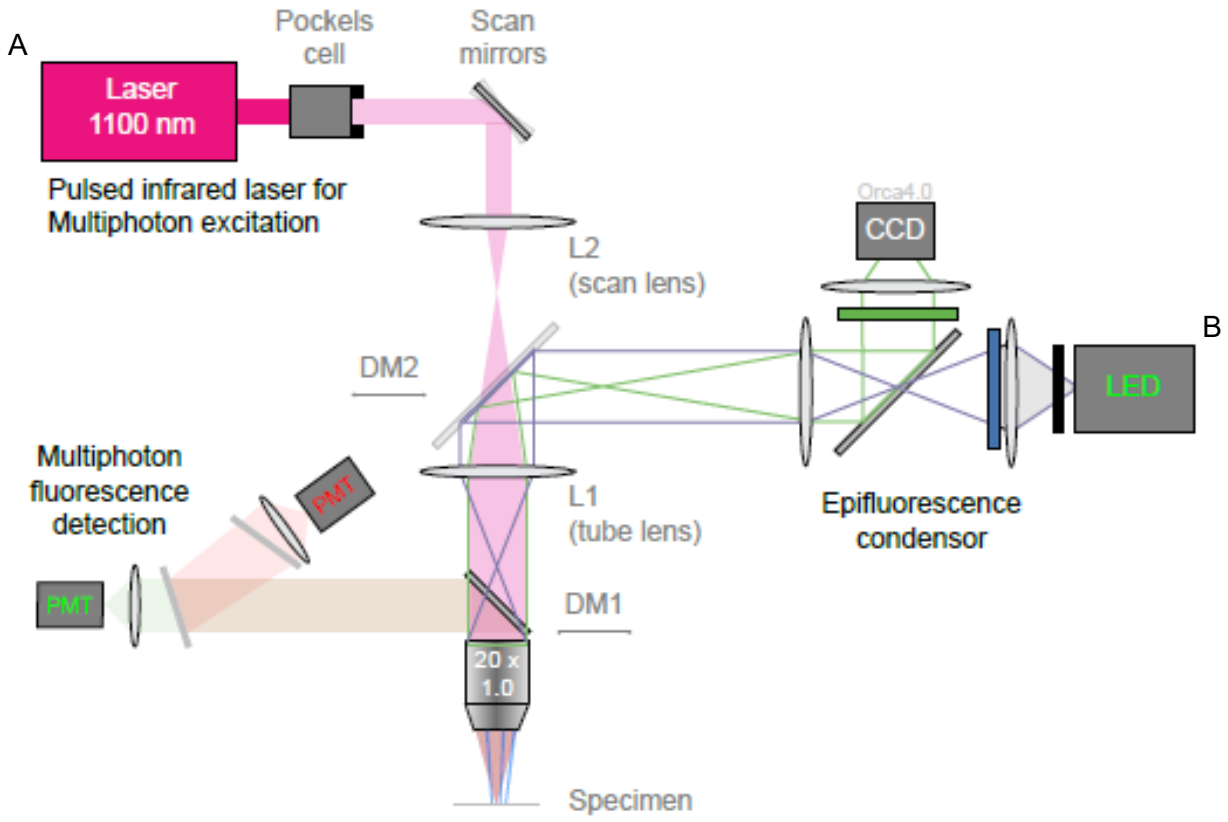


Figure 2.1: Two-photon microscope setup for viewing of zebrafish embryos. An upright microscope was used with two sources of excitation: (A) laser excitation, with the light path towards the specimen indicated in pink. The reflected light path (grey/red) is directed towards PMTs for conversion into a computer-readable signal. The scan lens (L2) is used to filter the wavelength of excitation. (B) LED excitation, with light path towards the specimen indicated in purple lines. The reflected light path (green) is directed towards an ORCA-Flash4.0 CCD camera. *CCD, charge-coupled device; PMT, photomultiplier tube.* Image courtesy of Dr Peixin Zhu (NIBR).

2.6.1.6 DNA extraction and sequencing

DNA was isolated from zebrafish embryos after imaging in order to assess the CRISPR activity in the knockout fish. Whole embryos were digested by incubation with 25 mM NaOH at 95°C for 15 mins, followed by neutralisation with 40 mM Tris-HCl. The CRISPR target regions were PCR amplified using the Q5® High-Fidelity PCR kit (NEB). Each reaction contained 1X Q5® High-Fidelity master mix, 4 µM of each primer (Table 2.7) and 1 µl DNA to a total volume of 50 µl. Reaction conditions consisted of a 2 min 94°C hold, followed by 30 cycles of 94°C for 15 s, 58°C for 15 s and 72°C for 30s. A final step of 72°C for 1 min was also included. NGS of the amplicons was performed at NIBR and was used to determine the rate of insertion or deletion at the point of gene disruption.

Table 2.7: Primers used for assessing CRISPR activity

Gene target	Forward primer (5` - 3`)	Reverse primer (5` - 3`)
myh7 (8:16543250-16543273)	CGCAATAGGCAATAGCATTCCGG	CCTTATAGTGCTCATAATGTTTC
pkp2 (4:17057073-17057096)	CCACAGCAGAGTCGATATCCGCA	CCACCAACATTTGACTGACATCT
polg (25:9339993-9340016)	CCGGTGGCAGACGTTCGTACGAT	CCTAGCTCGCAGTGTTTGGTCTT

2.6.1.7 Statistical analysis

Differences between experimental groups were investigated using R version 3.6.1 (https://www.r-project.org/) and plotted using the ggplot2 package (https://ggplot2.tidyverse.org/). Two-tailed independent t-tests and analysis of variance (ANOVA) tests were performed to compare the means between groups, as the values were normally distributed. Post-hoc statistical power analyses were performed using G*Power version 3.1.9.2 (Faul et al. 2007). Kaplan-Meier survival analysis was used to compare survival rates between larval groups. Survival plots were created using the survminer package of R (https://cran.r-project.org/web/packages/survminer/index.html). Pearson's correlation was used to assess potential relationships between numerical variables. In all analyses, p-values < 0.05 were considered statistically significant. Statistical outliers were defined as values 1.5 interquartile ranges (IQRs) from the first or third quartiles.

2.6.2 Mutant mRNA overexpression

To further explore the role of *POLG* in ACM, a second zebrafish model was employed. In this model, mRNA carrying the mutation of interest (c.2492A>G) was generated and introduced to zebrafish embryos in a transient gene overexpression analysis.

2.6.2.1 Isolation of human POLG

Human *POLG* was amplified using the high fidelity *Pfu* DNA polymerase (Promega). Primers were designed to amplify the entire coding region of *POLG* at an expected product size of 3.7 kb (Table 2.8). The primers were also used to introduce the restriction enzyme (RE) recognition sites for Cla1 and Xba1 into the 5` and 3` ends of the gene, respectively. Each amplification reaction consisted of 1X *Pfu* DNA polymerase reaction buffer (Promega), 200 µM dNTPs (Bioline, Taunton, MA USA), 0.1 µM of each cloning primer (Table 2.8), 0.5 µM *POLG* cDNA (OriGene, Rockville, MD USA) and 1.25 U *Pfu* DNA polymerase (Promega) to a final volume of 50 µl. Reactions were conducted in a Bio-Rad T100^{TM} thermal cycler (Bio-Rad Laboratories, Hercules,

CA USA). Thermal cycling conditions consisted of an initial denaturation step of 95°C for 2 mins, followed by 34 cycles of denaturation at 95°C for 1 min, annealing at 60°C for 30 s and extension at 72°C for 4 mins, followed by a final extension step of 72°C for 5 mins.

Table 2.8: Primers used for the amplification and site-directed mutagenesis of *POLG* cDNA

Primer function	Direction	Primer sequence (5` - 3`)
POLG amplification and introduction of Cla1 RE site	Forward	AACCATCGATATGAGCCGCCTGCTCTGGAG
POLG amplification and introduction of Xba1 RE site	Reverse	ACCCTCTAGACTATGGTCCAGGCTGGCTTC
Site-directed mutagenesis	Forward	TCAGGCACCCCGACTGTGATGAGGAAGGCCT
Confirmation of mutagenesis	Forward	GTGCTCTGGAAATCAACAAAATG
	Reverse	AACTCACTGCCTACTCGGTCA

RE, restriction enzyme

PCR products were visualised using agarose gel electrophoresis. The total PCR products from each reaction were separated on a 1% agarose gel (Appendix E) loaded with 1:10 (stain:agarose) SafeView™ nucleic acid stain (NBS Biologicals, Huntingdon, UK) and compared to Quick-Load® 1 kb DNA ladder (NEB) (Appendix E). Amplicons were then gel purified using the Isolate II PCR and Gel kit (Bioline) following the manufacturer's instructions (Appendix I).

2.6.2.2 Cloning of human POLG

Purified PCR products were digested using ClaI and XbaI, as was PCS2+ plasmid (Figure 2.2). Each digest contained 10 U ClaI (NEB), 20 U XbaI (NEB), 1X CutSmart® buffer (NEB) and 5 µl of the purified PCR product, in a total reaction volume of 50 µl. Reactions were incubated at 37°C for 15 mins, after which the DNA was recovered using the Isolate II PCR and Gel kit (Bioline) (Appendix I).

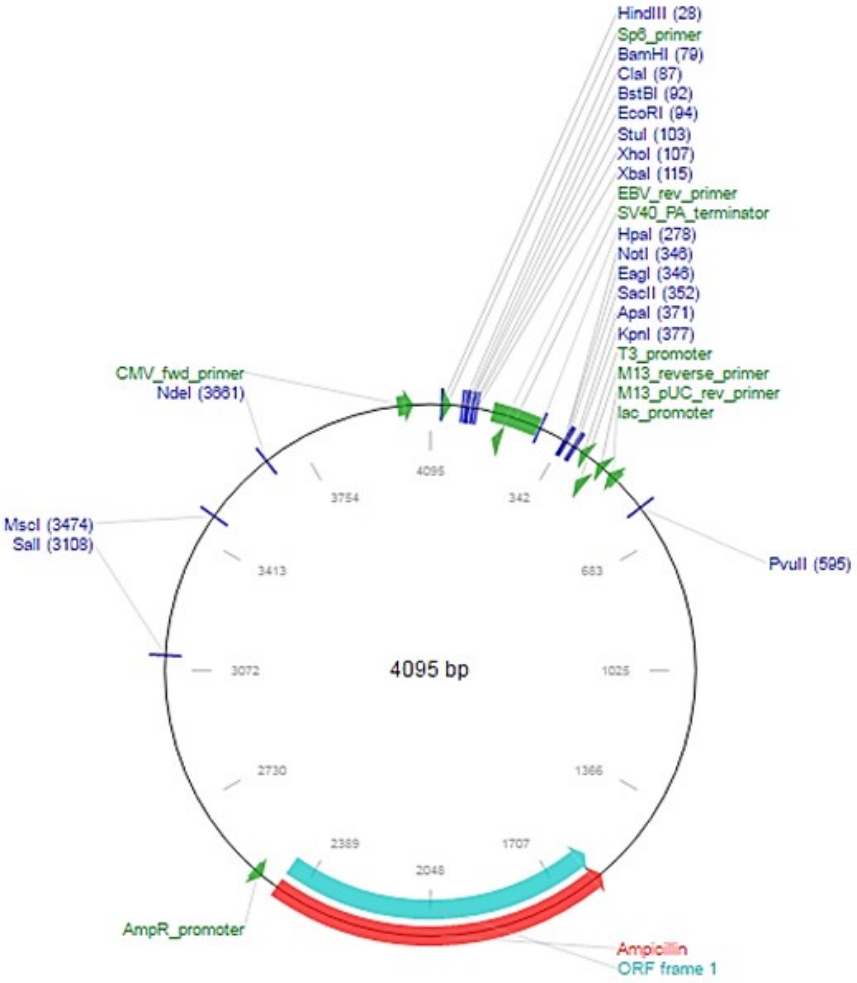


Figure 2.2: Plasmid map for the PCS2+ expression vector. Indicated are the positions of restriction enzyme sites (blue) and primer binding sites or promotor sequences (green). The plasmid also contains an ampicillin resistance gene (red) and is a total of 4,095 bp in size. Adapted from the Addgene Vector Database (Addgene).

The digested *POLG* coding region amplicons were then ligated into the PCS2+ vector. Ligation reactions consisted of 1X T4 DNA Ligase Reaction buffer (NEB), 10 µl *POLG* insert, 1.5 µl PSC2+ plasmid and 600 U T4 DNA Ligase (NEB). Ligation reactions were incubated at 4°C overnight.

2.6.2.3 E. coli *replication of* POLG *vector*

The recombinant *POLG*-carrying PCS2+ plasmids were introduced to *E. coli* cells for selection and amplification. Each cloning reaction was transferred to a vial of One Shot™ TOP10 chemically competent *E. coli* cells (Thermo Fisher Scientific). The reactions were left on ice for 30 mins before

heat shock at 42°C for 30 s. After this step, 1 mL S.O.C. medium (Thermo Fisher Scientific) was added to each reaction. The *E. coli* cells were incubated at 37°C for 1 hr in a bacterial shaker at 200 rpm and spread on an ampicillin plate (Appendix E). Plates were incubated at 37°C overnight in order to facilitate selection of plasmid-containing cells, as well as amplification of the DNA.

Five colonies were picked from each ampicillin plate and cultured in LB broth (Appendix E) and ampicillin at a concentration of 1:1000. The colonies were incubated at 37°C overnight in a bacterial shaker at 120 rpm. Plasmids were recovered from *E. coli* host cells using the Isolate II Plasmid Mini kit (Bioline), following the manufacturer's protocol (Appendix I).

Each recovered plasmid was digested using ClaI and XbaI to check for the presence of *POLG* insert. Digestion reactions contained 2.5 U ClaI (NEB), 5 U XbaI (NEB), 1X CutSmart® buffer (NEB) and 5 µl of the purified plasmid. Reactions were incubated at 37°C for 15 mins before separation and visualisation on a 1% (w/v) agarose gel. Samples were loaded with 1:10 (stain:agarose) SafeView™ nucleic acid stain (NBS Biologicals) and compared to Quick-Load® 1 kb DNA ladder (NEB) (Appendix E).

2.6.2.4 Site-directed mutagenesis of POLG

Plasmids that were confirmed to carry the *POLG* insert were used as template for site-directed mutagenesis. The QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent, Santa Clara, CA USA) was used, in which each reaction contained 1X QuikChange Lightning Multi reaction buffer, 10% QuikSolution, 50 ng mutagenic primer (Table 2.8), 1 µl dNTP mix, 100 ng *POLG*-containing PCS2+ plasmid and 1 µl QuikChange Lightning Multi enzyme blend, to a total volume of 25 µl.

Thermal cycling conditions consisted of a 95°C hold for 2 mins, followed by 30 cycles of 95°C for 20 s, 55°C for 30 s and 65°C for 4 min, followed by a final step at 65°C for 5 mins. The resulting plasmids were then transformed into *E. coli* cells, cultured overnight and recovered as described previously (Section 2.6.2.3).

2.6.2.5 Confirmation of POLG *mutation*

Sanger sequencing was conducted on each plasmid isolate to check for the presence of the c.2492A>G mutation in the vector. For each sequencing reaction, 0.4 µM of primer (Table 2.8) was added to 1 µl plasmid, in a total reaction volume of 10 µl. Sequencing reactions were conducted at the Genomic Technologies Core Facility, University of Manchester. Sequences were aligned to the expected *POLG* sequence to determine their similarity.

2.6.2.6 In vitro *transcription and mRNA recovery*

Plasmids that were confirmed to harbour the *POLG* c.2492A>G mutation (Section 2.6.2.5) were used as template for an *in vitro* transcription reaction to generate mRNA. A control plasmid that did not carry the mutation was also used to generate wild type *POLG* mRNA. Plasmids were first linearised through digestion with KpnI: each reaction contained 10 U KpnI-HF (NEB), 1X CutSmart® buffer (NEB) and 1 µg plasmid, in a total volume of 50 µl. Reactions were conducted at 37°C for 15 mins, and the linearised plasmids recovered using the Isolate II PCR and Gel kit following the manufacturer's protocol (Appendix I).

The linearised plasmids were then transcribed *in vitro* using the mMessage mMachine SP6 kit (Thermo Fisher Scientific), according to the manufacturer's instructions. Each reaction contained 1X Reaction Buffer, 1X NTP/CAP, 2 µl Enzyme Mix and up to 1 µg template DNA, and was incubated at 37°C for 2 hr before treatment with 1 µl TURBO DNase and further incubation at 37°C for 15 mins. The RNA was purified with the MEGAclear™ Transcription Clean-Up kit and eluted in RNase-free water (Thermo Fisher Scientific) (Appendix H).

The mRNA was quantified using a NanoDrop® 8000 spectrophotometer (NanoDrop Technologies Inc.), and visualised on a 1% (w/v) agarose gel (Appendix E). All mRNA was diluted to working concentrations of 100 ng/µl and stored at -80°C.

2.6.2.7 Zebrafish lines used for mRNA overexpression

Wild type AB zebrafish were used for mRNA overexpression experiments. Zebrafish were raised and maintained at The University of Manchester Biological Services Unit under standard conditions (Westerfield 2000). Adult zebrafish husbandry was approved by the University of

Manchester Animal Welfare and Ethical Review Board, and all experiments were performed in accordance with UK Home Office regulations.

2.6.2.8 Zebrafish breeding and embryo collection

Paired mating was set up as described previously (Section 2.6.1.3). Fertilised embryos were harvested and washed in E3 media containing methylene blue (0.002 g/l) (Appendix E) and used for RNA injection experiments.

2.6.2.9 Injection of mRNA into zebrafish embryos

Injections were performed using a PLI-90 pico-injector (Harvard Apparatus, Holliston, MA USA). The dose of mRNA was optimised by injecting 25 pg, 50 pg, 100 pg or 200 pg of wild type *POLG* mRNA into single-cell zebrafish embryos and monitoring them until 5 dpf. The heart rate of each injected fish was determined at 3 dpf (Section 2.6.2.10), and statistically analysed (Section 2.6.2.11). Experimental embryos were injected with 50 pg mutant *POLG* mRNA in a total volume of 2 nl (n = 135). Control fish included sibling embryos (from the same clutch) that were injected with 50 pg wild type *POLG* mRNA in a volume of 2 nl (n = 75). In all experiments, control groups also included fish injected with an equivalent volume of nuclease-free water (n = 75) and uninjected embryos (n = 75). Embryos from all groups were incubated at 28°C and were monitored regularly until visualisation at 3 dpf.

2.6.2.10 Imaging and phenotypic analysis

At 3 dpf, larvae from all experimental groups were mounted in 3% methylcellulose (Appendix E) and viewed on a Leica MZ9.5 stereomicroscope. The larvae were not treated with tricaine anaesthetic. Videos of 30 frames per second were generated at 10x magnification. The videos were analysed using DanioScope (Noldus) to measure heart rhythm, cardiac output and chamber size.

2.6.2.11 Statistical analysis

Differences between groups were investigated as described before (Section 2.6.1.7). Because the heart rate was not normally distributed, differences between experimental groups were

calculated using two-tailed Mann-Whitney U tests and Kruskal-Wallis tests, and the median from each group was reported. Statistical outliers (1.5 IQRs from the first or third quartiles) were removed from the analysis. Differences in ventricular measurements and flow rates, which were normally distributed, were investigated using two-tailed independent t-tests and ANOVA to compare means between groups. P-values < 0.05 were considered statistically significant. Post-hoc power analyses were performed using G*Power version 3.1.9.2 (Faul et al. 2007).

Chapter 3 Mutations in established cardiomyopathy genes as the cause of disease in South African cardiomyopathy patients

3.1 Introduction

Heritable cardiomyopathies are characterised by marked clinical and genetic heterogeneity. The genetic diversity displayed by cardiomyopathy includes both locus heterogeneity, whereby mutations in different genes can cause the same form of cardiomyopathy, and allelic heterogeneity, in which one gene may be associated with different types of cardiomyopathy (Mayo Clinic Laboratories 2015) (Figure 3.1).

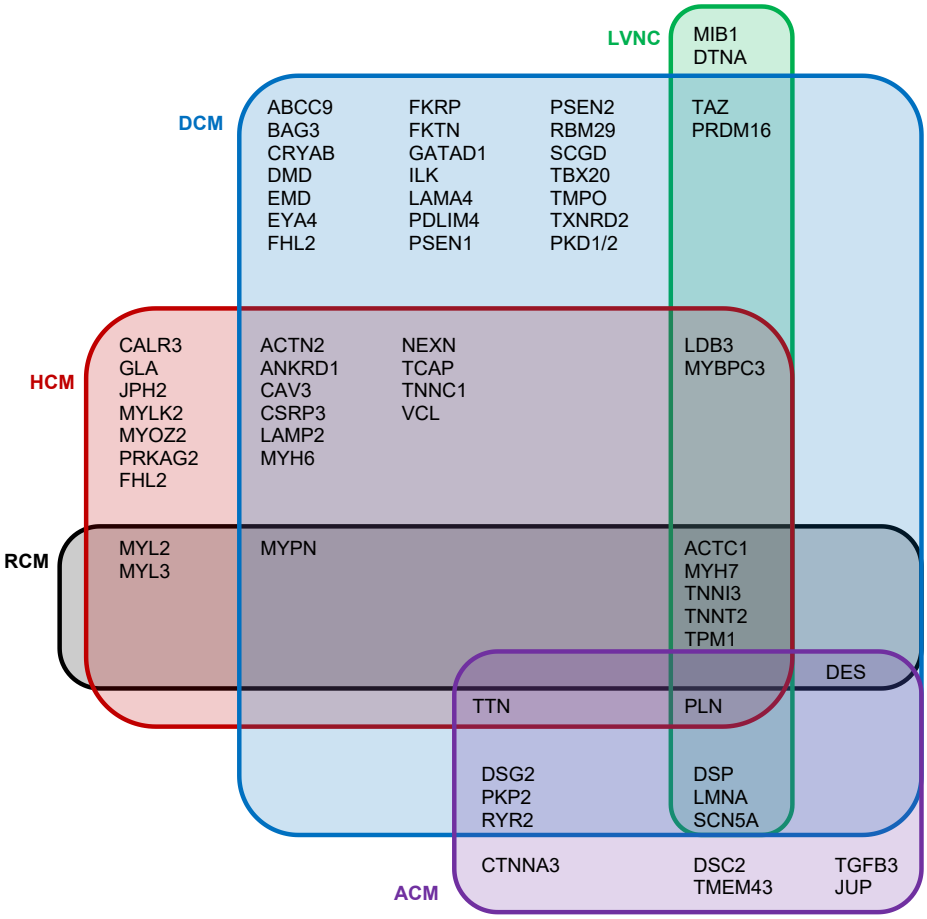


Figure 3.1: The genetic overlap between different types of cardiomyopathy. Adapted from Van Tintelen JP et al. *Cardiovasc Res*, **101**(4):572. *ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular non-compaction; RCM, restrictive cardiomyopathy*

Advances in sequencing technology have only emphasised the substantial genetic overlap amongst the different cardiomyopathy phenotypes. For instance, NGS gene panels have been used to describe novel associations of ACM, HCM and channelopathy gene mutations in DCM (Haas et al. 2015, Mahdieh et al. 2018), ACM gene mutations in LVNC and DCM (Klauke et al. 2017, Ramond et al. 2017) and HCM gene mutations in LVNC and RCM (Brodehl et al. 2017, Schaefer et al. 2014, Wu et al. 2015). Similarly, the genes *HCN4*, *FLNC* and *MTO1*, which were identified in exome sequencing studies, have since been demonstrated as mutation hotspots for numerous forms of heritable cardiomyopathy (Baruffini et al. 2013, Brun et al. 2020, Hall et al. 2020, Janin et al. 2017, Liu et al. 2019, Millat et al. 2015, Nozari et al. 2018, Ortiz-Genga et al. 2016, Schweizer et al. 2014, Taylor et al. 2014, Tucker et al. 2017).

To our knowledge, the genetic heterogeneity of cardiomyopathy is largely unexplored in the South African patient population. In this chapter, three families with cardiomyopathy are investigated and found to illustrate this genetic diversity of cardiomyopathy.

3.2 Family 1 (DCM 389)

3.2.1 Clinical history of Family 1

In 2014, a multi-generation family with AD DCM was recruited at GSH. The family consisted of three affected female siblings and their affected father (Figure 3.2). The proband, III:4, was diagnosed with severe DCM at the age of 44 years with an ejection fraction (EF) of 22%. Her father (II:2) underwent a heart transplant but died before he could be enrolled into the study. The proband's sisters (III:6 and III:8) were also diagnosed with DCM in their fifth decades of life, but with variable phenotypes: individual III:6 had similarly severe DCM to the proband (EF of 10-15%), while the youngest sister (III:8) had DCM that recovered to a normal EF during treatment.

DNA from all sisters and some of their children and grandchildren was available in this investigation. Although they were clinically examined and found to be unaffected, all the children (generation IV) and grandchildren (generation V, not shown) were below the age of disease onset in this family (range: 2 to 28 years). The proband's son (IV:1) had an end-diastolic left ventricular internal diameter of 56 mm when he was screened at the age of 19 years, which was within normal values but borderline dilated. Two of the proband's cousins (III:2 and III:3) were examined in their fifth decades of life and were found to be phenotypically normal.

Family 1

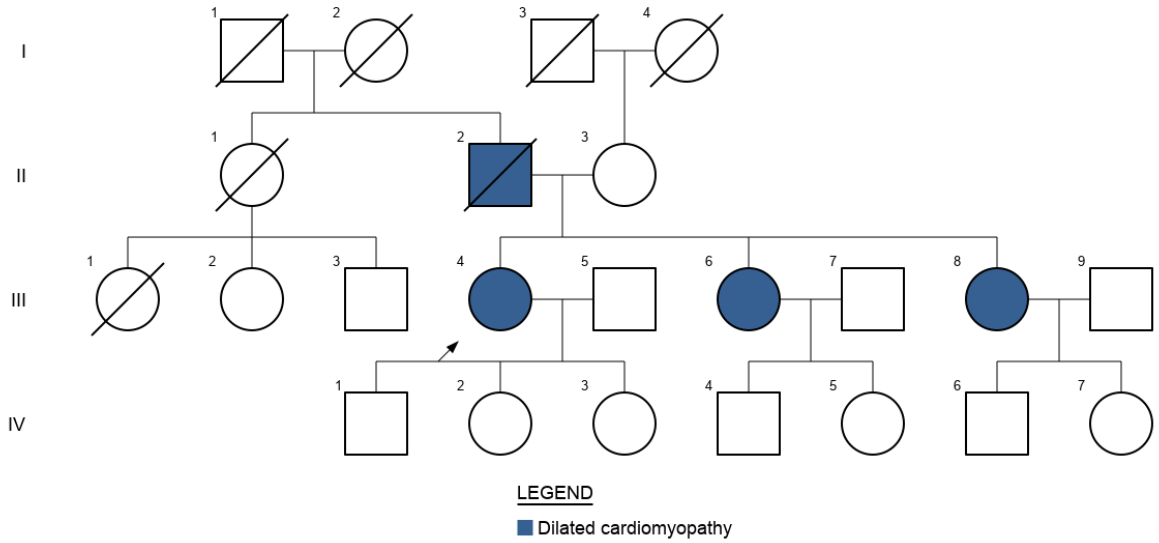


Figure 3.2: Pedigree indicating disease segregation in Family 1. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates DCM. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

3.2.2 Genetic analysis of Family 1

Exome sequencing was conducted on two affected individuals from the family: the proband and her sister, individual III:6. The third affected sibling (III:8) was recruited after the study began, and their father died before the study began, so these individuals were not included in the exome sequencing; however, DNA from individual III:8 was used for segregation analysis. The sequencing generated a total of 46,720 variants in both individuals, at a mean sequence coverage of 120x. Family history was indicative of AD inheritance. After filtering, 84 variants remained (Figure 3.3; Appendix J).

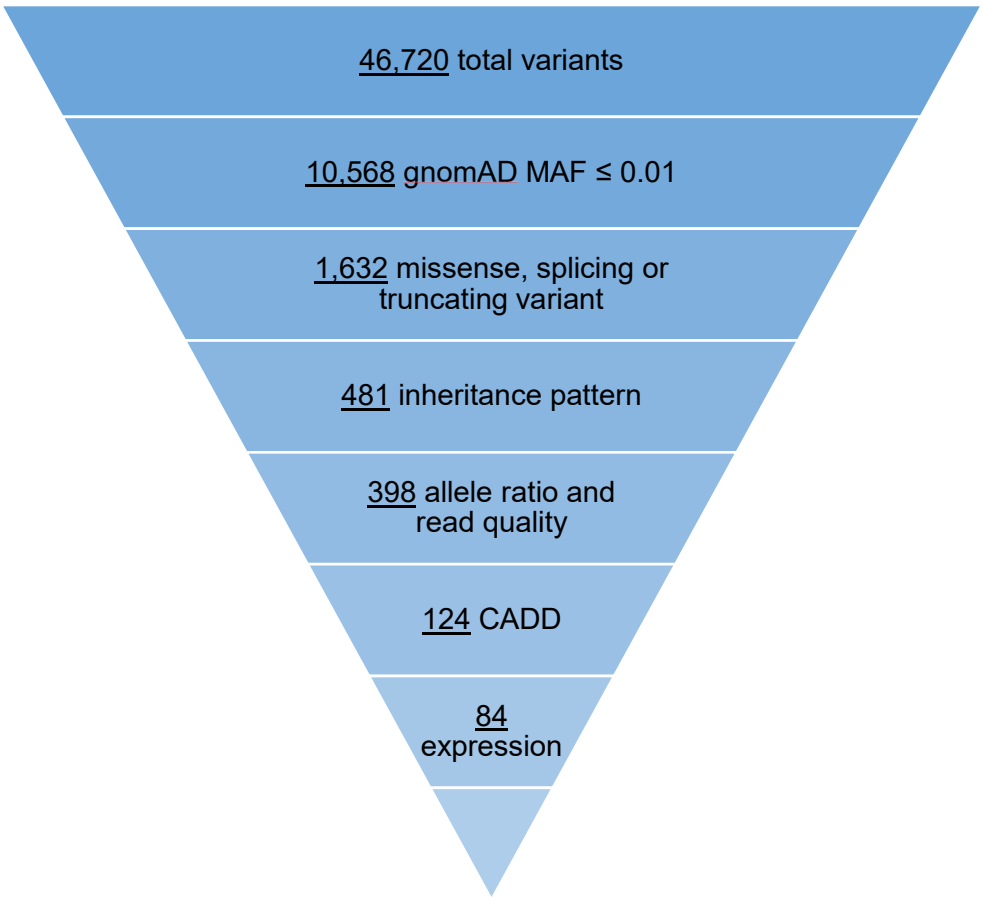


Figure 3.3: Filtering of exome sequencing data from Family 1. Exome sequencing produced a total of 46,720 variants, of which 84 met the filtering criteria. *CADD, Combined Annotation Dependent Depletion; gnomAD, Genome Aggregation Database; MAF, minor allele frequency*

Two of the variants occurred in genes on the cardiomyopathy panel (Table 3.1): missense mutations were found in the genes *MYOM1* and *MYH7*. Other genes not previously associated with cardiomyopathy were also considered, but many were excluded due to low expression in the heart or a lack of functional relevance to DCM (Appendix J). However, three genes were considered based on their function and pathogenicity predictions: *TJP1*, *LAMA3* and *ABCB4* (Table 3.1).

Table 3.1: Candidate variants in Family 1

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Cardiomyopathy panel														
MYH7	c.4394C>T	p.S1465L	0	0	0	0	D (0.581)	D (0.67)	D (0)	B (0.322)	25.4	4553.0	Yes	7/7
MYOM1	c.139A>G	p.S47G	0.003	0.005796	0.007187	0.00201	NS	D (0.76)	B (0.08)	B (0.079)	21.9	214.6	NA	5/7
Other genes														
ABCB4	c.523A>G	p.T175A	0.006	0.01062	0.01163	0.00846	NS	D (0.99)	D (0.01)	D (0.952)	24.2	1.2	Yes	6/7
LAMA3	c.4643A>G	p.D1548G	0.002	0.0008778	0.0009064	0.00240	B (0.012)	D (0.99)	B (0.06)	D (0.795)	25.1	1.2	No	4/7
TJP1	c.1412A>G	p.N471S	0.005	0.009164	0.009114	0.00975	NS	D (0.99)	D (0.02)	B (0.040)	23.2	14.9	Yes	7/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

The variants in *TJP1* and *LAMA3* were excluded due to lack of segregation with disease phenotype (Appendix K). The variant *ABCB4* c.523A>G (p.T175A) was of interest because the same variant was reported to track with disease in an Italian family with AD atrial fibrillation and/or atrial flutter (Maciag et al. 2015). Although it tracked with DCM in this family too (Appendix K), *ABCB4* p.T175A was too common in the ExAC and gnomAD databases to be considered the primary pathogenic variant in this family (Table 3.1).

The variant *MYOM1* c.139A>G was homozygous in individual III:6 and was excluded. The *MYH7* c.4394C>T (p.S1465L) variant was validated (Figure 3.4) and determined to be sufficiently rare in the population, as *MYH7* p.S1465L was absent in all population databases checked. It is a missense variant, and the affected amino acid was highly conserved in all species (Figure 3.5): all species had the same residue as humans (serine) or alanine or threonine, which are non-radical amino acid changes (Zhang 2000). Due to these factors, the *MYH7* tail region variant was considered the most likely disease-causing variant in this family (Figure 3.6).

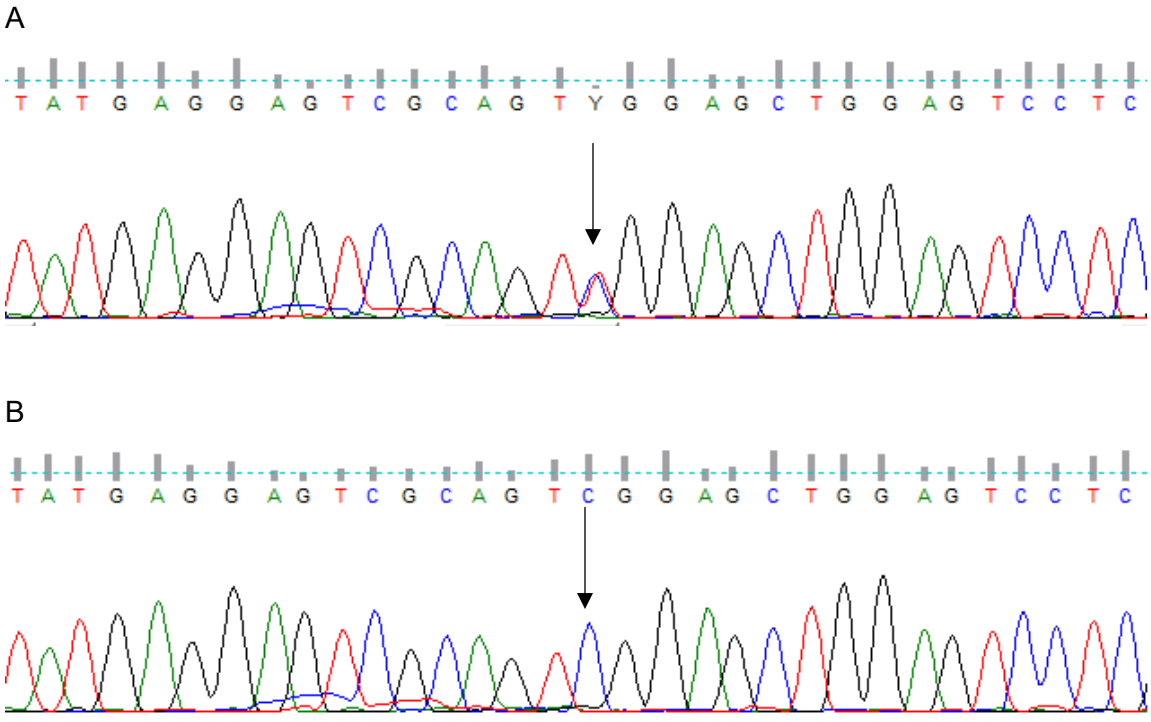
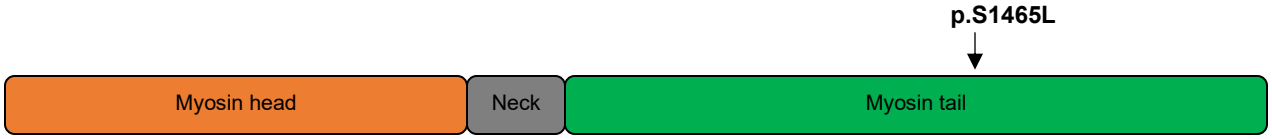


Figure 3.4: Validation of *MYH7* c.4394C>T identified in Family 1. (A) Sanger sequencing to confirm the presence of *MYH7* c.4394C>T in affected individual III:4 (indicated with an arrow), (B) Sanger sequencing confirms the absence of the variant in unaffected individual III:2 (indicated with an arrow).

A



B

H. sapiens	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
M. mulatta	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
P. troglodites	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
F. catus	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
M. musculus	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
G. gallus	S	E	W	K	Q	K	F	E	E	S	Q	**T**	E	L	E	S	S	Q	K	E	A	R	S
D. rerio	A	E	W	K	Q	K	Y	E	E	S	Q	**S**	E	L	E	S	S	Q	K	E	A	R	S
X. tropicalis	S	E	W	K	Q	K	F	E	E	S	Q	**A**	E	L	E	S	S	Q	K	E	A	R	S

Figure 3.5: *MYH7* c.4394C>T identified in Family 1. (A) Protein structure of MYH7 with position of the p.S1465L mutation marked with an arrow, (B) Multiple species conservation alignment of the p.S1465 residue of MYH7 (shaded in blue).

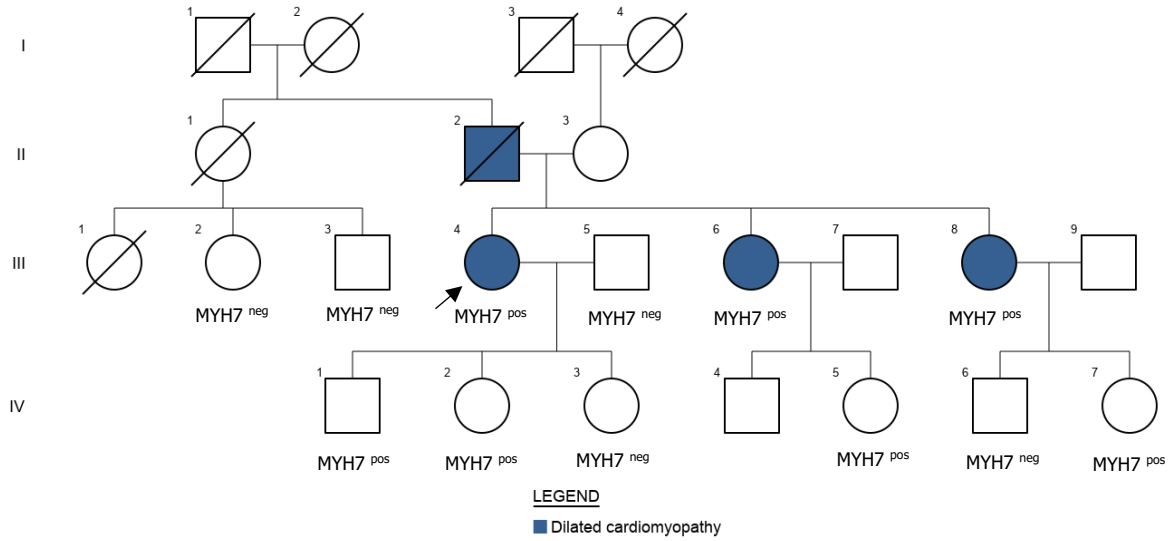


Figure 3.6: Pedigree indicating segregation of *MYH7* c.4394C>T in Family 1. Individuals marked 'MYH7 ^{pos}' were heterozygous for the variant, and those marked 'MYH7 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Shaded symbols indicate affected individuals. Blue shading indicates DCM. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

3.2.3 Discussion: MYH7 *mutations as a cause of DCM*

The sarcomere is composed of thick and thin filaments that together make up the contractile apparatus of muscle tissue. *MYH7* encodes the heavy chain component of cardiac muscle cells (Tajsharghi and Oldfors 2012). Myosin heavy chains are characterised by a globular actin-binding head region and a rod-like helical tail, which is used for dimerisation of the protein (Figure 3.5A).

Typically, mutations in *MYH7* are associated with HCM, a disorder characterised by marked ventricular hypertrophy as opposed to the dilatation observed in DCM. In fact, mutations in *MYH7* are the second most common cause of HCM, accounting for up to 20% of all cases (Marian and Braunwald 2017). However, genetic overlap between DCM and HCM is not uncommon, and approximately 3-4% of DCM cases have been attributed to mutations in *MHY7* already (McNally and Mestroni 2017, Merlo et al. 2013, Millat et al. 2011, Wasielewski et al. 2014, Zhao et al. 2015).

The mutation found in Family 1 occurred in the tail region of the MYH7 protein (Figure 3.5). The majority of reported pathogenic HCM *MYH7* mutations are in the head and neck domains, which are involved in actin binding and muscle contraction, but DCM-causing mutations have been observed to occur more frequently in the tail domain of *MYH7* (Walsh et al. 2010). A most notable example is the Dutch founder mutation *MYH7* c.5754C>G (p.N1918K) which was described in 15 families with DCM and/or congenital heart defects (van der Linde et al. 2017). Other tail region mutations include c.3942C>G (p.D1314E) and c.4377G>T (p.K1459N), reported in two Portuguese DCM patients (Sousa et al. 2019), and c.2770G>A (p.E924K), which was coinherited with a *LAMA4* mutation to cause severe paediatric cardiomyopathy (Abdallah et al. 2019).

MYH7 tail region mutations have also been described in infantile bi-ventricular noncompaction (Miura et al. 2019) and a unique adult-onset form of systolic cardiomyopathy characterised by heart failure and reduced EF in the absence of ventricular hypertrophy or dilation (Yang et al. 2018). In the latter case, the c.2543A>G (p.E848G) mutation was demonstrated to impair the interaction of MYH7 protein with MYBPC3 and cause contractile dysfunction of patient-derived heart tissue (Yang et al. 2018). The basis for how *MYH7* tail region mutations lead to cardiomyopathy is unclear, although this study indicates that it may involve disrupted protein-protein interactions in the sarcomere. It is also consistent with the present finding of a pathogenic mutation in the *MYH7* tail region in this family. The high sequence conservation of the p.S1465 residue and surrounding region in many species indicates its functional importance.

3.2.4 Assessment of the pathogenicity of MYH7 *c.4394C>T*

For genes that have been associated with disease already, established criteria of pathogenicity can be used to investigate their disease-causing potential. The most widely used variant pathogenicity guidelines are those published by the ACMG, which were discussed earlier (Chapter 1). However, missense mutations may be difficult to interpret as they seldom cause a complete loss of gene function. The greatest evidence is if the same mutation has been reported in other affected individuals or if the mutation occurs at or near the site of other known disease-causing mutations (Richards et al. 2015). Other lines of evidence that may be pursued are *in vitro* or *in vivo* functional investigations, or co-segregation in large family pedigrees.

The *MYH7* variant identified in Family 1 occurred in the tail region of the protein, a region with several observational links to DCM (Abdallah et al. 2019, Sousa et al. 2019, van der Linde et al. 2017, Walsh et al. 2010). The occurrence of *MYH7* c.4394C>T within this region is moderate evidence of pathogenicity. An adaptation of the ACMG variant classification guidelines for *MYH7*-associated cardiomyopathies emphasises segregation within seven or more family members as indicative of pathogenicity, or presence of the mutation in 15 different, unrelated probands (Kelly et al. 2018). Unfortunately, due to the small family size and the late onset of disease, segregation analysis was only possible in three affected individuals (Figure 3.6). Importantly, the variant did not occur in any unaffected individual who was the age of disease onset. The modified guidelines also consider variants that are significantly more prevalent in cases compared to controls. It is unfortunate that a South African control population could not be screened in this investigation. However, this variant was completely absent in all control databases checked (totalling 150,463 individuals), indicating that this is likely a private pathogenic mutation in this family.

Following the ACMG guidelines, this variant fulfils moderate and supporting criteria (such as segregation with disease, rarity and computational predictions of deleteriousness) sufficient for a classification of ‘likely pathogenic’. As the progeny of the affected siblings reach the age of disease onset, increasing co-segregation data may help to confirm the pathogenicity of this variant.

3.3 Family 2 (HCM 50)

3.3.1 Clinical history of Family 2

Family 2 was recruited at GSH in 2014 and encompasses two individuals who were diagnosed with HCM (Figure 3.7). The family consists of the proband (II:2) and his mother (I:2), who were 53 and 74 years of age respectively at the time of recruitment. Both individuals showed signs of ventricular hypertrophy on echocardiography and were described as having an arrhythmic heartbeat; they have both subsequently had pacemakers inserted.

Family 2

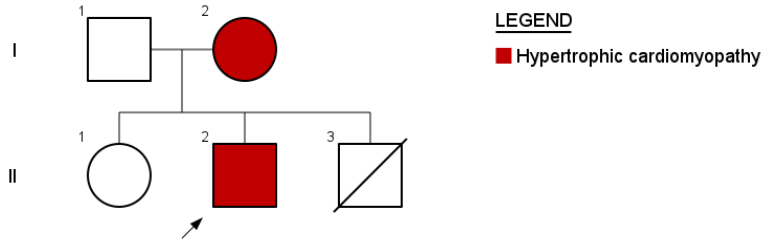


Figure 3.7: Pedigree indicating disease segregation in Family 2. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Red shading indicates HCM. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

The medical history of the extended family is unclear at this stage. The proband's mother had two additional children, although one died at the age of 12 years in a motor vehicle accident. The proband's sister has not been screened, although to the best of our knowledge she is asymptomatic and was 55 years old at the time of recruitment. DNA was therefore only available for individuals I:2 and II:2.

3.3.2 Genetic analysis of Family 2

After exome sequencing, 48,103 variants were observed and annotated in both individuals at a mean sequence coverage of 137x. Due to the family history, a dominant mode of inheritance was

considered when filtering the exome data. Of the total 48,103 variants, 92 met the filtering criteria and were considered for their disease-causing potential (Figure 3.8; Appendix L).

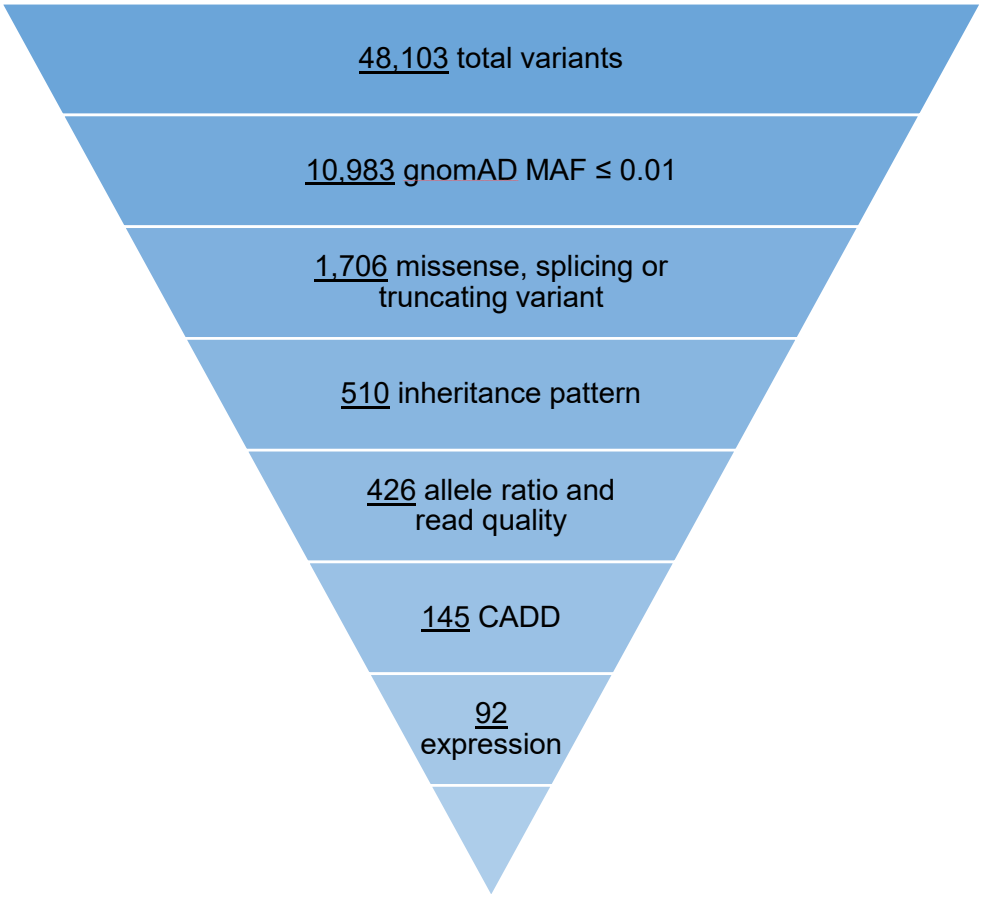


Figure 3.8: Filtering of exome sequencing data from Family 2. Exome sequencing produced a total of 48,103 variants, of which 92 met the filtering criteria. *CADD, Combined Annotation Dependent Depletion; gnomAD, Genome Aggregation Database; MAF, minor allele frequency*

Three variants occurred in the known cardiomyopathy genes *GLA*, *MYBPC3* and *PKP2*, as well as variants in the putative cardiomyopathy genes *GBE1* (glycogen storage disease) and *LAMA2* (muscular dystrophy) (Table 3.2). Although predicted deleterious by various pathogenicity prediction tools, the missense variants in *MYBPC3* and *PKP2* were reported as likely benign for cardiomyopathy and other cardiovascular abnormalities (ClinVar accession nos.: RCV000157319.1, RCV000244315.2, RCV000476215.3, RCV000172579.2, RCV000231874.4 and RCV000249816.1).

Table 3.2: Candidate variants in Family 2

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Cardiomyopathy panel														
GBE1	c.664A>G	p.N222D	0	0	0	0	D (0.040)	D (0.99)	B (0.37)	B (0.021)	22.4	24.5	Yes	3/7
GLA	c.774_775delAC	p.P259R*5	0	0	0	0	NS	D (1)	NS	NS	35.0	4.0	Yes	NS
LAMA2	c.7040G>T	p.G2347V	0.004	0.000157	0.0001156	0	D (0.148)	D (0.99)	D (0)	D (0.952)	32.0	33.1	No	5/7
MYBPC3	c.3392T>C	p.I1131T	0	0.001064	0.000840	0.05670	D (0.184)	D (0.99)	D (0.04)	B (0.400)	24.3	1369.0	Yes	7/7
PKP2	c.1592T>G	p.I531S	0.001	0.004722	0.004722	0.00323	NS	D (0.80)	D (0)	D (0.974)	24.8	84.1	Yes	6/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

The *GLA* variant, in contrast, was absent in all population databases checked. Other genes not present in the cardiomyopathy panel were also considered, but no candidates were identified (Appendix L). *GLA* c.774_775delAC (p.P259Rfs*5) is a two bp frameshift deletion. The variant, which was confirmed by Sanger sequencing (Figure 3.9), is predicted to result in protein truncation at residue 259 (Figure 3.10). The wild type GLA protein (UniProt ID P06280) is 429 amino acids long; the mutation results in loss of the terminal 171 amino acids (approximately 40% of the protein) and termination of the protein within 5 residues. Due to its disruptive effects on the GLA protein and the prior involvement of *GLA* in HCM phenocopies, *GLA* c.774_775delAC was considered the most likely disease-causing variant in this family. Because DNA was not available for additional family members, segregation analysis beyond these individuals was not possible; however, the *GLA* c.774_775delAC variant was present in both affected individuals tested.

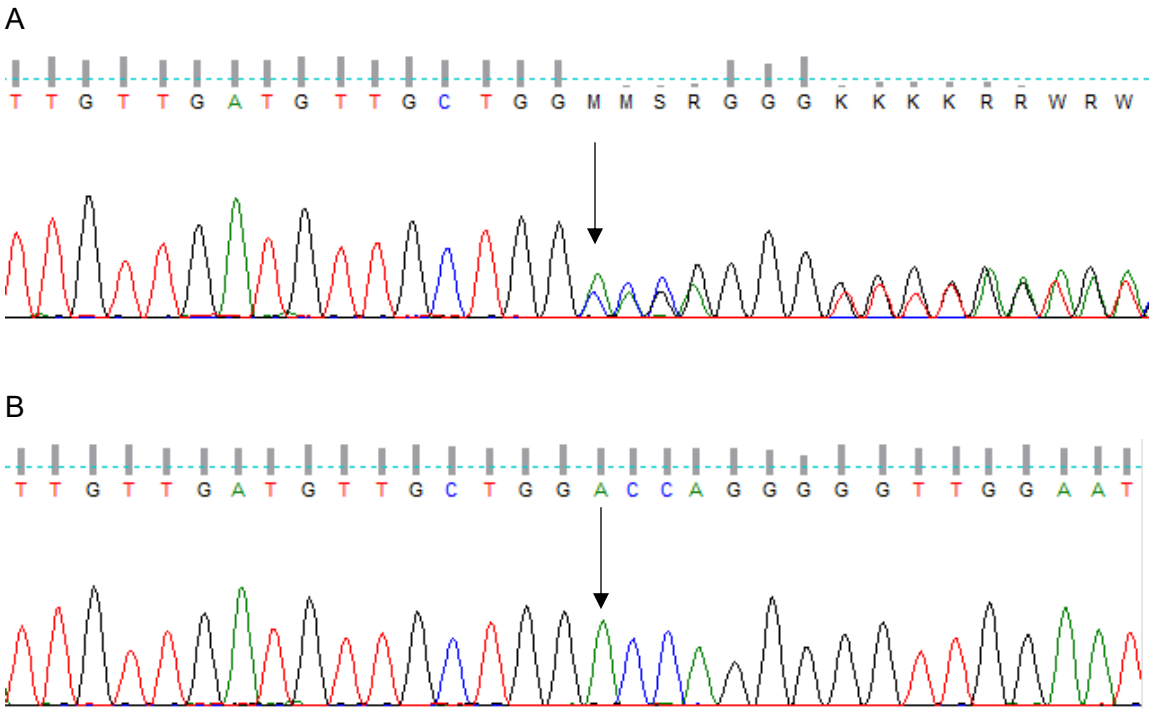


Figure 3.9: Validation of *GLA* c.774_775delAC identified in Family 2. (A) Sanger sequencing to confirm the presence of *GLA* c.774_775del in affected individual I:2 (point of disruption indicated with an arrow), (B) Sanger sequencing confirms the absence of the variant in an unaffected control individual, IV:1, from Family 1 (indicated with an arrow).

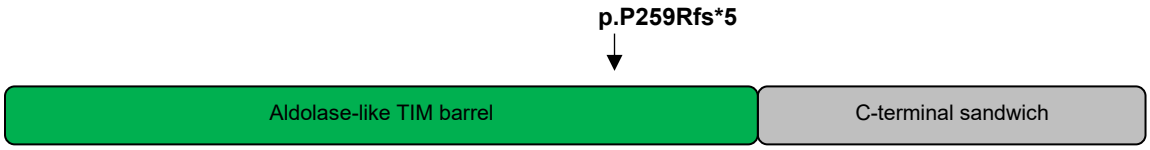


Figure 3.10: Protein structure of GLA. The position of the p.P259Rfs*5 mutation is indicated (marked with an arrow). *TIM, triosephosphate isomerase*

3.3.3 Discussion: GLA *mutations as a cause of HCM phenocopy*

GLA encodes the lysosomal α-galactosidase A, an enzyme which hydrolises glycosphingolipids such as globotriaosylceramide (GL-3) in various tissues and organs. Deficient or absent α-galactosidase A activity, caused by *GLA* mutations, results in Anderson-Fabry disease (Fabry disease), which is characterised by the accumulation of GL-3 in the lysosomes of cells throughout the body (Eng et al. 1993, Hagège et al. 2019). Although clinical signs of Fabry disease typically include kidney, dermatological, neurological and cardiac manifestations, many variants of the disorder have been reported, including predominantly cardiac variants (Havndrup et al. 2010, Linhart et al. 2007). These variants can manifest as a multisystemic disorder (Linhart et al. 2007) or may mimic the HCM phenotype (Havndrup et al. 2010). In fact, cardiovascular complication is considered the leading cause of mortality amongst Fabry disease patients: a systematic review of 4,185 patients found that 75% of deaths were attributable to cardiovascular causes such as SCD (Baig et al. 2018). Cardiac hypertrophy and myocardial fibrosis in Fabry disease have been ascribed to the deposition of GL-3 in cardiomyocytes, while accumulation of GL-3 in conduction tissue can lead to arrhythmias or short PR intervals in patients (Hagège et al. 2019). Because the underlying genetics and aetiology of the disorders are distinct, Fabry disease is regarded as an HCM phenocopy.

In European HCM cohorts, approximately 1-5% of patients have been found to have *GLA* mutations (Adalsteinsdottir et al. 2014, Azevedo et al. 2019, Cecchi et al. 2017, Cecconi et al. 2016, Havndrup et al. 2010, Jääskeläinen et al. 2019, Monserrat et al. 2007, Rubattu et al. 2016). Similarly, analysis of a cohort of Korean HCM patients revealed a *GLA* mutation rate of 4.6% (Seo et al. 2016). Fabry disease may therefore constitute an important contribution to HCM prevalence, particularly amongst individuals without known sarcomeric gene mutations (Havndrup et al. 2010).

The prevalence of Fabry disease or *GLA* mutations in South African patients is unknown, although targeted sequencing of *GLA* and 14 other HCM genes in 43 South African HCM patients did not identify any pathogenic *GLA* mutations (Ntusi et al. 2016). Other investigations of HCM in South African patients have focussed primarily on the known HCM genes *MYBPC3* and *MYH7* (Moolman, Brink, and Corfield 1993, 1995, Moolman-Smook et al. 1998, Moolman-Smook et al. 1999, Posen et al. 1995). To the best of our knowledge, therefore, this is the first report of a *GLA* mutation in South African patients with unexplained left ventricular hypertrophy.

The mutation identified in Family 2 was a 2 bp deletion resulting in premature protein termination and possible nonsense-mediated decay. This means that, rather than exerting a dominant negative effect, this mutation likely acts through haploinsufficiency of the *GLA* gene. Although missense mutations in *GLA* have been linked to the cardiac variant of Fabry disease (Adalsteinsdottir et al. 2014, Azevedo et al. 2019, Barman et al. 2019, Brito et al. 2014, Cecconi et al. 2016, Germain et al. 2018, Havndrup et al. 2010, Jääskeläinen et al. 2019, Oder et al. 2017, Pavlu et al. 2019, Valtola et al. 2020), similar truncating *GLA* mutations have been described in HCM previously, including splice site and nonsense mutations (Cecconi et al. 2016, Jääskeläinen et al. 2019, Juang et al. 2019, Militaru et al. 2018, Seo et al. 2016, Watanabe et al. 2013, Zhao et al. 2017, Liang et al. 2020), as well as the 2 bp deletion *GLA* c.718_719delAA (p.K240E*5) that was reported in an Italian HCM patient with atrial fibrillation and ventricular tachycardia (Cecconi et al. 2016). This mutation shows many similarities with the c.774_775delAC mutation found in Family 2; the same protein domain is disrupted, and the mutation carriers were noted to have arrhythmic phenotypes. Indeed, the cardiac variant of Fabry disease may frequently present with arrhythmias such as short PR interval, atrial fibrillation, ventricular tachycardia and sinus bradycardia (Azevedo et al. 2019, Brito et al. 2014, Cecchi et al. 2017, Havndrup et al. 2010, Monserrat et al. 2007, Poulin et al. 2015, Seo et al. 2016, Watanabe et al. 2013).

The cardiac variant of Fabry disease is typically late-onset, presenting as unexplained left ventricular hypertrophy in middle-aged or older patients (Brito et al. 2018, Nakao et al. 1995). Both global left ventricular hypertrophy and asymmetrical septal hypertrophy (such as in sarcomeric HCM) have been observed in Fabry disease (Cecchi et al. 2017, Hagège et al. 2019, Nakao et al. 1995). It is thought that Fabry disease may mimic HCM through *GLA* mutations that do not completely disrupt GLA protein function: individuals with ‘classical’ Fabry disease have no detectable GLA activity, while the GLA protein in patients with the cardiac variant retains residual (approximately 10%) enzymatic activity (Brito et al. 2014, Eng et al. 1993, Hagège et al. 2019,

Oder et al. 2017, Smid et al. 2015). Due to the vast clinical heterogeneity of Fabry disease, distinguishing it from HCM can be challenging without CMR T1 mapping (Karur et al. 2018) or clinical genetic testing (Brito et al. 2014, Brito et al. 2018). However, identifying pathogenic *GLA* mutations in HCM patients can lead to a subsequent diagnosis of Fabry disease in these individuals. This has important implications for patient management, as antiarrhythmic drugs are unsuitable for Fabry disease and will not prevent arrhythmias or SCD (Acharya, Doppalapudi, and Tallaj 2015). Enzyme replacement therapy, in contrast, is available to rescue GLA function and may improve left ventricular hypertrophy (Weidemann et al. 2009, Yamamoto et al. 2019, Yogasundaram et al. 2017).

3.3.4 Assessment of the pathogenicity of GLA *c.774_775delAC*

According to the ACMG guidelines, the strongest evidence of pathogenicity is placed upon truncating variants in genes in which loss of function is known to cause disease (Richards et al. 2015). The *GLA* frameshift mutation c.774_775delAC, identified in HCM patients in Family 2, is an example of such a variant. Similar truncating deletions have been reported in Fabry disease and HCM (Cecconi et al. 2016, Jääskeläinen et al. 2019, Juang et al. 2019, Militaru et al. 2018, Seo et al. 2016, Watanabe et al. 2013, Zhao et al. 2017); this, coupled with the absence of the variant in all population databases checked, provides very strong evidence that this is the causative mutation in this family. Although only two family members were screened in this investigation, this variant is classed as a pathogenic mutation using the ACMG criteria.

3.4 Family 3 (DCM 3)

3.4.1 Clinical history of Family 3

Identified at GSH in 1997, Family 3 is a large family with DCM spanning two known generations (Figure 3.11). The family consists of four affected individuals, three of whom were brothers (III:6, III:8 and III:9). One of the brothers, individual III:6, died before the study began and could not be recruited. The fourth affected individual (IV:1) was the nephew of the other patients. DNA from the two brothers (III:8 and III:9) and their nephew (IV:1) was available in this study, as was DNA from several of their relatives.

Family 3

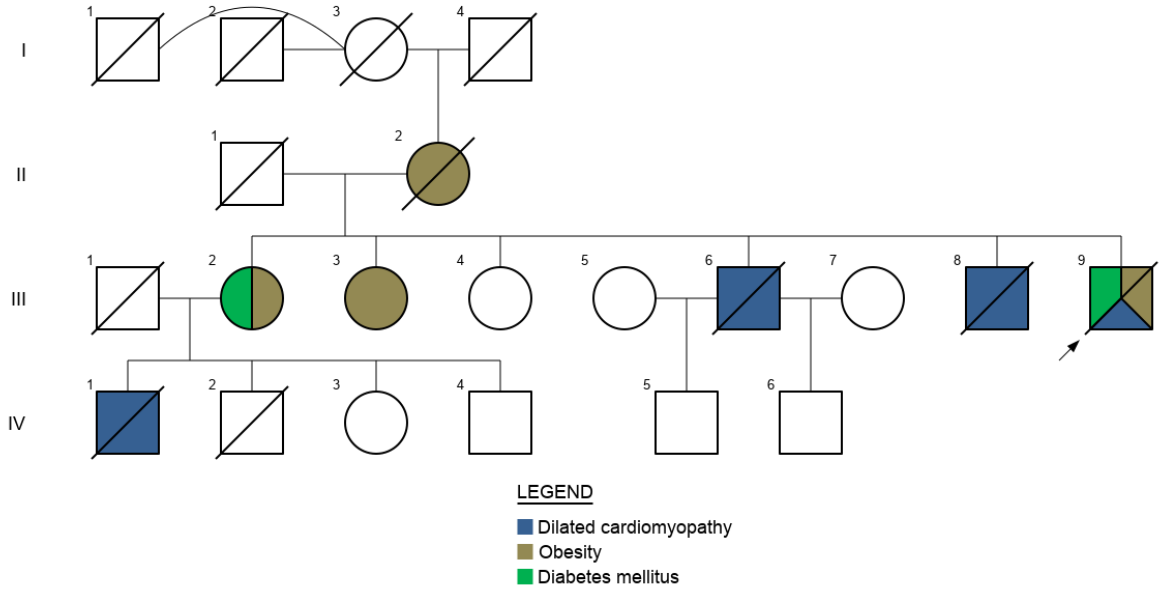


Figure 3.11: Pedigree indicating disease segregation in Family 3. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates DCM, green shading indicates diabetes mellitus and brown shading indicates obesity. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

The clinical features of the participants varied and may be complicated by the comorbidities diabetes mellitus and obesity that affect some individuals in the family (Figure 3.11). The proband (III:9) and his brother (III:6) and nephew (IV:1) were all diagnosed with severe DCM with marked systolic dysfunction and NYHA class II - IV heart failure (Table 3.3). The age of diagnosis in these individuals was between 25 and 31 years, and they all received heart transplants. Individual III:8 was diagnosed with milder DCM with a slightly dilated left ventricle and no heart failure at the age of 33 years. Despite this phenotypic variability, all affected individuals in the family have subsequently died due to their disease.

The proband had three sisters who were screened and found to be phenotypically unaffected, despite mild left ventricular enlargement in two of them (Table 3.3). The third sister (III:2) was the mother of affected individual IV:1. Although other offspring of the affected individuals were included in the study, their current disease status is unknown and they have been lost to follow-up.

Table 3.3: Clinical screening of Family 3

Individual	Age* (years)	Presenting symptom	ECG	Echocardiography	Clinical status
III:2	40	None	Sinus rhythm	LVEDD 4.48 cm, FS 44%	Normal
III:3	38	None	Sinus rhythm T wave inversion (V1)	LVEDD 5.9 cm, FS 30%, slightly dilated LA, LV	Mildly dilated, unaffected
III:4	37	None	Sinus rhythm T wave inversion (V1)	LVEDD 5.3 cm, FS 34%, slightly dilated LV	Borderline dilated, unaffected
III:6	**25**	**Dyspnoea NYHA class IV Palpitations Peripheral oedema**	**Sinus rhythm Anterior infarct pattern**	**LVEDD 7.7 cm, FS 11% LV dilated and poorly contracting, LA enlarged, mitral and tricuspid regurgitation**	**Severely impaired systolic function, DCM**
III:8	**33**	**Palpitations Chest pain**	**Sinus rhythm T wave inversion**	**LVEDD 5.4 cm, FS 24% No LV regional wall motion abnormalities, RV non-dilated**	**Mildly affected, DCM**
III:9	**31**	**Dyspnoea NYHA class IV Palpitations Peripheral oedema**	**Sinus rhythm Anterior infarct pattern**	**Not done prior to Batista surgery**	**Severely affected, DCM**
IV:1	**31**	**Dyspnoea NYHA class II**	**Sinus rhythm**	**LVEDD 7.9 cm, FS 14%, dilated poorly contracting LV**	**Severely impaired systolic function, DCM**
IV:5	18	None	Sinus rhythm	-	Unknown
IV:6	12	None	Sinus rhythm T wave inversion	LVEDD 5.26 cm, FS 39%, normal heart	Normal

* Age at the time of diagnosis and/or screening
DCM, dilated cardiomyopathy; ECG, electrocardiography; FS, fractional shortening; LA, left atrium; LV, left ventricle; LVEDD, left ventricular end-diastolic diameter; NYHA, New York Heart Association; RV, right ventricle

3.4.2 Genetic analysis of Family 3

Exome sequencing was conducted on individuals III:8, III:9 and IV:1, yielding a total of 69,110 variants in these individuals. A mean sequence coverage of 144x was achieved in the three sequencing reactions. Considering X-linked or AD inheritance patterns, 21 variants met the filtering criteria (Figure 3.12; Appendix M) and were analysed further. Many of the variants were excluded due to low pathogenicity prediction or a lack of functional relevance to the heart (Appendix M). However, other heterozygous variants that were not present in all affected individuals were considered, to allow for the analysis of more complex inheritance patterns such as genetic modifiers.

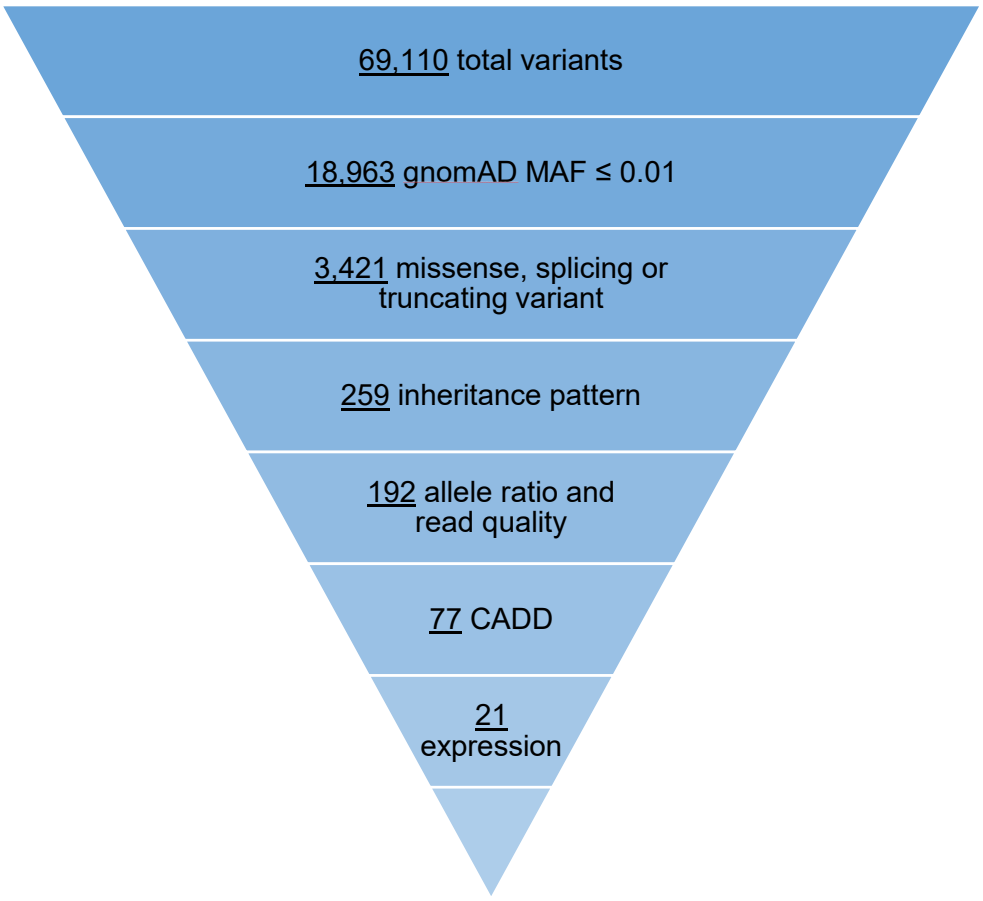


Figure 3.12: Filtering of exome sequencing data from Family 3. Exome sequencing produced a total of 69,110 variants, of which 21 met the filtering criteria. *CADD, Combined Annotation Dependent Depletion; gnomAD, Genome Aggregation Database; MAF, minor allele frequency*

3.4.2.1 Cardiomyopathy panel genes

On the cardiomyopathy panel, two *TTN* variants and a single *SYNE2* variant were found in all individuals (Table 3.4). The *TTN* variants were excluded due to a high MAF in ExAC and gnomAD sub-populations (East Asian population: MAF > 0.02 each), while the *SYNE2* variant was predicted benign by all prediction tools, and the affected amino acid was not conserved in any of the tested species (Table 3.4).

Table 3.4: Candidate variants in Family 3

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Cardiomyopathy panel – in all patients														
SYNE2	c.12001_12002inv	p.W4001Q	0	0	0	0	NS	B (0.99)	NS	B (0)	23.7	9.1	No	0/7
TTN	c.16529A>G	p.Y5510C	0.007	0.002240	0.002213	0.00025	NS	D (0.99)	NS	D (0.715)	21.0	67.7	Yes	3/7
TTN	c.76739C>T	p.T25580M	0.005	0.001661	0.001618	0.00917	NS	D (0.99)	NS	B (0.191)	22.2	67.7	Yes	5/7
Cardiomyopathy panel – other patterns														
DSC2	c.2642T>A	p.L881H	0	0	0	0	D (0.226)	D (0.99)	D (0)	D (0.999)	31.0	5.6	Yes	7/7
PKP2	c.2540T>C	p.L847P	0	0	0	0	D (0.139)	D (0.99)	D (0)	D (0.971)	26.8	84.1	Yes	7/7
X-linked variants														
FAM104B	c.331C>T	p.R111*	0	0.000461	0.000034	0	NS	B (0.98)	NS	NS	34.0	3.8	NA	NS
Other genes														
KCNK10	c.1052A>G	p.N351S	0	0	0	0	D (0.046)	D (0.99)	B (0.2)	D (0.997)	26.1	1.0	NA	7/7
TRPV1	c.860C>T	p.T287M	0.001	0.000504	0.000223	0.00086	D (0.035)	D (0.99)	D (0.1)	D (0.999)	24.4	4.0	Yes	6/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

When considering other inheritance patterns (i.e. variants not present in all three individuals), two variants of interest were identified on the cardiomyopathy panel. *PKP2* c.2540T>C (p.L847P) was found in individuals III:8 and III:9, while *DSC2* c.2642T>A (p.L881H) was detected in individual IV:1 only (Table 3.4). These variants were of interest because they were not detected in any population databases. *PKP2* p.L847P was identified in this family previously (Mbele 2008). The presence of *DSC2* p.L881H was confirmed (Figure 3.13) and segregation analysis of both variants performed (Figure 3.14). Because neither variant was present in all three affected individuals, they did not segregate with disease. The presence of *DSC2* p.L881H in individuals IV:1, IV:6 and III:7 (Figure 3.14) led to the conclusion that individuals III:1 and III:7 may be biologically related. A single individual (IV:6) tested positive for both *PKP2* p.L847P and *DSC2* p.L881H; however, he was lost to follow-up and could not be contacted during the investigation. *PKP2* p.L847P and *DSC2* p.L881H were also predicted deleterious by all the pathogenicity prediction tools and affected highly conserved protein regions: all species either matched the human leucine residue or had the conserved residue methionine at that point (Figure 3.15) (Zhang 2000).

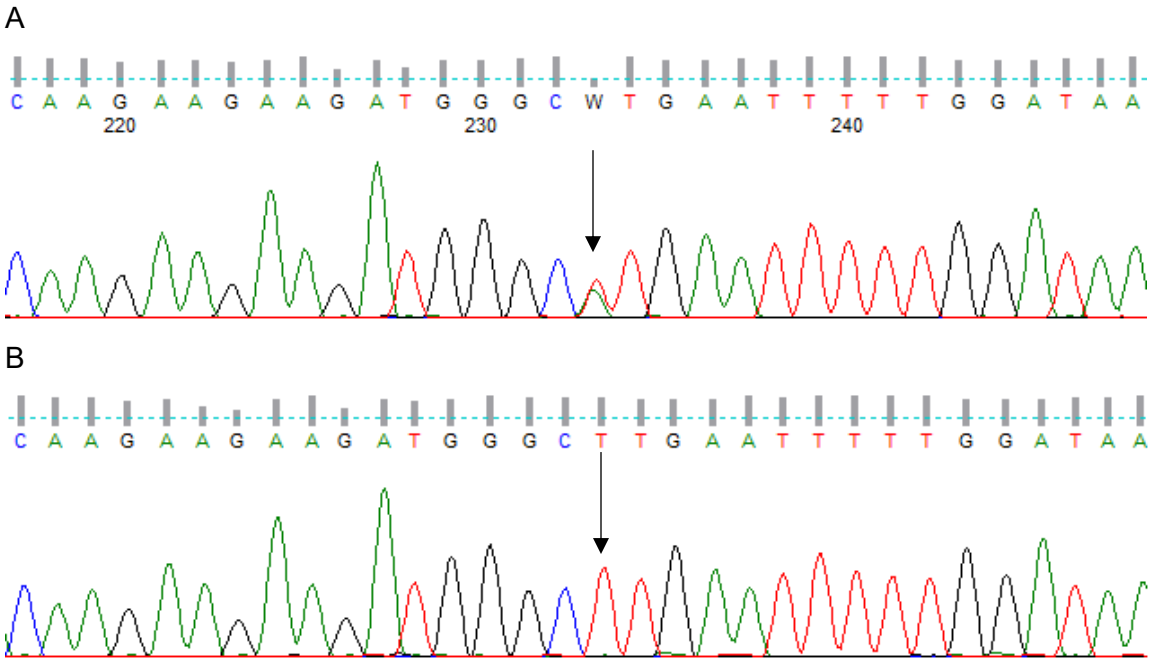


Figure 3.13: Validation of *DSC2* c.2642T>A identified in Family 3. (A) Sanger sequencing to confirm the presence of *DSC2* c.2642T>A in affected individual IV:1 (indicated with an arrow), (B) Sanger sequencing confirms the absence of the variant in unaffected individual III:2 (indicated with an arrow).

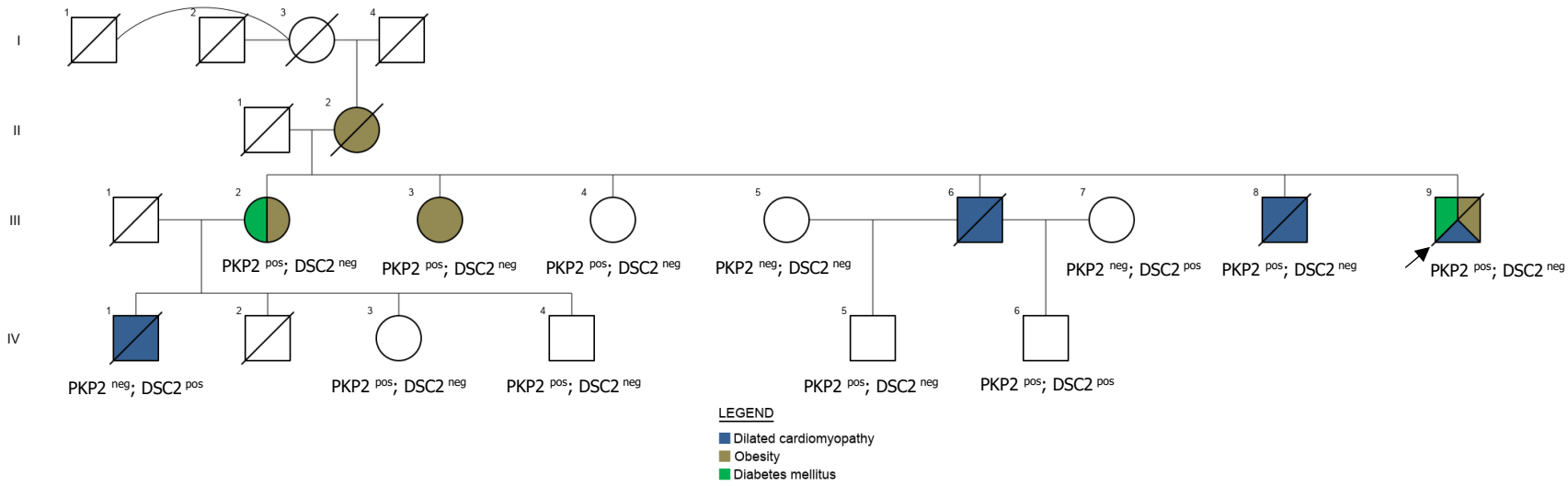


Figure 3.14: Pedigree indicating segregation of *DSC2* and *PKP2* variants in Family 3. Individuals marked 'PKP2 ^{pos}' were heterozygous for the c.2540T>C variant and those marked 'PKP2 ^{neg}' tested negative for it, while individuals marked 'DSC2 ^{pos}' were heterozygous for the c.2642T>A variant and those marked 'DSC2 ^{neg}' tested negative for it. Genotypes in brackets were inferred based on the offspring and partners' genotypes. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates DCM, green shading indicates diabetes mellitus and brown shading indicates obesity. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

A

H. sapiens	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	E	L	H	H	A	Y	K
M. mulatta	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	E	L	H	H	A	Y	K
P. troglodites	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	E	L	H	H	A	Y	K
F. catus	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	E	L	H	H	A	Y	K
M. musculus	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	E	L	H	H	A	Y	K
G. gallus	A	S	K	A	A	S	V	L	L	Y	S	L	W	A	H	T	D	L	S	H	A	Y	K
D. rerio	A	G	Q	A	A	C	V	L	L	H	T	L	W	R	H	S	E	L	H	S	S	F	K
X. tropicalis	T	G	K	A	A	S	I	V	L	Y	S	M	W	A	H	Q	D	L	H	S	T	Y	K

B

H. sapiens	G	C	C	S	E	R	Q	E	E	D	G	L	E	F	L	D	N	L	E	P	K	F	R
M. mulatta	G	C	C	S	E	R	Q	E	E	D	G	L	E	F	L	D	N	L	E	P	K	F	R
P. troglodites	G	C	C	S	E	R	Q	E	E	D	G	L	E	F	L	D	N	L	E	P	K	F	K
F. catus	G	C	C	S	E	R	Q	E	E	D	G	L	E	F	L	D	N	L	E	P	K	F	R
M. musculus	G	C	C	S	D	L	Q	E	E	D	G	L	E	F	L	D	H	L	E	P	K	F	R
G. gallus	G	C	C	S	D	Q	H	E	E	E	A	L	D	F	L	D	Q	L	E	P	K	F	R
D. rerio	G	C	C	S	I	L	G	E	Q	E	S	M	E	F	L	N	T	L	G	P	K	F	R
X. tropicalis	G	C	C	S	D	F	R	D	E	D	K	M	D	F	L	N	H	L	E	P	K	F	R

Figure 3.15: Residue conservation of the *DSC2* and *PKP2* missense variants in Family 3. Multiple species conservation alignment of (A) the p.L547 residue of PKP2 (shaded in blue), and (B) the p.L881 residue of DSC2 (shaded in blue).

3.4.2.2 Other genes

Two variants in genes outside the cardiomyopathy panel were also considered due to literature supporting their potential role in cardiomyopathy phenotypes (Table 3.4). The variants *TRPV1* c.860C>T (p.T287M) and *KCNK10* c.1052A>G (p.N351S) were selected as candidates. *TRPV1* has been implicated in cardiac protection from ischaemia/reperfusion myocardial injury (Gao et al. 2015, Jiang et al. 2018, Sexton et al. 2007, Sun et al. 2014, Zhong and Wang 2007), sepsis-induced dysfunction (Chen et al. 2018), overload-induced hypertrophy (Wang et al. 2014) and diabetes mellitus (Li et al. 2018, Wei et al. 2009, Zheng et al. 2015) in mouse and/or rat models. Expression of the gene has also been shown to be reduced in the hearts of rats with DCM (Smith et al. 2005). Given the potential role of *TRPV1* in DCM and protection of the heart from diabetic injury, the gene was of interest as diabetes mellitus is a comorbidity in this family. However, the *TRPV1* variant c.860C>T was not validated in the affected individuals (Figure 3.16). Interestingly though, the variant was found in the son of individual III:3 (not shown on pedigree) (Figure 3.16), in a homozygous state, meaning that this variant may be common in the South African population.

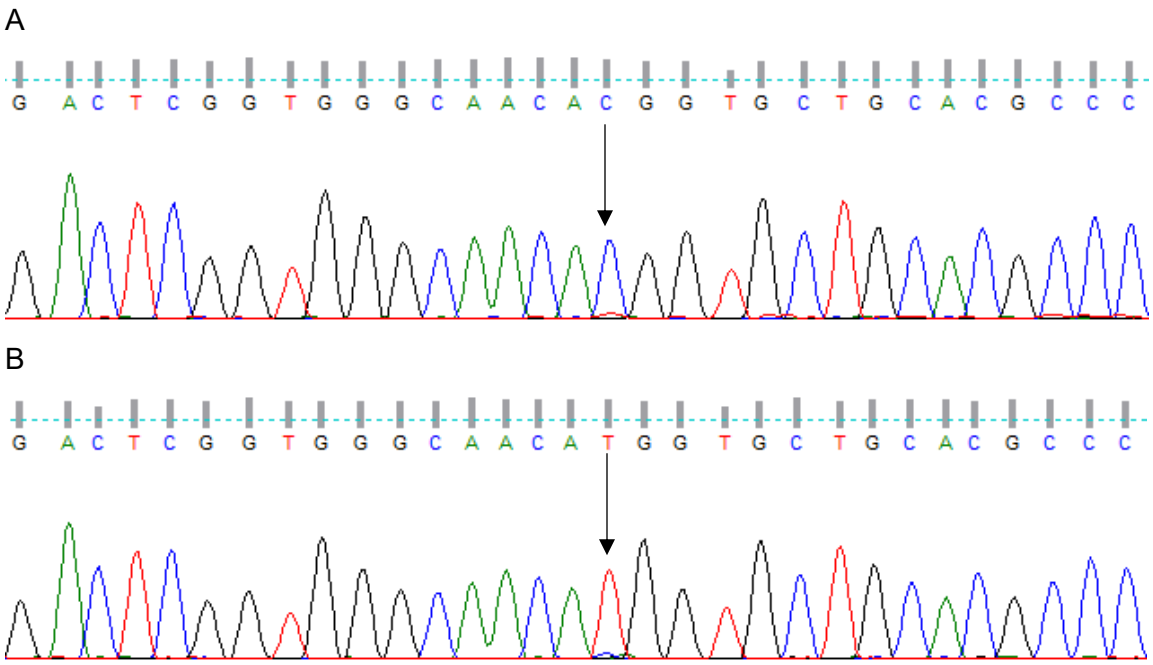


Figure 3.16: Lack of validation of *TRPV1* c.860C>T identified in Family 3. (A) Sanger sequencing shows the absence of *TRPV1* c.860C>T in affected individual IV:1 (indicated with an arrow), (B) Sanger sequencing shows the homozygous variant in an unaffected family member (indicated with an arrow).

KCNK10 encodes a stretch-activated two-pore-domain potassium channel, a class of proteins involved in the regulation of cellular excitability. Although not much is known about its function, *KCNK10* expression has been demonstrated in the hearts of mice, zebrafish and humans, as well as other tissues (Gierten et al. 2012, Schmidt et al. 2017, Staudacher et al. 2011), and was recently reported as the second-most abundant stretch-activated two-pore-domain potassium channel in the heart, with predominant expression in the atria (Schmidt et al. 2017). Expression of *KCNK10* was increased almost five-fold in patients with severe heart failure (Schmidt et al. 2017), while antiarrhythmic drugs and beta-blockers have been shown to inhibit KCNK10 protein function in human cells and zebrafish larvae (Gierten et al. 2012, Kisselbach et al. 2014). The *KCNK10* variant c.1052A>G (p.N351S) was absent in all population databases checked but was confirmed by Sanger sequencing to be present in the affected individuals in Family 3 (Figure 3.17). It may therefore be a private mutation occurring only within this family. Segregation analysis indicated the variant was present in all affected patients as well as some unaffected individuals in the family (Appendix N). The missense variant occurred at an amino acid residue that is highly conserved, indicating possible functional importance (Appendix N).

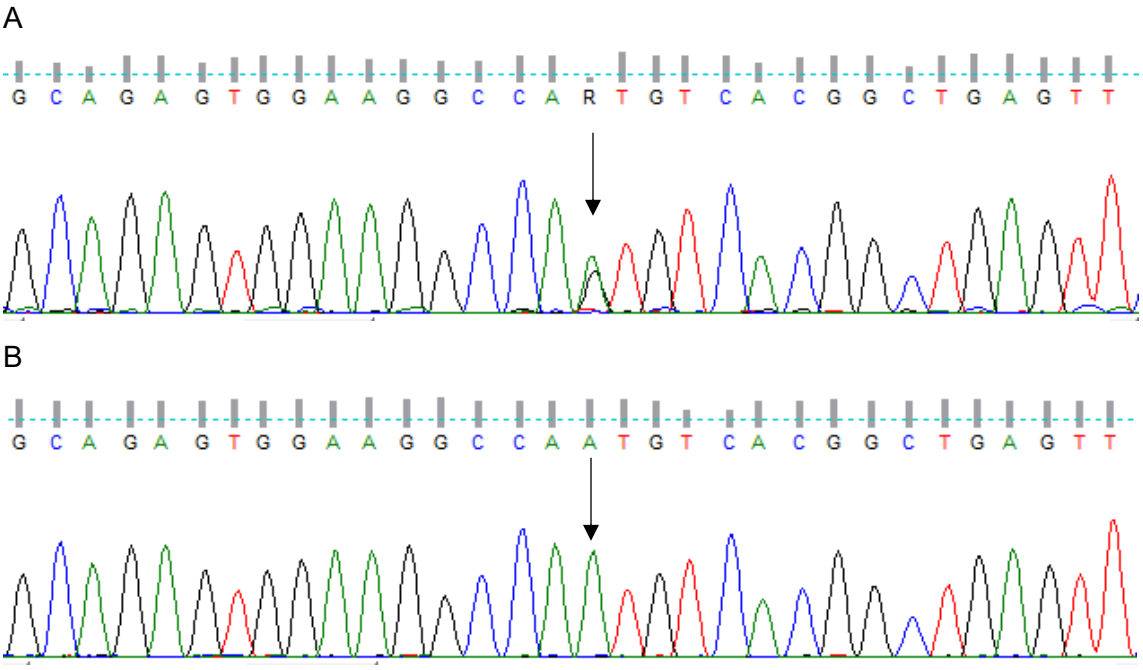


Figure 3.17: Validation of *KCNK10* c.1052A>G identified in Family 3. (A) Sanger sequencing to confirm the presence of *KCNK10* c.1052A>G in affected individual IV:1 (indicated with an arrow), (B) Sanger sequencing confirms the absence of the variant in unaffected individual IV:6 (indicated with an arrow).

3.4.2.3 X-linked variants

A single X-linked variant met the filtering criteria and was analysed (Table 3.4). The truncating variant *FAM104B* c.331C>T (p.R111*) was sufficiently rare to be considered disease-causing; however, the variant was predicted benign, probably because it only affects the terminal six amino acids of the protein. This variant was therefore not considered further. Because only males are known to be affected by disease in this family, X-linked inheritance is still a highly likely option.

3.4.3 Discussion: PKP2 *and* DSC2 *mutations as causes of DCM*

PKP2 and *DSC2* both encode components of the cardiac desmosome. The desmosome mediates cell-cell adhesion and electrical coupling in cardiomyocytes, and mutations in these genes have traditionally been associated with ACM. Up to 60% of ACM patients carry mutations in desmosomal genes (Ohno 2016).

The pathological basis of DCM is substantially more heterogeneous than ACM and may be at least partially attributed to desmosomal dysfunction. Several desmosomal mutations have been described in DCM patients who do not fulfil the Task Force criteria for ACM, including mutations in *PKP2* and *DSC2* (Elliott et al. 2010, García-Pavía et al. 2011, Klauke et al. 2017, Marston et al. 2015). In particular, a study of end-stage DCM heart transplant recipients reported that up to 13% of cases may harbour ACM-causing or novel desmosomal gene mutations (García-Pavía et al. 2011), while the novel mutation *PKP2* c.2035C>T (p.H697Y) was found in a large Turkish family with a history of DCM and heart transplant (Klauke et al. 2017). In the latter investigation, the *PKP2* variant alone could not explain the DCM phenotype, and the authors suggest that additional genetic factors may be involved.

These studies mirror the genetic findings in Family 3, in which novel missense mutations in *PKP2* and *DSC2* were found. Although neither variant independently tracked with disease, all affected individuals carried either mutation. Both variants were absent in all population databases checked, but *PKP2* p.L847P has been reported in another South African patient with ACM (Watkins et al. 2009). When it was first discovered in this family, *PKP2* p.L847P was considered benign for DCM although pathogenic for ACM (Mbele 2008). However, the rarity of the variant combined with high sequence conservation and the presence of *DSC2* p.L881H, a highly similar variant, in the family, means the significance of *PKP2* p.L847P in this family should be reconsidered. With the exception of segregation with disease, both variants meet all criteria for pathogenicity. It is therefore possible that *PKP2* p.L847P and *DSC2* p.L881H do contribute to the DCM phenotype in this family. Of interest is individual IV:6, who was found to carry both mutations. Phenotypic analysis of this individual should indicate if these variants are indeed disease-causing; unfortunately, this participant is currently lost to follow-up. It is also probable that other factors are involved in this family, such as comorbidities that may worsen the cardiac phenotype, or other genetic mutations.

Other genetic factors such as *KCNK10* p.N351S and *TRPV1* p.T287M were considered for their disease-causing potential in this family. Both genes were selected as candidates due to prior implications with cardiac function. However, the most likely inheritance pattern that would explain the disease segregation in this family is X-linked recessive inheritance. Because no candidates were identified on the X chromosome, any other inheritance pattern that is investigated will need to include incomplete penetrance. Therefore, while *KCNK10* may be an interesting disease candidate in this family, its role in DCM pathogenesis cannot be elucidated based on segregation

analysis alone. The role of potassium channels such as *KCNK10* in DCM in the absence of arrhythmic phenotypes is unclear at this stage. Due to the presence of two mutations in established cardiomyopathy genes in this family, the variation in *KCNK10* may best be considered as a potential modifier of disease.

3.4.4 Assessment of the pathogenicity of PKP2 *c.2540C>T and* DSC2 *c.2642T>A*

In Family 1, the difficulty associated with the interpretation of missense variants in small families was discussed (Section 3.2.4). Similar limitations were observed in Family 3. Although more relatives were available for genetic analysis, the family only contained three affected patients, limiting the ability of segregation analysis alone to provide clarity regarding pathogenicity. Two ultra-rare mutations in the cardiomyopathy genes *PKP2* and *DSC2* were identified in this family, although neither segregated entirely with the disease phenotype (Figure 3.14). While these variants did not occur in mutational hot spots for cardiomyopathy, the close proximity of these variants to other known or likely disease-causing mutations should be noted (Table 3.6). These include truncating and missense mutations within the affected protein domains in Family 3, meaning that disruption of these protein domains may be sufficient to cause disease. In particular, the *PKP2* c.2540T>C variant has been reported in two patients with ACM (Tayal et al. 2017, Watkins et al. 2009). However, nonsynonymous VUSs have been reported in these regions of *DSC2* and *PKP2* as well (data not shown); this indicates that interpretation of mutations in the affected domains of these proteins may be challenging.

The absence of the *PKP2* and *DSC2* variants in large population studies, alone, is insufficient proof for their potential to cause disease. However, combined with computational predictions indicating the deleteriousness of the mutations and the evolutionary conservation of affected protein residues, there is moderate and supporting evidence that these mutations could contribute to the disease phenotype in this family. Following the ACMG criteria, these variants should be classified as VUSs at this stage. The possibility that other genetic factors (such as *KCNK10* or other undetected variants, for example in intergenic or intronic regions) may be involved must be considered.

Table 3.5: Reported mutations in the vicinity of the *DSC2* and *PKP2* variants in Family 3

Mutation	Protein domain(s)	ClinVar status	Source
Reported *DSC2* variants			
c.2186del (p.729fs*4)	Desmosomal cadherin	Pathogenic	VCV000662906
c.2200C>T (p.Q734*)	Desmosomal cadherin	Pathogenic	VCV000568186
c.2463C>A (p.Y821*)	Desmosomal cadherin	Pathogenic	VCV000575762
c.2487del (p.F829fs*27)		Pathogenic	VCV000199779
c.2548delinsTT (p.A850fs*9)	Desmosomal cadherin, Catenin-binding domain	Likely pathogenic	VCV000199786
c.2642T>A (p.L881H)	Desmosomal cadherin, Catenin-binding domain	None	Family 3
Reported *PKP2* variants			
c.2443_8delinsGAAA (p.N815fs*11)	Armadillo repeat 5	Likely pathogenic	VCV000188663
c.2377del (p.S837fs*94)	Armadillo repeat 5	Pathogenic	VCV000177995
c.2522T>C (p.S841F)	Armadillo repeat 5	Likely pathogenic	VCV000201966
c.2531T>C (p.L844P)	Armadillo repeat 5	Likely pathogenic	VCV000201967
c.2540T>C (p.L847P)	Armadillo repeat 5	Uncertain significance	Family 3; VCV000419976
c.2554del (p.E852fs*79)	None	Pathogenic	VCV000464426

fs, frameshift; VCV, ClinVar accession number

3.5 Chapter summary

Although the allelic heterogeneity of familial cardiomyopathy has been characterised in international cohorts, this aspect of the disease has, to the best of our knowledge, not been reported in South Africa to date. In this chapter, three families are presented in which exome sequencing was used to identify potential causes of disease, and novel mutations were found in known cardiomyopathy genes. In particular, the finding of a pathogenic truncating *GLA* mutation in Family 2 resulted in the genetic re-diagnosis of those individuals with Fabry disease, an HCM phenocopy. Fabry disease and *GLA* mutations are frequently described in international HCM cohorts, representing between 1% and 5% of the total burden of HCM (Adalsteinsdottir et al. 2014, Cecchi et al. 2017, Monserrat et al. 2007, Seo et al. 2016). However, this is the first report of a Fabry disease-causing mutation in South African HCM patients. This indicates that, although rare, Fabry disease may contribute to HCM phenotypes in South African patients, similar to observations in other populations.

In Families 1 and 3, both of whom had DCM, missense mutations were identified in known ACM and HCM genes. These families may therefore illustrate the genetic overlap that is characteristic of heritable cardiomyopathy. The presence of both *MYH7* and desmosomal gene mutations in

DCM patients has been reported elsewhere (Abdallah et al. 2019, García-Pavía et al. 2011, Sousa et al. 2019, van der Linde et al. 2017). However, like the *GLA* mutation in Family 2, this is the first description of such genotype-phenotype correlations in South African cardiomyopathy patients. To date, *PKP2* mutations in South African patients have been limited to individuals diagnosed with ACM (Watkins et al. 2009), while *MYH7* mutations have only been reported in HCM patients (Moolman, Brink, and Corfield 1993, Ntusi et al. 2016, Posen et al. 1995). Based on the genetic analyses, the *MYH7* c.4394C>T mutation was classed as likely pathogenic, meaning that there is a high likelihood that this is the disease-causing mutation in Family 1 (Richards et al. 2015). The desmosomal gene mutations in Family 3, in contrast, are VUSs even though they fulfil many of the criteria for pathogenicity. Further phenotypic characterisation of mutation carriers in this family may help in the resolution of the genetic uncertainty surrounding these variants, but the possibility that other genetic variation may play a role cannot be ignored. Although genes outside the cardiac panel were considered in all cases, the analysis of this family in particular illustrates how exome sequencing data can be used to detect candidate cardiomyopathy genes. The variant *KCNK10* c.1052A>G was identified as a possible modifier of the disease phenotype in Family 3; nevertheless, the greatest pathogenic potential in this family, based on current evidence, still resides in the known cardiomyopathy gene mutations *DSC2* c.2642T>A and *PKP2* c.2540T>C.

The interrogation of exome sequencing data for potential new cardiomyopathy genes is most suitable in families where no mutations are found in the cardiac panel. The identification of candidate disease genes forms the basis of the next chapter, in which two such families are presented and are investigated using exome sequencing.

Chapter 4 Variants in novel cardiomyopathy genes as the possible cause of disease in South African cardiomyopathy patients

4.1 Introduction

NGS is used to screen numerous genes simultaneously for genetic variation. This technology can be used in cardiomyopathy research to identify mutations within known cardiomyopathy genes, as described earlier (Chapter 3). However, exome sequencing also allows the exploration of genes that have not yet been associated with cardiomyopathy; this is particularly useful in patients in whom investigation of known disease genes has yielded no results. In this section, two such families are described.

4.2 Family 4 (ACM 142)

4.2.1 Clinical history of Family 4

Family 4 was identified at GSH in 2015 (Figure 4.1). The proband, individual IV:2, was first diagnosed with definite ACM at the age of 50 years after being hospitalised following a traffic accident and suffering an in-hospital cardiac arrest (survived). He presented with classical features of arrhythmogenic right ventricular cardiomyopathy (ARVC), including major structural abnormalities of the right ventricle wall, repolarisation abnormalities, late potentials and recurrent ventricular arrhythmias, as well as marked left ventricular scarring on CMR. Gene panel testing conducted in the UK did not identify any pathogenic mutations in this patient.

Subsequently, his sister (IV:4) suffered a SCD at the age of 43 years. Post-mortem examination of her heart found non-specific features of cardiomyopathy, including mild right ventricular dilatation with some fatty infiltration and interstitial and focal areas of fibrosis involving the ventricular walls and septum (similar to the proband). She was found to fulfil criteria for early-stage ACM even though the macroscopic findings were not sufficient to confidently support a diagnosis of ARVC (morphometric analysis: adipose tissue 17.5%, fibrosis 10.2%, myocardium 72.3%). Other first-degree relatives were then screened at GSH. Due to the positive family history of disease, clinical abnormalities on signal-averaged ECG and 24 hr-Holter, and patchy/diffuse left ventricular fibrosis on CMR, individual IV:3 (46 years) and the proband's mother (III:2) (75 years), were later classified as affected by the diagnostic criteria.

The proband's father (III:1), older brother (IV:1) and uncle (III:4) were clinically evaluated and found to be clinically unaffected with no late gadolinium enhancement on CMR. This, together with a history of suspected cardiac events (e.g. SCD) in several of the mother's family members, suggests inheritance of the disease from the maternal side of the family. While ischaemic heart disease could not be convincingly excluded as a contributing factor in some of the sudden deaths in this family, ischaemic heart disease was excluded in all living affected family members with a cardiomyopathy phenotype.

Family 4

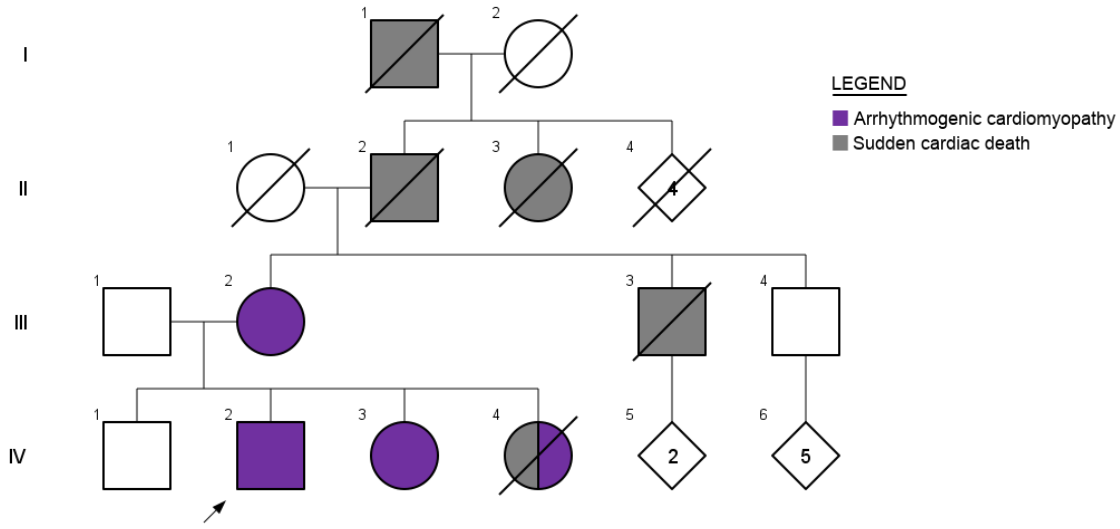


Figure 4.1: Pedigree indicating disease segregation in Family 4. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

4.2.2 Genetic analysis of Family 4

Exome sequencing was conducted on four affected family members (III:2, IV:2, IV:3, IV:4). In total, 51,341 variants were found in all four affected individuals and after applying additional filters and considering AD inheritance (Figure 4.2), 13 variants remained, occurring in 13 genes (Table 4.1). None of the variants occurred in a known cardiomyopathy gene, meaning that the cause of

this heritable condition may be due to a mutation in another gene or more complex inheritance patterns, including potential genetic modifiers.

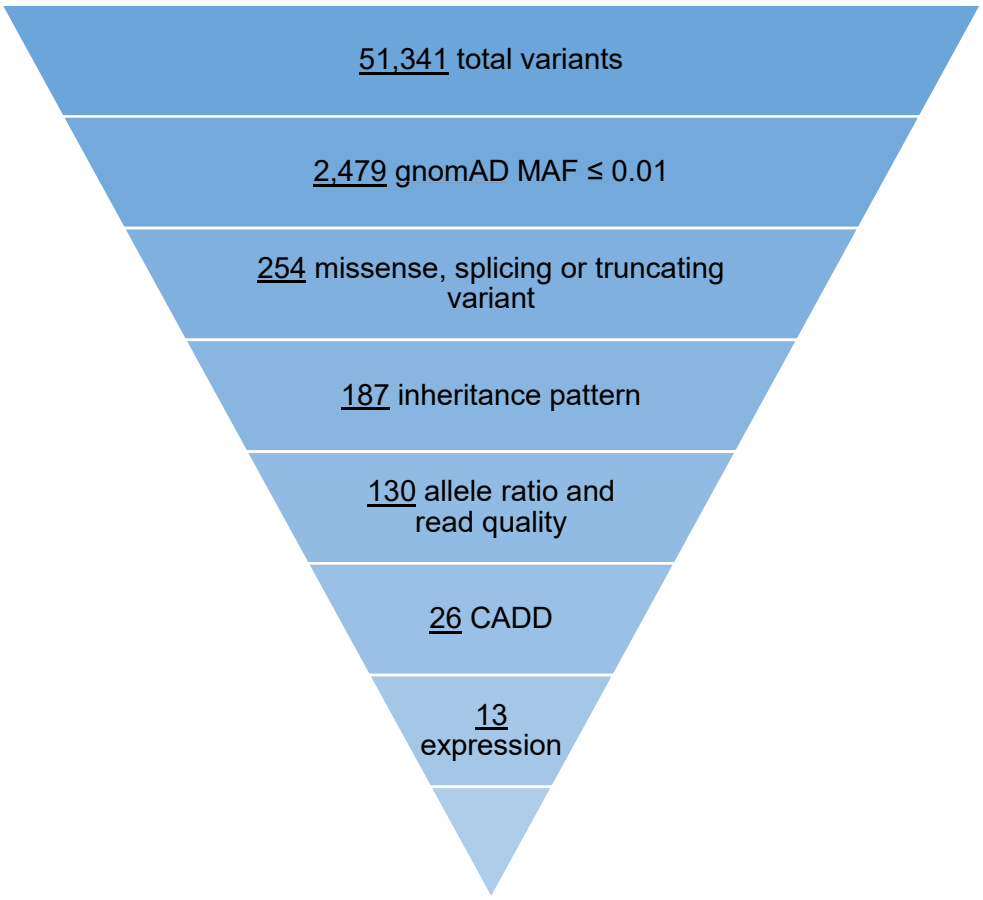


Figure 4.2: Filtering of exome sequencing data from Family 4. Exome sequencing produced a total of 51,341 variants, of which 13 met the filtering criteria. *CADD, Combined Annotation Dependent Depletion; gnomAD, Genome Aggregation Database; MAF, minor allele frequency*

Table 4.1: Candidate variants in Family 4

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Other genes														
ADAL	c.196T>G	p.C66G	0	0.000008	0.000008	0	D (0.448)	D (0.99)	D (0.01)	D (0.607)	29.4	6.2	No	6/7
ATG2A	c.2036G>T	p.R679L	0	0.000093	0.000051	0	B (0.009)	D (0.96)	D (0.02)	D (0.858)	23.3	6.5	NA	4/7
DHTKD1	c.1079T>C	p.V360A	0.001	0.003534	0.003349	0.00438	D (0.468)	D (0.83)	D (0)	D (0.744)	24.1	7.3	No	4/7
ENGASE	c.1190C>T	p.S397F	0	0.000351	0.000314	0.00031	D (0.098)	D (0.99)	D (0)	D (0.839)	26.3	11.6	NA	4/7
FPGS	c.253C>T	p.R85W	0.001	0.002434	0.002596	0.00246	D (0.123)	D (0.99)	D (0)	D (1.000)	27.5	13.5	No	5/7
GOLGA5	c.842G>A	p.R281Q	0.001	0.002675	0.002912	0.00400	D (0.028)	D (0.99)	D (0)	D (0.999)	31.0	10.7	No	7/7
HEXB	c.1250C>T	p.P417L	0.001	0.000659	0.000588	0.00046	D (0.184)	D (0.99)	D (0.02)	D (0.893)	22.2	27.6	No	5/7
NDUFB1	c.257G>C	p.R86P	0	0.000025	0.000047	0.00008	D (0.125)	D (0.92)	D (0.05)	D (1.000)	26.5	119.0	NA	4/7
POLG	c.2492A>G	p.Y831C	0.002	0.006277	0.007011	0.00785	NS	D (0.99)	B (0.17)	D (0.995)	23.3	23.7	Yes	7/7
RAD54L2	c.2845G>C	p.E949Q	0	0.000757	0.000483	0.00077	D (0.085)	D (0.99)	B (0.06)	D (0.979)	27.1	3.1	No	5/7
SRPRA	c.1448G>A	p.G483D	0.002	0.005052	0.005423	0.00900	NS	D (0.99)	B (0.08)	B (0.383)	28.0	41.8	NA	7/7
SYMPK	c.1844G>A	p.R615H	0.003	0.007678	0.007384	0.00769	NS	D (0.99)	D (0.05)	B (0.362)	23.7	10.5	No	6/7
USP36	c.598C>A	p.H200N	0.003	0.005148	0.004762	0.00646	NS	D (0.99)	B (0.11)	B (0.192)	23.2	4.5	No	5/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

The genes with the highest expression in the heart were *HEXB*, *NDUFB1*, *POLG* and *SRPR* (Table 4.1). Only the variants in *HEXB*, *NDUFB1* and *POLG* were predicted deleterious by MutationTaster, SIFT and/or Polyphen, and were investigated further. *HEXB* c.1250C>T (p.P417L) and *NDUFB1* c.257G>C (p.R86P) were both validated by Sanger sequencing, but these variants were also present in the unaffected older brother of the proband (IV:1) and therefore excluded (Appendix O). The *POLG* c.2492A>G (p.Y831C) variant (Figures 4.3 and 4.4), however, was present in all affected individuals within this family but was absent in the proband's older brother and uncle, both of whom were confirmed clinically to have no signs of disease, on both ECG and CMR (Figure 4.5). This missense variant affected a highly conserved protein residue, as all species carry either a tyrosine or a phenylalanine residue at that position (Figure 4.4B); both are large nonpolar amino acids (Zhang 2000). In addition, *POLG* was the only gene with a function that impacts the heart, with documented cardiac phenotypes in mouse models (Table 4.1).

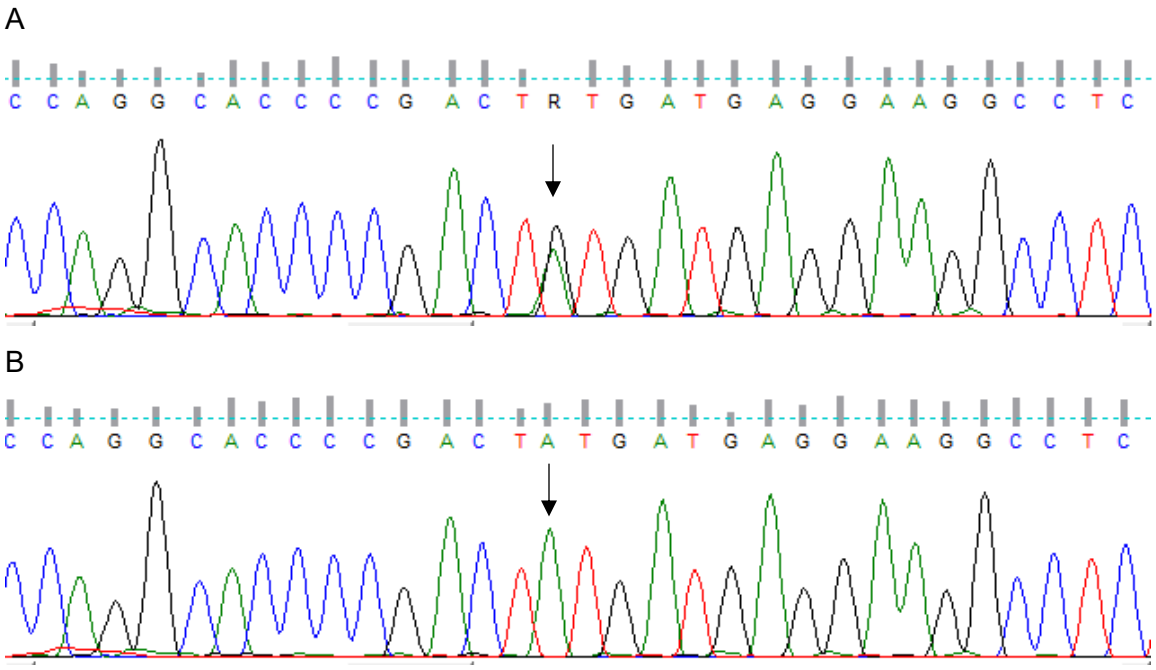


Figure 4.3: Validation of *POLG* c.2492A>G identified in Family 4. (A) Sanger sequencing to confirm the presence of *POLG* c.2492A>G in affected individual IV:4 (indicated with an arrow), (B) Sanger sequencing confirms the absence of the variant in unaffected individual IV:1 (indicated with an arrow).

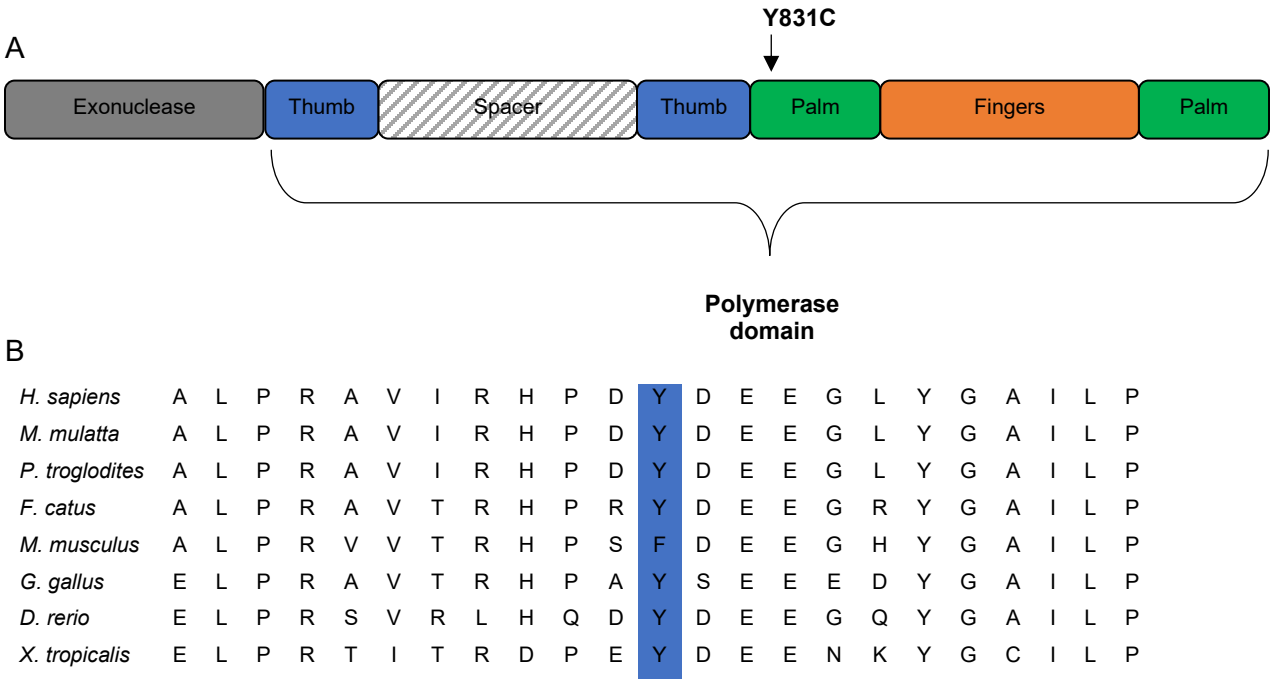


Figure 4.4: *POLG* c.2492A>G identified in Family 4. (A) Protein structure of POLG with position of the p.Y831C mutation marked with an arrow, (B) Multiple species conservation alignment of the p.Y831 residue of POLG (shaded in blue)

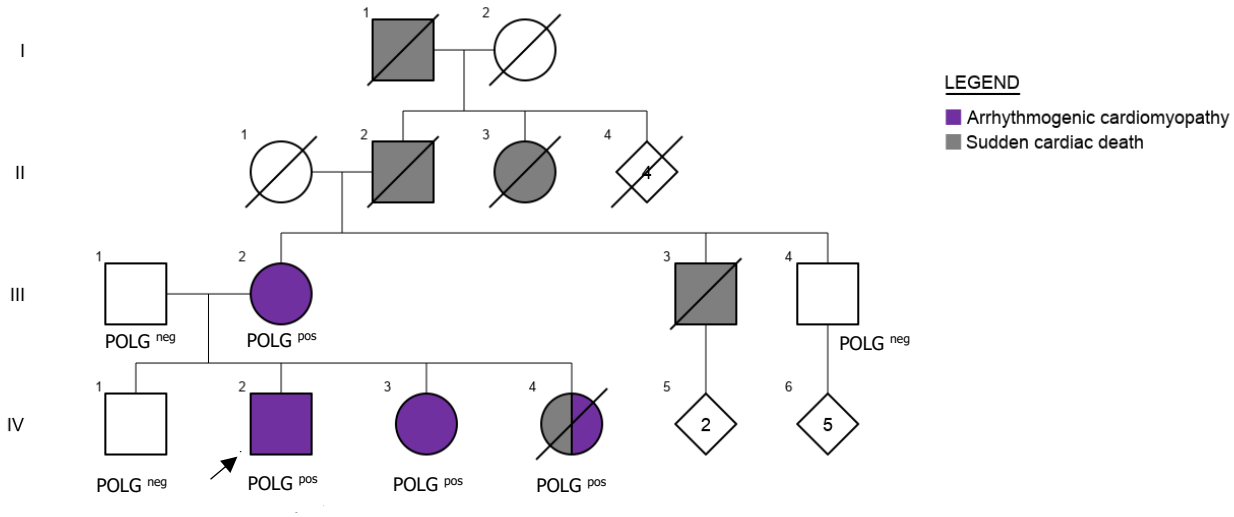


Figure 4.5: Pedigree indicating segregation of *POLG* c.2492A>G in Family 4. Individuals marked 'POLG ^{pos}' were heterozygous for the variant and those marked 'POLG ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

4.2.2.1 Possible genetic modifiers

Due to the phenotypic variability in Family 4, additional genetic factors were considered in the proband (the only family member with 'classical' features of ACM) and his sister (IV:4; the only family member to suffer an SCD). Seven potential genetic modifiers were identified in the siblings (Table 4.2), both in known cardiomyopathy genes and putative genes that were characterised using the same criteria as before (Section 2.4.4).

In the proband, three mutations were identified in the cardiomyopathy panel, including two truncating mutations in *KCNH2* and *KCNN3* that were false positives of exome sequencing. *TTN* mutations typically cause DCM, although overlaps with ACM and other arrhythmic phenotypes have been described (Brun et al. 2014, Corden et al. 2019, König et al. 2017, Kryczka et al. 2018, Lyu, Chen, and Xu 2018, Nielsen et al. 2018, Tayal et al. 2017, Taylor et al. 2011, Verdonschot et al. 2018). The missense mutation *TTN* c.24083G>C (p.G8028A) was identified in the proband and it was confirmed upon Sanger sequencing (Figure 4.6). However, this variant was also present in the older, unaffected brother of the proband (IV:1) (Appendix O).

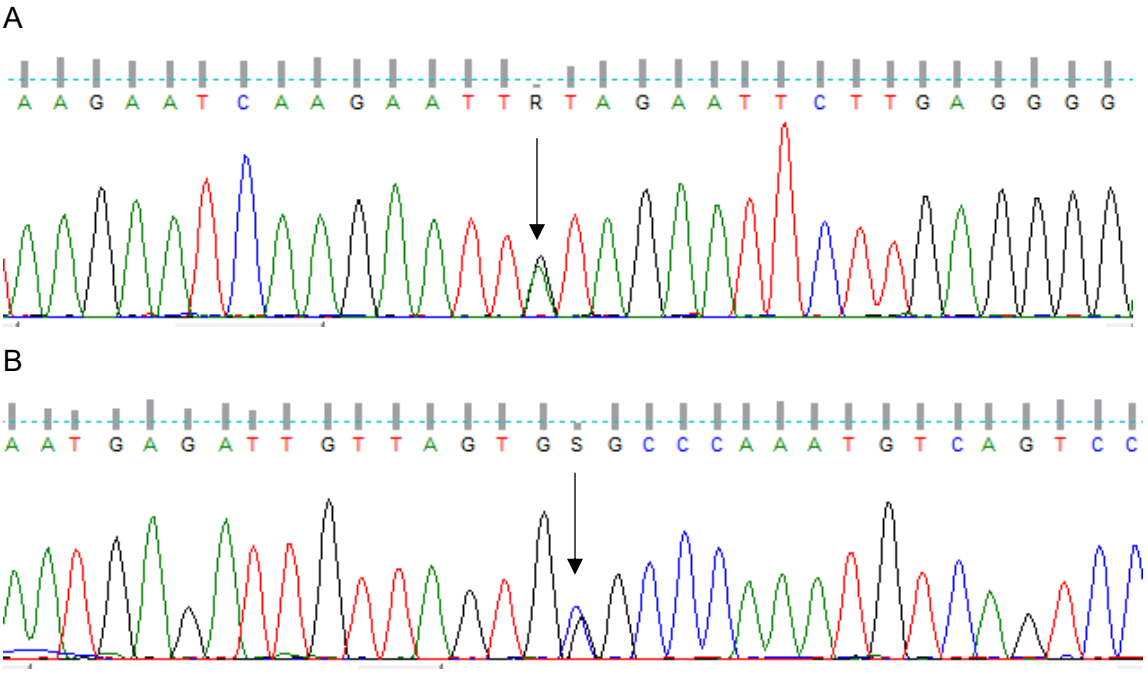


Figure 4.6: Validation of *TTN* variants identified in Family 4. (A) Sanger sequencing to confirm the presence of *TTN* c.3899A>G in affected individual IV:4 (indicated with an arrow), (B) Sanger sequencing to confirm the presence of *TTN* c.24083G>C in affected individual IV:2 (indicated with an arrow).

Table 4.2: Candidate genetic modifiers in Family 4

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Present in proband (IV:2)														
KCNH2	c.2805delC	p.S936P*38	0	0	0	0	NS	D (1)	NS	NS	35.0	26.6	Yes	3/7
KCNN3	c.202C>T	p.Q68*	0	0	0	0	NS	D (1)	NS	NS	36.0	0.4	No	4/7
TTN	c.24083G>C	p.G8028A	0	0	0.000032	0	D (0.049)	D (0.99)	NS	D (0.935)	23.1	66.8	Yes	1/7
Present in SCD individual (IV:4)														
PXDNL	c.2218C>T	p.P740S	0	0.006198	0.005093	0.00508	NS	D (0.89)	D (0.01)	D (0.995)	23.0	15.6	NA	2/7
SORBS2	c.322T>C	p.F108L	0.004	0.008899	0.009501	0.00754	NS	D (0.99)	B (0.49)	B (0.448)	24.7	67.1	Yes	7/7
TTN	c.3899A>G	p.Y1300C	0.0002	0.000017	0.000020	0	D (0.028)	D (0.99)	NS	D (1)	23.0	66.8	Yes	0/7
ZSCAN10	c.1253insC	p.T419D*168	0	0	0	0	NS	D (1)	NS	NS	22.5	0.01	Yes	3/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SCD, sudden cardiac death; SIFT, Sorting Intolerant From Tolerant

Four potential modifiers were present in individual IV:4 (Table 4.2). Like the proband, a missense *TTN* mutation was identified in this sibling. The other candidates were not in the cardiomyopathy panel but were prioritised based on their reported function or expression: *PXDNL*, a peroxidase-like protein that is expressed only in cardiomyocytes and localises to the cell-cell junctions (Péterfi et al. 2014); *SORBS2*, an actin-binding Z-disk protein and cardiac regulator which is highly expressed in the heart and may play a role in myofibril structure and functioning as well as cardiac hypertrophy and diabetic cardiomyopathy (Bang et al. 2014, Fu et al. 2018, Rönty et al. 2005, Sanger et al. 2010, Shao, Chen, and Zheng 2018, Wang, Golemis, and Kruh 1997); and *ZSCAN10*, a transcription factor, knockout of which did not have significant effects on mouse heart functioning (Kraus et al. 2014). *ZSCAN10* was therefore not investigated further. *SORBS2* c.322T>C was found only in the SCD individual and her mother (Figure 4.7; Appendix P). While the variants in *PXDNL*, *SORBS2* and *TTN* variants were all confirmed by Sanger sequencing to be real (Figures 4.6 and 4.8), *PXDNL* c.2218C>T and *TTN* c.3899A>G were both identified in the unaffected father and in the unaffected older brother, respectively (Appendix O).

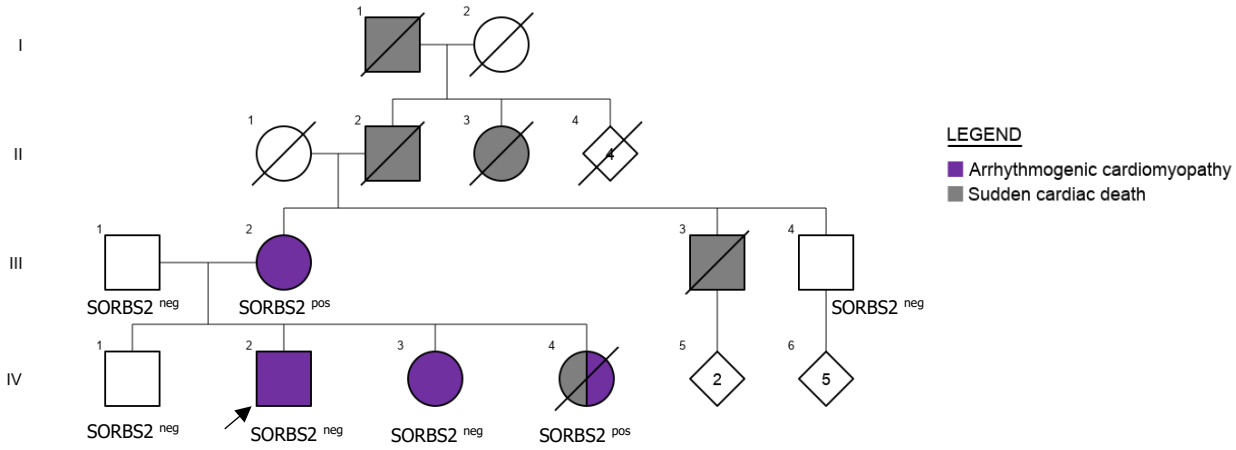


Figure 4.7: Pedigree indicating segregation of *SORBS2* c.322T>C in Family 4. Individuals marked 'SORBS2 ^{pos}' were heterozygous for the variant and those marked 'SORBS2 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

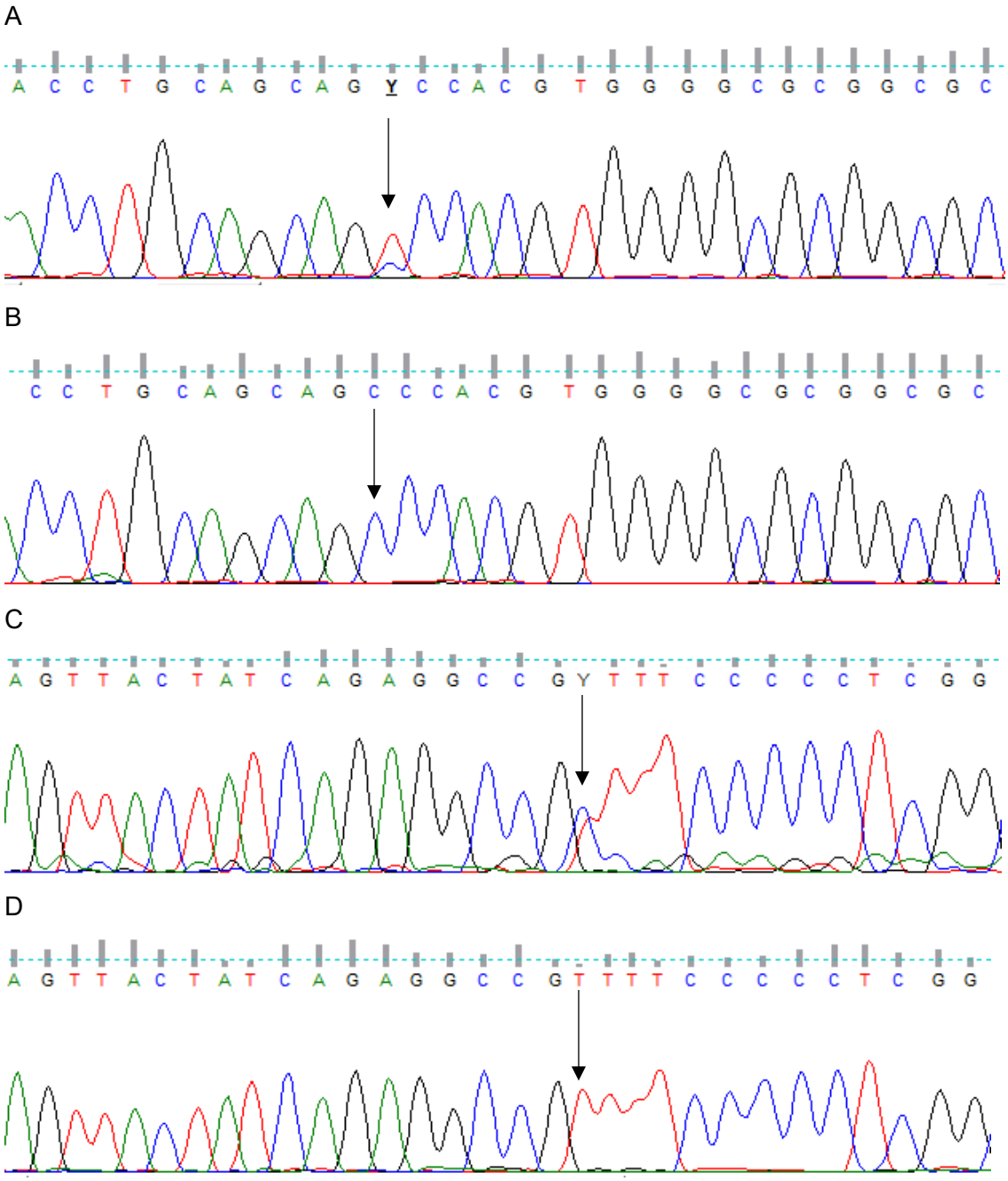


Figure 4.8: Validation of *PXDNL* and *SORBS2* variants identified in Family 4. (A) Sanger sequencing to confirm the presence of *PXDNL* c.2218C>T in affected individual IV:4 (indicated with an arrow), (B) Sanger sequencing confirms the absence of *PXDNL* c.2218C>T in individual IV:3 (indicated with an arrow), (C) Sanger sequencing to confirm the presence of *SORBS2* c.322T>C in affected individual IV:4 (indicated with an arrow), (D) Sanger sequencing confirms the absence of *SORBS2* c.322T>C in individual IV:3 (indicated with an arrow).

4.2.3 Discussion: POLG *mutations as a putative cause of ACM*

The proband in Family 4 met several diagnostic criteria for a definite diagnosis of ARVC. However, the prominent phenotypic features amongst the affected family members were the presence of ventricular fibrosis and ventricular ectopy or arrhythmias, and the subsequent diagnosis of ARVC in other affected family members was mainly driven by the positive family history. The lack of classical pathological and histological findings of ARVC in these individuals supports a non-specific ACM phenotype in the family.

The genetic mutations that most frequently cause ACM are found in *PKP2*, *DSC2*, *DSG2*, *JUP* and *DSP* (Ohno 2016). These five genes encode components of the desmosome, a structure that maintains cell-cell adhesion between cardiomyocytes (Haugaa et al. 2016, Ohno 2016). The recent identification of disease-causing *CDH2* mutations in ACM patients in South Africa means that the pathological site of ACM may extend beyond the desmosome to encompass the adherence junction of the intercalated disc (Mayosi et al. 2017). Mutations in other genes such as *DES*, *PLN*, *CTNNA3* and *RYR2* have also been implicated in ACM (Ohno 2016), expanding the scope of arrhythmia-causing mutations beyond the genes that code for the proteins of the intercalated disc. This indicates that genes that affect any aspect of cardiomyocyte functioning may play a role in the development of ACM. It was this class of genes that was considered for causative mutations in Family 4, as no mutations were found in desmosomal genes or other known ACM loci (Table 4.3). Instead, filtering of the exome sequencing data identified the *POLG* c.2492A>G variant as the possible disease-causing variant in this family.

POLG encodes the catalytic subunit of mitochondrial DNA polymerase. As a mitochondrial protein, mutations in the gene have been associated with several conditions such as chronic progressive external ophthalmoplegia, Parkinson's disease, skeletal myopathies and various syndromic phenotypes (Cohen, Chinnery, and Copeland 2018). In a case report, ultrastructural analysis of the heart tissue from a paediatric liver failure patient carrying two *POLG* mutations (p.A467T and p.K1191N) showed cardiac mitochondrial defects and an accumulation of lipids in the myocardium (Müller-Höcker et al. 2011). This was the first indication that germ-line mutations in *POLG* may have cardiac manifestations in humans.

Table 4.3: Coding region variation in known ACM genes in Family 4

Gene	Variant	Protein change	Carriers	gnomAD MAF	MT	CADD
CTNNA3	c.1787G>A	p.S596N	IV:2, IV:4	0.4130	B (0.99)	12.1
DSC2	c.111A>G	p.L37=	III:2, IV:2, IV:3, IV:4	0.1133	B (0.99)	0.1
DSC2	c.2326A>G	p.I776V	IV:4	0.1217	B (0.99)	10.1
DSG2	c.861C>T	p.N287=	III:2, IV:2, IV:3, IV:4	0.4183	B (8.0^{-6})	3.2
DSG2	c.887A>G	p.F296C	IV:2, IV:3	0	B (0.99)	23.0
DSG2	c.2318G>A	p.R773K	III:2, IV:2, IV:3, IV:4	0.2604	B (0.99)	16.4
DSG2	c.3321T>C	p.V1107=	III:2, IV:2, IV:3, IV:4	0.4296	B (1.6^{-9})	18.2
DSP	c.741T>G	p.A247=	III:2, IV:2, IV:3, IV:4	0.9993	B (1.6^{-7})	0.1
DSP	c.2091A>G	p.G697=	III:2, IV:2, IV:3, IV:4	0.7633	B (0.99)	7.9
DSP	c.2631G>A	p.R877=	III:2, IV:2, IV:3, IV:4	0.7702	B (1.6^{-24})	10.2
DSP	c.2862C>T	p.C954=	III:2, IV:2, IV:3, IV:4	0.2434	B (1.9^{-13})	10.4
DSP	c.5213G>A	p.R1738Q	III:2, IV:4	0.1750	B (0.99)	15.1
DSP	c.5498A>T	p.E1833V	IV:3, IV:4	0.009549	D (0.99)*	28.4
DSP	c.7122C>T	p.T2374=	IV:2, IV:3, IV:4	0.2862	B (9.5^{-16})	1.9
DSP	c.8472G>C	p.G2824=	IV:2, IV:3, IV:4	0.7095	B (3.7^{-18})	14.5
JUP	c.213T>C	p.D71=	IV:2, IV:3, IV:4	0.7176	B (4.6^{-22})	12.3
JUP	c.2089A>T	p.M697L	IV:2, IV:3, IV:4	0.6489	B (0.80)	15.7
PKP2	c.1097T>C	p.L366P	III:2, IV:2, IV:4	0.1820	B (0.99)	13.5
RYR2	c.2973A>G	p.S991=	III:2, IV:2, IV:3, IV:4	0.9240	B (1.7^{-15})	1.8
RYR2	c.6906T>C	p.L2302=	III:2, IV:2, IV:3, IV:4	0.9877	B (1.4^{-17})	9.6
RYR2	c.7806C>T	p.H2602=	IV:2, IV:3, IV:4	0.4778	B (2.8^{-17})	2.6
RYR2	c.8873A>G	p.Q2958R	III:2, IV:2	0.2139	B (0.004)	22.6
RYR2	c.9318T>G	p.S3106=	III:2, IV:2, IV:3, IV:4	0.9904	B (0.004)	0.7
RYR2	c.10503C>T	p.T3501=	III:2, IV:2, IV:3, IV:4	0.9874	B (2.4^{-15})	2.8
RYR2	c.10776C>T	p.S3592=	III:2, IV:2, IV:3, IV:4	0.9875	B (4.3^{-14})	0.04

* Although predicted deleterious by MT, this variant was benign in ClinVar

B, benign; CADD, Combined Annotation Dependent Depletion; D, deleterious; gnomAD, Genome Aggregation Database; MT, MutationTaster

Investigation into animal models showed records of well-phenotyped mouse models with *Polg* mutations in the MGI database. The documented phenotypic effects of deficient *Polg* included dilated heart ventricles, increased heart weight and abnormal myocardial fibre morphology in *Polg* point mutation mice (Ross et al. 2013, Trifunovic et al. 2004), while a targeted transgenic mouse model recorded several cardiovascular phenotypes including cardiomyopathy, congestive heart failure, thickened ventricular walls and an enlarged heart. These transgenic mice, in which the point mutation *Polg* p.Y955C was targeted to the heart, were shown to develop significant cardiac defects including cardiomegaly, increased left ventricular mass and cardiomyocyte hypertrophy (Lewis et al. 2007). The phenotypic manifestations were attributed to mitochondrial dysfunction and increased oxidative stress. The enlargement of the heart in these animals is in keeping with the observed cardiac phenotype in individuals IV:2 and IV:4. Mice expressing the same mutation were later shown to have marked cardiac fibrosis, with a three-fold increase in scar tissue formation when compared to wild-type animals (Koczor et al. 2013). Again, this phenotype reflects the diffuse ventricular fibrosis that was detected in all *POLG* variant carriers in this investigation.

The *POLG* c.2492A>G variant has been reported before. Mancuso et al. (2004) first associated the variant with mitochondrial DNA deletions in muscle sections from patients with familial Parkinson's disease, ophthalmoplegia and neuropathy, and it has since been identified in patients with similar mitochondrial phenotypes (Wong et al. 2008, Woodbridge et al. 2013, Ylönen et al. 2013). However, a contradictory study in 2008 detected the variant in 0.9% of the control population and considering certain European gnomAD sub-populations, it was reclassified as a VUS (Gui et al. 2012, Luoma et al. 2007, Wong et al. 2008). No other pathogenic mutation had been found in the family where the original *POLG* c.2492A>G variant was reported (Mancuso et al. 2004). In addition, no neurological or syndromic features were observed in Family 4.

Collectively, the results of prior animal model studies of *POLG* make it a compelling candidate gene for cardiomyopathy in humans. Two prominent phenotypes observed in the affected family members in this study (cardiomyopathy and diffuse ventricular fibrosis) were also present in the transgenic mice (Koczor et al. 2013, Lewis et al. 2007). A previous project linked to this investigation identified an ultra-rare *POLG* VUS, c.3077G>A (p.R1026H) in another South African ACM proband (Booi 2017). This individual presented with cardiac arrest (survived), recurrent right-bundle-branch-block morphology ventricular tachycardias, and a DCM phenotype with biventricular dilatation and systolic dysfunction. While CMR was not performed in this patient, interstitial fibrosis was observed on endomyocardial biopsy. Although this phenotypic presentation is similar to the *POLG* variant carriers in Family 4, the c.3077G>A variant was classified as a VUS because no family members were available for segregation analysis.

A possible limitation of this study is the choice of a MAF threshold of 0.01, as recent investigations have suggested that more stringent thresholds of up to 0.0001 may be beneficial due to the rarity of the disease (Tayal et al. 2017). The *POLG* c.2492A>G variant would not pass this MAF filter. However, *POLG* remains a viable candidate for several reasons, the most important of which is the work done in mice which supports the notion of *POLG* as a cardiomyopathy gene. With the exception of individual IV:4 who suffered a SCD, all carriers of the *POLG* c.2492A>G variant had sub-clinical forms of ACM, which was possibly exacerbated in the proband due to the car accident that led to his diagnosis with cardiomyopathy at the age of 50 years. The age of disease onset in this family ranged between 43 and 76 years, well above the average age of ACM onset of 33 years (Bhonsale et al. 2015). It is plausible then that the disease-causing variant in this family may occur at a higher frequency in the background population than more highly expressive, early-onset mutations.

Due to the identification of healthy *PKP2* mutation carriers in ACM families, it has been proposed that digenic inheritance may be more prevalent in ACM than previously thought (König et al. 2017). It may be that, in the present family, *POLG* c.2492A>G is coinherited with another pathogenic cardiomyopathy variant. However, despite rigorous screening of exome data from all four affected individuals, no other likely candidate was identified. Although it is possible that the true disease-causing variant may reside in intronic or intergenic regions of the genome (Harper et al. 2020), current knowledge of ACM genetics indicates that this is unlikely. Given that no other candidate mutation was evident upon analysis of all affected individuals, the potential role of *POLG* in the development of cardiomyopathy or arrhythmias was prioritised for functional investigation (Chapter 5).

4.3 Family 5 (DCM 435)

4.3.1 Clinical history of Family 5

Recruited at GSH in 2016, Family 5 is a small, two-generation family of self-reported Congolese descent (Figure 4.9). The proband, individual II:3, presented with severe paediatric DCM at 16 months and died by the age of 2 years. Both of his parents (I:1 and I:2) were screened in their fourth decades of life and found to be phenotypically normal. They had another son who was also affected with paediatric DCM and died before the study began (age 18 months), as well as a daughter who is unaffected but has not been recruited into the study as yet. The pattern of disease indicates that recessive or X-linked modes of inheritance are most likely: these and compound heterozygous inheritance were considered when analysing genetic data from this family.

Family 5

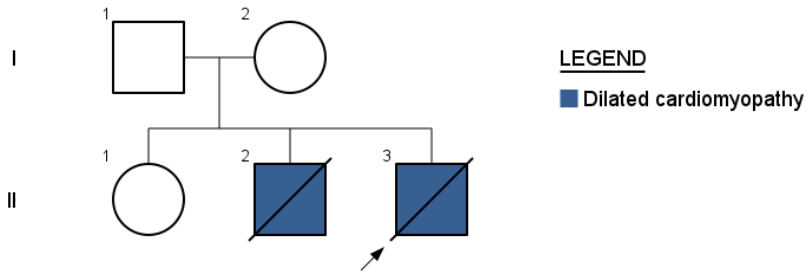


Figure 4.9: Pedigree indicating disease segregation in Family 5. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Blue shaded symbols indicate individuals diagnosed with DCM. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

4.3.2 Genetic analysis of Family 5

The trio of available individuals (I:1, I:2, II:3) was investigated using exome sequencing. The sequencing reaction generated 68,061 variants in total, at a mean sequence coverage of 136x. Of the 68,061 variants, 7 met the filtering criteria (Figure 4.10).

No candidate X-linked variants were identified, while five AR variants and two compound heterozygous variants were found (Table 4.4). The recessive variants were largely predicted benign; the exceptions were the 4 bp insertion in *ITPR2* that was much more common than VEP predicted, possibly due to annotation error, and the 4 bp deletion in *LFNG*. Although *LFNG* had reported cardiovascular phenotypes in knockouts, these were limited to the lung and/or retinal vasculature, meaning this gene has no discernible links to potential cardiomyopathy phenotypes.

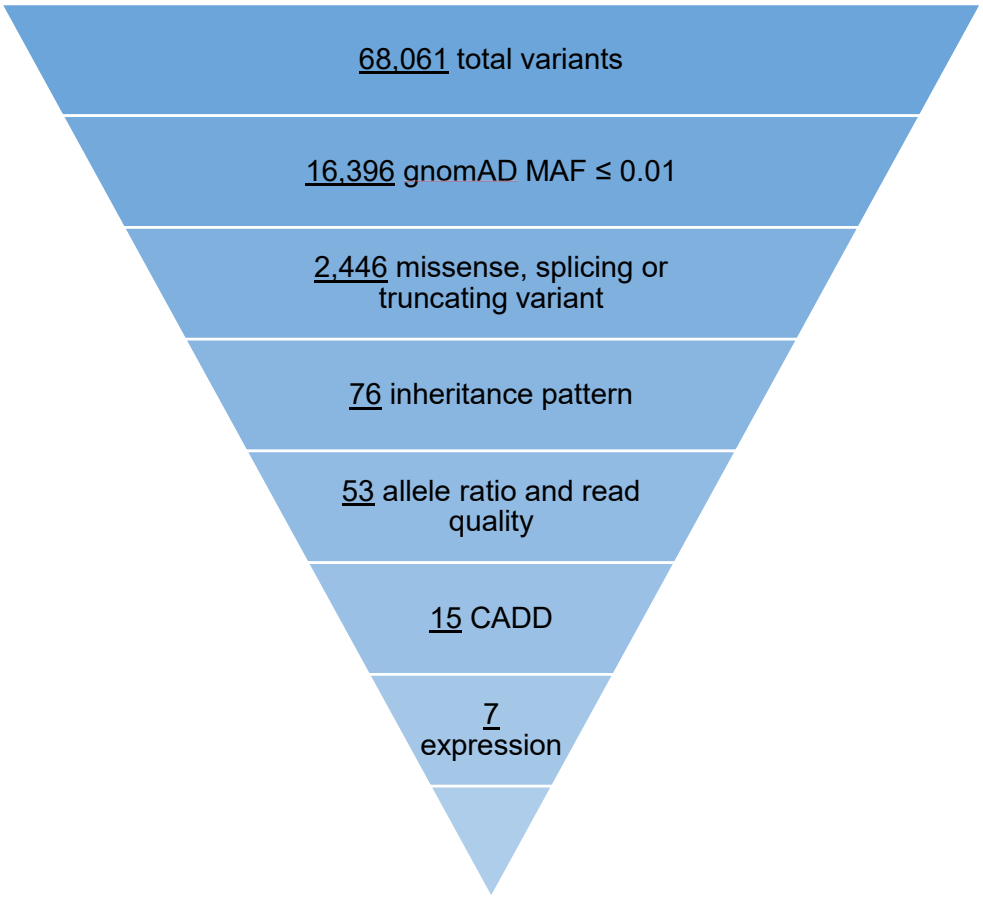


Figure 4.10: Filtering of exome sequencing data from Family 5. Exome sequencing produced a total of 68,061 variants, of which 7 met the filtering criteria. *CADD, Combined Annotation Dependent Depletion; gnomAD, Genome Aggregation Database; MAF, minor allele frequency*

Table 4.4: Candidate variants in Family 5

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
Autosomal recessive														
ITPR2	c.1409+2_3insGAGT	-	0.710	0.652000	0.655100	0.36753	NS	NS	NS	NS	35.0	1.5	No	NS
KIAA1671	c.1316A>G	p.K439R	0	0	0	0	B (0.017)	B (0.99)	B (0.6)	D (0.801)	22.3	2.5	No	3/7
LFNG	c.163_6del	p.D55S*141	0	0	0	0	NS	D (1)	NS	NS	21.6	2.3	Yes	NS
SH3BGR	c.693_5dup	p.E231dup	0	0	0	0	NS	B (0.99)	NS	NS	22.7	58.4	NA	NS
TRMT13	c.142_3delinsAT	p.A48M	0	0	0.000032	0	NS	B (0.59)	NS	D (0.935)	22.9	3.7	NA	4/7
Compound heterozygous														
ITGB5	c.146G>A	p.C49Y	0	0.000247	0.000265	0.00062	D (0.120)	D (0.99)	D (0)	D (1)	27.5	27.1	NA	6/7
ITGB5	c.1929G>C	p.E643D	0	0.000264	0.000272	0.00100	D (0.086)	D (0.99)	D (0.01)	D (0.983)	26.0	27.1	NA	6/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SCD, sudden cardiac death; SIFT, Sorting Intolerant From Tolerant

The two compound heterozygous variants occurred in the gene *ITGB5*. They were predicted disease-causing by all pathogenicity prediction tools and affected highly conserved residues (Table 4.4). Although mouse knockout models have not yet been described for *Itgb5*, the gene was prioritised due to its potential role in several KEGG pathways relevant to cardiomyopathy, including focal adhesion, regulation of the actin cytoskeleton, integrin-mediated cell adhesion, ARVC, HCM and DCM (Kanehisa et al. 2019). A genome-wide association study of 10,898 patients with coronary artery disease and 76,535 controls reported that variation in *ITGB5* was associated with disease status and, together with other genetic risk factors, contributed to the risk of myocardial infarction, heart failure and cardiomyopathy (Verweij et al. 2017). The gene was also implicated in myocardial fibrosis and injury in a rat model of menopause, when cardioprotective agents were demonstrated to act at least in part through the down-regulation of left ventricular *Itgb5* expression (Zhang et al. 2018).

Two *ITGB5* variants were identified in the proband. The maternally inherited c.146G>A (p.C49Y) and paternally inherited c.1929G>C (p.E643D) (Figure 4.11) both affect highly conserved protein residues in two distinct functional regions of the ITGB5 protein (Figure 4.12). Both variants were confirmed upon Sanger sequencing (Figure 4.13).

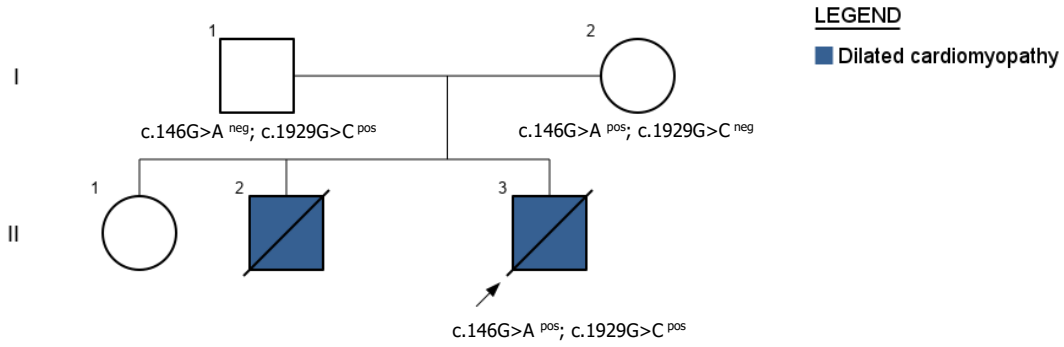
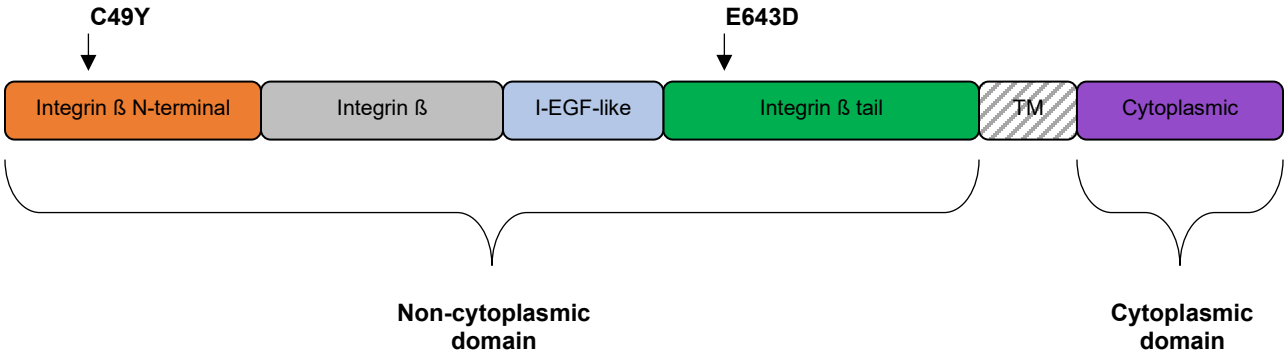


Figure 4.11: Pedigree indicating segregation of *ITGB5* c.146G>A and c.1929G>C in Family 5. Individuals marked 'c.146G>A ^{pos}' and 'c.1929G>C ^{pos}' were heterozygous for those variants and those marked 'c.146G>A ^{neg}' and 'c.1929G>C ^{neg}' tested negative for them. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Blue shaded symbols indicate individuals diagnosed with DCM. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

A



B

H. sapiens	E	C	L	L	I	H	P	K	C	A	W	C	S	K	E	D	F	G	S	P	R	S	I
M. mulatta	E	C	L	L	I	H	P	K	C	A	W	C	S	K	E	D	F	G	S	P	R	S	I
P. troglodites	E	C	L	L	I	H	P	K	C	A	W	C	S	K	E	D	F	G	S	P	R	S	I
F. catus	E	C	L	L	I	H	P	K	C	A	W	C	F	K	E	D	F	G	S	L	R	S	V
M. musculus	E	C	L	L	I	H	P	K	C	A	W	C	S	K	E	Y	F	G	N	P	R	S	I
G. gallus	E	C	L	L	I	H	P	K	C	A	W	C	S	K	E	E	F	G	S	T	K	S	V
D. rerio	E	C	L	L	I	H	P	S	C	A	W	C	A	Q	E	D	F	G	Q	A	R	T	L
X. tropicalis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

C

H. sapiens	P	D	A	C	S	T	K	D	R	C	V	E	C	L	L	L	H	S	G	K	P	D	N
M. mulatta	P	D	A	C	S	T	K	D	R	C	V	E	C	L	L	L	H	S	G	K	P	D	N
P. troglodites	P	D	A	C	S	T	K	D	R	C	V	E	C	L	L	L	H	S	G	K	P	D	N
F. catus	P	D	A	C	S	T	K	D	R	C	V	E	C	L	L	L	H	V	G	D	P	D	N
M. musculus	P	D	A	C	S	S	K	D	R	C	V	E	C	L	L	L	H	Q	G	K	P	D	N
G. gallus	P	G	V	C	S	T	K	D	R	C	I	E	C	K	L	F	N	S	G	R	L	D	N
D. rerio	P	D	A	C	G	T	K	R	E	C	I	E	C	R	L	F	N	T	G	R	L	D	N
X. tropicalis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 4.12: *ITGB5* c.146G>A and c.1929G>C identified in Family 5. (A) Protein structure of ITGB5 with position of the p.C49Y and p.E643D mutations marked with arrows, (B) Multiple species conservation alignment of the p.C49 residue of ITGB5 (shaded in blue), (C) Multiple species conservation alignment of the p.E643 residue of ITGB5 (shaded in blue). *I-EGF, integrin epidermal growth factor; TM, transmembrane*

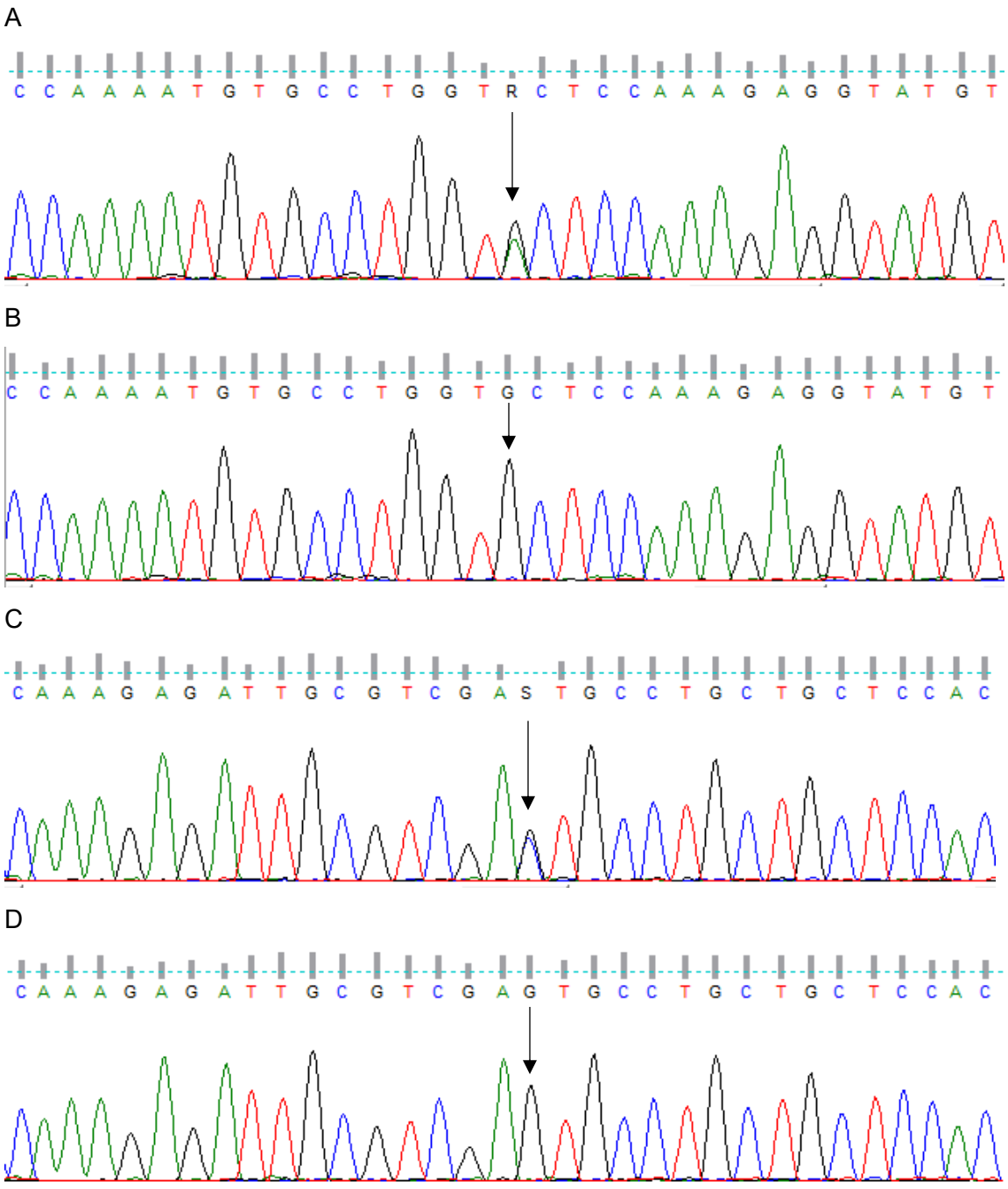


Figure 4.13: Validation of *ITGB5* variants identified in Family 5. (A) Sanger sequencing to confirm the presence of *ITGB5* c.146G>A in affected individual II:3 (indicated with an arrow), (B) Sanger sequencing confirms the absence of *ITGB5* c.146G>A in individual I:1 (indicated with an arrow), (C) Sanger sequencing to confirm the presence of *ITGB5* c.1929G>C in affected individual II:3 (indicated with an arrow), (D) Sanger sequencing confirms the absence of *ITGB5* c.1929G>C in individual I:2 (indicated with an arrow).

4.3.3 Discussion: ITGB5 *mutations as a putative cause of DCM*

Integrins are multifunctional transmembrane receptor proteins that are involved in cell signalling and cellular adhesion to the extracellular matrix (ECM). They are heterodimeric molecules, consisting of α and β subunits which can form up to 24 different combinations (Hynes 2002). In the myocardium, integrins play a role in focal adhesion, cell survival, mechanotransduction and remodelling of the actin cytoskeleton (Manso, Kang, and Ross 2009). They are also signal transducers, acting through downstream molecules such as integrin-linked kinase, extracellular-signal related kinases and small GTPases (Hynes 2002, Wu and Dedhar 2001). Perturbation of myocardial integrin-actin signalling pathways has been linked to cardiac hypertrophy, fibrosis and the development of cardiomyopathy (Chen et al. 2016, Kubota et al. 2019, Manso et al. 2017, Nishimura et al. 2014, Quang et al. 2015, Valiente-Alandi et al. 2018).

ITGB5 encodes the β_5 subunit of integrin. The $\alpha_V\beta_5$ integrin heterodimer is expressed in the heart, where it binds fibronectin, osteopontin and vitronectin; expression of this dimer is reduced amongst heart failure patients (Manso, Kang, and Ross 2009). Although *ITGB5* has not been associated with cardiomyopathy before, its expression in the heart, together with the documented role of integrins in normal cardiac function and development, means it is a plausible disease candidate for cardiomyopathy in Family 5. Indeed, the gene has been linked to other related phenotypes such as cardiomyopathy and heart failure in the context of coronary artery disease (Verweij et al. 2017), and cardiac fibrosis in the context of menopause (Zhang et al. 2018).

The ITGB5 protein, like other integrin subunits, is composed of a large extracellular domain coupled to smaller transmembrane and cytoplasmic domains (Figure 4.12A) (Ross and Borg 2001). The extracellular domain mediates contact with the ECM and other ligands; it is in this domain that both *ITGB5* mutations in the proband were found. Because the affected protein residues are highly conserved (Figure 4.12B-C) and occur in known integrin β subunit signatures, it is possible that the mutations may affect contacts between cardiac cells and the ECM. Although intriguing, and no other candidate mutations in known DCM genes were identified in this family (Appendix Q), further research will be needed to elucidate the role of *ITGB5* in the heart and the potential of mutations in the gene to cause paediatric cardiomyopathy.

4.4 Chapter summary

In this chapter, two families are described in which no disease-causing mutations were identified in known cardiomyopathy genes. Instead, variation in the putative new disease genes *ITGB5* and *POLG* was found. These genes have never been associated with heritable cardiomyopathy in humans before. The ACMG guidelines used in the previous chapter, or any derivations from them, grade mutations according to their position within known disease-causing genes or observation in several unrelated patients (Richards et al. 2015). Therefore, these criteria cannot be used to judge potential novel disease genes in single families, as is the case here, although the *in silico* prediction of deleteriousness and high degree of evolutionary conservation seen for all variants is supportive of their pathogenic potential.

The variants in *ITGB5* and *POLG* should therefore be considered VUSs until further proof supporting or refuting their pathogenicity becomes available. Similarly, the modifying impact of the *SORBS2* variant in ACM remains unclear at this stage. The utility of modelling cardiomyopathy in cell culture or animal models was discussed earlier (Chapter 1). These techniques may be particularly useful when variants have limited observations in small family settings where segregation analysis alone is not entirely informative. The functional exploration of *POLG* forms the basis of the next chapter, in which this gene and others are investigated using the *in vivo* zebrafish model (Chapter 5). Unfortunately, *ITGB5* and *SORBS2* were identified as possible disease genes after the zebrafish modelling and thus could not be investigated.

Chapter 5 Zebrafish as a model for familial cardiomyopathy

5.1 Introduction

When examining the genetics of familial cardiomyopathy using exome sequencing, interpreting the pathogenicity of variants can be challenging. The identification of known disease-causing mutations in patients requires little clarification; however, as demonstrated in Chapters 3 and 4, known cardiomyopathy-causing mutations may not be found in patients, even when disease is ostensibly heritable. In some instances, novel mutations or unique genotype-phenotype relationships may be uncovered (Chapter 3), while the absence of mutations in any of the cardiomyopathy genes may lead to the identification of candidate genes (Chapter 4). In these cases, variants should ideally be validated in functional assays.

As NGS becomes more available, it is likely that the need for functional assays will increase. This is particularly true in African patient populations, where the genetics of cardiomyopathy is largely unexplored. The Cardiovascular Genetics laboratory (HICRA, UCT) proposes to address the problem of variant validation using the zebrafish model. The zebrafish is an increasingly popular tool in cardiomyopathy research, offering the convenience of *in vitro* modelling but in a whole-organism environment (Asnani and Peterson 2014). Although physiologically a much simpler animal than mice, zebrafish have been used to study numerous forms of cardiomyopathy to date, offering unique insights into the genetics of DCM, HCM and LVNC (Bainbridge et al. 2015, Dhandapany et al. 2014, Sasagawa et al. 2016, Zou et al. 2015). Due to the high-throughput potential of the zebrafish model, it looks to be a powerful tool in cardiomyopathy research when combined with NGS unbiased screening of the patient genome.

5.2 Zebrafish CRISPR/Cas9 model

5.2.1 Guide RNA synthesis

SgRNAs were designed to target and disrupt the genes *myh7*, *pkp2* and *polg*. All sgRNAs disrupted genes within the first exon; however, all disrupted at least 75% of the target genes (Table 5.1).

Table 5.1: Single guide RNA designed points of gene disruption

Target gene	Target transcript accession no.	Target exon (total no. exons)	Transcript length (bp)	Proportion of CDS disrupted	Proportion of protein disrupted
myh7	ENSDART00000192445	1 (37)	5,817	88.5%	97.4%
pkp2	ENSDART00000035899	1 (13)	2,448	91.0%	91.0%
polg	ENSDART00000150129	1 (23)	3,621	87.0%	87.1%

bp, base pairs; CDS, coding sequence; no., number

The sgRNAs were determined to be of sufficient quality and quantity for use in injection experiments (Table 5.2).

Table 5.2: Single guide RNA concentration and purity

Gene target	Concentration (ng/µl)	260/280 ratio*
myh7 (8:16543250-16543273)	600.9	2.33
pkp2 (4:17057073-17057096)	304.5	2.34
polg (25:9339993-9340016)	2076.0	2.32

* The 260/280 ratio was determined by spectrophotometry and was used to determine the RNA purity.

µl, microlitre; ng, nanogram

5.2.2 Characterisation of uninjected control zebrafish

Controls in this experiment were uninjected zebrafish larvae from the same clutch as the embryos used for injection. The *cmlc2*:mCherry line was used, in which all cardiomyocytes express red fluorescence protein. Two-photon fluorescence image analysis of control zebrafish hearts at 3 dpf showed that the zebrafish ventricle is oval in shape, with prominent fluorescence along the ventricle walls (Figure 5.1). The atrium is not as visible due to the chosen plane of the microscope as well as lesser expression of the *cmlc2* promotor in these cells.

Figure 5.1: Multiphoton image of 3 dpf zebrafish uninjected control ventricle at end-diastole. Pictured is a *cmlc2*:mCherry transgenic zebrafish embryo at 32x magnification. Cardiomyocytes fluoresce red and were visualised at a laser wavelength of 720 nm.

The uninjected *cmlc2*:mCherry ventricles were morphologically similar and used as a basis of comparison for the *myh7*, *pkp2* and *polg* knockouts (Figure 5.2). There were overall significant differences in ventricle area (ANOVA, $p < 0.001$), length (ANOVA, $p = 0.004$) and thickness (ANOVA, $p = 0.009$) between the experimental groups (Figure 5.3), although only the area comparison was sufficiently powered (>80%) to detect significant differences as determined by post-hoc power calculations in which the observed effect size was assumed to reflect true variability (Appendix R). The ventricular area was directly correlated with chamber length for each group apart from the *polg* knockout larvae, but all groups displayed similar trends of association (Figure 5.4). Other numerical variables did not correlate (Appendix S). Each gene knockout will be discussed in more detail separately below.

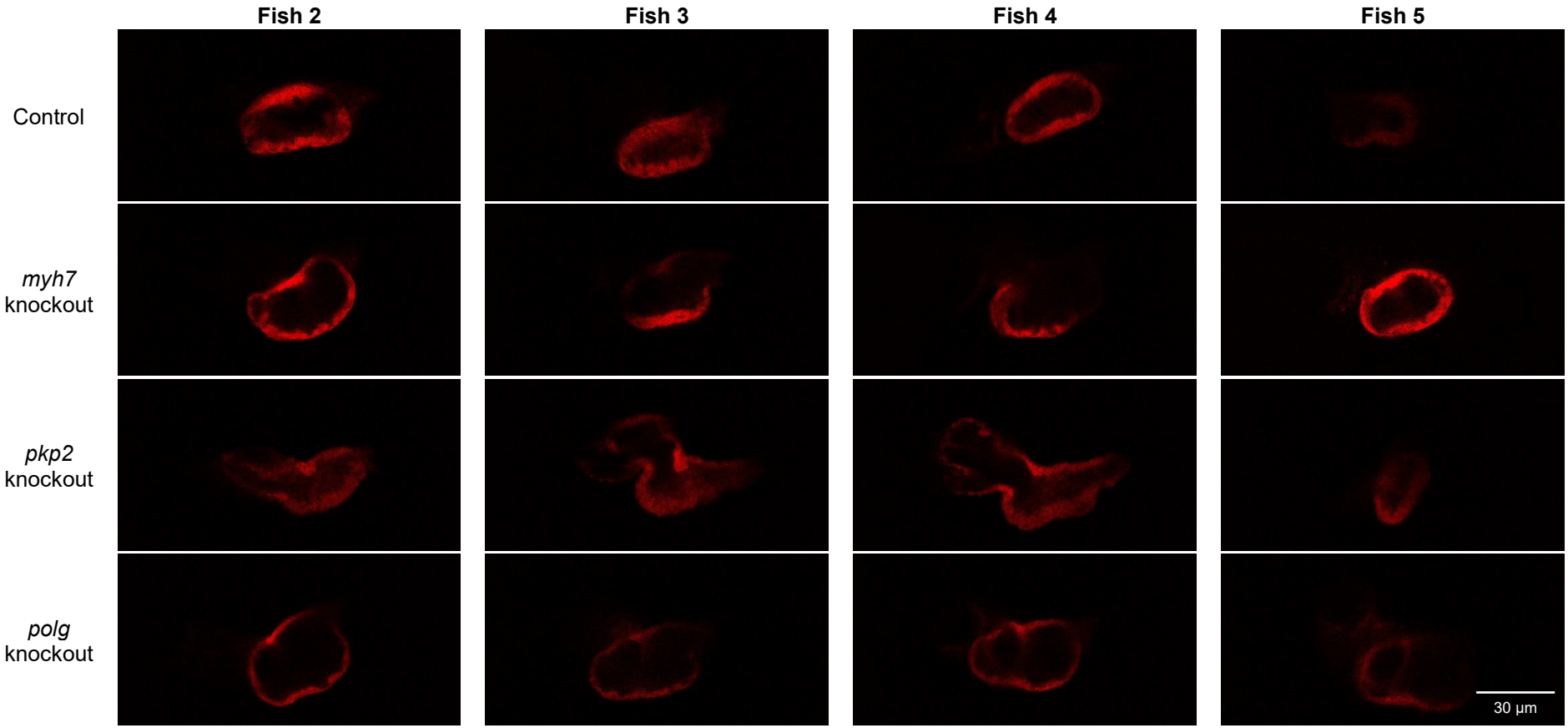


Figure 5.2: Multiphoton images of 3 dpf zebrafish knockout and control ventricles at end-diastole. Fish 2 to 5 are represented for each group; the image for Fish 1 is included in each gene knockout subsection. Pictured are *cmlc2*:mCherry transgenic zebrafish embryos in control and gene knockout groups, at 32x magnification. Cardiomyocytes fluoresce red and were visualised at a laser wavelength of 720 nm.

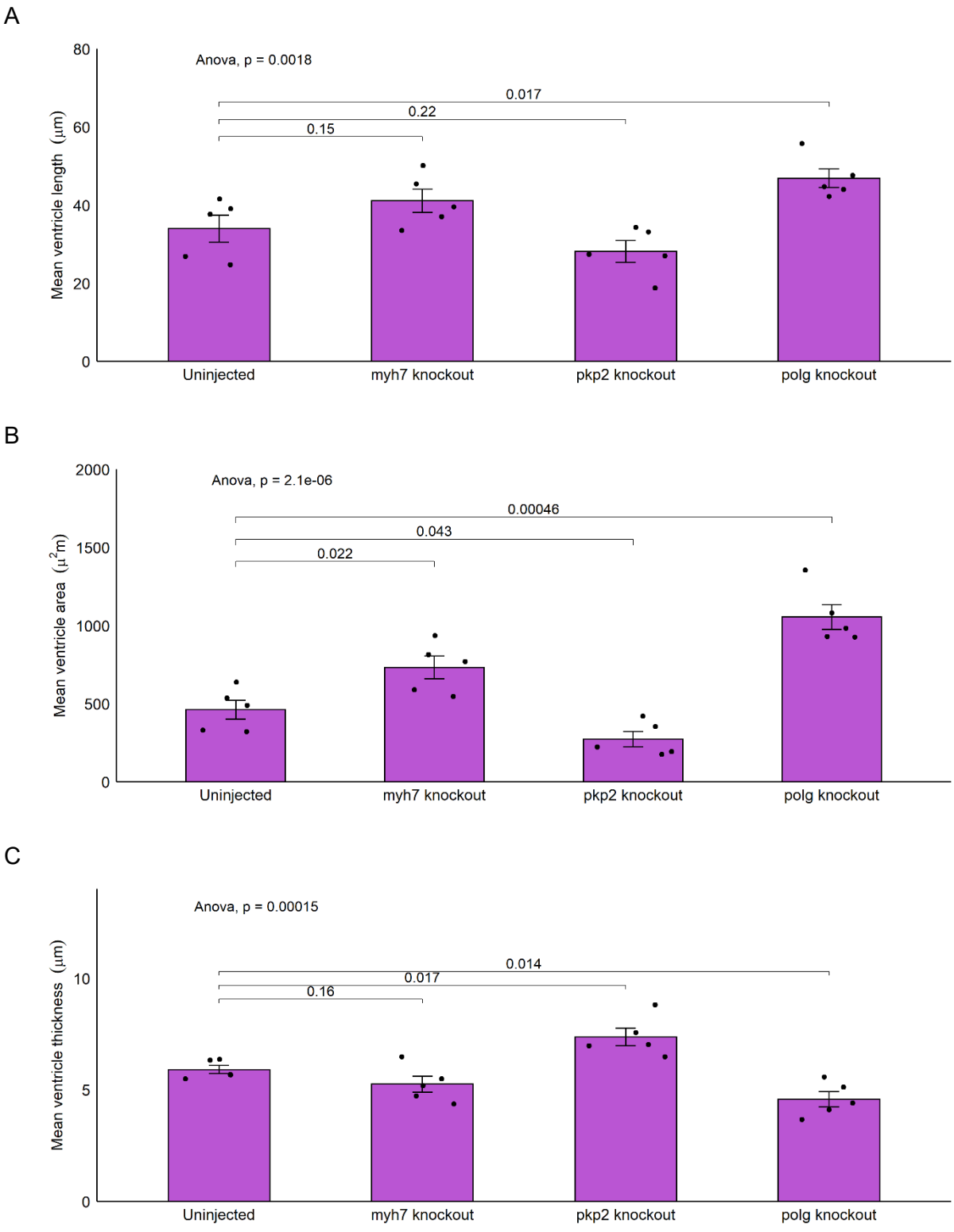


Figure 5.3: Comparison of ventricular length, area and thickness amongst the experimental groups. The length (A), area (B) and thickness (C) are represented on each y-axis, while experimental groups are on the x-axis. Error bars indicate the standard error in each group (n = 5). Significance bars show the p-value when comparing each experimental group to the uninjected control group (two-sided t-test).

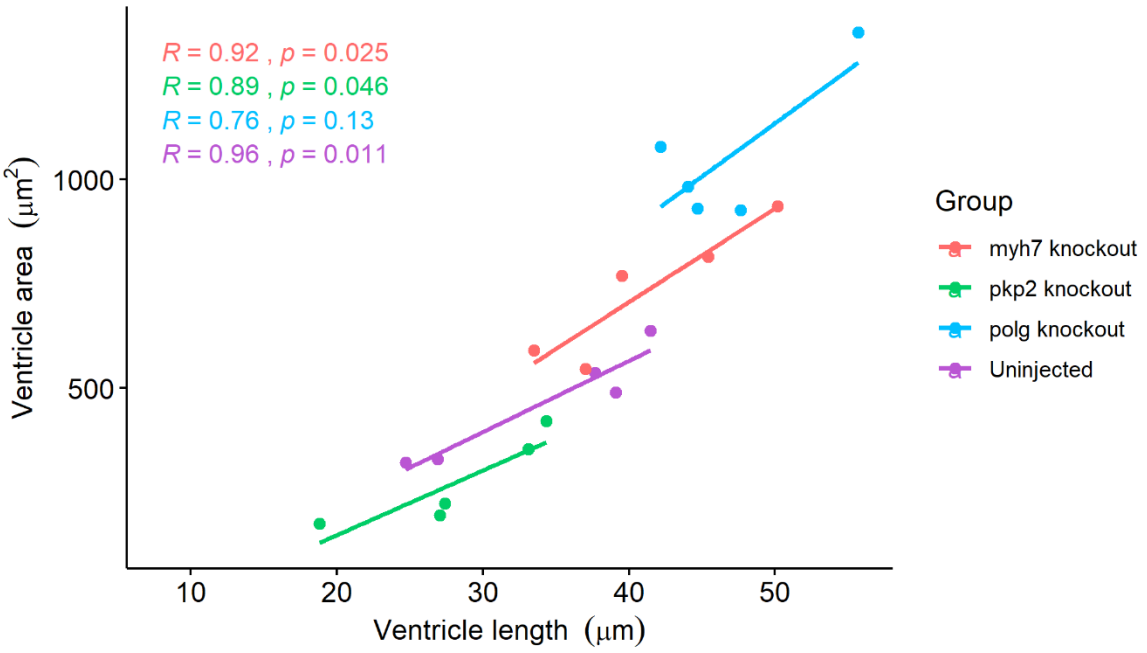


Figure 5.4: Correlation of ventricular length and area amongst the experimental groups. The length is represented on the x-axis, while area is on the y-axis (n = 5 per group). The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

5.2.3 Phenotyping and genotyping of myh7 *knockout larvae*

The *myh7* sgRNA achieved an average CRISPR activity of 48.4%, the lowest efficiency of the sgRNAs synthesised in this study (Appendix T). In addition, the majority of induced mutations at the cut site were non-frameshift mutations (Figure 5.5), meaning that these mutations were less likely to impair protein function.

The knockout zebrafish hearts were largely structurally intact by 5 dpf, closely resembling those of uninjected control larvae (Figures 5.2 and 5.6). Although the effects of the mutations were mild, there was a slight increase in overall ventricular size amongst *myh7* knockouts (Figure 5.3B). Human *MYH7* is typically associated with HCM, and these findings may be consistent with this although there was no significant change in ventricle wall thickness (Figure 5.3C). In fact, only one outlier was identified when compared to uninjected controls, with marginally (26.1%) thinner walls; other values were within the normal range.

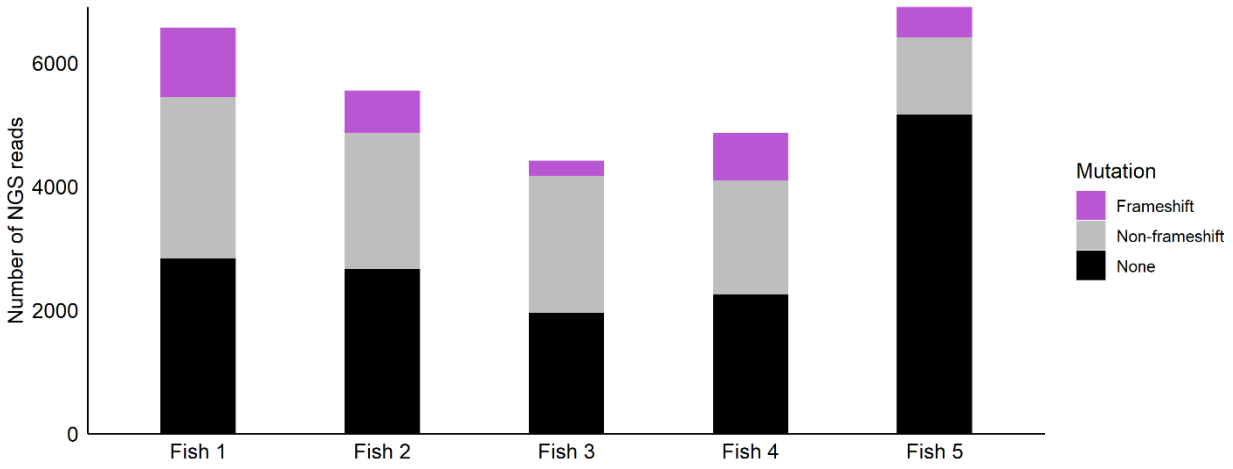


Figure 5.5: Genotyping of zebrafish larvae injected with *myh7* guide RNAs. The number of NGS reads with frameshift mutations (purple bars), non-frameshift mutations (grey bars) and no mutation (black bars) are depicted (y-axis) per larva (x-axis). Mutation rates are taken at the site of greatest insertion or deletion rate at that point in the larva. *NGS, next-generation sequencing*

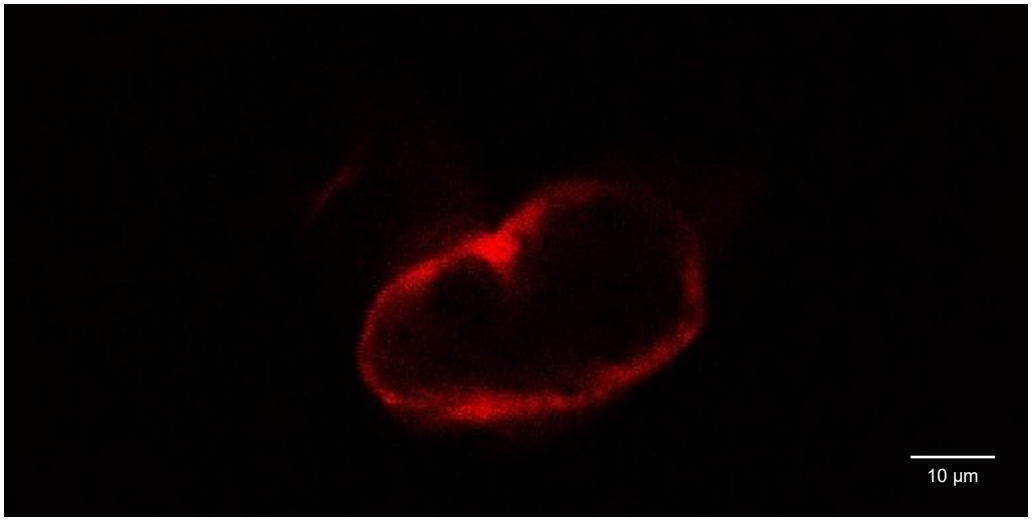


Figure 5.6: Multiphoton image of 3 dpf *myh7* knockout Fish 1 ventricle at end-diastole. Pictured is a *cmlc2*:mCherry transgenic zebrafish embryo at 32x magnification, after treatment with *myh7* knockout sgRNA and Cas9. Cardiomyocytes fluoresce red and were visualised at a laser wavelength of 720 nm.

5.2.4 Phenotyping and genotyping of pkp2 *knockout larvae*

A high CRISPR activity level was achieved with the *pkp2* knockout sgRNA, with a mean efficiency of 78.6% across the larvae (Appendix T). However, similar to the *myh7* knockout, a low level of frameshift mutations was observed, with non-frameshift insertions and deletions detected in the majority of genotyped calls (Figure 5.7). The most frequent genotype was a 9 bp non-frameshift deletion (Appendix T).

Despite the low frameshift mutation rate, the *pkp2* mutations had observable effects on the cardiac structure (Figures 5.2 and 5.8). The knockout led to a significant increase in ventricular thickness combined with a moderate, but significant, reduction in ventricle size (Figure 5.3). In the most extreme example of this (Figure 5.8), the ventricle was 62.2% smaller and 44.7% shorter, with a 49.1% thicker wall.

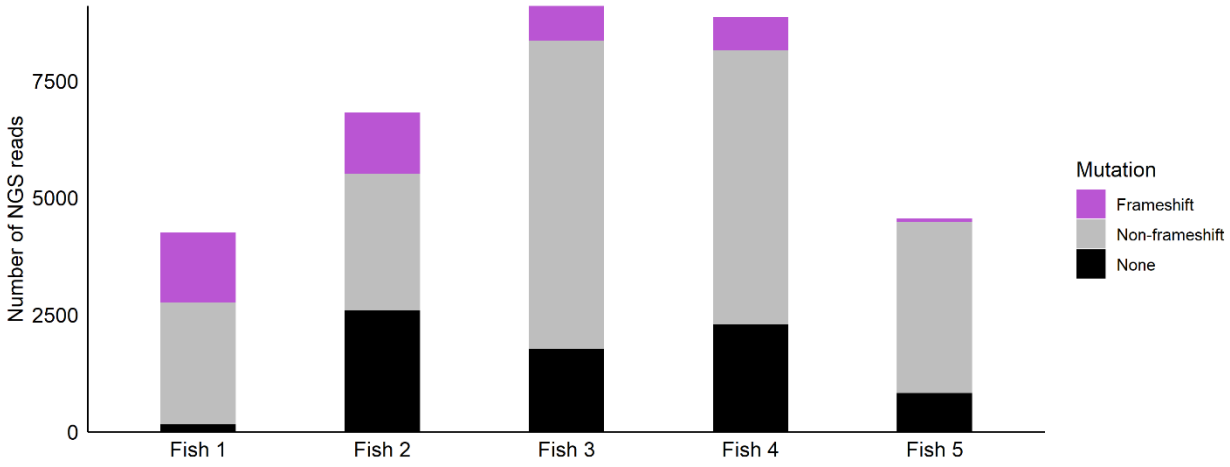


Figure 5.7: Genotyping of zebrafish larvae injected with *pkp2* guide RNAs. The number of NGS reads with frameshift mutations (purple bars), non-frameshift mutations (grey bars) and no mutation (black bars) are depicted (y-axis) per larva (x-axis). Mutation rates are taken at the site of greatest insertion or deletion rate at that point in the larva. *NGS, next-generation sequencing*

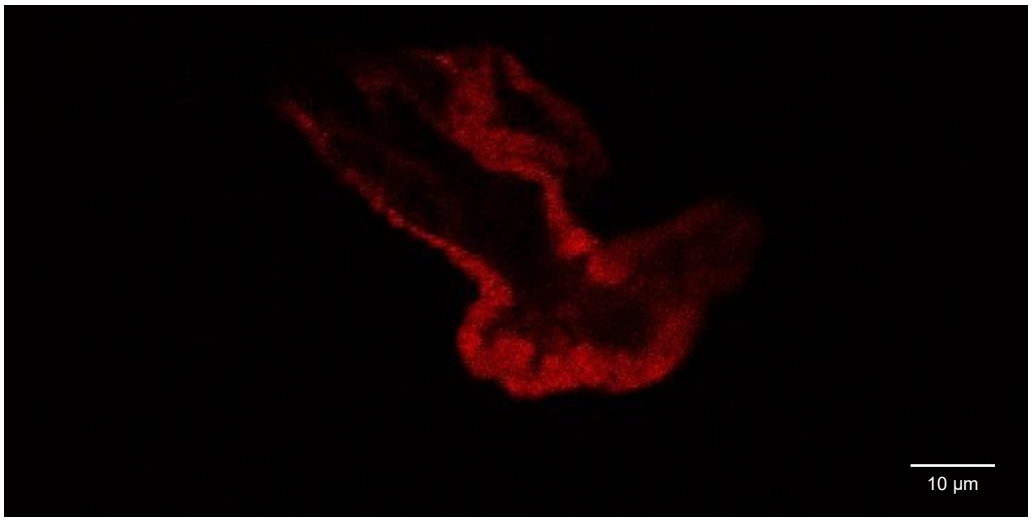


Figure 5.8: Multiphoton image of 3 dpf *pkp2* knockout Fish 1 ventricle at end-diastole. Pictured is a *cmlc2*:mCherry transgenic zebrafish embryo at 32x magnification, after treatment with *pkp2* knockout sgRNA and Cas9. Cardiomyocytes fluoresce red and were visualised at a laser wavelength of 720 nm.

5.2.5 Phenotyping and genotyping of polg knockout larvae

A mean CRISPR efficiency of 73.2% was achieved using the *polg* sgRNA (Appendix T). Very few non-frameshift mutations were detected; the majority of genotypes corresponded to frameshift insertions and deletions (Figure 5.9). The genotype that was most frequently observed was a 7 bp deletion in the gene (Appendix T).

The phenotypic effects of *polg* knockout were more consistent than in the other knockouts (Figures 5.2 and 5.10). The ventricles were significantly larger, longer and had thinner walls compared to uninjected control larvae (Figure 5.3). All the imaged larvae had ventricles that were outside the normal range for ventricle size, ranging from 100.3% to 192.5% greater than the mean uninjected control ventricle. In the most extreme case (Figure 5.10), in which the ventricle was 192.5% larger than the control values, the ventricle was also 63.9% longer than expected, with 37.9% thinner walls.

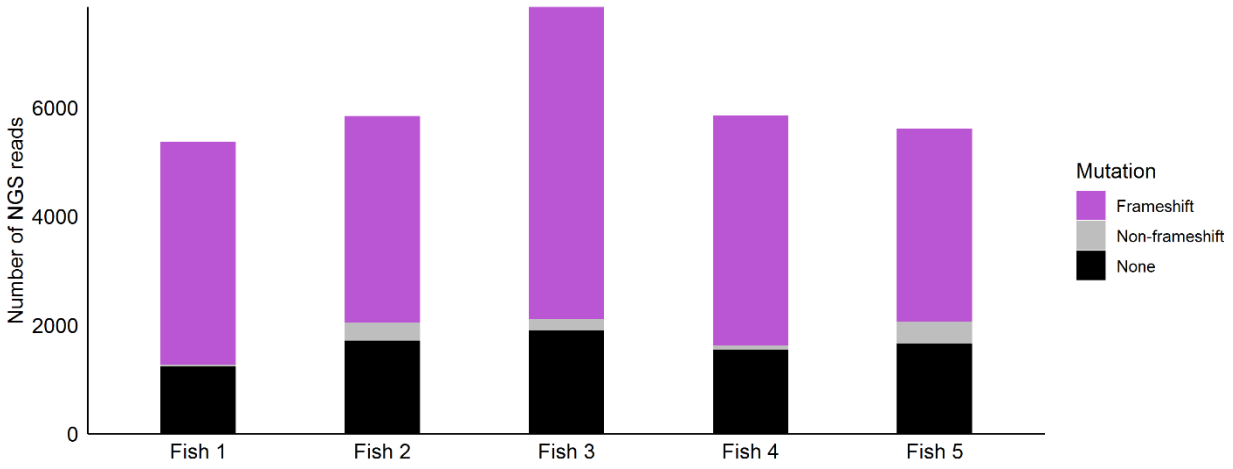


Figure 5.9: Genotyping of zebrafish larvae injected with *polg* guide RNAs. The number of NGS reads with frameshift mutations (purple bars), non-frameshift mutations (grey bars) and no mutation (black bars) are depicted (y-axis) per larva (x-axis). Mutation rates are taken at the site of greatest insertion or deletion rate at that point in the larva. *NGS, next-generation sequencing*

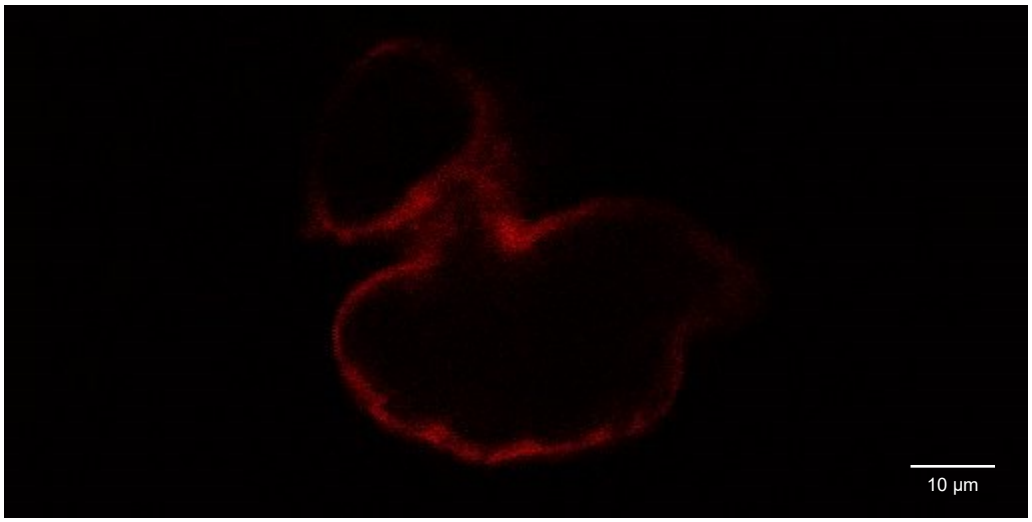


Figure 5.10: Multiphoton image of 3 dpf *polg* knockout Fish 1 ventricle at end-diastole. Pictured is a *cmlc2*:mCherry transgenic zebrafish embryo at 32x magnification, after treatment with *polg* knockout sgRNA and Cas9. Cardiomyocytes fluoresce red and were visualised at a laser wavelength of 720 nm.

5.3 mRNA overexpression model

5.3.1 Cloning and mutagenesis of POLG

The entire protein-coding region of *POLG* was amplified by high-fidelity PCR (Figure 5.11). A band size of > 3 kb was obtained, as expected (*POLG* amplicon: 3.7 kb). Although non-specific PCR products < 1.5 kb and > 10 kb were observed, the PCR was not optimised as gel purification was used to isolate the specific PCR products directly from the gel.

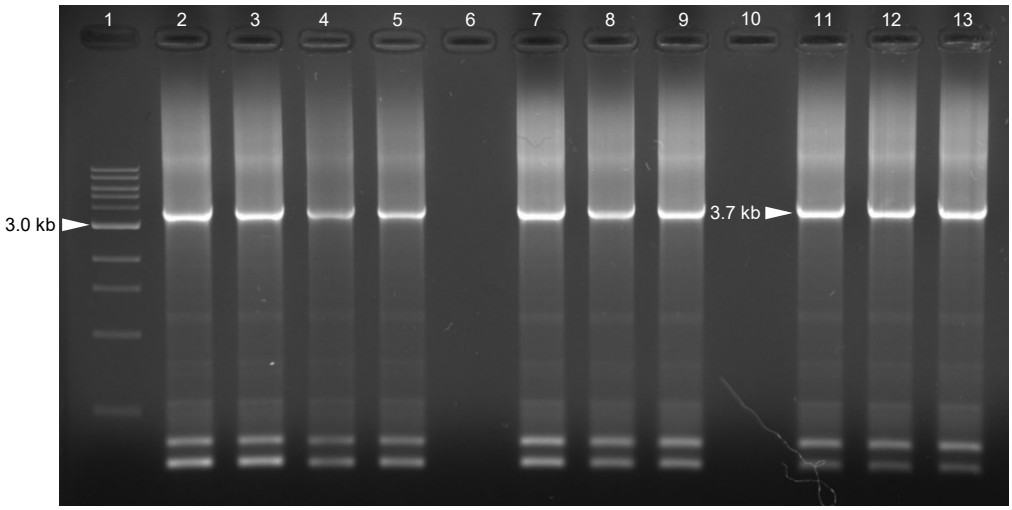


Figure 5.11: Image of PCR amplification of *POLG* cDNA. PCR products electrophoresed on a 1% (w/v) agarose gel. Lane 1 contains Quick-Load® 1 kb DNA ladder (NEB), and lanes 2-5, 7-9 and 11-13 contain PCR products. Lanes 6 and 10 are empty. The arrowheads denote the size of the DNA in the molecular ladder and of the *POLG* cDNA, in kilobases.

The *POLG* coding sequence was successfully cloned into two out of ten plasmids (Figure 5.12), indicating a ligation efficiency of 20%. This was determined by the presence of a band of the expected *POLG* cDNA size (3.7 kb) as well as a band the size of the insert and PCS2+ vector, undigested (7.8 kb), occurring on the gel (Figure 5.12, lanes A3 and B2). The absence of the vector (4.1 kb) on the gel in lanes A3 and B2 could be due to partial digestion. However, the presence of the vector and insert in the colonies was confirmed by Sanger sequencing. The other eight colonies were transformed with PCS2+ vector without the insert, as indicated by single bands the size of PCS2+ (4.1 kb) on the gel (Figure 5.12, lanes A1, A2, A4, A5, B1, B3-5).

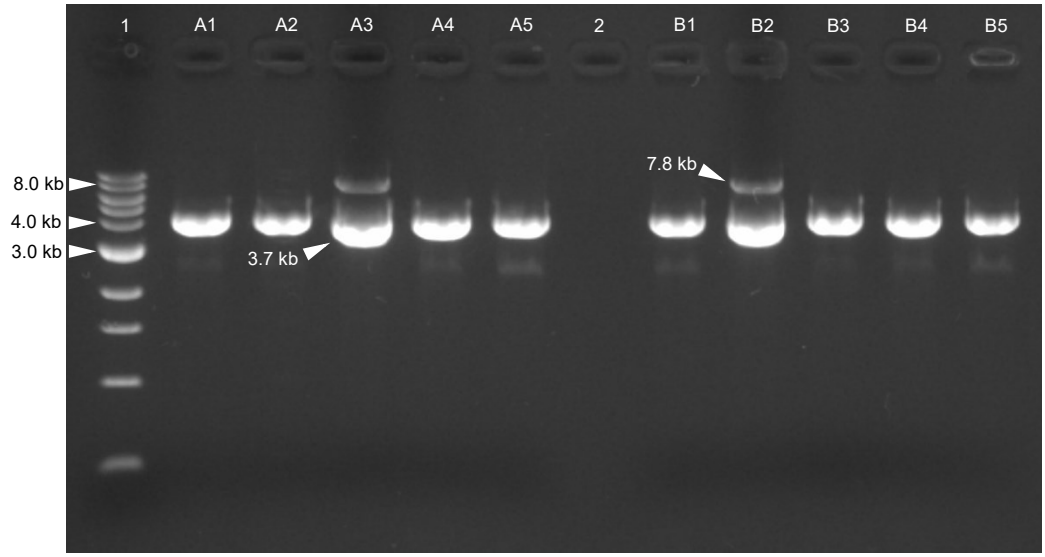


Figure 5.12: Image of restriction enzyme digestion of ligated plasmids. Digested products electrophoresed on a 1% (w/v) agarose gel. Lane 1 contains Quick-Load® 1 kb DNA ladder (NEB), and lanes A1-5 and B1-5 contain digested products. A and B refer to different plates from which colonies were picked. Lane 2 is empty. The arrowheads denote the size of the DNA in the molecular ladder and of the *POLG* cDNA, in kilobases.

Plasmid A3 (Figure 5.12) was used as a template for site-directed mutagenesis and plasmid B2 was kept as wild type *POLG*. The mutagenised plasmid A3 and wild type plasmid B2 were *in vitro* transcribed to produce mRNA for injection into zebrafish embryos (Figure 5.13). The site-directed mutagenesis reaction was confirmed to introduce the c.2492A>G single nucleotide change to the coding sequence, while plasmid B2 remained unchanged (Figure 5.14).

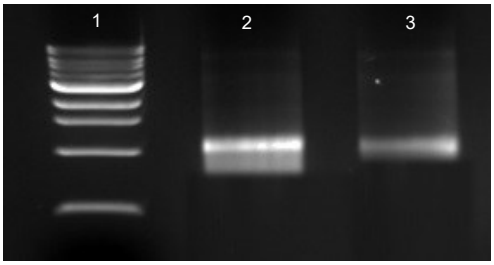


Figure 5.13: Image of mRNA for injection experiments. The *in vitro* transcription reaction products electrophoresed on a 1% (w/v) agarose gel. Lane 1 contains Quick-Load® 1 kb DNA ladder (NEB), lane 2 contains mRNA from mutagenised plasmid A3 and lane 3 contains mRNA from unmutated plasmid B2.

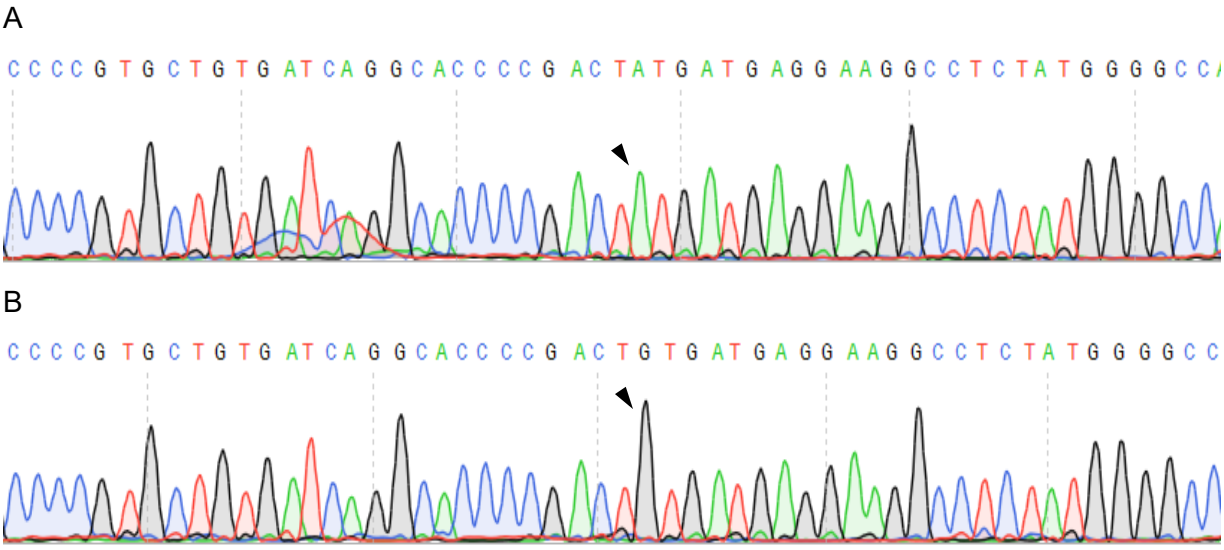


Figure 5.14: Sequencing of the *POLG* region of site-directed mutagenesis. The electropherograms show (A) plasmid B2 in which *POLG* was not mutated (the wild type control plasmid) and (B) plasmid A3 in which the c.2492 base was successfully changed to G. The position of the A>G change is indicated with arrowheads.

5.3.2 Injection optimisation

In order to determine the optimal dose for comparing wild type *POLG* to mutated *POLG*, zebrafish embryos were first injected with wild type *POLG* at doses ranging from 25 pg to 200 pg. Although fish at all doses were developmentally normal within the first 5 dpf, a slight but significant reduction in heart rate was observed in larvae injected with 200 pg wild type *POLG* (Figure 5.15). No significant differences were observed at other doses (Figure 5.15).

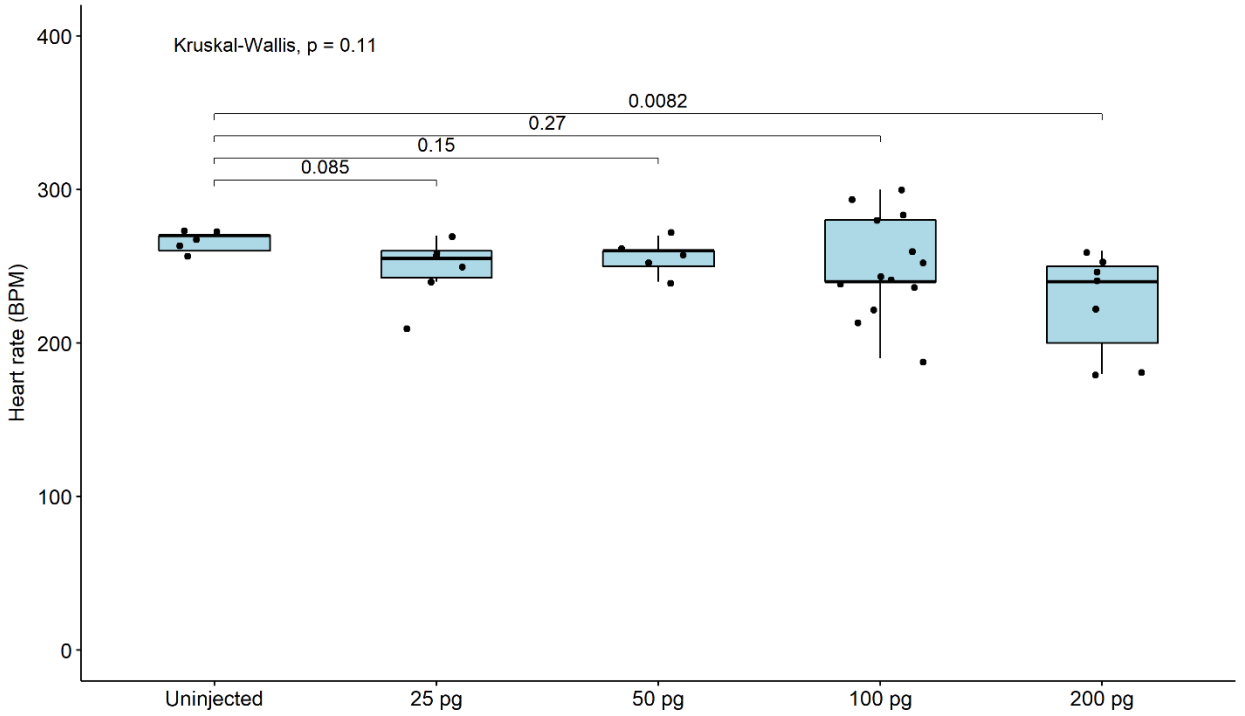


Figure 5.15: Wild type *POLG* mRNA dose analysis. Average heart rates in uninjected controls and injected larvae (at doses of 25 pg, 50 pg, 100 pg or 200 pg) are displayed. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate in each group are represented by box-and-whisker plots, in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test).

5.3.3 Injection of mutant POLG

A total of 135 zebrafish embryos were injected with mRNA carrying the *POLG* c.2492A>G variant. By 3 dpf, a proportion (16.3%) of the mutant *POLG*-expressing zebrafish had developmental deformities including cardiac oedema (14.8%), dorsalisation (5.9%) and/or body curvature (11.1%) (Figures 5.16 and 5.17). These were not observed in the injected or uninjected control zebrafish. The presence of developmental abnormalities was not the focus of this investigation, and cardiac oedema and dorsalisation may indicate non-specific effects of the injection itself on normal developmental processes, so the deformed fish were removed from subsequent analyses.

A significant reduction in heart rate and function in *POLG* mutant zebrafish was observed compared to all controls (Figure 5.18). The distributions of heart rate did not differ amongst the control larvae (Kruskal-Wallis test, p = 0.940), but the heart rate was significantly lower amongst

the mutant RNA-injected zebrafish (Figure 5.18). This was observed over five separate replicates (Appendix U). Overall, the median heart rate was reduced by 22.2% in fish expressing mutant *POLG* compared to control zebrafish ($p < 0.001$).

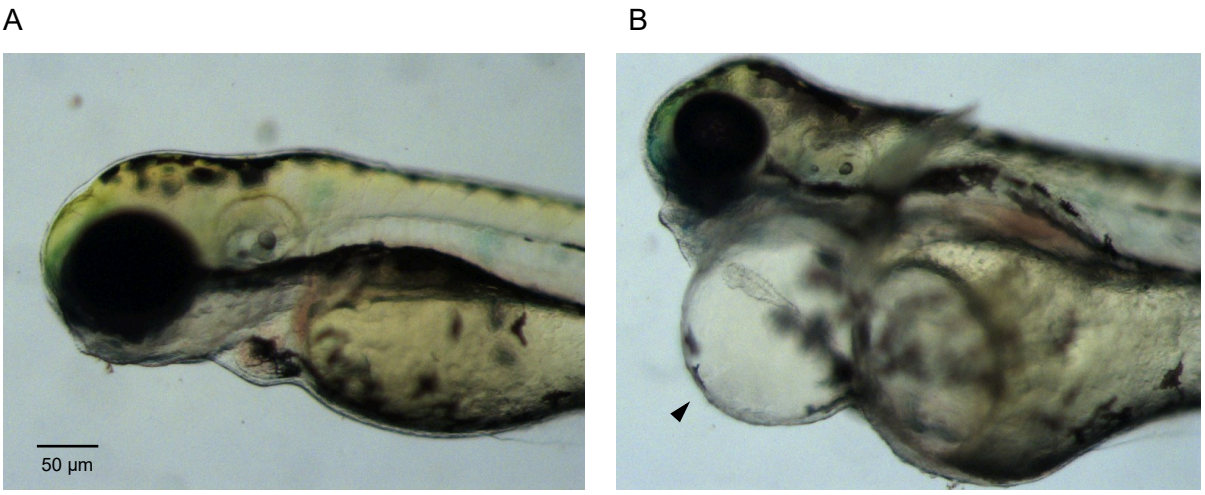


Figure 5.16: Close-up view of zebrafish following injection experiments. Shown are bright-field images of *POLG* mutant zebrafish larvae at 4 dpf following injection of (A) 10 pg wild type *POLG* mRNA and (B) 10 pg mutant *POLG* mRNA with resulting cardiac oedema (indicated with an arrowhead). Images were taken at 10X magnification.

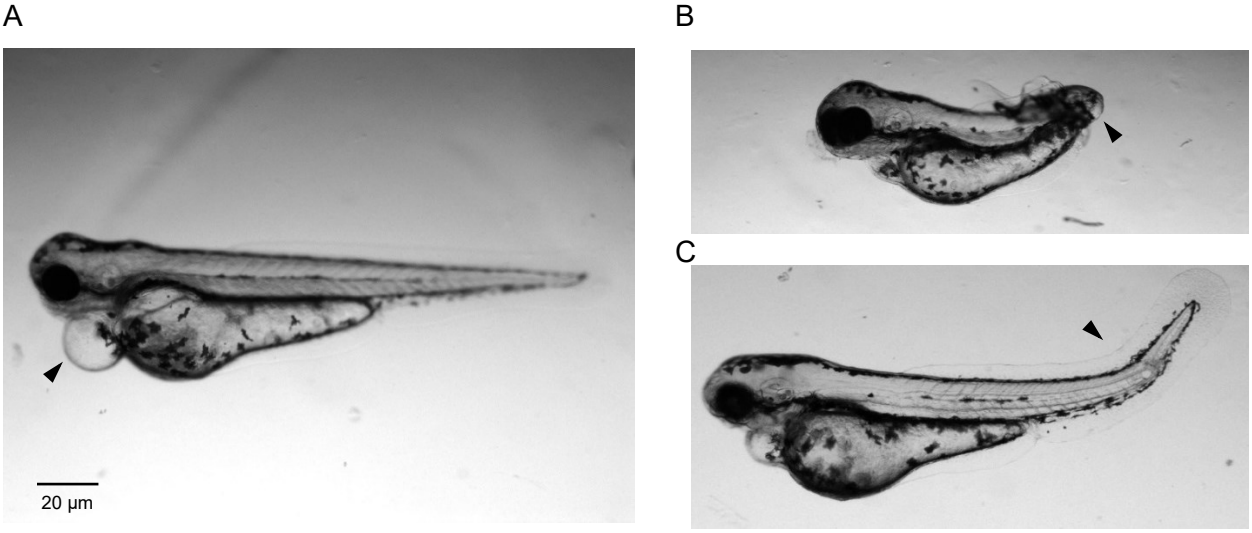


Figure 5.17: Whole-body view of zebrafish following injection experiments. Shown are bright-field images of zebrafish larvae at 4 dpf after injection with *POLG* mutant mRNA. Larvae were exhibiting (A) cardiac oedema (indicated with an arrowhead), (B) dorsalisation (indicated with an arrowhead) and (C) body curvature (indicated with an arrowhead). Images were taken at 4X magnification.

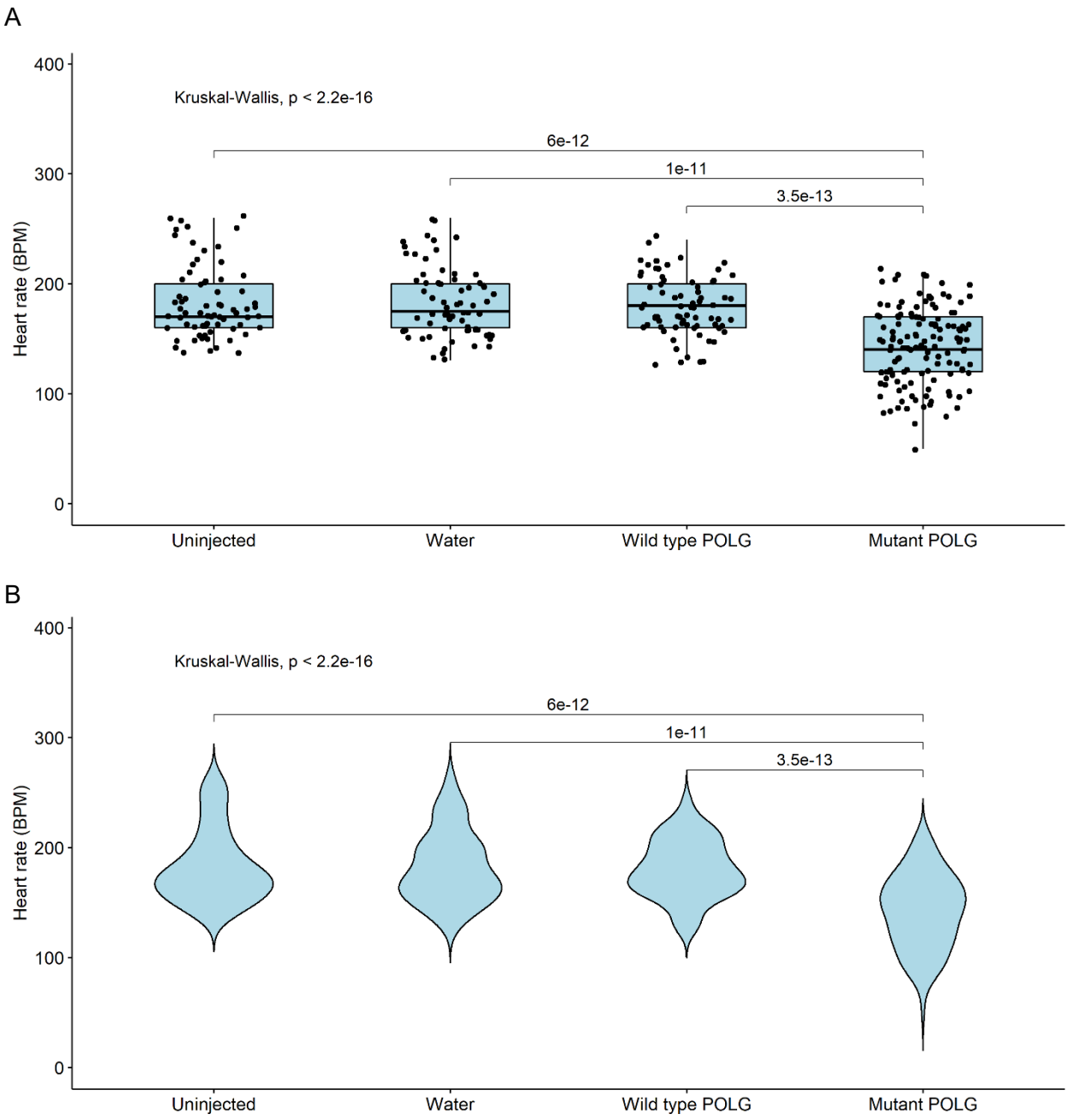


Figure 5.18: Average heart rates in each experimental group. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by (A) box-and-whisker plots and (B) violin plots. In A, the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. In B, the area of the plot indicates the kernel probabliity denisity. Significance bars show the p-value the mutant mRNA injected group (n = 135) to each control group (n = 75 per group) (two-sided Mann-Whitney U test).

In most of the experimental groups, the ventricle length and overall area were directly correlated, as may have been expected (Figure 5.19). Ventricular area also correlated with the width (Appendix V). Mutant zebrafish ventricles were significantly smaller compared to the controls (Figure 5.20A-C; Appendix W) (area: p = 0.002; length: p < 0.001), and the average blood flow activity was decreased in the mutant fish (p = 0.023) (Figure 5.20D). There were no significant differences in heart morphology or function in the water- or wild type mRNA-injected zebrafish larvae as compared to the uninjected control zebrafish.

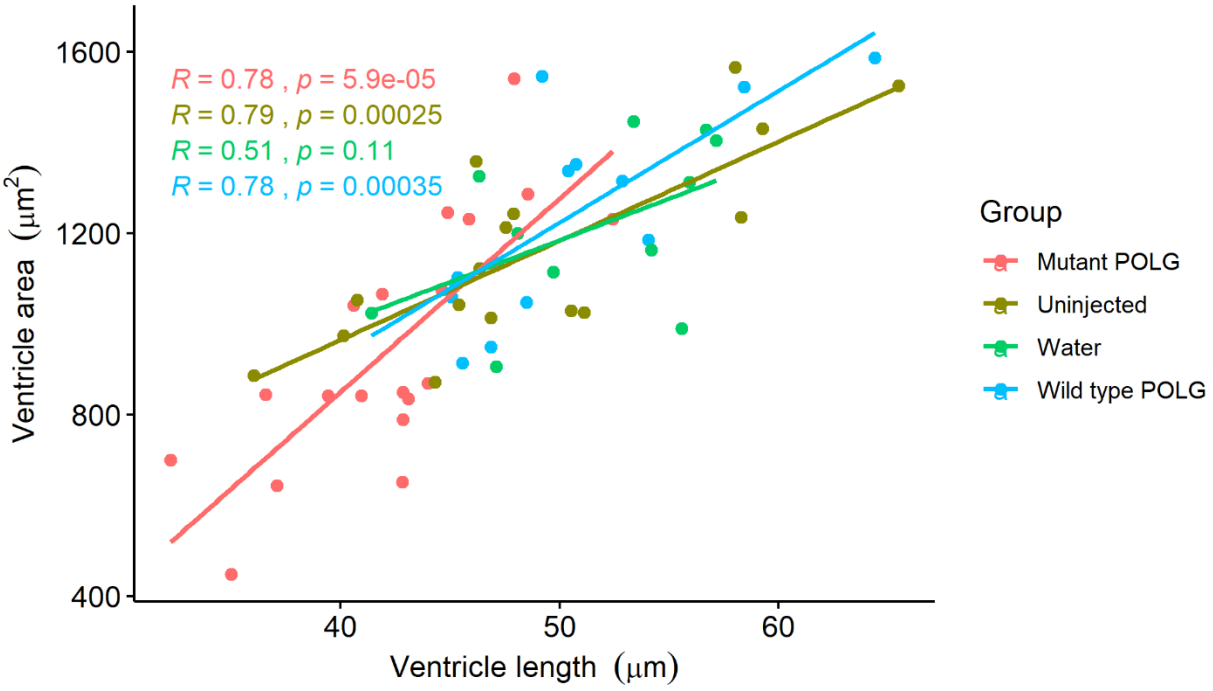


Figure 5.19: Correlation of ventricular length and area amongst the mRNA experimental groups. The length is represented on the x-axis, while area is on the y-axis. The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

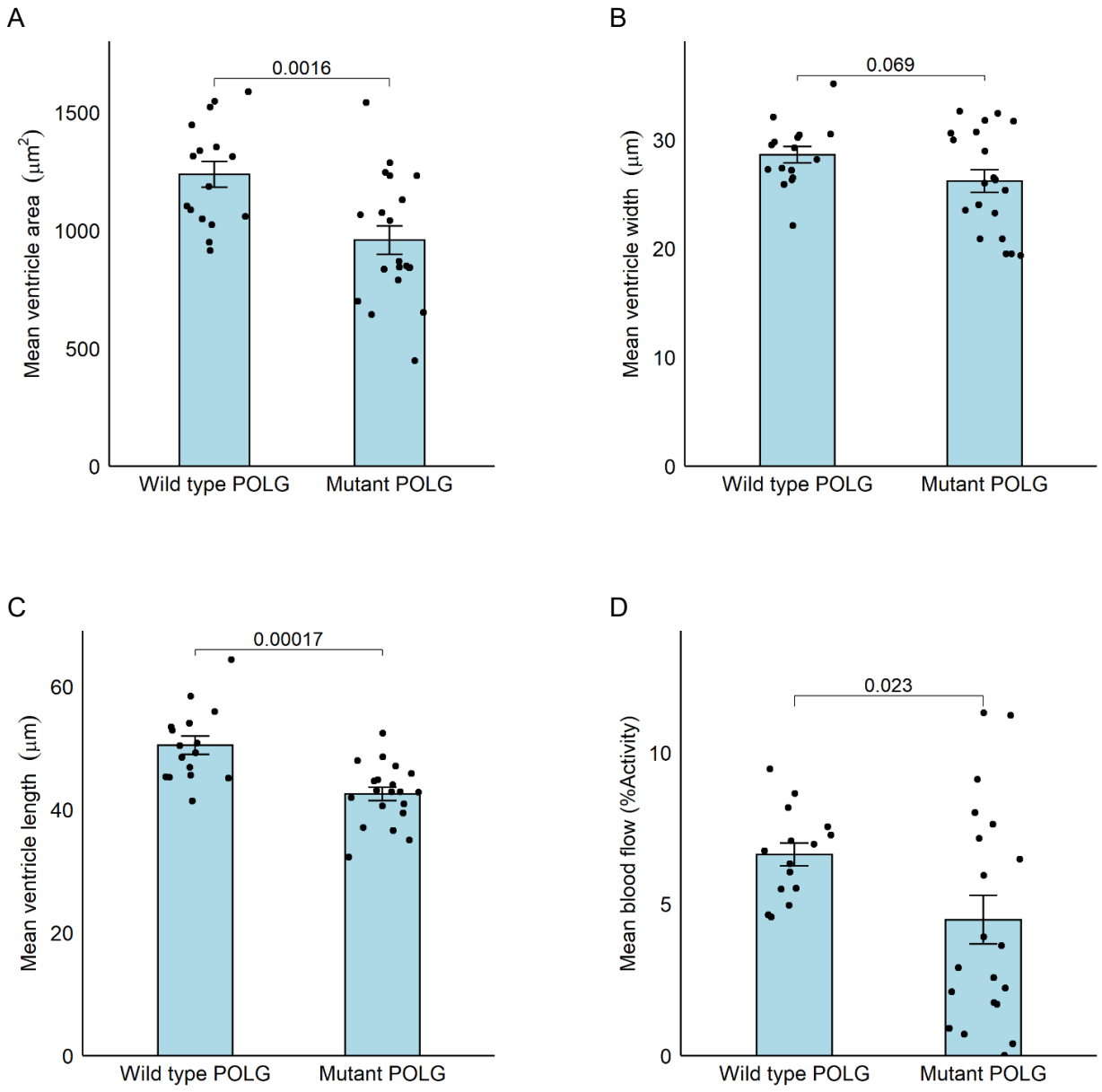


Figure 5.20: Morphological measurements of mRNA-injected zebrafish larvae at 4 dpf. (A) Mean ventricle area, (B) mean ventricle width, (C) mean ventricle length, (D) mean flow rate in both *POLG* mutant (n = 20) and control (n = 16) zebrafish. Only larvae with clear images of the heart were included in these analyses; one statistical outlier was removed from the control group in the analysis of blood flow (n = 15). Controls were embryos injected with wild-type *POLG* mRNA. The injection groups are on the x-axes while the morphological measurements are on each y-axis. Significance bars show the p-value when comparing measurements between the groups (two-sided t-tests).

5.4 Discussion: zebrafish models of cardiomyopathy

In this chapter, CRISPR was used to target and disrupt the zebrafish orthologues of two known cardiomyopathy genes (*myh7* and *pkp2*), as well as the candidate ACM gene *polg*. Due to lengthy image acquisition times, a total of five larvae were imaged per experimental group. As a result, this study was mostly underpowered to perform intergroup comparisons (Appendix R). However, despite the small sample sizes, cardiomyopathy phenotypes could be distinguished in some cases. It should be noted that correction for multiple statistical testing was not performed in this investigation, and that the CRISPR results still require replication in an independent experiment.

POLG was identified as a putative cardiomyopathy gene in a South African family with ACM (Chapter 4). The clinical significance of the gene is unknown, and although mutations in *POLG* have been associated with a wide variety of disorders including Parkinson's disease and skeletal myopathy (Cohen, Chinnery, and Copeland 2018), *POLG* genetic variation has not been associated with cardiomyopathy to date. All *polg* mutant larvae in this investigation had significantly larger ventricles compared to control zebrafish, with significantly thinner walls. The reduction in wall size may indicate myocyte cell death in the region, a hallmark feature of ACM; it may alternatively indicate developmental problems in the larval heart. Regardless of the cause, these deviations from normal in the mutant larvae indicate that *polg* may play a role in the morphological development of the zebrafish heart. Disruption of *polg* by TALENs has been described, although the cardiovascular effects were not investigated or reported on (Rahn et al. 2015). What is noteworthy is that these homozygous *polg* mutants survived up to three weeks, meaning that stable mutations could perhaps be studied for longer periods if desired.

Because of the uncertainty surrounding the interpretation of zebrafish phenotypes of cardiomyopathy, knockouts of the well-established cardiomyopathy genes *myh7* and *pkp2* were included. Human *MYH7* is traditionally an HCM gene, although can be associated with DCM in human patients, as demonstrated in Chapter 3. There were no drastic effects of *myh7* knockout in zebrafish, although the ventricular area was increased in all larvae, and one larva had thinner ventricular walls than expected; these are DCM characteristics. In a study by Auman et al, the described zebrafish mutant *haf* carries a stable, recessive truncating *myh7* mutation: the lack of *myh7* in this line was shown to cause ventricular distension and enlargement, as well as contractility defects (Auman et al. 2007). The relatively mild phenotypic effects in the present study are possibly due to the large number of non-frameshift insertions and deletions that this

CRISPR reaction produced (Appendix T), or may be due to other *MYH7* orthologues such as *myh7l* present in zebrafish that may compensate for the gene disruption (Shih et al. 2015). Similarly in mice, predominant expression of *myh6* in the ventricles means that it is not an appropriate model for studying *myh7*-related cardiomyopathies (Duncker et al. 2015, Lompré, Nadal-Ginard, and Mahdavi 1984); it may be that both *myh7* and *myh7l* need to be targeted in zebrafish to induce a myocardial phenotype, although the results of prior *myh7* disruption in zebrafish do not support this (Auman et al. 2007). Indeed, the *pkp2* CRISPR reaction also resulted in largely non-frameshift insertions and deletions generated in embryos; the phenotypic consequences, however, were more severe. *PKP2* only has one zebrafish orthologue and is a desmosomal gene associated with ACM in humans. Targeting the gene in this study led to thicker ventricle walls and smaller, misshapen ventricle chambers. In previous reports, morpholino knockdown of *pkp2* in zebrafish led to, in the severest case, cardiac oedema, blood pooling and small, abnormally developed ventricles (Moriarty et al. 2012). Because a fluorescent transgenic line was used in this experiment, blood flow and oedema were not measured in the mutant zebrafish; however, the finding of smaller, deformed ventricles is in keeping with the *pkp2* knockdown study (Moriarty et al. 2012).

There are two important limitations of the CRISPR technique used here. Firstly, the use of an uninjected control is not preferable as it does not eliminate the potential non-specific effects of the injection itself on embryonic development. Rather, an injection control would allow delineation that the observed effects may be due to the gene knockout itself. Although there is no gold standard for CRISPR injection controls as yet, examples that may be suitable include sgRNA without Cas9 (or vice versa), a scrambled sgRNA which is composed of the same nucleotides but in a different order, or sgRNAs that target different regions of the gene (Masselink 2021). A second limitation is that different genotypes were generated in the F0 generation. Ideally, mutant zebrafish should be bred over two or three generations to produce stable knockouts with no off-target effects. The benefit of this approach is that single mutations can be compared. However, this was not possible due to high embryonic lethality of the mutations, especially amongst the *polg* knockout group. Therefore, while some of the embryos were observed to develop signs of cardiomyopathy phenotypes in this pilot investigation, it cannot be said with certainty that these are not non-specific effects of the microinjection. As discussed above, conditional knockouts may be more suitable; other techniques that may remedy this are ‘knock-in’ CRISPR reactions or mRNA overexpression. CRISPR knock-in utilises homology-directed repair mechanisms to incorporate a repair template carrying a desired mutation into the zebrafish genome, as opposed

to the nonhomologous end-joining approach used in this study. The knock-in approach is currently limited by extremely low efficiency, although factors such as the donor template length and type of donor DNA may be optimised (Collery and Link 2019, Eschstruth, Schneider-Maunoury, and Giudicelli 2019); successful approaches may still require hundreds to thousands of zebrafish bred over several generations (Zhang, Zhang, and Ge 2018). The overexpression of mRNA is a comparatively simple technique and was explored in more detail in analysis of *POLG* c.2492A>G.

The *POLG* c.2492A>G variant was found in a South African family with ACM, after exome sequencing and the identification of no variants in known cardiomyopathy genes (Chapter 4). The variant was found in four affected individuals with a variable disease phenotype. Due to ambiguity in the literature regarding the role of *POLG* in cardiomyopathy, and the lack of any other genetic candidate in this family, this variant was selected for functional investigation to determine if the disease seen in Family 4 could be recapitulated variant in zebrafish.

The previous CRISPR/Cas9 knockout model was suggestive that zebrafish *polg* may play a role in heart development and/or functioning (Section 5.2.5). However, the model was restricted to five larvae due to the imaging technique used and was also limited due to genotypic heterogeneity between mutant zebrafish. In the mRNA overexpression model, however, light microscopy was used to visualise the hearts of the larvae, meaning that substantially more zebrafish could be analysed, reaching statistically significant numbers (Appendix R). Because mutagenised mRNA was used instead of CRISPR/Cas9-mediated gene disruption, the genotypic heterogeneity within experimental groups was reduced. Potential limitations of this technique are that gene overexpression may not mimic physiologically relevant levels of mRNA, and the endogenous expression of zebrafish *polg* is not abolished. In addition, when injecting human mRNA into zebrafish embryos, the stability and function of the resulting protein may be compromised (Becker et al. 2003), for example due to differing physiological conditions between the species. For these reasons, several control groups were established, and a dose-response analysis was performed to identify the dose at which wild type *POLG* will not induce significant cardiac dysfunction in zebrafish. Due to the reduction in mRNA concentrations over time, the model is limited to the first few days of development and therefore cannot be used to model adult zebrafish; however, other successful recapitulations of ACM phenotypes have been achieved within 2-4 dpf (Giuliodori et al. 2018, Heuser et al. 2006, Martin et al. 2009, Moriarty et al. 2012), so the analysis of larvae at 4 dpf may be appropriate in this investigation.

Expression of the mutation in zebrafish larvae induced several cardiac effects suggestive of arrhythmia and cardiomyopathy. Extracardiac phenotypes such as dorsalisation and body curvature were also observed, but these were not the focus of this investigation. In comparison to siblings expressing the wild type form of the gene, larvae expressing mutant *POLG* developed smaller hearts with a slower heartbeat and reduced cardiac output, suggesting an overall reduction in heart function. Similar results have been described following mutation of other ACM genes in zebrafish. A reduced heart rate and size were reported in *jup* mutant zebrafish (Martin et al. 2009), while *dsc2*, *dspa* and *dspb* null fish had lower heart rates (Heuser et al. 2006, Giuliodori et al. 2018), and morpholino knockdown of *pkp2* in zebrafish resulted in smaller ventricles, reduced blood flow as well as body curvature (Moriarty et al. 2012). Unfortunately, the ventricular thickness could not be measured because of the strain of zebrafish used.

In Chapter 4, the results of prior animal model studies of *POLG* were discussed. Two prominent phenotypes observed in the affected family members in this study (cardiomyopathy and diffuse ventricular fibrosis) were also present in transgenic mice expressing *Polg* point mutations (Koczor et al. 2013, Lewis et al. 2007). However, arrhythmias have never been reported in *Polg* mutant animal models. To our knowledge, this may be the first evidence that *POLG* could play a role in arrhythmia phenotypes in both humans and zebrafish. Zebrafish hearts are capable of regeneration following cardiac injury, and therefore undergo limited, transient fibrosis (Sánchez-Iranzo et al. 2018). The fact that expression of *POLG* p.Y831C in zebrafish could recapitulate, at least in part, the disease phenotype seen in Family 4 suggests that, although additional genetic factors may play a role in disease expression, *POLG* p.Y831C is a likely disease-contributing variant in this family.

Chapter 6 Conclusion and future perspectives

6.1 Principal findings

Cardiomyopathies are rare disorders of the myocardium that constitute an important cause of heart failure, arrhythmias and sudden death. They frequently present as heritable disease, but many causative genes are yet to be identified. The aims of this study were to investigate the genetics of familial cardiomyopathy using exome sequencing, and to establish methods of functional validation. Using these techniques, pathogenic or likely pathogenic mutations were identified in two out of five families studied, while three VUSs with moderate or strong pathogenic potential were identified in two other families (Figure 6.1). This study illustrates the benefits and challenges of using exome sequencing to study cardiomyopathy in Africa.

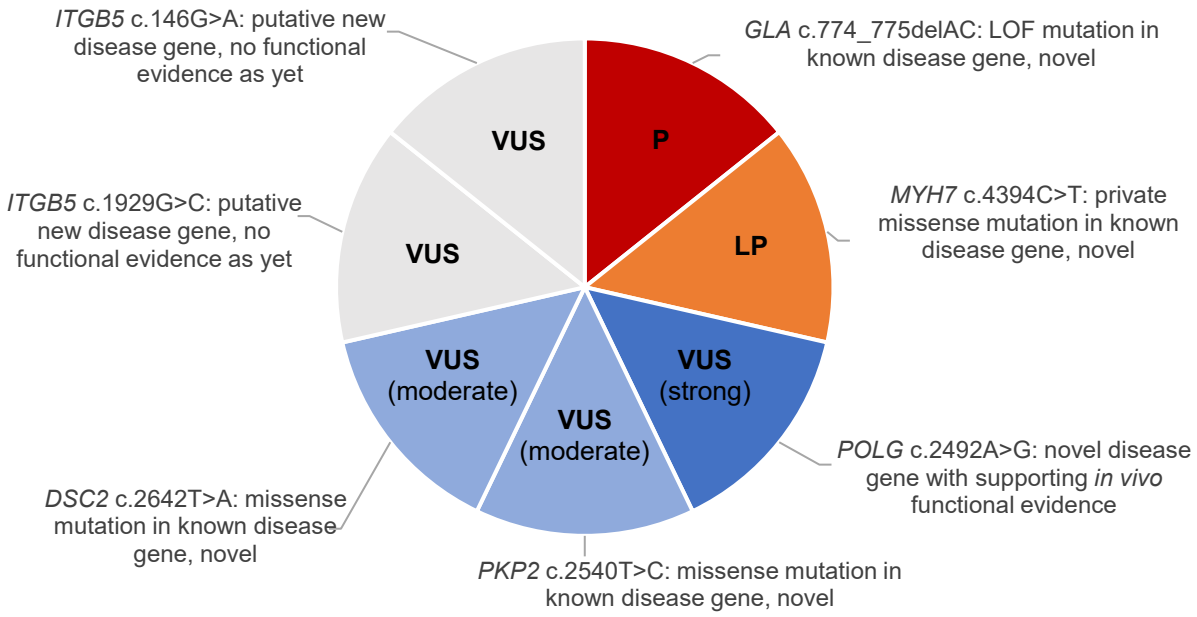


Figure 6.1: Summary of the genetic findings in this investigation. Pie chart in which red slices indicate pathogenic variants (P), orange slices indicate likely pathogenic variants (LP) and blue slices indicate VUSs with strong or moderate evidence of pathogenicity. Grey slices indicate VUSs. Pathogenicity as described by ACMG criteria (Richards et al. 2015). *ACMG, American College of Medical Genetics and Genomics; LOF, loss-of-function; VUS, variant of uncertain significance*

The pathogenic and likely pathogenic mutations, *GLA* c.774_775delAC (Family 2) and *MYH7* c.4394C>T (Family 1), were both novel mutations in known disease-causing genes. The identification of these variants as causative of HCM phenocopy and DCM, respectively, indicates an allelic heterogeneity of cardiomyopathy that has not been described in South African patients to date. In addition, two desmosomal gene VUSs were found in a family with DCM (Family 3): although desmosomal gene mutations have been reported in DCM patients before, their contribution to DCM is unclear at this stage, and neither variant segregated entirely with disease despite fulfilling many other pathogenicity criteria (absence in population databases, evolutionary conservation and *in silico* prediction of pathogenicity). Oligogenic or more complex inheritance patterns may underly the phenotypic manifestations in Family 3, and more research is needed to resolve the roles of the VUSs identified in these individuals.

Exome sequencing is well suited for the identification of candidate disease genes, which was necessary in Families 4 and 5 where no mutations were identified in known cardiomyopathy genes. Variants of interest were found in the candidate genes *ITGB5* and *POLG*; however, proving the disease-causing potential of candidate genes is challenging based on literature searches alone, especially in the absence of additional reported mutation carriers (as was the case for *ITGB5*), or conflicting reports of variant pathogenicity (as was the case for *POLG*). While *ITGB5* was identified as a candidate too late for functional investigation, *POLG* variation was modelled using CRISPR/Cas9 knockout and mRNA overexpression in zebrafish larvae. The results of CRISPR knockout suggested that *polg* may play a role in cardiac development, while overexpressing mutant *POLG* mRNA in zebrafish larvae led to signs of arrhythmia. As the cardiomyopathy seen in Family 4 was characterised by arrhythmia and left ventricular fibrosis, these results suggest that *POLG* variation could well contribute to the disease phenotype observed in this family.

This study also identified two potential genetic modifiers, *SORBS2* c.322T>C and *KCNK10* c.1052A>G, in ACM and DCM families respectively. Genetic modifiers are variants that, when inherited with a disease-causing mutation, can affect the expressivity or penetrance of the mutation, and may account for phenotypic variability that is observed in some families or between unrelated mutation carriers. As with candidate disease genes, proving their modifying potential may require functional investigation even though both *SORBS2* and *KCNK10* have documented roles in cardiac function. Genetic modifiers have been documented in heritable cardiomyopathies, particularly HCM and DCM, where they can comprise rare, deleterious variants that are co-

inherited with known disease-causing mutations (Long et al. 2015, Lopes et al. 2015) or common polymorphisms in cardiovascular genes that associate with disease severity or outcomes in mutation carriers (Kumar et al. 2018). Similar modifiers have been demonstrated in South African HCM patients, in which *AGTR1* and *MYBPH* polymorphisms were shown to associate with the degree of cardiac hypertrophy (Carstens et al. 2011, Mouton et al. 2016). Therefore, the role of modifying variants such as those in *SORBS2* and *KCNK10* could be further explored, as more evidence will be needed to determine the true role of these variants in disease. It should also be noted that in these families, only rare mutations were considered as modifiers due to the current bioinformatic challenges in analysing variants of low impact.

6.2 Practical implications

Current guidelines for the diagnosis and management of DCM, HCM and ACM from the ESC, American Heart Association and Heart Rhythm Society recommend the construction of a detailed family history (usually at least three generations), periodic clinical screening of first-degree relatives by echocardiography and ECG, and genetic testing in the case of familial disease (Bozkurt et al. 2016, Elliott et al. 2014, Gersh et al. 2011, Pinto et al. 2016, Towbin et al. 2019). This approach allows at-risk family members to be identified and, when disease-causing mutations are found, cardiac follow-up can be guided by the presence of the mutation in family members. While such international guidelines are typically subscribed to amongst African cardiology societies, limited resources as well as incomplete knowledge of the genetics of cardiomyopathy in Africa means that genetic testing is unlikely to be routinely applied in these settings. However, given the familial implications of cardiomyopathies, it may be advisable to obtain a family history from index cases and consider genetic testing in the instance of a family history of CVD.

Because targeted sequencing can be used to accurately screen numerous cardiomyopathy genes in a single experiment, NGS panels are becoming routinely incorporated into the diagnosis of heritable cardiomyopathy in some countries (Martin et al. 2019), although there is a degree of uncertainty about the disclosure of VUSs to patients (Vears, Sénécal, and Borry 2017). Nevertheless, the increasing availability of NGS panels means that it now may be considered more cost-effective to conduct genetic testing on asymptomatic family members than to phenotypically screen them (Catchpool et al. 2019, Ingles et al. 2012). This means that, if a patient is found to carry a disease-causing mutation, this could allow clinical follow-up to be limited to

relatives who also carry the mutation and thus release asymptomatic non-carriers from regular screening. However, in order to realise these benefits of genetic testing, a better understanding of the genetics of cardiomyopathy in Africa is crucial.

An objective of the IMHOTEP project is to test the applicability of an international 48-gene cardiomyopathy panel in South African and Mozambiquan cardiomyopathy patients (Kraus 2019). No large-scale genetic analysis of African cardiomyopathy patients has been conducted to date, and this ongoing research will be essential in determining, firstly, whether genetic testing should be routinely performed in South African cardiomyopathy patients and, secondly, if panel sequencing or exome sequencing should be the primary testing platform used in African cardiomyopathy. Both NGS platforms, while carrying increasingly similar costs (Mazzarotto, Olivotto, and Walsh 2020), have advantages and limitations (Chapter 1). If panel testing of the IMHOTEP cohort yields a mutation rate that is similar to the populations on which the panels were based, this will indicate that panel testing of African cardiomyopathy patients may be warranted in the clinical setting, especially when familial disease is observed. Two sequencing studies of African cohorts have been conducted using sarcomere gene panels, comprising a total of 80 HCM patients from South Africa and Tunisia; both studies reported mutation yields less than 30% (Jaafar et al. 2016, Ntusi et al. 2016). These figures are considerably lower than the yield of approximately 50% for HCM reported internationally (Marian and Braunwald 2017), although the larger patient cohort in IMHOTEP (target recruitment of 750 patients) and larger gene panel used may lead to greater mutation rates.

Previous research conducted in African patients suggests that the genetic basis of cardiomyopathy, while underexplored (Shaboodien et al. 2020), may differ slightly compared to other populations, with a high proportion of founder mutations as well as many novel, previously unreported variants occurring in established cardiomyopathy genes (Moolman-Smook et al. 1999, Moolman-Smook et al. 1998, Mouton et al. 2015, Ntusi et al. 2016, Watkins et al. 2009). Given the lower yield of genetic testing reported so far, it is likely, too, that additional disease-causing genes are yet to be discovered in the African patient population. This is reflected by the results of this study, in which novel mutations were found in three out of five families, and candidate genes identified in the other two families. There is therefore much scope for future research into the genetics of cardiomyopathy in Africa.

6.3 Future perspectives

6.3.1 Genetic studies of African cardiomyopathy

Although the field of genetic research into cardiomyopathy has been considerably broadened through the advent of NGS technologies such as exome sequencing (Chapter 1), NGS has had limited utility in Africa so far. To our knowledge, there have only been six reports of NGS in African cardiomyopathy patients or families: four studies used targeted sequencing (Adadi et al. 2018, Choung et al. 2017, Jaafar et al. 2016, Ntusi et al. 2016), and two used exome sequencing (Lahrouchi, Behr, and Bezzina 2016, Mayosi et al. 2017), one of which reported the novel ACM gene *CDH2* (Mayosi et al. 2017). Aside from NGS, few candidate gene investigations have been described in Africa (Shaboodien et al. 2020). It is perhaps unsurprising then that little is known about the genetics of inherited cardiac diseases in African patients. This may be remedied through NGS studies, as discussed above, although variant interpretation is a matter for some consideration.

Internationally accepted pathogenicity criteria emphasise loss-of-function or *de novo* mutations in known disease-causing genes when previously established pathogenic variants are not identified (Richards et al. 2015). Given the genetically unique African population, one may expect that many cardiomyopathy patients will not harbour reported pathogenic mutations; the results of this investigation support this notion. The next levels of evidence for pathogenicity, according to ACMG guidelines, are functional studies and segregation in multiple (preferably more than seven) family members (Kelly et al. 2018, Richards et al. 2015). This stresses the need for large family studies and/or well-established functional models of cardiomyopathy, which may be challenging to perform in Africa. In this study, the families were relatively small, ranging in size from one to four affected individuals. It may be that different pathogenicity criteria are needed in Africa, although the stringency of international guidelines reflects the recent finding that previously published mutations may not be disease-causing in light of modern, large-scale NGS population databases such as gnomAD in which these variants can now be interrogated (Tayal et al. 2017).

Using the ACMG guidelines as a framework and taking disease prevalence and penetrance into account, a MAF of < 0.01 was used in this investigation. We evaluated variants using evidence from population and gene/disease-specific databases, *in silico* prediction tools, our in-house variant database, and the relevant scientific literature. To this end, established criteria were used

to evaluate variants for potential pathogenicity, with evidence from the relevant databases and literature as the foundation for classification. While it is recognised that other studies are using MAFs that are more stringent (for example, a maximum MAF of 0.00004 has been advocated for pathogenic HCM variants (Whiffin et al. 2017)), the choice of a relatively low MAF for filtering in this study was motivated by the unique African population and the need to strike a balance between retaining variants of potential interest compared to excluding potentially causative ones. Our robust filtering approach led to identification of the *POLG* variant c.2492A>G, which initial functional work indicates may indeed play a role in disease presentation, as well as pathogenic and likely pathogenic mutations in known cardiomyopathy genes. That said, combining all of these factors along with comprehensive patient clinical information and history with recent powerful advances in bioinformatics (afforded by large population sequencing databases such as gnomAD and ExAC) (Karczewski et al. 2020)), may be key in identifying rare African mutations in genes otherwise unknown to cause disease.

To ascertain the role of common variation in cardiomyopathy in Africa, genome-wide association studies (GWASs) may be considered. Two recent GWASs have demonstrated the association of single-nucleotide polymorphisms with cardiac parameters and the risks of DCM and HCM in largely European populations (Harper et al. 2021, Pirruccello et al. 2020). Both studies used the genetic results to reliably predict the presence of DCM and HCM in validation cohorts, implying an additive effect in which accumulation of common variants with subtle phenotypic impacts can ultimately influence disease risk. These studies indicate the valuable insights that can be gleaned by genotyping large cohorts of cardiomyopathy patients and controls, and point to potential future avenues for research into cardiomyopathy in Africa, providing the logistical challenges can be overcome.

The utility of exome sequencing in identifying new genotype-phenotype correlations, as well as potential new disease genes for cardiomyopathy, means that it should be considered for researching the genetics of cardiac disease in Africa. Implications of genotype-phenotype correlations have been realised in international cohorts, where mutations can have consequences on individual prognosis and treatment depending on the affected gene (Paldino et al. 2018). Whether these or similar prognostic mutations exist in African patients remains to be determined. The zebrafish model system may also be of use when conducting future genetic studies of this population, as there is a need to validate variants with unclear clinical consequences (discussed below). It may well be that NGS coupled with zebrafish modelling could be used to close the gap

in our knowledge of the genetics of cardiomyopathy and, most importantly, provide answers to the families that are affected by heritable cardiac conditions.

6.3.2 Further investigation of POLG

Exome sequencing of Family 4 led to the identification of *POLG* c.2492A>G as a candidate variant, and functional investigation demonstrated that the variant could induce signs of arrhythmia when introduced to zebrafish larvae. The potential role of *POLG* in human cardiomyopathy and arrhythmic phenotypes is a finding that should be explored further. Although 39 ACM patients were screened for additional *POLG* mutations in this study, with the identification of the VUS c.3077G>A (p.R1026H) as a result, continued screening of the gene in an extended cardiomyopathy cohort may yield more variants of interest. Future functional work should be considered, including validation of the zebrafish as a model of *POLG* variation, for example by modelling additional *POLG* variants, and determination of the mechanism whereby mutations in *POLG* may lead to features of cardiomyopathy.

The mRNA overexpression approach used in this investigation has much scope for further refinement. One chief concern is the need to confirm that *POLG* mRNA is taken up and expressed in the heart during development. The incorporation of a fused fluorescent tag in the vector is an experiment that may address this, and would allow quantification of the POLG protein, determination of its expression patterns, and the ability to perform direct comparisons of embryos expressing equal amounts of the mutant *POLG*. An experimental protocol of this nature may be of interest, as this would serve to confirm (or refute) the cardiovascular effects of *POLG* mutation at the protein level. Confirmation of pathogenicity at this level would be strong evidence for the role of *POLG* in human cardiomyopathy; indeed, any other genes that are identified through future NGS studies may be examined in this way.

6.3.3 Functional modelling of variants in zebrafish

One of the greatest challenges to arise from exome sequencing is the identification of VUSs. In the literature review (Chapter 1), it was suggested that by screening more genes, NGS experiments may simply give rise to more unclear genetic findings. The results here certainly seem to support this notion, as VUSs were detected in three out of five families. When studying the genetic material of living participants, there is a profound need to resolve the uncertainty

surrounding these variants, as the results will have consequences on patients and their relatives. Sometimes, genetic results can also improve their treatment, as seen in Family 2 where the identification of a pathogenic *GLA* mutation indicates that the participants have Fabry disease, not HCM as originally indicated (Chapter 3). VUSs may be resolved through large family studies but when these are not possible, *in vivo* functional assays can be used to support or prove their disease-causing potential. An example of this is the *POLG* variant identified in Family 4, where exome sequencing combined with analysis in zebrafish demonstrated that mutations in *POLG* may contribute to ACM in that family.

The zebrafish is an emerging model for heart disease and heart development, due to numerous advantages which include the ease of genetic manipulation, large clutch sizes and larval transparency. In this investigation, two genetic models were used in zebrafish: CRISPR-mediated gene knockout, and mRNA overexpression. While CRISPR was used to successfully knock out three genes, the technique was limited due to small sample sizes and inconsistent genotypes. What is clear from this research is that future CRISPR projects will need to focus on the generation of stable gene knockout or 'knock-in' lines, propagated over at least two generations; this. An alternative model, mRNA overexpression, was explored. This model allowed investigation of a stable genotype expressed over the first five days of zebrafish development. It also enabled the analysis of substantially more larvae than were possible in the CRISPR investigation, and statistically significant numbers could be reached (Chapter 5). Although this model is limited to the first few days of development, this approach may prove useful in the first-line investigation of VUSs and other variants that are detected using NGS studies of cardiomyopathy, especially when modified to allow fluorescent protein production as described above (Section 6.3.2) or combined with mRNA rescue experiments (Badrock et al. 2020).

Alternatively, CRISPR genome editing can be used in some instances to create stable zebrafish mutants to recapitulate cardiovascular disease to great effect (Farr et al. 2018). In particular, the mutations *PKP2* c.2540T>C, *DSC2* c.2642T>A, *ITGB5* c.146G>A, and *ITGB5* c.1929G>C, which were identified during the course of this investigation, may be explored in this manner. Similar zebrafish models may also be used to resolve the role of genetic modifiers in cardiomyopathy, an underexplored research area that may warrant further investigation. This study identified two potential genetic modifiers, *SORBS2* c.322T>C and *KCNK10* c.1052A>G, and the expression of these variants in zebrafish, both alone and in combination with other disease-causing mutations, may be useful in delineating their contributions to disease phenotypes. Functional data of this

nature, supplemented with bioinformatic analysis and familial segregation, may provide clarity regarding the true disease-causing potential of these variants, as well as other variants that will be identified through further screening of African cardiomyopathy patients.

References

Abdallah A. M., Carlus S. J., Al-Mazroea A. H., Alluqmani M., Almohammadi Y., Bhuiyan Z. A., et al. 2019. "Digenic inheritance of LAMA4 and MYH7 mutations in patient with infantile dilated cardiomyopathy." *Medicina (Kaunas)* 55 (1). doi: 10.3390/medicina55010017.

Abegaz B. 1990. "The impact of echocardiography in the diagnosis of hypertrophic cardiomyopathy." *East Afr Med J* 67 (8):556-67.

Acharya D., Doppalapudi H., and Tallaj J. A. 2015. "Arrhythmias in Fabry cardiomyopathy." *Card Electrophysiol Clin* 7 (2):283-91. doi: 10.1016/j.ccep.2015.03.014.

Adadi N., Radi F. Z., Lahrouchi N., Hara L., Ratbi I., Elalaoui S. C., et al. 2018. "Inherited dilated cardiomyopathy in a large Moroccan family caused by LMNA mutation." *Anatol J Cardiol* 20 (1):65-8. doi: 10.14744/AnatolJCardiol.2018.69639.

Adalsteinsdottir B., Teekakirikul P., Maron B. J., Burke M. A., Gudbjartsson D. F., Holm H., et al. 2014. "Nationwide study on hypertrophic cardiomyopathy in Iceland: evidence of a MYBPC3 founder mutation." *Circulation* 130 (14):1158-67. doi: 10.1161/CIRCULATIONAHA.114.011207.

Addgene. "Vector Database." https://www.addgene.org/vector-database/2295/.

Adebayo A. K., Adebiyi A. A., Oladapo O. O., Ogah O. S., Aje A., Ojji D. B., et al. 2009. "Characterisation of heart failure with normal ejection fraction in a tertiary hospital in Nigeria." *BMC Cardiovasc Disord* 9:52. doi: 10.1186/1471-2261-9-52.

Ader F., De Groote P., Réant P., Rooryck-Thambo C., Dupin-Deguine D., Rambaud C., et al. 2019. "FLNC pathogenic variants in patients with cardiomyopathies: Prevalence and genotype-phenotype correlations." *Clin Genet* 96 (4):317-29. doi: 10.1111/cge.13594.

Adzhubei I. A., Schmidt S., Peshkin L., Ramensky V. E., Gerasimova A., Bork P., et al. 2010. "A method and server for predicting damaging missense mutations." *Nat Methods* 7 (4):248-9. doi: 10.1038/nmeth0410-248.

Al-Yacoub N., Shaheen R., Awad S. M., Kunhi M., Dzimiri N., Nguyen H. C., et al. 2016. "FBXO32, encoding a member of the SCF complex, is mutated in dilated cardiomyopathy." *Genome Biol* 17:2. doi: 10.1186/s13059-015-0861-4.

Alfares A. A., Kelly M. A., McDermott G., Funke B. H., Lebo M. S., Baxter S. B., et al. 2015. "Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity." *Genet Med* 17 (11):880-8. doi: 10.1038/gim.2014.205.

Alimadadi A., Munroe P. B., Joe B., and Cheng X. 2020. "Meta-analysis of dilated cardiomyopathy using cardiac RNA-seq transcriptomic datasets." *Genes (Basel)* 11 (1):60. doi: 10.3390/genes11010060.

Almomani R., Herkert J. C., Posafalvi A., Post J. G., Boven L. G., van der Zwaag P. A., et al. 2020. "Homozygous damaging SOD2 variant causes lethal neonatal dilated cardiomyopathy." *J Med Genet* 57 (1):23-30. doi: 10.1136/jmedgenet-2019-106330.

Almomani R., Verhagen J. M., Herkert J. C., Brosens E., van Spaendonck-Zwarts K. Y., Asimaki A., et al. 2016. "Biallelic truncating mutations in ALPK3 cause severe pediatric cardiomyopathy." *J Am Coll Cardiol* 67 (5):515-25. doi: 10.1016/j.jacc.2015.10.093.

Ang Y. S., Rivas R. N., Ribeiro A. J. S., Srivas R., Rivera J., Stone N. R., et al. 2016. "Disease model of GATA4 mutation reveals transcription factor cooperativity in human cardiogenesis." *Cell* 167 (7):1734-1749 e22. doi: 10.1016/j.cell.2016.11.033.

Aoyagi T., and Matsui T. 2011. "Phosphoinositide-3 kinase signaling in cardiac hypertrophy and heart failure." *Curr Pharm Des* 17 (18):1818-24. doi: 10.2174/138161211796390976.

Asnani A., and Peterson R. T. 2014. "The zebrafish as a tool to identify novel therapies for human cardiovascular disease." *Dis Model Mech* 7 (7):763-7. doi: 10.1242/dmm.016170.

Auman H. J., Coleman H., Riley H. E., Olale F., Tsai H. J., and Yelon D. 2007. "Functional modulation of cardiac form through regionally confined cell shape changes." *PLoS Biol* 5 (3):e53. doi: 10.1371/journal.pbio.0050053.

Azevedo O., Gal A., Faria R., Gaspar P., Miltenberger-Miltenyi G., Gago M. F., et al. 2019. "Founder effect of Fabry disease due to p.F113L mutation: Clinical profile of a late-onset phenotype." *Mol Genet Metab* 129 (2):150-60. doi: 10.1016/j.ymgme.2019.07.012.

Badrock A. P., Uggenti C., Wacheul L., Crilly S., Jenkinson E. M., Rice G. I., et al. 2020. "Analysis of U8 snoRNA variants in zebrafish reveals how bi-allelic variants cause leukoencephalopathy with calcifications and cysts." *Am J Hum Genet* 106 (5):694-706. doi: 10.1016/j.ajhg.2020.04.003.

Baig M. K., Goldman J. H., Caforio A. L. P., Coonar A. S., Keeling P. J., and McKenna W. J. 1998. "Familial dilated cardiomyopathy: cardiac abnormalities are common in asymptomatic relatives and may represent early disease." *Journal of the American College of Cardiology* 31 (1):195-201. doi: 10.1016/s0735-1097(97)00433-6.

Baig S., Edward N. C., Kotecha D., Liu B., Nordin S., Kozor R., et al. 2018. "Ventricular arrhythmia and sudden cardiac death in Fabry disease: a systematic review of risk factors in clinical practice." *Europace* 20 (FI2):f153-f161. doi: 10.1093/europace/eux261.

Bainbridge M. N., Davis E. E., Choi W. Y., Dickson A., Martinez H. R., Wang M., et al. 2015. "Loss of function mutations in NNT are associated with left ventricular noncompaction." *Circ Cardiovasc Genet* 8 (4):544-52. doi: 10.1161/CIRCGENETICS.115.001026.

Bamshad M. J., Ng S. B., Bigham A. W., Tabor H. K., Emond M. J., Nickerson D. A., et al. 2011. "Exome sequencing as a tool for Mendelian disease gene discovery." *Nat Rev Genet* 12 (11):745-55. doi: 10.1038/nrg3031.

Bang C., Batkai S., Dangwal S., Gupta S. K., Foinquinos A., Holzmann A., et al. 2014. "Cardiac fibroblast-derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy." *J Clin Invest* 124 (5):2136-46. doi: 10.1172/JCI70577.

Bao J., Wang J., Yao Y., Wang Y., Fan X., Sun K., et al. 2013. "Correlation of ventricular arrhythmias with genotype in arrhythmogenic right ventricular cardiomyopathy." *Circ Cardiovasc Genet* 6 (6):552-6. doi: 10.1161/CIRCGENETICS.113.000122.

Barefield D. Y., Puckelwartz M. J., Kim E. Y., Wilsbacher L. D., Vo A. H., Waters E. A., et al. 2017. "Experimental modeling supports a role for MyBP-HL as a novel myofilament component in arrhythmia and dilated cardiomyopathy." *Circulation* 136 (16):1477-1491. doi: 10.1161/CIRCULATIONAHA.117.028585.

Barman H. A., Ikitimur B., Kilickiran Avci B., Durmaz E., Atici A., Aslan S., et al. 2019. "The prevalence of Fabry disease among Turkish patients with non-obstructive hypertrophic cardiomyopathy: insights from a screening study." *Balkan Med J* 36 (6):354-358. doi: 10.4274/balkanmedj.galenos.2019.2019.5.125.

Baruffini E., Dallabona C., Invernizzi F., Yarham J. W., Melchionda L., Blakely E. L., et al. 2013. "MTO1 mutations are associated with hypertrophic cardiomyopathy and lactic acidosis and cause respiratory chain deficiency in humans and yeast." *Hum Mutat* 34 (11):1501-9. doi: 10.1002/humu.22393.

Baruscotti M., Bottelli G., Milanesi R., DiFrancesco J. C., and DiFrancesco D. 2010. "HCN-related channelopathies." *Pflugers Arch* 460 (2):405-15. doi: 10.1007/s00424-010-0810-8.

Becker L., Kling E., Schiller E., Zeh R., Schrewe A., Hölter S. M., et al. 2014. "MTO1-deficient mouse model mirrors the human phenotype showing complex I defect and cardiomyopathy." *PLoS One* 9 (12):e114918. doi: 10.1371/journal.pone.0114918.

Becker T., Bossenz M., Tursun B., Schlüter A., Peters M. A., Becker C. G., et al. 2003. "Comparing protein stabilities during zebrafish embryogenesis." *Methods Cell Sci* 25 (1-2):85-9. doi: 10.1023/B:MICS.0000006895.03556.9b.

Bedell V. M., and Ekker S. C. 2015. "Using engineered endonucleases to create knockout and knockin zebrafish models." *Methods Mol Biol* 1239:291-305. doi: 10.1007/978-1-4939-1862-1_17.

Begay R. L., Tharp C. A., Martin A., Graw S. L., Sinagra G., Miani D., et al. 2016. "FLNC gene splice mutations cause dilated cardiomyopathy." *JACC Basic Transl Sci* 1 (5):344-359. doi: 10.1016/j.jacbts.2016.05.004.

Bendig G., Grimmler M., Huttner I. G., Wessels G., Dahme T., Just S., et al. 2006. "Integrin-linked kinase, a novel component of the cardiac mechanical stretch sensor, controls contractility in the zebrafish heart." *Genes Dev* 20 (17):2361-72. doi: 10.1101/gad.1448306.

Bhonsale A., Groeneweg J. A., James C. A., Dooijes D., Tichnell C., Jongbloed J. D., et al. 2015. "Impact of genotype on clinical course in arrhythmogenic right ventricular dysplasia/cardiomyopathy-associated mutation carriers." *Eur Heart J* 36 (14):847-55. doi: 10.1093/eurheartj/ehu509.

Bill B. R., Petzold A. M., Clark K. J., Schimmenti L. A., and Ekker S. C. 2009. "A primer for morpholino use in zebrafish." *Zebrafish* 6 (1):69-77. doi: 10.1089/zeb.2008.0555.

Bioconductor - SNPRelate. n.d., accessed 11/09/2019. https://bioconductor.org/packages/release/bioc/html/SNPRelate.html.

BioEdit biological sequence alignment editor for Windows 95/98/NT/2K/XP/7 v7.1.11. n.d., accessed 10/07/2014. http://www.mbio.ncsu.edu/bioedit/bioedit.html.

Blue Print Genetics. 2019. "Cardiomyopathy panel." accessed 02/07/2019. https://blueprintgenetics.com/tests/panels/cardiology/cardiomyopathy-panel/.

Booi Z. 2017. "Investigating the genetic basis of familial arrhythmogenic right ventricular cardiomyopathy." BSc(Med) Honours, Department of Medicine, University of Cape Town.

Bozkurt B., Colvin M., Cook J., Cooper L. T., Deswal A., Fonarow G. C., et al. 2016. "Current diagnostic and treatment strategies for specific dilated cardiomyopathies: a scientific statement from the American Heart Association." *Circulation* 134 (23):e579-e646. doi: 10.1161/CIR.0000000000000455.

Brito D., Cardim N., Lopes L. R., Belo A., Mimoso J., Gonçalves L., et al. 2018. "Awareness of Fabry disease in cardiology: A gap to be filled." *Rev Port Cardiol* 37 (6):457-466. doi: 10.1016/j.repc.2018.03.010.

Brito D., Miltenberger-Miltenyi G., Moldovan O., Navarro C., and Madeira H. C. 2014. "Cardiac Anderson-Fabry disease: lessons from a 25-year-follow up." *Rev Port Cardiol* 33 (4):247 e1-7. doi: 10.1016/j.repc.2013.10.014.

Brodehl A., Ferrier R. A., Hamilton S. J., Greenway S. C., Brundler M. A., Yu W., et al. 2016. "Mutations in FLNC are associated with familial restrictive cardiomyopathy." *Hum Mutat* 37 (3):269-79. doi: 10.1002/humu.22942.

Brodehl A., Gaertner-Rommel A., Klauke B., Grewe S. A., Schirmer I., Peterschröder A., et al. 2017. "The novel alphaB-crystallin (CRYAB) mutation p.D109G causes restrictive cardiomyopathy." *Hum Mutat* 38 (8):947-952. doi: 10.1002/humu.23248.

Brun F., Barnes C. V., Sinagra G., Slavov D., Barbati G., Zhu X., et al. 2014. "Titin and desmosomal genes in the natural history of arrhythmogenic right ventricular cardiomyopathy." *J Med Genet* 51 (10):669-76. doi: 10.1136/jmedgenet-2014-102591.

Brun F., Gigli M., Graw S. L., Judge D. P., Merlo M., Murray B., et al. 2020. "FLNC truncations cause arrhythmogenic right ventricular cardiomyopathy." *J Med Genet*:[epub ahead of print]. doi: 10.1136/jmedgenet-2019-106394.

Buchanan J., Blair E., Thomson K. L., Ormondroyd E., Watkins H., Taylor J. C., et al. 2019. "Do health professionals value genomic testing? A discrete choice experiment in inherited cardiovascular disease." *Eur J Hum Genet* 27 (11):1639-48. doi: 10.1038/s41431-019-0452-z.

Burns C., Bagnall R. D., Lam L., Semsarian C., and Ingles J. 2017. "Multiple gene variants in hypertrophic cardiomyopathy in the era of next-generation sequencing." *Circ Cardiovasc Genet* 10 (4). doi: 10.1161/CIRCGENETICS.116.001666.

CADD - Combined Annotation Dependent Depletion. n.d., accessed 24/02/2020. https://cadd.gs.washington.edu/.

Çağlayan A. O., Sezer R. G., Kaymakçalan H., Ulgen E., Yavuz T., Baranoski J. F., et al. 2017. "ALPK3 gene mutation in a patient with congenital cardiomyopathy and dysmorphic features." *Cold Spring Harb Mol Case Stud* 3 (5). doi: 10.1101/mcs.a001859.

Captur G., and Nihoyannopoulos P. 2010. "Left ventricular non-compaction: genetic heterogeneity, diagnosis and clinical course." *Int J Cardiol* 140 (2):145-53. doi: 10.1016/j.ijcard.2009.07.003.

Carstens N., van der Merwe L., Revera M., Heradien M., Goosen A., Brink P. A., et al. 2011. "Genetic variation in angiotensin II type 2 receptor gene influences extent of left ventricular hypertrophy in hypertrophic cardiomyopathy independent of blood pressure." *J Renin Angiotensin Aldosterone Syst* 12 (3):274-80. doi: 10.1177/1470320310390725.

Catchpool M., Ramchand J., Martyn M., Hare D. L., James P. A., Trainer A. H., et al. 2019. "A cost-effectiveness model of genetic testing and periodical clinical screening for the evaluation of families with dilated cardiomyopathy." *Genet Med* 21 (12):2815-2822. doi: 10.1038/s41436-019-0582-2.

Cecchi F., Iascone M., Maurizi N., Pezzoli L., Binaco I., Biagini E., et al. 2017. "Intraoperative diagnosis of Anderson-Fabry disease in patients with obstructive hypertrophic cardiomyopathy undergoing surgical myectomy." *JAMA Cardiol* 2 (10):1147-1151. doi: 10.1001/jamacardio.2017.2353.

Cecconi M., Parodi M. I., Formisano F., Spirito P., Autore C., Musumeci M. B., et al. 2016. "Targeted next-generation sequencing helps to decipher the genetic and phenotypic heterogeneity of hypertrophic cardiomyopathy." *Int J Mol Med* 38 (4):1111-24. doi: 10.3892/ijmm.2016.2732.

Chang N., Sun C., Gao L., Zhu D., Xu X., Zhu X., et al. 2013. "Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos." *Cell Res* 23 (4):465-72. doi: 10.1038/cr.2013.45.

Chen C., Li R., Ross R. S., and Manso A. M. 2016. "Integrins and integrin-related proteins in cardiac fibrosis." *J Mol Cell Cardiol* 93:162-74. doi: 10.1016/j.yjmcc.2015.11.010.

Chen J., Hamers A. J. P., Finsterbusch M., Massimo G., Zafar M., Corder R., et al. 2018. "Endogenously generated arachidonate-derived ligands for TRPV1 induce cardiac protection in sepsis." *FASEB J* 32 (7):3816-3831. doi: 10.1096/fj.201701303R.

Choi M., Scholl U. I., Ji W., Liu T., Tikhonova I. R., Zumbo P., et al. 2009. "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." *Proc Natl Acad Sci U S A* 106 (45):19096-101.

Choi Y., Sims G. E., Murphy S., Miller J. R., and Chan A. P. 2012. "Predicting the functional effect of amino acid substitutions and indels." *PLoS One* 7 (10):e46688. doi: 10.1371/journal.pone.0046688.

Choung H. Y. G., Vyas M., Jacoby D., and West B. 2017. "Arrhythmogenic right ventricular cardiomyopathy (ARVC) in a young female athlete at 36 weeks gestation: a case report." *Pathol Res Pract* 213 (10):1302-5. doi: 10.1016/j.prp.2017.07.015.

Cohen B. H., Chinnery P. F., and Copeland W. C. 2018. POLG-related disorders. University of Washington, Seattle: GeneReviews.

Collery R. F., and Link B. A. 2019. "Precise short sequence insertion in zebrafish using a CRISPR/Cas9 approach to generate a constitutively soluble Lrp2 protein." *Front Cell Dev Biol* 7:167. doi: 10.3389/fcell.2019.00167.

Corden B., Jarman J., Whiffin N., Tayal U., Buchan R., Sehmi J., et al. 2019. "Association of titin-truncating genetic variants with life-threatening cardiac arrhythmias in patients with dilated cardiomyopathy and implanted defibrillators." *JAMA Netw Open* 2 (6):e196520. doi: 10.1001/jamanetworkopen.2019.6520.

Corrado D., Basso C., and Judge D. P. 2017. "Arrhythmogenic cardiomyopathy." *Circ Res* 121 (7):784-802. doi: 10.1161/CIRCRESAHA.117.309345.

Corrado D., Basso C., and Thiene G. 2009. "Arrhythmogenic right ventricular cardiomyopathy: an update." *Heart* 95 (9):766-73. doi: 10.1136/hrt.2008.149823.

Coyan G. N., Zinn M. D., West S. C., and Sharma M. S. 2019. "Heart transplantation from biventricular support in infant with novel SMYD1 mutation." *Pediatric Cardiology* 40 (8):1745-1747. doi: 10.1007/s00246-019-02139-7.

CRAN - Package survminer. n.d., accessed 07/09/2019. https://cran.r-project.org/web/packages/survminer/index.html.

Create custom map - MapChart. n.d., accessed 06/02/2020. https://mapchart.net/.

Csányi B., Popoiu A., Hategan L., Hegedűs Z., Nagy V., Rácz K., et al. 2016. "Identification of two novel LAMP2 gene mutations in Danon disease." *Can J Cardiol* 32 (11):1355 e23-1355 e30. doi: 10.1016/j.cjca.2016.02.071.

D'Souza A., Pearman C. M., Wang Y., Nakao S., Logantha Sjrj, Cox C., et al. 2017. "Targeting miR-423-5p reverses exercise training-induced HCN4 channel remodeling and sinus bradycardia." *Circ Res* 121 (9):1058-1068. doi: 10.1161/CIRCRESAHA.117.311607.

Dagan T., Talmor Y., and Graur D. 2002. "Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection." *Mol Biol Evol* 19 (7):1022-5.

Damasceno A., Mayosi B. M., Sani M., Ogah O. S., Mondo C., Ojji D., et al. 2012. "The causes, treatment, and outcome of acute heart failure in 1006 Africans from 9 countries." *Arch Intern Med* 172 (18):1386-94. doi: 10.1001/archinternmed.2012.3310.

De Bortoli M., Postma A. V., Poloni G., Calore M., Minervini G., Mazzotti E., et al. 2018. "Whole-exome sequencing identifies pathogenic variants in TJP1 gene associated with arrhythmogenic cardiomyopathy." *Circ Genom Precis Med* 11 (10):e002123. doi: 10.1161/CIRCGEN.118.002123.

Desmet F. O., Hamroun D., Lalande M., Collod-Béroud G., Claustres M., and Béroud C. 2009. "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." *Nucleic Acids Res* 37 (9):e67. doi: 10.1093/nar/gkp215.

Dewar L. J., Alcaide M., Fornika D., D'Amato L., Shafaatalab S., Stevens C. M., et al. 2017. "Investigating the genetic causes of sudden unexpected death in children through targeted next-generation sequencing analysis." *Circ Cardiovasc Genet* 10 (4). doi: 10.1161/CIRCGENETICS.116.001738.

Dhandapany P. S., Razzaque M. A., Muthusami U., Kunnoth S., Edwards J. J., Mulero-Navarro S., et al. 2014. "RAF1 mutations in childhood-onset dilated cardiomyopathy." *Nat Genet* 46 (6):635-639. doi: 10.1038/ng.2963.

Dong X., Fan P., Tian T., Yang Y., Xiao Y., Yang K., et al. 2017. "Recent advancements in the molecular genetics of left ventricular noncompaction cardiomyopathy." *Clin Chim Acta* 465:40-4. doi: 10.1016/j.cca.2016.12.013.

Duncker D. J., Bakkers J., Brundel B. J., Robbins J., Tardiff J. C., and Carrier L. 2015. "Animal and in silico models for the study of sarcomeric cardiomyopathies." *Cardiovasc Res* 105 (4):439-48. doi: 10.1093/cvr/cvv006.

Elliott P., Andersson B., Arbustini E., Bilinska Z., Cecchi F., Charron P., et al. 2008. "Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases." *Eur Heart J* 29 (2):270-6. doi: 10.1093/eurheartj/ehm342.

Elliott P. M., Anastasakis A., Borger M. A., Borggrefe M., Cecchi F., Charron P., et al. 2014. "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 (39):2733-79. doi: 10.1093/eurheartj/ehu284.

Elliott P. M., Gimeno J. R., Tome M. T., Shah J., Ward D., Thaman R., et al. 2006. "Left ventricular outflow tract obstruction and sudden death risk in patients with hypertrophic cardiomyopathy." *Eur Heart J* 27 (16):1933-41. doi: 10.1093/eurheartj/ehl041.

Elliott P., O'Mahony C., Syrris P., Evans A., Rivera Sorensen C., Sheppard M. N., et al. 2010. "Prevalence of desmosomal protein gene mutations in patients with dilated cardiomyopathy." *Circ Cardiovasc Genet* 3 (4):314-22. doi: 10.1161/CIRCGENETICS.110.937805.

Eng C. M., Resnick-Silverman L. A., Niehaus D.J., Astrin K.H., and Desnick R. J. 1993. "Nature and frequency of mutations in the alpha-galactosidase A gene that cause Fabry disease." *Am J Hum Genet* 53 (6):1186-97.

Ensembl genome browser 99. n.d., accessed 24/02/2020. https://www.ensembl.org/index.html.

Eschstruth A., Schneider-Maunoury S., and Giudicelli F. 2019. "Creation of zebrafish knock-in reporter lines in the nefma gene by Cas9-mediated homologous recombination." *Genesis*:e23340. doi: 10.1002/dvg.23340.

Esslinger U., Garnier S., Korniat A., Proust C., Kararigas G., Muller-Nurasyid M., et al. 2017. "Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy." *PLoS One* 12 (3):e0172995. doi: 10.1371/journal.pone.0172995.

ExAC browser. n.d., accessed 12/10/2019. http://exac.broadinstitute.org/.

Exome Variant Server. n.d., accessed 12/10/2019. http://evs.gs.washington.edu/EVS/.

Falase A. O., and Ogah O. S. 2012. "Cardiomyopathies and myocardial disorders in Africa: present status and the way forward." *Cardiovasc J Afr* 23 (10):552-62. doi: 10.5830/CVJA-2012-046.

Fan L. L., Ding D. B., Huang H., Chen Y. Q., Jin J. Y., Xia K., et al. 2019. "A de novo mutation of SMYD1 (p.F272L) is responsible for hypertrophic cardiomyopathy in a Chinese patient." *Clin Chem Lab Med* 57 (4):532-539. doi: 10.1515/cclm-2018-0578.

Farr G. H., 3rd, Imani K., Pouv D., and Maves L. 2018. "Functional testing of a human PBX3 variant in zebrafish reveals a potential modifier role in congenital heart defects." *Dis Model Mech* 11 (10). doi: 10.1242/dmm.035972.

Faul F., Erdfelder E., Lang A. G., and Buchner A. 2007. "G*Power 3:A flexible statistical power analysis program for the social, behavioral, and biomedical sciences." *Behav Res Methods* 39 (2):175-91.

FinchTV software. n.d., accessed 10/07/2014. http://www.geospiza.com/Products/finchtv.shtml.

Fish M. 2010. "Analysis of desmoplakin in arrhythmogenic right ventricular cardiomyopathy." MSc, Department of Medicine, University of Cape Town.

Fish M. 2016. "Analysis of genetic variations associated with arrhythmogenic right ventricular cardiomyopathy." PhD, Department of Medicine, University of Cape Town.

Fish M., Shaboodien G., Kraus S., Sliwa K., Seidman C. E., Burke M. A., et al. 2016. "Mutation analysis of the phospholamban gene in 315 South Africans with dilated, hypertrophic, peripartum and arrhythmogenic right ventricular cardiomyopathies." *Sci Rep* 6:22235. doi: 10.1038/srep22235.

Forleo C., D'Erchia A. M., Sorrentino S., Manzari C., Chiara M., Iacoviello M., et al. 2017. "Targeted next-generation sequencing detects novel gene-phenotype associations and expands the mutational spectrum in cardiomyopathies." *PLoS One* 12 (7):e0181842. doi: 10.1371/journal.pone.0181842.

Franklin S., Kimball T., Rasmussen T. L., Rosa-Garrido M., Chen H., Tran T., et al. 2016. "The chromatin-binding protein Smyd1 restricts adult mammalian heart growth." *Am J Physiol Heart Circ Physiol* 311 (5):H1234-H1247. doi: 10.1152/ajpheart.00235.2016.

Fu K., Nakano H., Morselli M., Chen T., Pappoe H., Nakano A., et al. 2018. "A temporal transcriptome and methylome in human embryonic stem cell-derived cardiomyocytes identifies novel regulators of early cardiac development." *Epigenetics* 13 (10-11):1013-1026. doi: 10.1080/15592294.2018.1526029.

Gao Y., Song J., Chen H., Cao C., and Lee C. 2015. "TRPV1 activation is involved in the cardioprotection of remote limb ischemic postconditioning in ischemia-reperfusion injury rats." *Biochem Biophys Res Commun* 463 (4):1034-9. doi: 10.1016/j.bbrc.2015.06.054.

García-Pavía P., Syrris P., Salas C., Evans A., Mirelis J. G., Cobo-Marcos M., et al. 2011. "Desmosomal protein gene mutations in patients with idiopathic dilated cardiomyopathy undergoing cardiac transplantation: a clinicopathological study." *Heart* 97 (21):1744-52. doi: 10.1136/hrt.2011.227967.

Germain D. P., Brand E., Burlina A., Cecchi F., Garman S. C., Kempf J., et al. 2018. "Phenotypic characteristics of the p.Asn215Ser (p.N215S) GLA mutation in male and female patients with Fabry disease: A multicenter Fabry Registry study." *Mol Genet Genomic Med* 6 (4):492-503. doi: 10.1002/mgg3.389.

Gersh B. J., Maron B. J., Bonow R. O., Dearani J. A., Fifer M. A., Link M. S., et al. 2011. "2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines." *Circulation* 124 (24):e783-831. doi: 10.1161/CIR.0b013e318223e2bd.

Ghezzi D., Baruffini E., Haack T. B., Invernizzi F., Melchionda L., Dallabona C., et al. 2012. "Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis." *Am J Hum Genet* 90 (6):1079-87. doi: 10.1016/j.ajhg.2012.04.011.

Gierten J., Hassel D., Schweizer P. A., Becker R., Katus H. A., and Thomas D. 2012. "Identification and functional characterization of zebrafish K(2P)10.1 (TREK2) two-pore-domain K(+) channels." *Biochim Biophys Acta* 1818 (1):33-41. doi: 10.1016/j.bbamem.2011.09.015.

Giuliodori A., Beffagna G., Marchetto G., Fornetto C., Vanzi F., Toppo S., et al. 2018. "Loss of cardiac Wnt/beta-catenin signalling in desmoplakin-deficient AC8 zebrafish models is rescuable by genetic and pharmacological intervention." *Cardiovasc Res* 114 (8):1082-1097. doi: 10.1093/cvr/cvy057.

Global Burden of Disease Collaborators. 2018. "Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017." *Lancet* 392 (10159):1736-88. doi: 10.1016/S0140-6736(18)32203-7.

Glotov A. S., Kazakov S. V., Zhukova E. A., Alexandrov A. V., Glotov O. S., Pakin V. S., et al. 2015. "Targeted next-generation sequencing (NGS) of nine candidate genes with custom AmpliSeq in patients and a cardiomyopathy risk group." *Clin Chim Acta* 446:132-40. doi: 10.1016/j.cca.2015.04.014.

gnomAD. n.d., accessed 12/10/2019. https://gnomad.broadinstitute.org/.

Gómez J., Lorca R., Reguero J. R., Morís C., Martín M., Tranche S., et al. 2017. "Screening of the Filamin C Gene in a large cohort of hypertrophic cardiomyopathy patients." *Circ Cardiovasc Genet* 10 (2):e001584. doi: 10.1161/CIRCGENETICS.116.001584.

Gramlich M., Michely B., Krohne C., Heuser A., Erdmann B., Klaassen S., et al. 2009. "Stress-induced dilated cardiomyopathy in a knock-in mouse model mimicking human titin-based disease." *J Mol Cell Cardiol* 47 (3):352-8. doi: 10.1016/j.yjmcc.2009.04.014.

graphics Create elegant data visualisations using the grammar of, and ggplot2. n.d., accessed 05/09/2019. https://ggplot2.tidyverse.org/.

Grimm D. G., Azencott C. A., Aicheler F., Gieraths U., MacArthur D. G., Samocha K. E., et al. 2015. "The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity." *Hum Mutat* 36 (5):513-23. doi: 10.1002/humu.22768.

Gruner C., Ivanov J., Care M., Williams L., Moravsky G., Yang H., et al. 2013. "Toronto hypertrophic cardiomyopathy genotype score for prediction of a positive genotype in hypertrophic cardiomyopathy." *Circ Cardiovasc Genet* 6 (1):19-26. doi: 10.1161/CIRCGENETICS.112.963363.

GTex portal. n.d., accessed 24/02/2020. https://gtexportal.org/home/.

Gui Y. X., Xu Z. P., Lv W., Liu H. M., Zhao J. J., and Hu X. Y. 2012. "Association of mitochondrial DNA polymerase gamma gene POLG1 polymorphisms with parkinsonism in Chinese populations." *PLoS One* 7 (12):e50086. doi: 10.1371/journal.pone.0050086.

Guo X., Fan C., Tian L., Liu Y., Wang H., Zhao S., et al. 2017. "The clinical features, outcomes and genetic characteristics of hypertrophic cardiomyopathy patients with severe right ventricular hypertrophy." *PLoS One* 12 (3):e0174118. doi: 10.1371/journal.pone.0174118.

Haas J., Frese K. S., Peil B., Kloos W., Keller A., Nietsch R., et al. 2015. "Atlas of the clinical genetics of human dilated cardiomyopathy." *Eur Heart J* 36 (18):1123-35a. doi: 10.1093/eurheartj/ehu301.

Hagège A., Réant P., Habib G., Damy T., Barone-Rochette G., Soulat G., et al. 2019. "Fabry disease in cardiology practice: Literature review and expert point of view." *Arch Cardiovasc Dis* 112 (4):278-287. doi: 10.1016/j.acvd.2019.01.002.

Hall C. L., Akhtar M. M., Sabater-Molina M., Futema M., Asimaki A., Protonotarios A., et al. 2020. "Filamin C variants are associated with a distinctive clinical and immunohistochemical arrhythmogenic cardiomyopathy phenotype." *Int J Cardiol* 307:101-8. doi: 10.1016/j.ijcard.2019.09.048.

Hall C. L., Gurha P., Sabater-Molina M., Asimaki A., Futema M., Lovering R. C., et al. 2019. "RNA sequencing-based transcriptome profiling of cardiac tissue implicates novel putative disease mechanisms in FLNC-associated arrhythmogenic cardiomyopathy." *Int J Cardiol* 302:124-130. doi: 10.1016/j.ijcard.2019.12.002.

Harper A. R., Bowman M., Hayesmoore J. B. G., Sage H., Salatino S., Blair E., et al. 2020. "Reevaluation of the South Asian MYBPC3(delta25bp) intronic deletion in hypertrophic cardiomyopathy." *Circ Genom Precis Med* 13 (3):e002783. doi: 10.1161/CIRCGEN.119.002783.

Harper A. R., Goel A., Grace C., Thomson K. L., Petersen S. E., Xu X., et al. 2021. "Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity." *Nat Genet* 53 (2):135-142. doi: 10.1038/s41588-020-00764-0.

Haugaa K. H., Haland T. F., Leren I. S., Saberniak J., and Edvardsen T. 2016. "Arrhythmogenic right ventricular cardiomyopathy, clinical manifestations, and diagnosis." *Europace* 18 (7):965-72. doi: 10.1093/europace/euv340.

Havndrup O., Christiansen M., Stoevring B., Jensen M., Hoffman-Bang J., Andersen P. S., et al. 2010. "Fabry disease mimicking hypertrophic cardiomyopathy: genetic screening needed for establishing the diagnosis in women." *Eur J Heart Fail* 12 (6):535-40. doi: 10.1093/eurjhf/hfq073.

Haywood M. E., Cocciolo A., Porter K. F., Dobrinskikh E., Slavov D., Graw S. L., et al. 2020. "Transcriptome signature of ventricular arrhythmia in dilated cardiomyopathy reveals increased fibrosis and activated TP53." *J Mol Cell Cardiol* 139:124-34. doi: 10.1016/j.yjmcc.2019.12.010.

Heather J. M., and Chain B. 2016. "The sequence of sequencers: The history of sequencing DNA." *Genomics* 107 (1):1-8. doi: 10.1016/j.ygeno.2015.11.003.

Hedley P. L., Kanters J. K., Dembic M., Jespersen T., Skibsbye L., Aidt F. H., et al. 2013. "The role of CAV3 in long-QT syndrome: clinical and functional assessment of a caveolin-3/Kv11.1 double heterozygote versus caveolin-3 single heterozygote." *Circ Cardiovasc Genet* 6 (5):452-61. doi: 10.1161/CIRCGENETICS.113.000137.

Hershberger R. E., Hedges D. J., and Morales A. 2013. "Dilated cardiomyopathy: the complexity of a diverse genetic architecture." *Nat Rev Cardiol* 10 (9):531-47. doi: 10.1038/nrcardio.2013.105.

Hershberger R. E., and Siegfried J. D. 2011. "Update 2011: clinical and genetic issues in familial dilated cardiomyopathy." *J Am Coll Cardiol* 57 (16):1641-9. doi: 10.1016/j.jacc.2011.01.015.

Heuser A., Plovie E. R., Ellinor P. T., Grossmann K. S., Shin J. T., Wichter T., et al. 2006. "Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy." *Am J Hum Genet* 79 (6):1081-8. doi: 10.1086/509044.

Home - PubMed - NCBI. n.d., accessed 22/02/2020. https://www.ncbi.nlm.nih.gov/pubmed/.

Howe K., Clark M. D., Torroja C. F., Torrance J., Berthelot C., Muffato M., et al. 2013. "The zebrafish reference genome sequence and its relationship to the human genome." *Nature* 496 (7446):498-503. doi: 10.1038/nature12111.

Huang L., Wu K. H., Zhang L., Wang Q., Tang S., Wu Q., et al. 2018. "Critical roles of Xirp proteins in cardiac conduction and their rare variants identified in sudden unexplained nocturnal death syndrome and Brugada syndrome in Chinese Han population." *J Am Heart Assoc* 7 (1):e006320. doi: 10.1161/JAHA.117.006320.

Hynes R. O. 2002. "Integrins: bidirectional, allosteric signaling machines." *Cell* 110 (6):673-87.

Index of /vol1/ftp/release/20130502/. n.d., accessed 11/09/2019. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/.

Ingles J., McGaughran J., Scuffham P. A., Atherton J., and Semsarian C. 2012. "A cost-effectiveness model of genetic testing for the evaluation of families with hypertrophic cardiomyopathy." *Heart* 98 (8):625-30. doi: 10.1136/heartjnl-2011-300368.

Ito E., Miyagawa S., Fukushima S., Yoshikawa Y., Saito S., Saito T., et al. 2017. "Histone modification is correlated with reverse left ventricular remodeling in nonischemic dilated cardiomyopathy." *Ann Thorac Surg* 104 (5):1531-1539. doi: 10.1016/j.athoracsur.2017.04.046.

Iuso A., Wiersma M., Schüller H. J., Pode-Shakked B., Marek-Yagel D., Grigat M., et al. 2018. "Mutations in PPCS, encoding phosphopantothenoylcysteine synthetase, cause autosomal-recessive dilated cardiomyopathy." *Am J Hum Genet* 102 (6):1018-1030. doi: 10.1016/j.ajhg.2018.03.022.

Ivanov A., Dabiesingh D. S., Bhumireddy G. P., Mohamed A., Asfour A., Briggs W. M., et al. 2017. "Prevalence and prognostic significance of left ventricular noncompaction in patients referred for cardiac magnetic resonance imaging." *Circ Cardiovasc Imaging* 10 (9):e006174. doi: 10.1161/CIRCIMAGING.117.006174.

Jaafar N., Gómez J., Kammoun I., Zairi I., Amara W. B., Kachboura S., et al. 2016. "Spectrum of mutations in hypertrophic cardiomyopathy genes among Tunisian patients." *Genet Test Mol Biomarkers* 20 (11):674-679. doi: 10.1089/gtmb.2016.0187.

Jääskeläinen P., Vangipurapu J., Raivo J., Kuulasmaa T., Heliö T., Aalto-Setälä K., et al. 2019. "Genetic basis and outcome in a nationwide study of Finnish patients with hypertrophic cardiomyopathy." *ESC Heart Failure* 6 (2):436-445. doi: 10.1002/ehf2.12420.

Jagadeesh K. A., Wenger A. M., Berger M. J., Guturu H., Stenson P. D., Cooper D. N., et al. 2016. "M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity." *Nat Genet* 48 (12):1581-1586. doi: 10.1038/ng.3703.

James O. O., Efosa J. D., Romokeme A. M., Zuobemi A., and Sotonye D. M. 2012. "Dominance of hypertensive heart disease in a tertiary hospital in southern Nigeria: an echocardiographic study." *Ethn Dis* 22 (2):136-9.

Janin A., N'Guyen K., Habib G., Dauphin C., Chanavat V., Bouvagnet P., et al. 2017. "Truncating mutations on myofibrillar myopathies causing genes as prevalent molecular explanations on patients with dilated cardiomyopathy." *Clin Genet* 92 (6):616-623. doi: 10.1111/cge.13043.

Jensen M. K., Havndrup O., Christiansen M., Andersen P. S., Diness B., Axelsson A., et al. 2013. "Penetrance of hypertrophic cardiomyopathy in children and adolescents: a 12-year follow-up study of clinical screening and predictive genetic testing." *Circulation* 127 (1):48-54. doi: 10.1161/CIRCULATIONAHA.111.090514.

Jiang X. X., Liu G. Y., Lei H., Li Z. L., Feng Q. P., and Huang W. 2018. "Activation of transient receptor potential vanilloid 1 protects the heart against apoptosis in ischemia/reperfusion injury through upregulating the PI3K/Akt signaling pathway." *Int J Mol Med* 41 (3):1724-1730. doi: 10.3892/ijmm.2017.3338.

Jingi A. M., Noubiap J. J., Kamdem P., Wawo Yonta E., Temfack E., Kouam C. K., et al. 2013. "The spectrum of cardiac disease in the West Region of Cameroon: a hospital-based cross-sectional study." *Int Arch Med* 6 (1):44. doi: 10.1186/1755-7682-6-44.

Juang J. J., Shun C. T., Chen Y. S., Hwu W. L., Lee N. C., Tsai W. H., et al. 2019. "Fabry disease cardiac variant IVS4+919 G>A is associated with multiple cardiac gene variants in patients with severe cardiomyopathy and fatal arrhythmia." *Genet Med* 21 (8):1890-1. doi: 10.1038/s41436-019-0436-y.

Kalyva A., Parthenakis F. I., Marketou M. E., Kontaraki J. E., and Vardas P. E. 2014. "Biochemical characterisation of Troponin C mutations causing hypertrophic and dilated cardiomyopathies." *J Muscle Res Cell Motil* 35 (2):161-78. doi: 10.1007/s10974-014-9382-0.

Kanehisa M., Sato Y., Furumichi M., Morishima K., and Tanabe M. 2019. "New approach for understanding genome variations in KEGG." *Nucleic Acids Res* 47 (D1):D590-D595. doi: 10.1093/nar/gky962.

Karczewski K. J., Francioli L. C., Tiao G., Cummings B. B., Alföldi J., Wang Q., et al. 2020. "The mutational constraint spectrum quantified from variation in 141,456 humans." *Nature* 581 (7809):434-443. doi: 10.1038/s41586-020-2308-7.

Karczewski K. J., Francioli L. C., Tiao G., Cummings B. B., Alföldi J., Wang Q., et al. 2019. "Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes." *bioRxiv*:531210. doi: 10.1101/531210.

Karur G. R., Robison S., Iwanochko R. M., Morel C. F., Crean A. M., Thavendiranathan P., et al. 2018. "Use of myocardial T1 mapping at 3.0 T to differentiate Anderson-Fabry disease from hypertrophic cardiomyopathy." *Radiology* 288 (2):398-406. doi: 10.1148/radiol.2018172613.

Kassem HSh, Azer R. S., Saber-Ayad M., Moharem-Elgamal S., Magdy G., Elguindy A., et al. 2013. "Early results of sarcomeric gene screening from the Egyptian National BA-HCM Program." *J Cardiovasc Transl Res* 6 (1):65-80. doi: 10.1007/s12265-012-9425-0.

Kelly M. A., Caleshu C., Morales A., Buchan J., Wolf Z., Harrison S. M., et al. 2018. "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* 20 (3):351-9. doi: 10.1038/gim.2017.218.

Kiselev A., Vaz R., Knyazeva A., Khudiakov A., Tarnovskaya S., Liu J., et al. 2018. "De novo mutations in FLNC leading to early-onset restrictive cardiomyopathy and congenital myopathy." *Hum Mutat* 39 (9):1161-1172. doi: 10.1002/humu.23559.

Kisselbach J., Seyler C., Schweizer P. A., Gerstberger R., Becker R., Katus H. A., et al. 2014. "Modulation of K2P 2.1 and K2P 10.1 K(+) channel sensitivity to carvedilol by alternative mRNA translation initiation." *Br J Pharmacol* 171 (23):5182-94. doi: 10.1111/bph.12596.

Klauke B., Gaertner-Rommel A., Schulz U., Kassner A., Zu Knyphausen E., Laser T., et al. 2017. "High proportion of genetic cases in patients with advanced cardiomyopathy including a novel homozygous Plakophilin 2-gene mutation." *PLoS One* 12 (12):e0189489. doi: 10.1371/journal.pone.0189489.

Knöll R., Postel R., Wang J., Krätzner R., Hennecke G., Vacaru A. M., et al. 2007. "Laminin-alpha4 and integrin-linked kinase mutations cause human cardiomyopathy via simultaneous defects in cardiomyocytes and endothelial cells." *Circulation* 116 (5):515-25. doi: 10.1161/CIRCULATIONAHA.107.689984.

Koczor C. A., Torres R. A., Fields E., Qin Q., Park J., Ludaway T., et al. 2013. "Transgenic mouse model with deficient mitochondrial polymerase exhibits reduced state IV respiration and enhanced cardiac fibrosis." *Lab Invest* 93 (2):151-8. doi: 10.1038/labinvest.2012.146.

König E., Volpato C. B., Motta B. M., Blankenburg H., Picard A., Pramstaller P., et al. 2017. "Exploring digenic inheritance in arrhythmogenic cardiomyopathy." *BMC Med Genet* 18 (1):145. doi: 10.1186/s12881-017-0503-7.

Kozasa Y., Nakashima N., Ito M., Ishikawa T., Kimoto H., Ushijima K., et al. 2018. "HCN4 pacemaker channels attenuate the parasympathetic response and stabilize the spontaneous firing of the sinoatrial node." *J Physiol* 596 (5):809-825. doi: 10.1113/JP275303.

Kraus P., V S., Yu H. B., Xing X., Lim S. L., Adler T., et al. 2014. "Pleiotropic functions for transcription factor zscan10." *PLoS One* 11 (9):e104568. doi: 10.1371/journal.pone.0104568.

Kraus S. 2019. "The rationale, design and implementation of the African Cardiomyopathy and Myocarditis Registry." PhD, Department of Medicine, University of Cape Town.

Kryczka K. E., Dzielińska Z., Franaszczyk M., Wojtkowska I., Henzel J., Śpiewak M., et al. 2018. "Severe course of peripartum cardiomyopathy and subsequent recovery in a patient with a novel TTN gene-truncating mutation." *Am J Case Rep* 19:820-4. doi: 10.12659/AJCR.909601.

Kubota A., Juanola-Falgarona M., Emmanuele V., Sanchez-Quintero M. J., Kariya S., Sera F., et al. 2019. "Cardiomyopathy and altered integrin-actin signaling in Fhl1 mutant female mice." *Hum Mol Genet* 28 (2):209-219. doi: 10.1093/hmg/ddy299.

Kumar A., Rani B., Sharma R., Kaur G., Prasad R., Bahl A., et al. 2018. "ACE2, CALM3 and TNNI3K polymorphisms as potential disease modifiers in hypertrophic and dilated cardiomyopathies." *Mol Cell Biochem* 438 (1-2):167-174. doi: 10.1007/s11010-017-3123-9.

Kumar P., Henikoff S., and Ng P. C. 2009. "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." *Nat Protoc* 4 (7):1073-81. doi: 10.1038/nprot.2009.86.

Lahrouchi N., Behr E. R., and Bezzina C. R. 2016. "Next-generation sequencing in post-mortem genetic testing of young sudden cardiac death cases." *Front Cardiovasc Med* 3:13. doi: 10.3389/fcvm.2016.00013.

Lee T. M., Hsu D. T., Kantor P., Towbin J. A., Ware S. M., Colan S. D., et al. 2017. "Pediatric cardiomyopathies." *Circ Res* 121 (7):855-873. doi: 10.1161/CIRCRESAHA.116.309386.

Lek M., Karczewski K. J., Minikel E. V., Samocha K. E., Banks E., Fennell T., et al. 2016. "Analysis of protein-coding genetic variation in 60,706 humans." *Nature* 536 (7616):285-91. doi: 10.1038/nature19057.

Lewis W. , Day B. J., Kohler J. J., Hosseini S. H., Chan S. S., Green E. C., et al. 2007. "Decreased mtDNA, oxidative stress, cardiomyopathy, and death from transgenic cardiac targeted human mutant polymerase gamma." *Lab Invest* 87 (4):326-35. doi: 10.1038/labinvest.3700523.

Li D., Czernuszewicz G. Z., Gonzalez O., Tapscott T., Karibe A., Durand J. B., et al. 2001. "Novel cardiac troponin T mutation as a cause of familial dilated cardiomyopathy." *Circulation* 104 (18):2188-93. doi: 10.1161/hc4301.098285.

Li R. G., Li L., Qiu X. B., Yuan F., Xu L., Li X., et al. 2013. "GATA4 loss-of-function mutation underlies familial dilated cardiomyopathy." *Biochem Biophys Res Commun* 439 (4):591-6. doi: 10.1016/j.bbrc.2013.09.023.

Li X., Hou J., Du J., Feng J., Yang Y., Shen Y., et al. 2018. "Potential protective mechanism in the cardiac microvascular injury." *Hypertension* 72 (1):116-127. doi: 10.1161/HYPERTENSIONAHA.118.11035.

Lian X., Zhang J., Azarin S. M., Zhu K., Hazeltine L. B., Bao X., et al. 2013. "Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions." *Nat Protoc* 8 (1):162-75. doi: 10.1038/nprot.2012.150.

Liang K.-H., Lu Y.-H., Niu C.-W., Chang S.-K., Chen Y.-R., Cheng C.-Y., et al. 2020. "The Fabry disease-causing mutation, GLA IVS4+919G>A, originated in Mainland China more than 800 years ago." *J Hum Genet* 65 (7):619-25. doi: 10.1038/s10038-020-0745-7.

Lieschke G. J., and Currie P. D. 2007. "Animal models of human disease: zebrafish swim into view." *Nat Rev Genet* 8 (5):353-67. doi: 10.1038/nrg2091.

Linhart A., Kampmann C., Zamorano J. L., Sunder-Plassmann G., Beck M., Mehta A., et al. 2007. "Cardiac manifestations of Anderson-Fabry disease: results from the international Fabry outcome survey." *Eur Heart J* 28 (10):1228-35. doi: 10.1093/eurheartj/ehm153.

Liu S., Xie Y., Zhang H., Feng Z., Huang J., Huang J., et al. 2019. "Multiple genetic variants in adolescent patients with left ventricular noncompaction cardiomyopathy." *Int J Cardiol* 302:117-23. doi: 10.1016/j.ijcard.2019.12.001.

Lompré A. M., Nadal-Ginard B., and Mahdavi V. 1984. "Expression of the cardiac ventricular alpha- and beta-Myosin Heavy Chain genes is developmentally and hormonally regulated." *J Biol Chem* 259 (10):6437-46.

Long P. A., Larsen B. T., Evans J. M., and Olson T. M. 2015. "Exome sequencing identifies pathogenic and modifier mutations in a child with sporadic dilated cardiomyopathy." *J Am Heart Assoc* 4 (12):e002443. doi: 10.1161/JAHA.115.002443.

Long P. A., Theis J. L., Shih Y. H., Maleszewski J. J., Abell Aleff P. C., Evans J. M., et al. 2017. "Recessive TAF1A mutations reveal ribosomopathy in siblings with end-stage pediatric dilated cardiomyopathy." *Hum Mol Genet* 26 (15):2874-2881. doi: 10.1093/hmg/ddx169.

Lopes L. R., Syrris P., Guttmann O. P., O'Mahony C., Tang H. C., Dalageorgou C., et al. 2015. "Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy." *Heart* 101 (4):294-301. doi: 10.1136/heartjnl-2014-306387.

Lopez A. D., and Adair T. 2019. "Is the long-term decline in cardiovascular-disease mortality in high-income countries over? Evidence from national vital statistics." *Int J Epidemiol* 48 (6):1815-23. doi: 10.1093/ije/dyz143.

Louw J. J., Nunes Bastos R., Chen X., Verdood C., Corveleyn A., Jia Y., et al. 2018. "Compound heterozygous loss-of-function mutations in KIF20A are associated with a novel lethal congenital cardiomyopathy in two siblings." *PLoS Genet* 14 (1):e1007138. doi: 10.1371/journal.pgen.1007138.

Luoma P. T., Eerola J., Ahola S., Hakonen A. H., Hellström O., Kivistö K. T., et al. 2007. "Mitochondrial DNA polymerase gamma variants in idiopathic sporadic Parkinson disease." *Neurology* 69 (11):1152-9.

Lyu Y., Chen J., and Xu H. 2018. "The pathogenic gene screening in a Chinese familial dilated cardiomyopathy pedigree from Hubei." *Gene* 642:159-162. doi: 10.1016/j.gene.2017.11.001.

M-CAP: Mendelian Clinically Applicable Pathogenicity score. n.d., accessed 24/02/2020. http://bejerano.stanford.edu/mcap/.

Maciag A., Villa F., Ferrario A., Spinelli C. C., Carrizzo A., Malovini A., et al. 2015. "Exome sequencing of a family with lone, autosomal dominant atrial flutter identifies a rare variation in ABCB4 significantly enriched in cases." *BMC Genet* 16:15. doi: 10.1186/s12863-015-0177-0.

Magi A., Semeraro R., Mingrino A., Giusti B., and D'Aurizio R. 2018. "Nanopore sequencing data analysis: state of the art, applications and challenges." *Brief Bioinform* 19 (6):1256-1272. doi: 10.1093/bib/bbx062.

Mahdieh N., Hosseini Moghaddam M., Motavaf M., Rabbani A., Soveizi M., Maleki M., et al. 2018. "Genotypic effect of a mutation of the MYBPC3 gene and two phenotypes with different patterns of inheritance." *J Clin Lab Anal* 32 (6):e22419. doi: 10.1002/jcla.22419.

Makubi A., Hage C., Lwakatare J., Kisenge P., Makani J., Ryden L., et al. 2014. "Contemporary aetiology, clinical characteristics and prognosis of adults with heart failure observed in a tertiary hospital in Tanzania: the prospective Tanzania Heart Failure (TaHeF) study." *Heart* 100 (16):1235-41. doi: 10.1136/heartjnl-2014-305599.

Mancuso M., Filosto M., Oh S. J., and DiMauro S. 2004. "A novel polymerase gamma mutation in a family with ophthalmoplegia, neuropathy, and Parkinsonism." *Arch Neurol* 61 (11):1777-9. doi: 10.1001/archneur.61.11.1777.

Manso A. M., Kang S., and Ross R. S. 2009. "Integrins, focal adhesions and cardiac fibroblasts." *J Invest Med* 57 (8):856-60. doi: 10.231/JIM.0b013e3181c5e61f.

Manso A. M., Okada H., Sakamoto F. M., Moreno E., Monkley S. J., Li R., et al. 2017. "Loss of mouse cardiomyocyte talin-1 and talin-2 leads to beta-1 integrin reduction, costameric instability, and dilated cardiomyopathy." *Proc Natl Acad Sci U S A* 114 (30):E6250-E6259. doi: 10.1073/pnas.1701416114.

Marcus F. I., McKenna W. J., Sherrill D., Basso C., Bauce B., Bluemke D. A., et al. 2010. "Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria." *Circulation* 121 (13):1533-41. doi: 10.1161/CIRCULATIONAHA.108.840827.

Marian A. J., and Braunwald E. 2017. "Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy." *Circ Res* 121 (7):749-770. doi: 10.1161/CIRCRESAHA.117.311059.

Maro E. E. , Janabi M., and Kaushik R. 2006. "Clinical and echocardiographic study of hypertrophic cardiomyopathy in Tanzania." *Trop Doct* 36 (4):225-7.

Maron B. J., Gardin J. M., Flack J. M., Gidding S. S., Kurosaki T. T., and Bild D. E. 1995. "Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults." *Circulation* 92 (4):785-9.

Marston S., Montgiraud C., Munster A. B., Copeland O., Choi O., Dos Remedios C., et al. 2015. "OBSCN mutations associated with dilated cardiomyopathy and haploinsufficiency." *PLoS One* 10 (9):e0138568. doi: 10.1371/journal.pone.0138568.

Martin A. R., Williams E., Foulger R. E., Leigh S., Daugherty L. C., Niblock O., et al. 2019. "PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels." *Nat Genet* 51 (11):1560-1565. doi: 10.1038/s41588-019-0528-2.

Martin E. D., Moriarty M. A., Byrnes L., and Grealy M. 2009. "Plakoglobin has both structural and signalling roles in zebrafish development." *Dev Biol* 327 (1):83-96. doi: 10.1016/j.ydbio.2008.11.036.

Masselink W. 2021. "Crispants take the spotlight." *Lab Anim (NY)* 50 (4):95-96. doi: 10.1038/s41684-021-00739-6.

Matthijs G., Souche E., Alders M., Corveleyn A., Eck S., Feenstra I., et al. 2016. "Guidelines for diagnostic next-generation sequencing." *Eur J Hum Genet* 24 (10):1515. doi: 10.1038/ejhg.2016.63.

Mayo Clinic Laboratories. 2015. "Inherited cardiomyopathies: understanding genetic causes [Communiqué]." accessed 20/04/2019. https://news.mayocliniclabs.com/2015/07/01/inherited-cardiomyopathies-understanding-genetic-causes-communique/.

Mayosi B. M., Fish M., Shaboodien G., Mastantuono E., Kraus S., Wieland T., et al. 2017. "Identification of cadherin 2 (CDH2) mutations in arrhythmogenic right ventricular cardiomyopathy." *Circ Cardiovasc Genet* 10 (2):e001605. doi: 10.1161/CIRCGENETICS.116.001605.

Mayosi B. M., Flisher A. J., Lalloo U. G., Sitas F., Tollman S. M., and Bradshaw D. 2009. "The burden of non-communicable diseases in South Africa." *The Lancet* 374 (9693):934-947. doi: 10.1016/s0140-6736(09)61087-4.

Mazzarotto F., Olivotto I., and Walsh R. 2020. "Advantages and perils of clinical whole-exome and whole-genome sequencing in cardiomyopathy." *Cardiovasc Drugs Ther*. doi: 10.1007/s10557-020-06948-4.

Mbele M. 2008. "The prevalence of Plakophilin-2 (PKP-2) mutations in South African patients with cardiomyopathy." MSc (Med), Department of Medicine, University of Cape Town.
McCauley M. D., and Wehrens X. H. 2009. "Animal models of arrhythmogenic cardiomyopathy." *Dis Model Mech* 2 (11-12):563-70. doi: 10.1242/dmm.002840.
McLaren W., Gil L., Hunt S. E., Riat H. S., Ritchie G. R., Thormann A., et al. 2016. "The Ensembl Variant Effect Predictor." *Genome Biol* 17 (1):122. doi: 10.1186/s13059-016-0974-4.
McNally E. M., Golbus J. R., and Puckelwartz M. J. 2013. "Genetic mutations and mechanisms in dilated cardiomyopathy." *J Clin Invest* 123 (1):19-26. doi: 10.1172/JCI62862.
McNally E. M., and Mestroni L. 2017. "Dilated cardiomyopathy: genetic determinants and mechanisms." *Circ Res* 121 (7):731-748. doi: 10.1161/CIRCRESAHA.116.309396.
McNally E. M., and Puckelwartz M. J. 2015. "Genetic variation in cardiomyopathy and cardiovascular disorders." *Circ J* 79 (7):1409-15. doi: 10.1253/circj.CJ-15-0536.
Mendes de Almeida R., Tavares J., Martins S., Carvalho T., Enguita F. J., Brito D., et al. 2017. "Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy." *PLoS One* 12 (8):e0182946. doi: 10.1371/journal.pone.0182946.
Merlo M., Sinagra G., Carniel E., Slavov D., Zhu X., Barbati G., et al. 2013. "Poor prognosis of rare sarcomeric gene variants in patients with dilated cardiomyopathy." *Clin Transl Sci* 6 (6):424-8. doi: 10.1111/cts.12116.
Mestroni L., Brun F., Spezzacatene A., Sinagra G., and Taylor M. R. G. 2014. "Genetic causes of dilated cardiomyopathy." *Prog Pediatr Cardiol* 37 (1-2):13-8. doi: 10.1016/j.ppedcard.2014.10.003.
Meynert A. M. , Ansari M., FitzPatrick D. R., and Taylor M. S. 2014. "Variant detection sensitivity and biases in whole genome and exome sequencing." *BMC Bioinformatics* 15:247. doi: 10.1186/1471-2105-15-247.
MGI-Mouse Genome Informatics. n.d., accessed 24/02/2020. http://www.informatics.jax.org/.
Milano A., Vermeer A. M., Lodder E. M., Barc J., Verkerk A. O., Postma A. V., et al. 2014. "HCN4 mutations in multiple families with bradycardia and left ventricular noncompaction cardiomyopathy." *J Am Coll Cardiol* 64 (8):745-56. doi: 10.1016/j.jacc.2014.05.045.
Militaru S., Saftoiu A., Streubel B., and Jurcut R. 2018. "New Fabry disease mutation confirms cardiomyopathy aetiology: a case report." *Eur Heart J Case Rep* 2 (4):yty133. doi: 10.1093/ehjcr/yty133.
Millat G., Bouvagnet P., Chevalier P., Sebbag L., Dulac A., Dauphin C., et al. 2011. "Clinical and mutational spectrum in a cohort of 105 unrelated patients with dilated cardiomyopathy." *Eur J Med Genet* 54 (6):e570-5. doi: 10.1016/j.ejmg.2011.07.005.
Millat G., Janin A., de Tauriac O., Roux A., and Dauphin C. 2015. "HCN4 mutation as a molecular explanation on patients with bradycardia and non-compaction cardiomyopathy." *Eur J Med Genet* 58 (9):439-42. doi: 10.1016/j.ejmg.2015.06.004.
Miura F., Shimada J., Kitagawa Y., Otani K., Sato T., Toki T., et al. 2019. "MYH7 mutation identified by next-generation sequencing in three infant siblings with bi-ventricular noncompaction presenting with restrictive hemodynamics: A report of three siblings with a severe phenotype and poor prognosis." *J Cardiol Cases* 19 (4):140-3. doi: 10.1016/j.jccase.2018.12.017.
Monserrat L., Gimeno-Blanes J. R., Marín F., Hermida-Prieto M., García-Honrubia A., Pérez I., et al. 2007. "Prevalence of Fabry disease in a cohort of 508 unrelated patients with hypertrophic cardiomyopathy." *J Am Coll Cardiol* 50 (25):2399-403. doi: 10.1016/j.jacc.2007.06.062.
Moolman-Smook J. C., De Lange W. J., Bruwer E. C., Brink P. A., and Corfield V. A. 1999. "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 (5):1308-20.
Moolman-Smook J. C., Mayosi B. M., Brink P. A., and Corfield V. A. 1998. "Identification of a new missense mutation in MyBP-C associated with hypertrophic cardiomyopathy." *J Med Genet* 35 (3):253-4.
Moolman J. C., Brink P. A., and Corfield V. A. 1993. "Identification of a new missense mutation at Arg403, a CpG mutation hotspot, in exon 13 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy." *Hum Mol Genet* 2 (10):1731-2.
Moolman J. C., Brink P. A., and Corfield V. A. 1995. "Identification of a novel Ala797Thr mutation in exon 21 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy." *Hum Mutat* 6 (2):197-8.
Moolman J. C., Corfield V. A., Posen B., Ngumbela K., Seidman C. E., Brink P. A., et al. 1997. "Sudden death due to Troponin T mutations." *J Am Coll Cardiol* 29 (3):549-555. doi: 10.1016/s0735-1097(96)00530-x.

Morey M., Fernández-Marmiesse A., Castiñeiras D., Fraga J. M., Couce M. L., and Cocho J. A. 2013. "A glimpse into past, present, and future DNA sequencing." *Mol Genet Metab* 110 (1-2):3-24. doi: 10.1016/j.ymgme.2013.04.024.

Moriarty M. A., Ryan R., Lalor P., Dockery P., Byrnes L., and Grealy M. 2012. "Loss of plakophilin 2 disrupts heart development in zebrafish." *Int J Dev Biol* 56 (9):711-8. doi: 10.1387/ijdb.113390mm.

Mouton J. M., Pellizzon A. S., Goosen A., Kinnear C. J., Herbst P. G., Brink P. A., et al. 2015. "Diagnostic disparity and identification of two TNNI3 gene mutations, one novel and one arising de novo, in South African patients with restrictive cardiomyopathy and focal ventricular hypertrophy." *Cardiovasc J Afr* 26 (2):63-9. doi: 10.5830/CVJA-2015-019.

Mouton J. M., van der Merwe L., Goosen A., Revera M., Brink P. A., Moolman-Smook J. C., et al. 2016. "MYBPH acts as modifier of cardiac hypertrophy in hypertrophic cardiomyopathy (HCM) patients." *Hum Genet* 135 (5):477-483. doi: 10.1007/s00439-016-1649-7.

Muchtar E., Blauwet L. A., and Gertz M. A. 2017. "Restrictive cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy." *Circ Res* 121 (7):819-837. doi: 10.1161/CIRCRESAHA.117.310982.

Muhammad E., Levitas A., Singh S. R., Braiman A., Ofir R., Etzion S., et al. 2015. "PLEKHM2 mutation leads to abnormal localization of lysosomes, impaired autophagy flux and associates with recessive dilated cardiomyopathy and left ventricular noncompaction." *Hum Mol Genet* 24 (25):7227-40. doi: 10.1093/hmg/ddv423.

Müller-Höcker J., Horvath R., Schäfer S., Hessel H., Müller-Felber W., Kühr J., et al. 2011. "Mitochondrial DNA depletion and fatal infantile hepatic failure due to mutations in the mitochondrial polymerase gamma (POLG) gene: a combined morphological/enzyme histochemical and immunocytochemical/biochemical and molecular genetic study." *J Cell Mol Med* 15 (2):445-56. doi: 10.1111/j.1582-4934.2009.00819.x.

MutationTaster. n.d., accessed 24/02/2020. http://www.mutationtaster.org/.

Nagandla H., Lopez S., Yu W., Rasmussen T. L., Tucker H. O., Schwartz R. J., et al. 2016. "Defective myogenesis in the absence of the muscle-specific lysine methyltransferase SMYD1." *Dev Biol* 410 (1):86-97. doi: 10.1016/j.ydbio.2015.12.005.

Nakao S., Takenaka T., Maeda M., Kodama C., Tanaka A., Tahara M., et al. 1995. "An atypical variant of Fabry's disease in men with left ventricular hypertrophy." *N Eng J Med* 333 (5):288-93.

New England Biolabs. "100 bp DNA Ladder." https://international.neb.com/products/n3231-100-bp-dna-ladder#Product%20Information.

New England Biolabs. "Quick-Load® 1 kb DNA Ladder."

Nielsen J. B., Fritsche L. G., Zhou W., Teslovich T. M., Holmen O. L., Gustafsson S., et al. 2018. "Genome-wide study of atrial fibrillation identifies seven risk loci and highlights biological pathways and regulatory elements involved in cardiac development." *Am J Hum Genet* 102 (1):103-115. doi: 10.1016/j.ajhg.2017.12.003.

Nishimura M., Kumsta C., Kaushik G., Diop S. B., Ding Y., Bisharat-Kernizan J., et al. 2014. "A dual role for integrin-linked kinase and beta1-integrin in modulating cardiac aging." *Aging Cell* 13 (3):431-40. doi: 10.1111/acel.12193.

Norton N., Li D., Rieder M. J., Siegfried J. D., Rampersaud E., Zuchner S., et al. 2011. "Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy." *Am J Hum Genet* 88 (3):273-82. doi: 10.1016/j.ajhg.2011.01.016.

Nozari A., Aghaei-Moghadam E., Zeinaloo A., Mollazadeh R., Majnoon M. T., Alavi A., et al. 2018. "A novel splicing variant in FLNC gene responsible for a highly penetrant familial dilated cardiomyopathy in an extended Iranian family." *Gene* 659:160-7. doi: 10.1016/j.gene.2018.03.044.

Ntusi N. A., Shaboodien G., Badri M., Gumedze F., and Mayosi B. M. 2016. "Clinical features, spectrum of causal genetic mutations and outcome of hypertrophic cardiomyopathy in South Africans." *Cardiovasc J Afr* 27 (3):152-8. doi: 10.5830/CVJA-2015-075.

O'Byrne J. J., Tarailo-Graovac M., Ghani A., Champion M., Deshpande C., Dursun A., et al. 2018. "The genotypic and phenotypic spectrum of MTO1 deficiency." *Mol Genet Metab* 123 (1):28-42. doi: 10.1016/j.ymgme.2017.11.003.

Oder D., Liu D., Hu K., Uceyler N., Salinger T., Müntze J., et al. 2017. "Apha-galactosidase A genotype N215S induces a specific cardiac variant of Fabry disease." *Circ Cardiovasc Genet* 10 (5):e001691. doi: 10.1161/CIRCGENETICS.116.001691.

Ogah O. S., Adegbite G. D., Akinyemi R. O., Adesina J. O., Alabi A. A., Udofia O. I., et al. 2008. "Spectrum of heart diseases in a new cardiac service in Nigeria: an echocardiographic study of 1441 subjects in Abeokuta." *BMC Res Notes* 1:98. doi: 10.1186/1756-0500-1-98.

Ohno S. 2016. "The genetic background of arrhythmogenic right ventricular cardiomyopathy." *J Arrhythm* 32 (5):398-403. doi: 10.1016/j.joa.2016.01.006.

Oliveira T. G., Mitne-Neto M., Cerdeira L. T., Marsiglia J. D., Arteaga-Fernandez E., Krieger J. E., et al. 2015. "A variant detection pipeline for inherited cardiomyopathy-associated genes using next-generation sequencing." *J Mol Diagn* 17 (4):420-30. doi: 10.1016/j.jmoldx.2015.02.003.

Ortiz-Genga M. F., Cuenca S., Dal Ferro M., Zorio E., Salgado-Aranda R., Climent V., et al. 2016. "Truncating FLNC mutations are associated with high-risk dilated and arrhythmogenic cardiomyopathies." *J Am Coll Cardiol* 68 (22):2440-2451. doi: 10.1016/j.jacc.2016.09.927.

Ouellette A. C., Mathew J., Manickaraj A. K., Manase G., Zahavich L., Wilson J., et al. 2018. "Clinical genetic testing in pediatric cardiomyopathy: Is bigger better?" *Clin Genet* 93 (1):33-40. doi: 10.1111/cge.13024.

Paldino A., De Angelis G., Merlo M., Gigli M., Dal Ferro M., Severini G. M., et al. 2018. "Genetics of dilated cardiomyopathy: clinical implications." *Curr Cardiol Rep* 20 (10):83. doi: 10.1007/s11886-018-1030-7.

Pan Y. A., Freundlich T., Weissman T. A., Schoppik D., Wang X. C., Zimmerman S., et al. 2013. "Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish." *Development* 140 (13):2835-46. doi: 10.1242/dev.094631.

Pavlu L., Kocourkova L., Taborsky M., and Petrkova J. 2019. "Ventricular tachycardia: a presentation of Fabry disease case report." *Eur Heart J Case Rep* 3 (1):yty154. doi: 10.1093/ehjcr/yty154.

Perk J., De Backer G., Gohlke H., Graham I., Reiner Z., Verschuren M., et al. 2012. "European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts)." *Eur Heart J* 33 (13):1635-701. doi: 10.1093/eurheartj/ehs092.

Péterfi Z., Tóth Z. E., Kovács H. A., Lázár E., Sum A., Donkó A., et al. 2014. "Peroxidasin-like protein: a novel peroxidase homologue in the human heart." *Cardiovasc Res* 101 (3):393-9. doi: 10.1093/cvr/cvt256.

Peters F., Khandheria B. K., dos Santos C., Matioda H., Maharaj N., Libhaber E., et al. 2012. "Isolated left ventricular noncompaction in sub-Saharan Africa: a clinical and echocardiographic perspective." *Circ Cardiovasc Imaging* 5 (2):187-93. doi: 10.1161/CIRCIMAGING.111.966937.

Phelan D. G., Anderson D. J., Howden S. E., Wong R. C., Hickey P. F., Pope K., et al. 2016. "ALPK3-deficient cardiomyocytes generated from patient-derived induced pluripotent stem cells and mutant human embryonic stem cells display abnormal calcium handling and establish that ALPK3 deficiency underlies familial cardiomyopathy." *Eur Heart J* 37 (33):2586-90. doi: 10.1093/eurheartj/ehw160.

Pilichou K., Thiene G., Bauce B., Rigato I., Lazzarini E., Migliore F., et al. 2016. "Arrhythmogenic cardiomyopathy." *Orphanet J Rare Dis* 11:33. doi: 10.1186/s13023-016-0407-1.

Pinto Y. M., Elliott P. M., Arbustini E., Adler Y., Anastasakis A., Bohm M., et al. 2016. "Proposal for a revised definition of dilated cardiomyopathy, hypokinetic non-dilated cardiomyopathy, and its implications for clinical practice: a position statement of the ESC working group on myocardial and pericardial diseases." *Eur Heart J* 37 (23):1850-8. doi: 10.1093/eurheartj/ehv727.

Pirruccello J. P., Bick A., Wang M., Chaffin M., Friedman S., Yao J., et al. 2020. "Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy." *Nat Commun* 11 (1):2254. doi: 10.1038/s41467-020-15823-7.

Poloni G., Calore M., Rigato I., Marras E., Minervini G., Mazzotti E., et al. 2019. "A targeted next-generation gene panel reveals a novel heterozygous nonsense variant in the TP63 gene in patients with arrhythmogenic cardiomyopathy." *Heart Rhythm* 16 (5):773-780. doi: 10.1016/j.hrthm.2018.11.015.

PolyPhen-2: prediction of functional effects of human nsSNPs. n.d., accessed 24/02/2020. http://genetics.bwh.harvard.edu/pph2/.

Posen B., Moolman J. C., Corfield V. A., and Brink P. A. 1995. "Clinical and prognostic evaluation of familial hypertrophic cardiomyopathy in two South African families with different cardiac beta myosin heavy chain gene mutations." *Br Heart J* 74 (1):40-6.

Poulin M. F., Shah A., Trohman R. G., and Madias C. 2015. "Advanced Anderson-Fabry disease presenting with left ventricular apical aneurysm and ventricular tachycardia." *World J Clin Cases* 3 (6):519-24. doi: 10.12998/wjcc.v3.i6.519.

Pugh T. J., Kelly M. A., Gowrisankar S., Hynes E., Seidman M. A., Baxter S. M., et al. 2014. "The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing." *Genet Med* 16 (8):601-8. doi: 10.1038/gim.2013.204.

Quang K. L., Maguy A., Qi X. Y., Naud P., Xiong F., Tadevosyan A., et al. 2015. "Loss of cardiomyocyte integrin-linked kinase produces an arrhythmogenic cardiomyopathy in mice." *Circ Arrhythm Electrophysiol* 8 (4):921-32. doi: 10.1161/CIRCEP.115.001668.

Quarta G., Muir A., Pantazis A., Syrris P., Gehmlich K., Garcia-Pavia P., et al. 2011. "Familial evaluation in arrhythmogenic right ventricular cardiomyopathy: impact of genetics and revised task force criteria." *Circulation* 123 (23):2701-9. doi: 10.1161/CIRCULATIONAHA.110.976936.

R: The R project for statistical computing. n.d., accessed 05/09/2019. https://www.r-project.org/.

Rabbani B., Tekin M., and Mahdieh N. 2014. "The promise of whole-exome sequencing in medical genetics." *J Hum Genet* 59 (1):5-15. doi: 10.1038/jhg.2013.114.

Rahn J. J., Bestman J. E., Stackley K. D., and Chan S. S. 2015. "Zebrafish lacking functional DNA polymerase gamma survive to juvenile stage, despite rapid and sustained mitochondrial DNA depletion, altered energetics and growth." *Nucleic Acids Res* 43 (21):10338-52. doi: 10.1093/nar/gkv1139.

Ramond F., Janin A., Di Filippo S., Chanavat V., Chalabreysse L., Roux-Buisson N., et al. 2017. "Homozygous PKP2 deletion associated with neonatal left ventricle noncompaction." *Clin Genet* 91 (1):126-130. doi: 10.1111/cge.12780.

Raphael D. M., Roos L., Myovela V., McHomvu E., Namamba J., Kilindimo S., et al. 2018. "Heart diseases and echocardiography in rural Tanzania: Occurrence, characteristics, and etiologies of underappreciated cardiac pathologies." *PLoS One* 13 (12):e0208931. doi: 10.1371/journal.pone.0208931.

Rehm H. L., Bale S. J., Bayrak-Toydemir P., Berg J. S., Brown K. K., Deignan J. L., et al. 2013. "ACMG clinical laboratory standards for next-generation sequencing." *Genet Med* 15 (9):733-47. doi: 10.1038/gim.2013.92.

Reinhardt P., Schmid B., Burbulla L. F., Schöndorf D. C., Wagner L., Glatza M., et al. 2013. "Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression." *Cell Stem Cell* 12 (3):354-67. doi: 10.1016/j.stem.2013.01.008.

Reinstein E., Gutierrez-Fernandez A., Tzur S., Bormans C., Marcu S., Tayeb-Fligelman E., et al. 2016. "Congenital dilated cardiomyopathy caused by biallelic mutations in Filamin C." *Eur J Hum Genet* 24 (12):1792-6. doi: 10.1038/ejhg.2016.110.

Rentzsch P., Witten D., Cooper G. M., Shendure J., and Kircher M. 2019. "CADD: predicting the deleteriousness of variants throughout the human genome." *Nucleic Acids Res* 47 (D1):D886-D894. doi: 10.1093/nar/gky1016.

Reuter J. A., Spacek D. V., and Snyder M. P. 2015. "High-throughput sequencing technologies." *Mol Cell* 58 (4):586-97. doi: 10.1016/j.molcel.2015.05.004.

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., et al. 2015. "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 (5):405-24. doi: 10.1038/gim.2015.30.

Roberts J. D. 2018. "TJP1 mutations in arrhythmogenic cardiomyopathy." *Circ Genom Precis Med* 11 (10):e002337. doi: 10.1161/CIRCGEN.118.002337.

Rojnueangnit K., Sirichongkolthong B., Wongwandee R., Khetkham T., Noojarern S., Khongkraparn A., et al. 2019. "Identification of gene mutations in primary pediatric cardiomyopathy by whole exome sequencing." *Pediatr Cardiol* 41 (1):165-74. doi: 10.1007/s00246-019-02240-x.

Rönty M., Taivainen A., Moza M., Kruh G. D., Ehler E., and Carpen O. 2005. "Involvement of palladin and alpha-actinin in targeting of the Abl/Arg kinase adaptor ArgBP2 to the actin cytoskeleton." *Exp Cell Res* 310 (1):88-98. doi: 10.1016/j.yexcr.2005.06.026.

Ross J. M., Stewart J. B., Hagstrom E., Brene S., Mourier A., Coppotelli G., et al. 2013. "Germline mitochondrial DNA mutations aggravate ageing and can impair brain development." *Nature* 501 (7467):412-5. doi: 10.1038/nature12474.

Ross R. K., and Borg T. K. 2001. "Integrins and the myocardium." *Circ Res* 88 (11):1112-9.
Rubattu S., Bozzao C., Pennacchini E., Pagannone E., Musumeci B. M., Piane M., et al. 2016. "A next-generation sequencing approach to identify gene mutations in early- and late-onset hypertrophic cardiomyopathy patients of an Italian cohort." *Int J Mol Sci* 17 (8):E1239. doi: 10.3390/ijms17081239.
Russell M. W., Raeker M. O., Geisler S. B., Thomas P. E., Simmons T. A., Bernat J. A., et al. 2014. "Functional analysis of candidate genes in 2q13 deletion syndrome implicates FBLN7 and TMEM87B deficiency in congenital heart defects and FBLN7 in craniofacial malformations." *Hum Mol Genet* 23 (16):4272-84. doi: 10.1093/hmg/ddu144.
Saito Y., Nakamura K., Nishi N., Igawa O., Yoshida M., Miyoshi T., et al. 2018. "TRPM4 mutation in patients with ventricular noncompaction and cardiac conduction disease." *Circ Genom Precis Med* 11 (5):e002103. doi: 10.1161/CIRCGEN.118.002103.
Sánchez-Iranzo H., Galardi-Castilla M., Sanz-Morejón A., González-Rosa J. M., Costa R., Ernst A., et al. 2018. "Transient fibrosis resolves via fibroblast inactivation in the regenerating zebrafish heart." *Proc Natl Acad Sci U S A* 115 (16):4188-4193. doi: 10.1073/pnas.1716713115.
Sanger J. M., Wang J., Gleason L. M., Chowrashi P., Dube D. K., Mittal B., et al. 2010. "Arg/Abl-binding protein, a Z-body and Z-band protein, binds sarcomeric, costameric, and signaling molecules." *Cytoskeleton (Hoboken)* 67 (12):808-23. doi: 10.1002/cm.20490.
Sanghvi R. V., Buhay C. J., Powell B. C., Tsai E. A., Dorschner M. O., Hong C. S., et al. 2018. "Characterizing reduced coverage regions through comparison of exome and genome sequencing data across 10 centers." *Genet Med* 20 (8):855-866. doi: 10.1038/gim.2017.192.
Sasagawa S., Nishimura Y., Okabe S., Murakami S., Ashikawa Y., Yuge M., et al. 2016. "Downregulation of GSTK1 is a common mechanism underlying hypertrophic cardiomyopathy." *Front Pharmacol* 7:162. doi: 10.3389/fphar.2016.00162.
Schaefer E., Helms P., Marcellin L., Desprez P., Billaud P., Chanavat V., et al. 2014. "Next-generation sequencing (NGS) as a fast molecular diagnosis tool for left ventricular noncompaction in an infant with compound mutations in the MYBPC3 gene." *Eur J Med Genet* 57 (4):129-32. doi: 10.1016/j.ejmg.2014.02.015.
Schmidt C., Wiedmann F., Kallenberger S. M., Ratte A., Schulte J. S., Scholz B., et al. 2017. "Stretch-activated two-pore-domain (K2P) potassium channels in the heart: Focus on atrial fibrillation and heart failure." *Prog Biophys Mol Biol* 130 (Pt B):233-243. doi: 10.1016/j.pbiomolbio.2017.05.004.
Schwarz J. M., Cooper D. N., Schuelke M., and Seelow D. 2014. "MutationTaster2: mutation prediction for the deep-sequencing age." *Nat Methods* 11 (4):361-2. doi: 10.1038/nmeth.2890.
Schweizer P. A., Schröter J., Greiner S., Haas J., Yampolsky P., Mereles D., et al. 2014. "The symptom complex of familial sinus node dysfunction and myocardial noncompaction is associated with mutations in the HCN4 channel." *J Am Coll Cardiol* 64 (8):757-67. doi: 10.1016/j.jacc.2014.06.1155.
Sehnert A. J., Huq A., Weinstein B. M., Walker C., Fishman M., and Stainier D. Y. 2002. "Cardiac troponin T is essential in sarcomere assembly and cardiac contractility." *Nat Genet* 31 (1):106-10. doi: 10.1038/ng875.
Seo J., Kim M., Hong G. R., Kim D. S., Son J. W., Cho I. J., et al. 2016. "Fabry disease in patients with hypertrophic cardiomyopathy: a practical approach to diagnosis." *J Hum Genet* 61 (9):775-80. doi: 10.1038/jhg.2016.52.
Sexton A., McDonald M., Cayla C., Thiemermann C., and Ahluwalia A. 2007. "12-Lipoxygenase-derived eicosanoids protect against myocardial ischemia/reperfusion injury via activation of neuronal TRPV1." *FASEB J* 21 (11):2695-703. doi: 10.1096/fj.06-7828com.
Shaboodien G., Spracklen T. F., Kamuli S., Ndibangwi P., Van Niekerk C., and Ntusi N. A. B. 2020. "Genetics of inherited cardiomyopathies in Africa." *Cardiovasc Diagn Ther* 10 (2):262-278. doi: 10.21037/cdt.2019.10.03.
Shao M., Chen J., and Zheng S. 2018. "Comparative proteomics analysis of myocardium in mouse model of diabetic cardiomyopathy using the iTRAQ technique." *Adv Clin Exp Med* 27 (11):1469-1475. doi: 10.17219/acem/74539.
Shehata B. M., Cundiff C. A., Lee K., Sabharwal A., Lalwani M. K., Davis A. K., et al. 2015. "Exome sequencing of patients with histiocytoid cardiomyopathy reveals a de novo NDUFB11 mutation that plays a role in the pathogenesis of histiocytoid cardiomyopathy." *Am J Med Genet A* 167A (9):2114-21. doi: 10.1002/ajmg.a.37138.

Shih Y. H., Zhang Y., Ding Y., Ross C. A., Li H., Olson T. M., et al. 2015. "Cardiac transcriptome and dilated cardiomyopathy genes in zebrafish." *Circ Cardiovasc Genet* 8 (2):261-9. doi: 10.1161/CIRCGENETICS.114.000702.
SIFT - Predict effects of nonsynonymous / missense variants. n.d., accessed 24/02/2020. https://sift.bii.a-star.edu.sg/.
Sim N. L., Kumar P., Hu J., Henikoff S., Schneider G., and Ng P. C. 2012. "SIFT web server: predicting effects of amino acid substitutions on proteins." *Nucleic Acids Res* 40 (Web Server issue):W452-7. doi: 10.1093/nar/gks539.
Sliwa K., Damasceno A., and Mayosi B. M. 2005. "Epidemiology and etiology of cardiomyopathy in Africa." *Circulation* 112 (23):3577-83. doi: 10.1161/CIRCULATIONAHA.105.542894.
Sliwa Karen, Wilkinson David, Hansen Craig, Ntyintyane Lucas, Tibazarwa Kemi, Becker Anthony, et al. 2008. "Spectrum of heart disease and risk factors in a black urban population in South Africa (the Heart of Soweto Study): a cohort study." *The Lancet* 371 (9616):915-922. doi: 10.1016/s0140-6736(08)60417-1.
Smid B. E., Hollak C. E., Poorthuis B. J., van den Bergh Weerman M. A., Florquin S., Kok W. E., et al. 2015. "Diagnostic dilemmas in Fabry disease: a case series study on GLA mutations of unknown clinical significance." *Clin Genet* 88 (2):161-6. doi: 10.1111/cge.12449.
Smith S. A., Williams M. A., Mitchell J. H., Mammen P. P., and Garry M. G. 2005. "The capsaicin-sensitive afferent neuron in skeletal muscle is abnormal in heart failure." *Circulation* 111 (16):2056-65. doi: 10.1161/01.CIR.0000162473.10951.0A.
Soldner F., and Jaenisch R. 2012. "iPSC disease modeling." *Science* 338 (6111):1155-6.
Sousa A., Canedo P., Azevedo O., Lopes L., Pinho T., Baixia M., et al. 2019. "Molecular characterization of Portuguese patients with dilated cardiomyopathy." *Rev Port Cardiol* 38 (2):129-139. doi: 10.1016/j.repc.2018.10.010.
Staudacher K., Baldea I., Kisselbach J., Staudacher I., Rahm A. K., Schweizer P. A., et al. 2011. "Alternative splicing determines mRNA translation initiation and function of human K(2P)10.1 K+ channels." *J Physiol* 589 (Pt 15):3709-20. doi: 10.1113/jphysiol.2011.210666.
Sturm A. C., and Hershberger R. E. 2013. "Genetic testing in cardiovascular medicine: current landscape and future horizons." *Curr Opin Cardiol* 28 (3):317-25. doi: 10.1097/HCO.0b013e32835fb728.
Sun Z., Han J., Zhao W., Zhang Y., Wang S., Ye L., et al. 2014. "TRPV1 activation exacerbates hypoxia/reoxygenation-induced apoptosis in H9C2 cells via calcium overload and mitochondrial dysfunction." *Int J Mol Sci* 15 (10):18362-80. doi: 10.3390/ijms151018362.
Sweet M., Taylor M. R., and Mestroni L. 2015. "Diagnosis, prevalence, and screening of familial dilated cardiomyopathy." *Expert Opin Orphan Drugs* 3 (8):869-876. doi: 10.1517/21678707.2015.1057498.
Tajsharghi Homa, and Oldfors Anders. 2012. "Myosinopathies: pathology and mechanisms." *Acta Neuropathologica* 125 (1):3-18. doi: 10.1007/s00401-012-1024-2.
Takahashi K., and Yamanaka S. 2006. "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." *Cell* 126 (4):663-76. doi: 10.1016/j.cell.2006.07.024.
Tang V. T., Arscott P., Helms A. S., and Day S. M. 2018. "Whole-exome sequencing reveals GATA4 and PTEN mutations as a potential digenic cause of left ventricular noncompaction." *Circ Genom Precis Med* 11 (1):e001966. doi: 10.1161/CIRCGEN.117.001966.
Tayal U., Newsome S., Buchan R., Whiffin N., Halliday B., Lota A., et al. 2017. "Phenotype and clinical outcomes of titin cardiomyopathy." *J Am Coll Cardiol* 70 (18):2264-2274. doi: 10.1016/j.jacc.2017.08.063.
Taylor M., Graw S., Sinagra G., Barnes C., Slavov D., Brun F., et al. 2011. "Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes." *Circulation* 124 (8):876-85. doi: 10.1161/CIRCULATIONAHA.110.005405.
Taylor R. W., Pyle A., Griffin H., Blakely E. L., Duff J., He L., et al. 2014. "Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies." *JAMA* 312 (1):68-77. doi: 10.1001/jama.2014.7184.
The 1000 Genomes Project Consortium. 2015. "A global reference for human genetic variation." *Nature* 526 (7571):68-74. doi: 10.1038/nature15393.
Theis J. L., Sharpe K. M., Matsumoto M. E., Chai H. S., Nair A. A., Theis J. D., et al. 2011. "Homozygosity mapping and exome sequencing reveal GATAD1 mutation in autosomal recessive dilated

cardiomyopathy." *Circ Cardiovasc Genet* 4 (6):585-94. doi: 10.1161/CIRCGENETICS.111.961052.
Thomson K. L., Ormondroyd E., Harper A. R., Dent T., McGuire K., Baksi J., et al. 2019. "Analysis of 51 proposed hypertrophic cardiomyopathy genes from genome sequencing data in sarcomere negative cases has negligible diagnostic yield." *Genet Med* 21 (7):1576-84. doi: 10.1038/s41436-018-0375-z.
Tobin D. M., Vary J. C., Jr., Ray J. P., Walsh G. S., Dunstan S. J., Bang N. D., et al. 2010. "The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans." *Cell* 140 (5):717-30. doi: 10.1016/j.cell.2010.02.013.
Towbin J. A. 2014. "Inherited cardiomyopathies." *Circ J* 78 (10):2347-56.
Towbin J. A., McKenna W. J., Abrams D. J., Ackerman M. J., Calkins H., Darrieux F. C. C., et al. 2019. "2019 HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy: Executive summary." *Heart Rhythm* 16 (11):e373-e407. doi: 10.1016/j.hrthm.2019.09.019.
Trifunovic A., Wredenberg A., Falkenberg M., Spelbrink J. N., Rovio A. T, Bruder C. E., et al. 2004. "Premature ageing in mice expressing defective mitochondrial DNA polymerase." *Nature* 429 (6990):413-23. doi: 10.1038/nature02544.
Truszkowska G. T., Bilińska Z. T., Muchowicz A., Pollak A., Biernacka A., Kozar-Kamińska K., et al. 2017. "Homozygous truncating mutation in NRAP gene identified by whole exome sequencing in a patient with dilated cardiomyopathy." *Sci Rep* 7 (1):3362. doi: 10.1038/s41598-017-03189-8.
Tucker N. R., McLellan M. A., Hu D., Ye J., Parsons V. A., Mills R. W., et al. 2017. "Novel mutation in FLNC (Filamin C) causes familial restrictive cardiomyopathy." *Circ Cardiovasc Genet* 10 (6):e001780. doi: 10.1161/CIRCGENETICS.117.001780.
Turkowski K. L., Tester D. J., Bos J. M., Haugaa K. H., and Ackerman M. J. 2017. "Whole exome sequencing with genomic triangulation implicates CDH2-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy." *Congenit Heart Dis* 12 (2):226-235. doi: 10.1111/chd.12462.
Turner T. N., Hormozdiari F., Duyzend M. H., McClymont S. A., Hook P. W., Iossifov I., et al. 2016. "Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA." *Am J Hum Genet* 98 (1):58-74. doi: 10.1016/j.ajhg.2015.11.023.
Valdés-Mas R., Gutiérrez-Fernández A., Gómez J., Coto E., Astudillo A., Puente D. A., et al. 2014. "Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy." *Nat Commun* 5:5326. doi: 10.1038/ncomms6326.
Valiente-Alandi I., Potter S. J., Salvador A. M., Schafer A. E., Schips T., Carrillo-Salinas F., et al. 2018. "Inhibiting fibronectin attenuates fibrosis and improves cardiac function in a model of heart failure." *Circulation* 138 (12):1236-1252. doi: 10.1161/CIRCULATIONAHA.118.034609.
Valtola K., Nino-Quintero J., Hedman M., Lottonen-Raikaslehto L., Laitinen T., Maria M., et al. 2020. "Cardiomyopathy associated with the Ala143Thr variant of the alpha-galactosidase A gene." *Heart*:[epub ahead of print]. doi: 10.1136/heartjnl-2019-315933.
van der Linde I. H. M., Hiemstra Y. L., Bokenkamp R., van Mil A. M., Breuning M. H., Ruivenkamp C., et al. 2017. "A Dutch MYH7 founder mutation, p.(Asn1918Lys), is associated with early onset cardiomyopathy and congenital heart defects." *Neth Heart J* 25 (12):675-681. doi: 10.1007/s12471-017-1037-5.
Van Tintelen J. P., Pieper P. G., Van Spaendonck-Zwarts K. Y., and Van Den Berg M. P. 2014. "Pregnancy, cardiomyopathies, and genetics." *Cardiovasc Res* 101 (4):571-8. doi: 10.1093/cvr/cvu014.
Vears D. F., Sénécal K., and Borry P. 2017. "Reporting practices for variants of uncertain significance from next generation sequencing technologies." *Eur J Med Genet* 60 (10):553-8. doi: 10.1016/j.ejmg.2017.07.016.
Verdonschot J. A. J., Hazebroek M. R., Derks K. W. J., Barandiarán Aizpurua A., Merken J. J., Wang P., et al. 2018. "Titin cardiomyopathy leads to altered mitochondrial energetics, increased fibrosis and long-term life-threatening arrhythmias." *Eur Heart J* 39 (10):864-873. doi: 10.1093/eurheartj/ehx808.
Verdonschot J. A. J., Robinson E. L., James K. N., Mohamed M. W., Claes G. R. F., Casas K., et al. 2019. "Mutations in PDLIM5 are rare in dilated cardiomyopathy but are emerging as potential disease modifiers." *Mol Genet Genomic Med*:e1049. doi: 10.1002/mgg3.1049.

Verhagen J. M. A., van den Born M., van der Linde H. C., P G. J. Nikkels, Verdijk R. M., Kivlen M. H., et al. 2019. "Biallelic variants in ASNA1, encoding a cytosolic targeting factor of tail-anchored proteins, cause rapidly progressive pediatric cardiomyopathy." *Circ Genom Precis Med* 12 (9):397-406. doi: 10.1161/CIRCGEN.119.002507.

Vermij S. H., Abriel H., and van Veen T. A. 2017. "Refining the molecular organization of the cardiac intercalated disc." *Cardiovasc Res* 113 (3):259-275. doi: 10.1093/cvr/cvw259.

Verweij N., Eppinga R. N., Hagemeijer Y., and van der Harst P. 2017. "Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure." *Sci Rep* 7 (1):2761. doi: 10.1038/s41598-017-03062-8.

Waldmüller S., Schroeder C., Sturm M., Scheffold T., Imbrich K., Junker S., et al. 2015. "Targeted 46-gene and clinical exome sequencing for mutations causing cardiomyopathies." *Mol Cell Probes* 29 (5):308-14. doi: 10.1016/j.mcp.2015.05.004.

Walsh R., Rutland C., Thomas R., and Loughna S. 2010. "Cardiomyopathy: a systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic manifestations." *Cardiology* 115 (1):49-60. doi: 10.1159/000252808.

Walsh R., Thomson K. L., Ware J. S., Funke B. H., Woodley J., McGuire K. J., et al. 2017. "Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples." *Genet Med* 19 (2):192-203. doi: 10.1038/gim.2016.90.

Wang B., Golemis E. A., and Kruh G. D. 1997. "Arg/Abl-interacting protein, is phosphorylated in v-Abl-transformed cells and localized in stress fibers and cardiocyte Z-disks." *J Biol Chem* 272 (28):17542-50.

Wang Q., Ma S., Li D., Zhang Y., Tang B., Qiu C., et al. 2014. "Dietary capsaicin ameliorates pressure overload-induced cardiac hypertrophy and fibrosis through the transient receptor potential vanilloid type 1." *Am J Hypertens* 27 (12):1521-9. doi: 10.1093/ajh/hpu068.

Wasielewski M., van Spaendonck-Zwarts K. Y., Westerink N. D., Jongbloed J. D., Postma A., Gietema J. A., et al. 2014. "Potential genetic predisposition for anthracycline-associated cardiomyopathy in families with dilated cardiomyopathy." *Open Heart* 1 (1):e000116. doi: 10.1136/openhrt-2014-000116.

Watanabe T., Hanawa H., Suzuki T., Jiao S., Yoshida K., Ogura M., et al. 2013. "A mutant mRNA expression in an endomyocardial biopsy sample obtained from a patient with a cardiac variant of Fabry disease caused by a novel acceptor splice site mutation in the invariant AG of intron 5 of the alpha-galactosidase A gene." *Intern Med* 52 (7):777-780. doi: 10.2169/internalmedicine.52.9213.

Watkins D. A., Hendricks N., Shaboodien G., Mbele M., Parker M., Vezi B. Z., et al. 2009. "Clinical features, survival experience, and profile of plakophylin-2 gene mutations in participants of the arrhythmogenic right ventricular cardiomyopathy registry of South Africa." *Heart Rhythm* 6 (11 Suppl):S10-7. doi: 10.1016/j.hrthm.2009.08.018.

Wei Z., Wang L., Han J., Song J., Yao L., Shao L., et al. 2009. "Decreased expression of transient receptor potential vanilloid 1 impaires the postischemic recovery of diabetic mouse hearts." *Circ J* 73 (6):1127-32.

Weidemann F., Niemann M., Breunig F., Herrmann S., Beer M., Stork S., et al. 2009. "Long-term effects of enzyme replacement therapy on fabry cardiomyopathy: evidence for a better outcome with early treatment." *Circulation* 119 (4):524-9. doi: 10.1161/CIRCULATIONAHA.108.794529.

Westerfield M. 2000. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4 ed: Univ. of Oregon Press, Eugene.

Whiffin N., Minikel E., Walsh R., O'Donnell-Luria A. H., Karczewski K., Ing A. Y., et al. 2017. "Using high-resolution variant frequencies to empower clinical genome interpretation." *Genet Med* 19 (10):1151-1158. doi: 10.1038/gim.2017.26.

Whiffin N., Walsh R., Govind R., Edwards M., Ahmad M., Zhang X., et al. 2018. "CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation." *Genet Med* 20 (10):1246-1254. doi: 10.1038/gim.2017.258.

WHO. 2017. "Cardiovascular diseases (CVDs)." accessed 25/08/2017. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

WHO. 2018. "Noncommunicable diseases." https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases.

Wong L. J., Naviaux R. K., Brunetti-Pierri N., Zhang Q., Schmitt E. S., Truong C., et al. 2008. "Molecular and clinical genetics of mitochondrial diseases due to POLG mutations." *Hum Mutat* 29 (9):E150-72. doi: 10.1002/humu.20824.

Woodbridge P., Liang C., Davis R. L., Vandebona H., and Sue C. M. 2013. "POLG mutations in Australian patients with mitochondrial disease." *Intern Med J* 43 (2):150-6. doi: 10.1111/j.1445-5994.2012.02847.x.

Wu C., and Dedhar S. 2001. "Integrin-linked kinase (ILK) and its interactors: a new paradigm for the coupling of extracellular matrix to actin cytoskeleton and signaling complexes." *J Cell Biol* 155 (4):505-10. doi: 10.1083/jcb.200108077.

Wu W., Lu C. X., Wang Y. N., Liu F., Chen W., Liu Y. T., et al. 2015. "Novel phenotype-genotype correlations of restrictive cardiomyopathy with myosin-binding protein C (MYBPC3) gene mutations tested by next-generation sequencing." *J Am Heart Assoc* 4 (7). doi: 10.1161/JAHA.115.001879.

Xiao F., Wei Q., Wu B., Liu X., Mading A., Yang L., et al. 2020. "Clinical exome sequencing revealed that FLNC variants contribute to the early diagnosis of cardiomyopathies in infant patients." *Transl Pediatr* 9 (1):21-33. doi: 10.21037/tp.2019.12.02.

Yadeta D., Guteta S., Alemayehu B., Mekonnen D., Gedlu E., Benti H., et al. 2017. "Spectrum of cardiovascular diseases in six main referral hospitals of Ethiopia." *Heart Asia* 9 (2):e010829. doi: 10.1136/heartasia-2016-010829.

Yamamoto S., Nagasawa T., Sugimura K., Kanno A., Tatebe S., Aoki T., et al. 2019. "Clinical diversity in patients with Anderson-fabry disease with the R301Q mutation." *Intern Med* 58 (4):603-607. doi: 10.2169/internalmedicine.0959-18.

Yang J., Shah S., Olson T. M., and Xu X. 2016. "Modeling GATAD1-associated dilated cardiomyopathy in adult zebrafish." *J Cardiovasc Dev Dis* 3 (1):6. doi: 10.3390/jcdd3010006.

Yang K. C., Breitbart A., De Lange W. J., Hofsteen P., Futakuchi-Tsuchida A., Xu J., et al. 2018. "Novel adult-onset systolic cardiomyopathy due to MYH7 E848G mutation in patient-derived induced pluripotent stem cells." *JACC Basic Transl Sci* 3 (6):728-740. doi: 10.1016/j.jacbts.2018.08.008.

Ylönen S., Ylikotila P., Siitonen A., Finnilä S., Autere J., and Majamaa K. 2013. "Variations of mitochondrial DNA polymerase gamma in patients with Parkinson's disease." *J Neurol* 260 (12):3144-9. doi: 10.1007/s00415-013-7132-7.

Yogasundaram H., Kim D., Oudit O., Thompson R. B., Weidemann F., and Oudit G. Y. 2017. "Clinical features, diagnosis, and management of patients with Anderson-Fabry cardiomyopathy." *Can J Cardiol* 33 (7):883-897. doi: 10.1016/j.cjca.2017.04.015.

Yu H. C., Coughlin C. R., Geiger E. A., Salvador B. J., Elias E. R., Cavanaugh J. L., et al. 2016. "Discovery of a potentially deleterious variant in TMEM87B in a patient with a hemizygous 2q13 microdeletion suggests a recessive condition characterized by congenital heart disease and restrictive cardiomyopathy." *Cold Spring Harb Mol Case Stud* 2 (3):a000844. doi: 10.1101/mcs.a000844.

Yuan H. X., Yan K., Hou D. Y., Zhang Z. Y., Wang H., Wang X., et al. 2017. "Whole exome sequencing identifies a KCNJ12 mutation as a cause of familial dilated cardiomyopathy." *Medicine (Baltimore)* 96 (33):e7727. doi: 10.1097/MD.0000000000007727.

Yusuf S., Hawken S., Ôunpuu S., Dans T., Avezum A., Lanas F., et al. 2004. "Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study." *The Lancet* 364 (9438):937-952. doi: 10.1016/s0140-6736(04)17018-9.

Zdebik A. A., Mahmood F., Stanescu H. C., Kleta R., Bockenhauer D., and Russell C. 2013. "Epilepsy in kcnj10 morphant zebrafish assessed with a novel method for long-term EEG recordings." *PLoS One* 8 (11):e79765. doi: 10.1371/journal.pone.0079765.

Zhang J. 2000. "Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes." *J Mol Evol* 50 (1):56-68. doi: 10.1007/s002399910007.

Zhang S. B., Liu Y. X., Fan L. L., Huang H., Li J. J., Jin J. Y., et al. 2019. "A novel heterozygous variant p.(Trp538Arg) of SYNM is identified by whole-exome sequencing in a Chinese family with dilated cardiomyopathy." *Ann Hum Genet* 83 (2):95-99. doi: 10.1111/ahg.12287.

Zhang Y., Zhang Z., and Ge W. 2018. "An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing." *J Biol Chem* 293 (17):6611-6622. doi: 10.1074/jbc.RA117.001080.

Zhang Z., Xiang L., Zhao L., Jiao H., Wang Z., Li Y., et al. 2018. "The protective effect of Er-Xian decoction against myocardial injury in menopausal rat model." *BMC Complement Altern Med* 18 (1):245. doi: 10.1186/s12906-018-2311-9.

Zhao Y., Cao H., Song Y., Feng Y., Ding X., Pang M., et al. 2016. "Identification of novel mutations including a double mutation in patients with inherited cardiomyopathy by a targeted sequencing approach using the Ion Torrent PGM system." *Int J Mol Med* 37 (6):1511-20. doi: 10.3892/ijmm.2016.2565.

Zhao Y., Feng Y., Ding X., Dong S., Zhang H., Ding J., et al. 2017. "Identification of a novel hypertrophic cardiomyopathy-associated mutation using targeted next-generation sequencing." *Int J Mol Med* 40 (1):121-9. doi: 10.3892/ijmm.2017.2986.

Zhao Y., Feng Y., Zhang Y. M., Ding X. X., Song Y. Z., Zhang A. M., et al. 2015. "Targeted next-generation sequencing of candidate genes reveals novel mutations in patients with dilated cardiomyopathy." *Int J Mol Med* 36 (6):1479-86. doi: 10.3892/ijmm.2015.2361.

Zheng L. R., Zhang Y. Y., Han J., Sun Z. W., Zhou S. X., Zhao W. T., et al. 2015. "Nerve growth factor rescues diabetic mice heart after ischemia/reperfusion injury via up-regulation of the TRPV1 receptor." *J Diabetes Complications* 29 (3):323-8. doi: 10.1016/j.jdiacomp.2015.01.006.

Zhong B., and Wang D. H. 2007. "TRPV1 gene knockout impairs preconditioning protection against myocardial injury in isolated perfused hearts in mice." *Am J Physiol Heart Circ Physiol* 293 (3):H1791-8. doi: 10.1152/ajpheart.00169.2007.

Zou J., Tran D., Baalbaki M., Tang L. F., Poon A., Pelonero A., et al. 2015. "An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of Titin in zebrafish." *Elife* 4:e09406. doi: 10.7554/eLife.09406.

Zou Y., Song L., Wang Z., Ma A., Liu T., Gu H., et al. 2004. "Prevalence of idiopathic hypertrophic cardiomyopathy in China: a population-based echocardiographic analysis of 8080 adults." *Am J Med* 116 (1):14-8. doi: 10.1016/S0002-9343(03)00597-7.

Appendix A New HCM genes and variants identified through next-generation sequencing

Table A.1: Expanded list of HCM disease genes identified through NGS

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
FLNC	c.19del (p.Y7Tfs*51)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.115C>T (p.Q39*)	0	Heterozygous	AD DCM	Not specified; Caucasian family	*Targeted sequencing*	Janin et al. (2017)
	c.230_234del (p.L77Pfs*73)	0	Heterozygous	AD DCM	Not specified; Caucasian family	*Targeted sequencing*	Janin et al. (2017)
	c.241delC (p.R81Afs*15)	0	Heterozygous	AD ACM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.249C>G (p.Y83*)	0	Heterozygous	AD ACM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.318C>G (p.F106L) c.2971C>T (p.R991*)	0.000051 0	Compound heterozygous	AR DCM	Non-consanguineous Ashkenazi Jewish family	*WES*	Reinstein et al. (2016)
	c.322G>T (p.E108*)	0	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.368>C (p.V123A)	0	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.581_599delTGGTGGACAACTGCGCCCC (p.L194Pfs*52)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.601+1G>T	0	Heterozygous	Sporadic DCM	French patient	*Targeted sequencing*	Ader et al. (2019)
	c.602-716_1010delins TGCCCCGGGAGGGGTGCCTCAGTCTCCCTGTCCCTCTG (p.G201Vfs*36)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.711del (p.E238Rfs*14)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.870C>A (p.N290K)	0	Not specified	HCM	Not specified	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.1322G>T (p.R441I)	0	Heterozygous	Sporadic DCM	Chinese patient	*Exome sequencing*	Xiao et al. (2020)
	c.1412-1G>A	0.000001	Heterozygous	Sporadic DCM	French patient	*Targeted sequencing*	Ader et al. (2019)

Continued on next page

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Table A.1 continued

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
	c.1444C>T (p.R482*)	0	Not specified	ACM	Not specified; non-consanguineous family	*Exome sequencing*	Hall et al. (2020)
	c.1890C>A (p.Y630*)	0	Heterozygous	AD DCM	Not specified; Caucasian family	*Targeted sequencing*	Janin et al. (2017)
	c.2041_2047dup (p.I683Rfs*9)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.2208delT (p.K737Sfs*11)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.2265+4del	0	Heterozygous	Sporadic DCM	Chinese patient	*Exome sequencing*	Xiao et al. (2020)
	c.2375G>T (p.S792I)	0.000007	Not specified	Sporadic HCM	Tunisian patients	*Targeted sequencing*	Jaafar et al. (2016)
	c.2971C>T (p.R911*)	0	Not specified	ACM	Not specified; non-consanguineous family	*Exome sequencing*	Hall et al. (2020)
	c.2784C>G (p.Y928*)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
FLNC	c.2888delC (p.P963Rfs*26)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.3592dup (p.V1198Gfs*64)	0	Heterozygous	Sporadic DCM	French patient	*Targeted sequencing*	Ader et al. (2019)
	c.3402_3403dupGCGAGTACACCA TCAACATCCTG (p.F1135Afs*62)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.3547GC>CT (p.A1183L)	0	Heterozygous	Sporadic RCM	Swedish patient	*Exome sequencing*	Kiselev et al. (2018)
	c.3557C>T (p.A1186V)	0	*De novo*	Sporadic RCM	Swedish patients	*Exome sequencing*	Kiselev et al. (2018)
	c.3557C>T (p.A1186V)	0	*De novo*	Sporadic RCM	Chinese patient	*Exome sequencing*	Xiao et al. (2020)
	c.3581C>T (p.S1194L)	0.000001	Heterozygous	Sporadic HCM	French patient	*Targeted sequencing*	Ader et al. (2019)
	c.4106_4107dupA (p.N1369Kfs*36)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.4271G>T (p.G1424V)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)

Continued on next page

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Table A.1 continued

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
FLNC	c.4824G>A (p.A1539T)	0	Heterozygous	AD HCM	Spanish family	**Exome sequencing**	Valdés-Mas et al. (2014)
	c.4824G>A (p.A1539T)	0	Heterozygous	AD DCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.4651G>A (p.A1551T)	0.000112	Not specified	Sporadic HCM	Tunisian patients	*Targeted sequencing*	Jaafar et al. (2016)
	c.4718T>A (p.L1573*)	0	Not specified	ACM	Not specified; non-consanguineous family	*Exome sequencing*	Hall et al. (2020)
	c.4871C>T (p.S1624L)	0	Heterozygous	AD RCM	Not specified; Caucasian family	*Exome sequencing*	Brodehl et al. (2016)
	c.4871C>T (p.S1624L)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.4916G>A (p.C1639Y)	0	*De novo*	Sporadic RCM	Chinese patient	*Exome sequencing*	Xiao et al. (2020)
	c.5398G>T (p.G1800*)	0	Heterozygous	AD ACM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.5836_5841dup (p.I1946_T1947dup)	0	Heterozygous	AD RCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.5996G>A (p.R1999Q)	0.000016	Not specified	Sporadic HCM	Tunisian patients	*Targeted sequencing*	Jaafar et al. (2016)
	c.6032G>A (p.G2011E)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6115G>A (p.G2039R)	0.000001	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6208G>A (p.G2070S)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.6231delT (p.S2077Rfs*50)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.6240_6259delCCCAAGCAAGGTGGACATCA (p.P2081Lfs*2)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.6291_6299dupAGTCACCTA (p.Y2100*)	0	Heterozygous	DCM	Thai patient	*Exome sequencing*	Rojnueangnit et al. (2019)
	c.6398G>A (p.R2133H)	0	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)

Continued on next page

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Table A.1 continued

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
FLNC	c.6419G>A (p.R2140Q)	0.000016	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Targeted sequencing*	Gómez et al. (2017)
	c.6419G>A (p.R2140Q)	0.000016	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6451G>A (p.G2151S)	0	Not specified	HCM	Not specified	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.6478A>T (p.I2160F)	0	Heterozygous	AD RCM	Not specified; Caucasian family	*Exome sequencing*	Brodehl et al. (2016)
	c.6565G>T (p.E2189*)	0	Not specified	ACM	Caucasian patient	*Targeted sequencing*	Brun et al. (2020)
	c.6889G>A (p.V2297M)	0	Heterozygous	AD RCM	Not specified; non-consanguineous family	*Exome sequencing*	Tucker et al. (2017)
	c.6889G>A (p.V2297M)	0	Heterozygous	AD RCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6892C>T (p.P2298S)	0	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Targeted sequencing*	Gómez et al. (2017)
	c.6893C>T (p.P2298L)	0	Heterozygous	AD RCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6895G>A (p.G2299S)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.6943C>A (p.H2315N)	0	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.6952C>T (p.R2318W)	0	Heterozygous	Sporadic HCM	French patient	*Targeted sequencing*	Ader et al. (2019)
	c.6976C>T (p.R2326*)	0.000004	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.6989dupG (p.V2331Rfs*25)	0	Heterozygous	AD DCM	Not specified; Caucasian family	*Targeted sequencing*	Janin et al. (2017)
	c.7076T>C (p.I2359T)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7118_7119del (p.Y2373Cfs*7)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7123G>C (p.V2375L)	0	Heterozygous	Sporadic HCM	French patient	*Targeted sequencing*	Ader et al. (2019)

Continued on next page

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Table A.1 continued

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
FLNC	c.7228C>T (p.R2410C)	0.000002	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7250A>C (p.Q2417P)	0	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7252-1G>A	0	Not specified	ACM	Not specified; non-consanguineous family	*Exome sequencing*	Hall et al. (2020)
	c.7750C>T (p.A2430V)	0.000098	Heterozygous	AD HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Valdés-Mas et al. (2014)
	c.7484G>A (p.R2495H)	0.000002	Heterozygous	AD HCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7645C>T (p.Q2549*)	0	Heterozygous	Sporadic DCM	French patient	*Targeted sequencing*	Ader et al. (2019)
	c.7665T>A (p.C2555*)	0	Heterozygous	AD DCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.7927_7935del (p.P2643_L2645del)	0	Heterozygous	AD RCM	French family	*Targeted sequencing*	Ader et al. (2019)
	c.8107delG (p.D2703Tfs*69)	0	Heterozygous	AD DCM	Not specified	*Targeted sequencing*	Ortiz-Genga et al. (2016)
	c.8107delG (p.D2703Tfs*69)	0	Possible digenic	LVNC	Chinese patient	*Exome sequencing*	Liu et al. (2019)
	c.8107delG (p.D2703Tfs*69)	0	Not specified	Sporadic ACM	Caucasian patient	*Targeted sequencing*	Brun et al. (2020)
MTO1	c.122T>G (p.V41G) c.1282G>A (p.A428T)	0 0.000051	Compound heterozygous	HCM	Caucasian non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.402_403delTA (p.Y134*) c.1282G>A (p.A428T)	0 0.000051	Compound heterozygous	HCM	Caucasian non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.631_631delG (p.G211Dfs*3) c.1282G>A (p.A428T)	0.000008 0.000051	Compound heterozygous	HCM	Croatian patient	*Exome sequencing*	Taylor et al. (2014)
	c.1222A>T (p.I408F)	0	Homozygous	HCM	Syrian consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.1232C>T (p.T411I)	0.000024	Homozygous	AR HCM	Not specified; consanguineous family	*Candidate gene screen*	Baruffini et al. (2013)
	c.1232C>T (p.T411I)	0.000024	Homozygous	HCM	Two British Pakistani patients	*Exome sequencing*	Taylor et al. (2014)

Continued on next page

* Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Table A.1 continued

Gene	Mutation	gnomAD MAF	Inheritance	Disease	Cohort	Technology*	Reference
MTO1	c.1261-5T>G (p.V422Ifs*18) c.1430G>A (p.R477H)	0 0.000191	Compound heterozygous	HCM	Caucasian non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.1273G>A (p.G425R) c.1417C>T (p.R473C)	0.000016 0	Compound heterozygous	HCM	Caucasian non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.1282G>A (p.A428T) c.1858dup (p.R620Kfs*8)	0.000051 0	Compound heterozygous	AR HCM	Italian family	**Exome sequencing**	Ghezzi et al. (2012)
	c.1282G>A (p.A428T)	0.000051	Homozygous	Sporadic HCM	Not specified	*Candidate gene screen*	Ghezzi et al. (2012)
	c.1282G>A (p.A428T) c.1430G>A (p.R477H)	0.000051 0.000191	Compound heterozygous	AR HCM	Not specified; non-consanguineous family	*Candidate gene screen*	Baruffini et al. (2013)
	c.1390C>T (p.R464C)	0.000004	Homozygous	HCM	European non-consanguineous families	*Not specified*	O'Byrne et al. (2018)
	c.1390C>T (p.R464C) c.1430G>A (p.R477H)	0.000004 0.000191	Compound heterozygous	HCM	Turkish non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.1430G>A (p.R477H) c.1450C>T (p.R484W)	0.000191 0.000018	Compound heterozygous	HCM	Caucasian non-consanguineous family	*Not specified*	O'Byrne et al. (2018)
	c.1436A>T (p.D479V)	0	Homozygous	HCM	Indian consanguineous family	*Not specified*	O'Byrne et al. (2018)
SMYD1	c.675delA (p.K225Nfs*8)	0	Not specified	LVNC	Chinese patient	*Exome sequencing*	Liu et al. (2019)
	c.814T>C (p.F272L)	0.000004	*De novo*	HCM	Chinese patient	**Exome sequencing**	Fan et al. (2019)

***** Technology used for the original discovery is indicated in bold print, and subsequent screens are indicated in italics

ACM, arrhythmogenic cardiomyopathy; AD, autosomal dominant; AR, autosomal recessive; DCM, dilated cardiomyopathy; gnomAD, Genome Aggregation Database; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; MAF, minor allele frequency; RCM, restrictive cardiomyopathy

Appendix B Letter of ethical approval

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room G 50 Old Main Building
Groote Schuur Hospital
Observatory 7925
Email: hrec-enquiries@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

07 February 2020

HREC REF: 111/2019

Prof N Ntusi/ A/Prof G Shaboodien
Department of Medicine
OMB GSH

Dear Prof Ntusi

PROJECT TITLE: WHOLE EXOME SEQUENCING OF CASES WITH FAMILIAL CARDIOMYOPATHY (SUB-STUDY:766/2014) (PHD CANDIDATE: MR. T SPRACKLEN)

Thank you for submitting your new study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study subject.

Approval is granted for one year until the 28 February 2021.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period. (Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

The HREC acknowledges that the student: Mr. T Spracklen will also be involved in this study.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC REF in all your correspondence

Yours sincerely

pp **PROFESSOR M BLOCKMAN**
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical

HREC Ref 111/2019
OL

Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HREC Ref 111/2019
OL

Appendix C Informed consent form for molecular analysis

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory
Division of Cardiovascular Genetics
4th floor, Chris Barnard Building,
UCT Medical School, Observatory 7925
Tel: (021) 406 6615 Fax: (021) 4478789

Blood should be drawn in plastic EDTA Tubes (Purple top)
20ml of blood in total is required in adults (2-4 EDTA tubes)
Please label blood tubes with the patient's name and DOB.
Place blood in fridge at 4°C until able to send to laboratory
Please DO NOT send specimens on ice or frozen.

Patient details: (or hospital sticker here)

Surname:_____ First Name (S): _____

Hospital folder number: _____

Sex: M ☐ F ☐ Date of Birth (DD/MM/YYYY): _____

Patient address: _____

Contact numbers: _____

Email address: _____

Referring hospital and doctor: _____

Referring doctor's contact details: _____

Clinical Information: (PLEASE COMPLETE A FAMILY PEDIGREE OVER THE PAGE)

New Family: Yes ☐ No ☐ **Family name:** _____

Ethnic Origin: Black African ☐, Mixed race ☐, Caucasian ☐, Asian ☐, Other _____

Provisional Diagnosis:

Clinically affected ☐ **At Risk (unaffected clinically) ☐** **Spouse ☐** **Query ☐**

Morphofunctional phenotype	Specific diagnosis
☐ Dilated cardiomyopathy	☐Familial ☐Idiopathic ☐PPCM ☐Secondary (specify) Specify (e.g. myocarditis)
☐ Hypertrophic cardiomyopathy	☐HCM ☐HCM – phenocopy (specify) Specify (e.g. Noonans)
☐ ARVC	☐Definite ☐Borderline ☐Possible ☐Unconfirmed
☐ Restrictive cardiomyopathy	☐Familial ☐Idiopathic ☐Specific aetiology (specify) Specify (e.g. EMF)
☐ Unspecified cardiomyopathy	☐LVNC ☐Takusubu ☐Other (specify) Specify:
☐ Arrhythmia (with no clinical evidence of CMO)	☐LQTS ☐VT ☐VF ☐SCD ☐Other Specify
☐Marfan Syndrome	☐Confirmed clinically ☐Suspected

Additional disorders (apparent or previously treated): _____

Have samples from this patient been sent to a DNA lab before? ☐YES ☐ NO ☐Don't Know
If yes, please specify lab: _____

For Laboratory use only:

DNA number: _____ *Vol.Blood:* _____*(ml)* *Other:* _____

Date Taken: _____ *Date Received (DD/MM/YYYY):* _____

Informed Consent

I understand that I have agreed to participate in genetics research that will be conducted at the molecular genetics laboratory at the Hatter Insititute for Cardiovascular Research in Africa, situated at the University of Cape Town in South Africa. This laboratory, in collaboration with other local and international laboratories, is dedicated to doing research related to genetic causes of cardiovascular disease in people living in Africa.

I understand that a blood sample (5-25ml or 1-5 teaspoons of blood) will be collected from me and my genetic material will be extracted from this sample for analysis. In some instances, other samples may be collected depending on the circumstances (please specify if collected): _____

I understand that a portion of my genetic material will be stored at the Hatter Institute for additional research projects approved by the University of Cape Town Human Research Ethics Committee (HREC). My genetic material/information, together with other relevent medical information, may be shared with other researchers and institutions involved in HREC approved studies but my personal identifying information will not be shared. I authorize that my doctors can provide relevent clinical information (medical records) to researchers.

I understand that the nature of research means that I may or may not receive a result from studies performed on my DNA. Although the laboratory will do its best to confirm that the findings relate to my condition, results received from a research laboratory should be confirmed diagnostically. If a genetic cause for my condition is found, the researchers from the Hatter Insitute will do their best to inform me of the results, either via my doctor, a genetics counsellor or in writing, depending on the available resources. If my contact details change, it is my responsibility to inform the laboratory. In the event that I am unavailable or incapacitated, **I do / do not** *(please delete where not applicable)* want my immediate family members to be informed of the results.

I understand that the genetics molecular laboratory is under obligation to respect my confidentiality. I understand that my participation in genetics research is entirely voluntary and that I may withdraw my consent at any time without it affecting my future medical care.

_____	_____	_____
Particpant name	Signature	Date
_____	_____	_____
Doctor/nurse/genetics counsellor	Signature	Date

Study: _____ **HREC REF:** _____ _____

(Study number)

Appendix D DNA extraction protocols

Note: exceptions to the protocols are written in **bold**

Gentra Puregene Blood Kit protocol – DNA purification from buffy coat

1. If the buffy coat preparation contains red blood cells, continue with step 2. Otherwise, pipet 3 ml Cell Lysis Solution into a 15 ml centrifuge tube, add 150–250 µl sample, and continue with step 8.
2. Dispense 3 volumes RBC Lysis Solution into a 15 ml centrifuge tube (e.g., if processing 250 µl buffy coat, dispense 750 µl RBC Lysis Solution). Add 150–250 µl buffy coat preparation.
3. Invert to mix and incubate for 10 min at room temperature (15–25°C). Invert again at least once during the incubation.
4. Centrifuge for 5 min at 2000 x g. **(Samples were centrifuged for 10 min)**
5. Carefully discard the supernatant by pipetting or pouring, leaving approximately 100–200 µl of the residual liquid and the pellet.
6. Vortex the tube vigorously to resuspend the pellet in the residual liquid.
7. Add 3 ml Cell Lysis Solution and pipet up and down or vortex vigorously to lyse the cells. Usually no incubation is required; however, if cell clumps are visible, incubate at 37°C until the solution is homogeneous. Samples are stable in Cell Lysis Solution for at least 2 years at room temperature.
8. Optional: If RNA-free DNA is required, add 15 µl RNase A Solution and mix by inverting 25 times. Incubate for 15 min at 37°C. Then incubate for 3 min on ice to quickly cool the sample. **(not done)**
9. Add 1 ml Protein Precipitation Solution and vortex vigorously for 20 s at high speed. **(1.5 ml Protein Precipitation Solution was added)**
10. Centrifuge for 5 min at 2000 x g. **(Samples were centrifuged for 10 min)**
11. Pipet 3 ml isopropanol into a clean 15 ml centrifuge tube and add the supernatant from the previous step by pouring carefully. **(6 ml isopropanol was used)**
12. Mix by inverting gently 50 times. **(samples were inverted until DNA was visible)**
13. Centrifuge for 3 min at 2000 x g. **(the visible DNA was removed by pipette and centrifuged for 1 min in a new 2 ml microcentrifuge tube)**
14. Carefully discard the supernatant and drain the tube by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube.
15. Add 3 ml of 70% ethanol and invert several times to wash the DNA pellet. **(1 ml of 70% ethanol was added)**
16. Centrifuge for 1 min at 2000 x g.
17. Carefully discard the supernatant. Drain the tube on a clean piece of absorbent paper, taking care that the pellet remains in the tube. Allow to air dry for 5–10 min.
18. Add 300 µl DNA Hydration Solution and vortex for 5 s at medium speed to mix. **(200 µl DNA Hydration Solution was added)**
19. Incubate at 65°C for 1h to dissolve the DNA. **(not done)**
20. Incubate at room temperature overnight.

PAXgene® Blood DNA Kit protocol

Before starting:

- Thaw frozen PAXgene Blood DNA Tubes in a wire rack at ambient temperature (18–25°C) for approximately 2 hr or at 37°C in a water bath for approximately 15 minutes. Carefully invert the thawed PAXgene Blood DNA Tubes 10 times before beginning the procedure.
- Heat a heating block or water bath to 65°C for use in steps 8 and 17.
- Add 1.4 ml Buffer BG4 (resuspension buffer) to lyophilised PreAnalytiX Protease. Dissolved PreAnalytiX Protease should be stored at 2–8°C or in aliquots at –20°C.
- For every sample, mix 5 ml Buffer BG3 (digestion buffer) and 50 µl reconstituted PreAnalytiX Protease. For example, to process 10 samples, mix 50 ml Buffer BG3 with 500 µl PreAnalytiX Protease. The Buffer BG3–PreAnalytiX Protease mixture should be prepared immediately before the start of the procedure.

Procedure:

1. Pour all the blood from one PAXgene Blood DNA Tube into a Processing Tube containing 25 ml Buffer BG1. Close the tube. To avoid cracking the blue lids of the Processing Tubes, do not overtighten them. Tighten the lid only until the first sign of resistance is felt. Mix by inverting the tube 5 times.
2. Centrifuge for 5 min at 2500 x g in a swing-out rotor.
3. Carefully discard the supernatant and place the tube in a rack.
4. Add 5 ml Buffer BG2, close the tube and wash the pellet by vortexing vigorously for 5 s.
5. Centrifuge for 3 min at 2500 x g in a swing-out rotor.
6. Carefully discard the supernatant and place the tube back in the rack.
7. Add 5 ml Buffer BG3/PreAnalytiX Protease, close the tube and vortex for 20 s at high speed.
8. Place the tube in a heating block or water bath and incubate at 65°C for 10 min.
9. Vortex again for 5 s at high speed.
10. Add 5 ml isopropanol (100%) and mix by inverting the tube at least 20 times until the white DNA strands clump visibly together.
11. Centrifuge for 3 min at 2500 x g.
12. Discard the supernatant and leave the tube inverted on a clean piece of absorbent paper for 1 min.
13. Add 5 ml 70% (v/v) ethanol and vortex for 1 s at high speed.
14. Centrifuge for 3 min at 2500 x g.
15. Discard the supernatant and leave the tube inverted on a clean piece of absorbent paper for at least 5 min.
16. Carefully dab the tube onto absorbent paper to remove ethanol from the rim and leave it inverted for a further 5 min to allow the DNA pellet to dry.
17. Add 1 ml Buffer BG4 and dissolve the DNA by incubating for 1 h at 65°C in a heating block or water bath, followed by incubation overnight at room temperature.

Appendix E Buffers and reagents

DNA resuspension

1 M Tris-HCl (pH 8.0):

- 121.1 g Tris base (Glentham Life Sciences, Corsham, UK) in 800 ml distilled water
- Adjust the pH to 8.0 with concentrated HCl
- Make to a total volume of 1 L with sterile distilled water

0.5 M EDTA (pH 8.0):

- 186.1 g EDTA (Glentham Life Sciences) in 800 ml distilled water
- Adjust the pH to 8.0 with NaOH
- Make to a total volume of 1 L with sterile distilled water

1X Tris/EDTA (TE) buffer:

- 1 ml 1 M Tris-HCl (final concentration: 10 mM)
- 0.2 ml 0.5 M EDTA (final concentration: 1 mM)
- 98.8 ml sterile distilled water

Agarose gel electrophoresis

Tris/Borate/EDTA (TBE) buffers

10X TBE (stock):

- 121.1 g Tris (Glentham Life Sciences) (final concentration: 1 M)
- 61.8 g Boric acid (AMRESCO, Solon, OH USA) (final concentration: 1 M)
- 7.4 g EDTA (Glentham Life Sciences) (final concentration: 0.02 M)
- Made to a total volume of 1 L with sterile distilled water

1X TBE (working):

- Made by a 1:10 dilution of stock 10X TBE buffer, with sterile distilled water

Agarose gels

Loading dye:

- 1 ml 5X Green GoTaq® Flexi buffer (Promega)
- 5 µl GelRed® nucleic acid gel stain (Biotium)
- Mixed with DNA samples at a ratio of 3:5 (dye:sample) before loading

1% agarose gel:

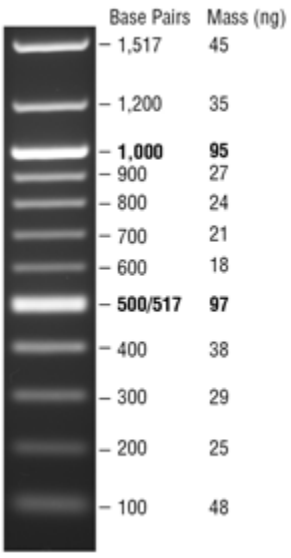
- 1 g SeaKem® LE agarose (Lonza, Basel, Switzerland)
- 100 ml 1X TBE

2% agarose gel:

- 2 g SeaKem® LE agarose (Lonza)
- 100 ml 1X TBE

DNA molecular weight markers

A



B

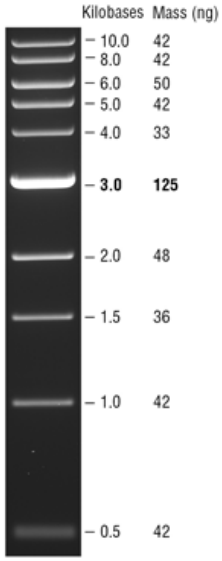


Figure E.1: DNA molecular weight markers. Shown are (A) the 100 bp ladder (NEB) and (B) the Quick-load® 1kb DNA ladder (NEB). Adapted from New England BioLabs product listing (New England Biolabs , New England Biolabs).

Zebrafish buffers and stock solutions

Water solutions

E3 60X stock solution:

- 17.5 g NaCl (Sigma-Aldrich, St Louis, MO USA)
- 0.76 g KCl (Sigma-Aldrich)
- 2.9 g $CaCl_2$ (Sigma-Aldrich)
- 4.88 g $MgSO_4$ (Sigma-Aldrich)
- Made to a total volume of 1 L with distilled water

Methylene blue E3 water:

- 1 g methylene blue (Sigma-Aldrich)
- Added to 1 L of 60X stock E3 water

Tricaine

Tricaine 0.4% stock solution:

- 400 mg tricaine powder (Sigma-Aldrich)
- 97.9 ml double-distilled water
- 2.1 ml 1 M Tris (pH 9)
- Adjust pH to 7
- Store at -20°C

Cloning and site-directed mutagenesis

LB broth:

- 10 g tryptone
- 5 g yeast extract
- 10 g NaCl (Sigma-Aldrich)
- Made to a total volume of 1 L with distilled water
- Autoclave

Ampicillin plates:

- Add 37 g LB-agar powder per L of sterile water and swirl to form a colloid
- Autoclave
- Submerge the molten gel mix in a 60°C water bath and add ampicillin while swirling
- Pour agar into petri dishes next to an open flame
- Leave on the bench to cool and solidify

3% methylcellulose:

- 3 g methylcellulose
- 100 ml E3 water or distilled water
- Cool the solution to 4°C and agitate until solubilised

Appendix F Variant annotation and analysis pipeline

Note: Protocol courtesy of Dr Simon Williams (University of Manchester, UK)

This protocol describes how to go from GRCh37 VCF files to annotated files containing variants that can be merged samples and/or filtered on gene candidate lists. These instructions are for running on the UCT HPC cluster NOT the UCT HEX cluster.

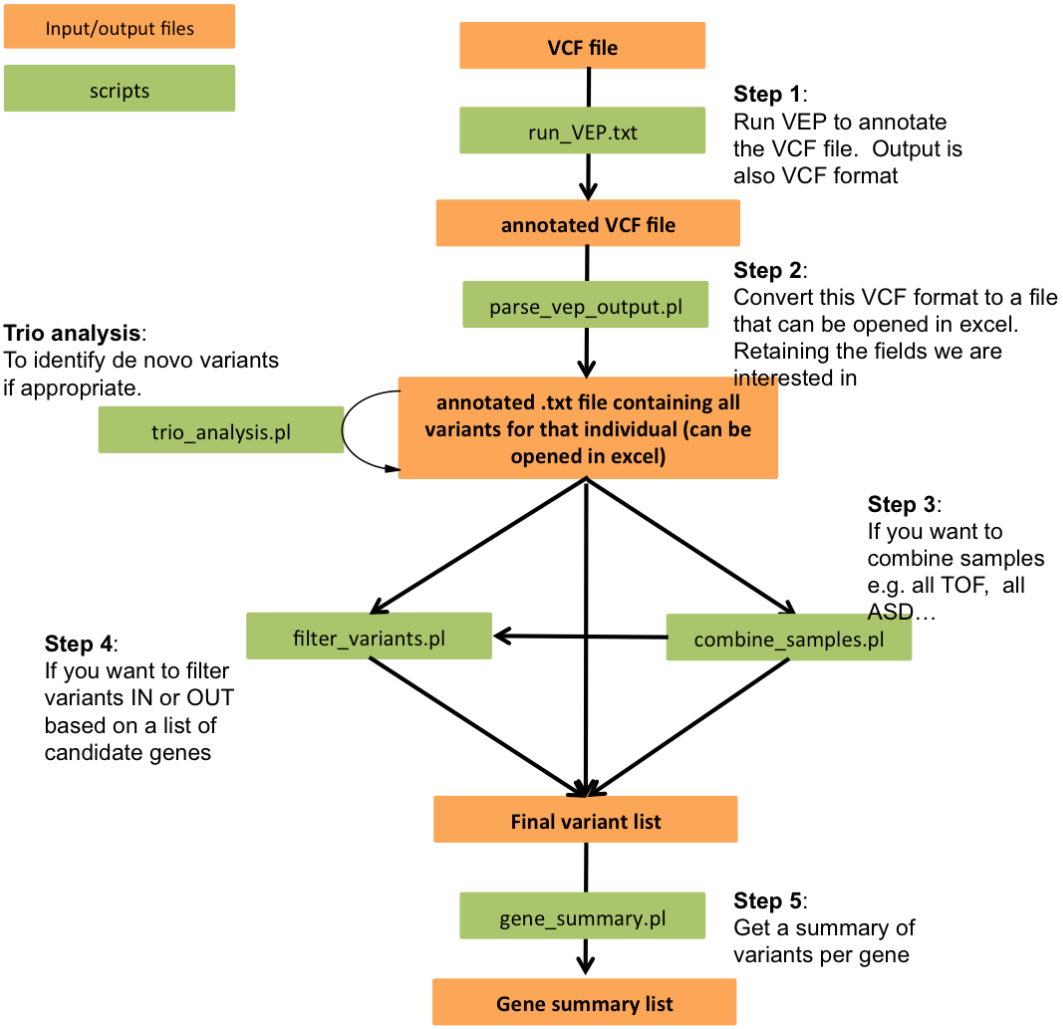


Figure F.1: Schematic representation of the annotation and filtering pipeline. Image courtesy of Dr Simon Williams (University of Manchester).

Procedure:

1. Run VEP to annotate the VCF file:
    a. Get the VCF files for each individual. These should be in gzipped format with the suffix '.vcf.gz'
    b. The script to run VEP on the UCT cluster is /opt/exp_soft/vep/run_VEP.sh. Copy this script to your own directory
    c. Once copied to your local directory, to run this on a single VCF file:

        sbatch run_VEP.sh file1.vcf.gz

    d. This will produce 'file1.annotated.vcf'
    e. To check the progress of your job type **squeue –u <username>**
2. Convert this VCF formatted output file into a file that can be read by excel:

    perl /opt/exp_soft/vep/scripts/parse_vep_output.pl file.annotated.vcf
3. (optional) Combine selected samples:
    a. Make a list of all the files to be combined:

    ls file1.annotated.txt file 2.annotated.txt file3.annotated.txt > input_list
    b. Run the script:

    perl /opt/exp_soft/vep/scripts/combine_samples/pl -sample_list input_list > combined_samples.annotated.txt
4. (optional) Filter variants against a list of candidate genes:
    a. Compile a list of candidate genes of interest (cardiomyopathy_panel; Appendix G)
    b. Run the script:

    perl /opt/exp_soft/vep/scripts/filter_variants.pl -sample file1.annotated.txt -gene_list cardiomyopathy_panel.txt
    c. To return results all variants in genes not on the list (i.e. previously unknown genes), add the **-exclude** flag to the command
5. (optional) Create a gene-based summary of variants:

    perl /opt/exp_soft/vep/scripts/gene_summary.pl –input file.annotated.txt > file1.gene_summary.txt

Trio analysis:

This is an optional step if you have trio files and you want to look for *de novo* or homozygous recessive variants. It should be run after stage 2 has been run for each of the trio samples and you have annotated.txt files for each.

perl /opt/exp_soft/vep/scripts/trio_analysis.pl –proband file_proband.annotated.txt –parent1 file_parent1.annotated.txt –parent2 file_parent2.annotated.vcf > file_proband_trio_analysis_annotated.txt

Appendix G Cardiomyopathy gene panel

A cardiomyopathy gene panel was compiled, consisting of 250 known and putative cardiomyopathy genes according to two sources:

1. The cardiomyopathy panel by Blueprint Genetics (n = 155) (Blue Print Genetics 2019)
2. A custom panel of known or predicted cardiac disease genes (n = 160) (Zou et al. 2015)

Table G.1: Cardiomyopathy gene panel used in NGS data analysis

Gene name	Associated heart disease(s)	Source*
AARS2	Mitochondrial cardiomyopathy	1,2
ABCC6	None (candidate)	1
ABCC9	DCM, atrial fibrillation	1,2
ACAD9	ACAD9 deficiency	1
ACADVL	Acyl-CoA dehydrogenase, very long chain, deficiency	1,2
ACTA1	None (candidate gene)	1,2
ACTC1	DCM, HCM, RCM, LVNC	1,2
ACTN2	DCM, HCM	1,2
AGK	Mitochondrial cardiomyopathy	1,2
AGL	Glycogen storage disease	1,2
AKAP1	None (candidate gene)	2
AKAP10	None (candidate gene)	2
AKAP12	None (candidate gene)	2
AKAP13	None (candidate gene)	2
AKAP5	None (candidate gene)	2
AKAP6	None (candidate gene)	2
AKAP7	None (candidate gene)	2
AKAP9	Long QT syndrome	2
ALMS1	Alström syndrome	1
ALPK3	Paediatric cardiomyopathy	1
ANK2	Long QT syndrome	2
ANKRD1	DCM, HCM	2
ANO5	None (candidate gene)	1
APOA1	Amyloidosis	1
ASPH	None (candidate gene)	2
BAG3	DCM	1,2
BRAF	Noonan syndrome, Cardiofaciocutaneous syndrome	1,2

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
CACNA1C	Long QT syndrome, Brugada syndrome	2
CACNA2D1	Brugada syndrome	2
CACNA2D2	None (candidate gene)	2
CACNA2D3	None (candidate gene)	2
CACNB2	Brugada syndrome	2
CALM1	Catecholaminergic polymorphic ventricular tachycardia	2
CALR3	HCM	1,2
CAPN3	None (candidate gene)	1
CASQ2	Catecholaminergic polymorphic ventricular tachycardia	1,2
CAV3	HCM, Long QT syndrome	2
CBL	Noonan syndrome-like disorder	1
CDH2	ACM	1
CMYA5	None (candidate gene)	2
COX15	Leigh syndrome, Cardioencephalomyopathy	1,2
CPT2	Carnitine palmitoyltransferase II deficiency	1
CRYAB	DCM, HCM	1,2
CSRP3	DCM, HCM	1,2
CTF1	DCM	2
CTNNA3	ACM	1,2
DBH	None (candidate gene)	1
DES	DCM, RCM	1,2
DMD	DCM	1,2
DNAJC19	DCM	1,2
DOLK	None (candidate gene)	1
DPP6	Ventricular fibrillation	2
DSC2	ACM	1,2
DSG2	DCM, ACM	1,2
DSP	DCM, ACM	1,2
DTNA	LVNC	1,2
DYSF	None (candidate gene)	1
EEF1A2	None (candidate gene)	1
ELAC2	Combined oxidative phosphorylation deficiency 17	1
EMD	DCM	1,2
EPG5	Vici syndrome	1
ETFA	None (candidate gene)	1
ETFB	None (candidate gene)	1

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
ETFDH	None (candidate gene)	1
EYA4	DCM	2
FBXO32	DCM	1
FHL1	HCM	1,2
FHL2	DCM	2
FKRP	None (candidate gene)	1
FKTN	DCM	1,2
FLNC	DCM, HCM, ACM, RCM	1,2
FOXD4	DCM	1,2
FOXRED1	Mitochondrial cardiomyopathy, Leigh syndrome	1,2
FXN	None (candidate gene)	1,2
GAA	Glycogen storage disease	1,2
GATA6	Congenital heart defects	1
GATAD1	DCM	1
GBE1	Glycogen storage disease	1
GFM1	Combined oxidative phosphorylation deficiency	1
GJA1	None (candidate gene)	2
GJA5	None (candidate gene)	2
GLA	Fabry disease	1,2
GLB1	GM1-gangliosidosis	1
GMPPB	None (candidate gene)	1
GPD1L	Brugada syndrome	2
GTPBP3	Combined oxidative phosphorylation deficiency 23	1
GUSB	None (candidate gene)	1
HADHA	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency, Trifunctional protein deficiency	1
HAND1	DCM, Congenital heart defects	1
HCN4	LVNC, Brugada syndrome	1,2
HFE	None (candidate gene)	1
HRAS	None (candidate gene)	1
HRC	None (candidate gene)	2
HSPB2	None (candidate gene)	2
HSPB6	None (candidate gene)	2
HSPB7	None (candidate gene)	2
HSPB8	None (candidate gene)	2
ILK	DCM	1,2
ISPD	None (candidate gene)	1

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
JPH2	HCM	1,2
JUP	DCM, ACM	1,2
KCNA5	None (candidate gene)	2
KCND3	None (candidate gene)	2
KCNE1	Long QT syndrome	2
KCNE1L	None (candidate gene)	2
KCNE2	Long QT syndrome	2
KCNE3	Brugada syndrome	2
KCNE4	None (candidate gene)	2
KCNH2	Long QT syndrome	2
KCNIP2	None (candidate gene)	2
KCNJ11	None (candidate gene)	2
KCNJ2	Long QT syndrome	2
KCNJ4	None (candidate gene)	2
KCNJ8	None (candidate gene)	2
KCNK1	None (candidate gene)	2
KCNN2	None (candidate gene)	2
KCNN3	None (candidate gene)	2
KCNQ1	Long QT syndrome	2
KLF10	HCM	2
KRAS	Noonan syndrome, Cardiofaciocutaneous syndrome	1
LAMA2	None (candidate gene)	1
LAMA4	DCM	2
LAMP2	Danon disease	1,2
LARGE	None (candidate gene)	1
LDB3	DCM, LVNC	1,2
LMNA	DCM	1,2
LMOD2	None (candidate gene)	2
LRRC10	DCM	1
LZTR1	Noonan syndrome	1
MAP2K1	Cardiofaciocutaneous syndrome	1
MAP2K2	Cardiofaciocutaneous syndrome	1
MIB1	LVNC	2
MLYCD	Malonyl-CoA decarboxylase deficiency	1
MRPL3	Mitochondrial cardiomyopathy	2
MTO1	HCM, Combined oxidative phosphorylation deficiency	1

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
MURC	HCM	2
MYBPC3	DCM, HCM, LVNC	1,2
MYH6	DCM, HCM	1,2
MYH7	DCM, HCM, LVNC	1,2
MYL2	HCM	1,2
MYL3	HCM	1,2
MYL4	Atrial fibrillation	1
MYLK2	None (candidate gene)	2
MYOM1	None (candidate gene)	2
MYOT	None (candidate gene)	1,2
MYOZ2	HCM	2
MYPN	DCM, HCM, RCM	1,2
NDUFAF1	Mitochondrial cardiomyopathy	2
NDUFAF2	Mitochondrial cardiomyopathy, Leigh syndrome	1
NDUFS2	Mitochondrial cardiomyopathy	2
NEBL	DCM	2
NEXN	DCM, HCM	1,2
NF1	Neurofibromatosis-Noonan syndrome	1
NOS1AP	None (candidate gene)	2
NPPA	Atrial fibrillation	2
NRAS	Noonan syndrome	1
OBSCN	HCM	2
PCCA	None (candidate gene)	1
PCCB	None (candidate gene)	1
PDE3A	None (candidate gene)	2
PDE4A	None (candidate gene)	2
PDE4B	None (candidate gene)	2
PDE4D	None (candidate gene)	2
PDE4DIP	None (candidate gene)	2
PDLIM3	DCM	2
PKP2	ACM	1,2
PLEC	None (candidate gene)	1
PLEKHM2	DCM, LVNC	1
PLN	DCM, HCM	1,2
PNPLA2	Neutral lipid storage disease with myopathy	1
PPA2	Sudden cardiac failure	1

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
PPP1CB	Noonan syndrome-like disorder with loose anagen hair 2	1
PRDM16	DCM, LVNC	1
PRKAG2	HCM	1,2
PSEN1	DCM	2
PSEN2	DCM, PPCM	2
PTPN11	None (candidate gene)	1
RAF1	DCM	1
RANGRF	Brugada syndrome	2
RASA2	Noonan syndrome	1
RBCK1	None (candidate gene)	1
RBM20	DCM	1,2
RBM24	None (candidate gene)	2
RIT1	Noonan syndrome	1
RMND1	Combined oxidative phosphorylation deficiency	1
RRAS	Noonan-syndrome like phenotype	1
RYR2	ACM, Catecholaminergic polymorphic ventricular tachycardia	1,2
SCN10A	None (candidate gene)	2
SCN1B	Brugada syndrome	2
SCN2B	None (candidate gene)	2
SCN3B	Brugada syndrome	2
SCN4B	Long QT syndrome	2
SCN5A	DCM, atrial/ventricular fibrillation, Long QT syndrome, Brugada syndrome	1,2
SCN7A	None (candidate gene)	2
SCNN1B	None (candidate gene)	1
SCNN1G	None (candidate gene)	1
SCO2	HCM, Mitochondrial cardiomyopathy	1,2
SDHA	DCM, Mitochondrial cardiomyopathy, Leigh syndrome	1,2
SELENON	None (candidate gene)	1
SGCA	None (candidate gene)	1
SGCB	None (candidate gene)	1
SGCD	DCM	1,2
SGCG	None (candidate gene)	1
SHOC2	Noonan-like syndrome with loose anagen hair	1
SLC22A5	Carnitine deficiency	1
SLC25A20	Carnitine-acylcarnitine translocase deficiency	1
SLC25A3	Mitochondrial cardiomyopathy	2

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
SLC25A4	Mitochondrial cardiomyopathy	1,2
SLMAP	Brugada syndrome	2
SMCHD1	None (candidate gene)	1
SNTA1	Long QT syndrome	2
SOS1	Noonan syndrome	1
SOS2	Noonan syndrome	1
SPRED1	None (candidate gene)	1
SVIL	None (candidate gene)	2
SYNE1	DCM	2
SYNE2	DCM	2
SYNM	DCM	2
TAB2	Congenital heart defects	1
TAZ	DCM, Mitochondrial cardiomyopathy, LVNC	1,2
TBX20	Atrial septal defect	1
TBX5	Holt-Oram syndrome	1
TCAP	DCM, HCM	1,2
TGFB3	ACM	1,2
TLN1	None (candidate gene)	2
TMEM43	ACM	1,2
TMEM70	Mitochondrial cardiomyopathy	1,2
TMOD1	None (candidate gene)	2
TMPO	DCM	2
TNNC1	DCM, HCM	1,2
TNNI3	DCM, HCM, RCM	1,2
TNNI3K	Cardiac conduction disease with or without DCM	1
TNNT2	DCM, HCM, RCM, LVNC	1,2
TOR1AIP1	None (candidate gene)	1
TPM1	DCM, HCM	1,2
TRDN	Catecholaminergic polymorphic ventricular tachycardia	2
TRIM32	None (candidate gene)	1
TRIM63	None (candidate gene)	2
TSFM	Combined oxidative phosphorylation deficiency	1
TTN	DCM, HCM, ACM	1,2
TTR	HCM	1,2
USP13	None (candidate gene)	2
UTRN	None (candidate gene)	2

Continued on next page

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Table G.1 continued

Gene name	Associated heart disease(s)	Source*
VCL	DCM, HCM	1,2
VCP	None (candidate gene)	1
VIM	None (candidate gene)	2
VPS13A	None (candidate gene)	1
XIRP1	None (candidate gene)	2
XK	McLeod syndrome	1
ZYX	None (candidate gene)	2

** 1: Blueprint Genetics cardiomyopathy gene panel; 2: known and predicted cardiomyopathy genes*
ACM, arrhythmogenic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; LVNC, left ventricular noncompaction; PPCM, peripartum cardiomyopathy; RCM, restrictive cardiomyopathy

Appendix H CRISPR guide RNA synthesis and purification protocols

Note: exceptions to the protocols are written in **bold**

QIAquick PCR purification kit protocol

Before starting:

- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature.
- Add 1:250 volume pH indicator I to Buffer PB (i.e., add 120 µl pH indicator I to 30 ml Buffer PB or add 600 µl pH indicator I to 150 ml Buffer PB). The yellow colour of Buffer PB with pH indicator I indicates a pH of ≤ 7.5. **(not done)**
- Add pH indicator I to entire buffer contents. Do not add pH indicator I to buffer aliquots. **(not done)**
- If the purified PCR product is to be used in sensitive microarray applications, it may be beneficial to use Buffer PB without the addition of pH indicator I.

Procedure:

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix. It is not necessary to remove mineral oil or kerosene. **(6 volumes of Buffer PB were added)**
2. If pH indicator I has been added to Buffer PB, check that the colour of the mixture is yellow. **(not done)**
3. Place a QIAquick spin column in a provided 2 ml collection tube.
4. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30–60 s.
5. Discard flow-through. Place the QIAquick column back into the same tube.
6. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30–60 s.
7. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min.
8. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.
9. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) to the centre of the QIAquick membrane and centrifuge the column for 1 min. Alternatively, for increased DNA concentration, add 30 µl elution buffer to the centre of the QIAquick membrane, let the column stand for 1 min and then centrifuge. **(elutions were performed using 30 µl of nuclease-free water)**
10. If the purified DNA is to be analysed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel. **(not done)**

MEGAclear™ Transcription Clean-Up kit protocol

Before starting and other notes:

- Add 20 ml of ACS grade 100% ethanol to the bottle labelled Wash Solution Concentrate. Mix well. Place a check in the box on the label to indicate that the ethanol was added. With the ethanol, this solution will be referred to as Wash Solution.
- Before working with RNA, it is always a good idea to clean the lab bench and pipettors with an RNase decontamination solution (e.g. Ambion® RNaseZap® Solution).
- Wear laboratory gloves at all times during this procedure and change them frequently. They will protect you from the reagents, and they will protect the RNA from nucleases that are present on skin.
- Use RNase-free pipette tips to handle the Wash Solution and the Elution Solution and avoid putting used tips into the reagent containers.
- Use the Collection and Elution Tubes supplied with the kit; they have been tested for RNase contamination and are certified RNase-free.

Procedure:

1. Bring the RNA sample to 100 µL with Elution Solution. Mix gently but thoroughly.
2. Add 350 µL of Binding Solution Concentrate to the sample. Mix gently by pipetting.
3. Add 250 µL of 100% ethanol to the sample. Mix gently by pipetting.
4. Apply the sample to the filter:
    a. Insert a Filter Cartridge into 1 of the Collection and Elution Tubes supplied.
    b. Pipet the RNA mixture onto the Filter Cartridge.
    c. Centrifuge for ~15 sec to 1 min, or until the mixture has passed through the filter. Centrifuge at RCF 10,000–15,000 × g (typically 10,000–14,000 rpm). Spinning harder than this may damage the filters.
    d. Discard the flow-through and reuse the Collection and Elution Tube for the washing steps.
5. Wash with 2 × 500 µL Wash Solution:
    a. Apply 500 µL Wash Solution. Draw the Wash Solution through the filter as in the previous step.
    b. Repeat with a second 500 µL aliquot of Wash Solution.
    c. After discarding the Wash Solution, continue centrifugation for 10–30 sec to remove the last traces of Wash Solution.
6. Elute RNA from the filter with 50 µL Elution Solution:
    a. Place the Filter Cartridge into a new Collection/Elution Tube.
    b. Apply 50 µL of Elution Solution to the centre of the Filter Cartridge. Close the cap of the tube and incubate in a heat block set to 65–70°C for 5–10 min. **(50 µl nuclease-free water was used instead of Elution Solution)**
    c. Recover eluted RNA by centrifuging for 1 min at RT (RCF 10,000–15,000 × g).
    d. To maximise RNA recovery, repeat this elution procedure with a second 50 µL aliquot of Elution Solution. Collect the eluate into the same tube. **(to maximise the concentration, the flow-through was used instead of a second aliquot)**

Appendix I Bioline Isolate II purification protocols

Note: exceptions to the protocols are written in **bold**

Isolate II PCR and gel kit – gel purification protocol

1. Excise and dissolve gel slice:
    a. Using a clean scalpel, excise DNA fragment from gel.
    b. Remove excess agarose, determine weight of gel slice and transfer into a clean tube.
    c. Add 200 µl Binding Buffer CB per 100 mg of 2% agarose gel*
    ** For gels containing > 2% agarose, double the volume of Binding Buffer CB.*
    d. Incubate sample at 50°C for 5-10 min, vortexing sample briefly every 2-3 min until gel slice is completely dissolved.
2. Bind DNA:
    a. Place ISOLATE II PCR and Gel Column in a 2 ml Collection Tube and load sample.
    b. Centrifuge 30 s at 11,000 x g and discard flow-through.
    c. Reuse collection tube for step 3.
3. Wash silica membrane:
    a. Add 700 µl Wash Buffer CW to ISOLATE II PCR and Gel Column.
    b. Centrifuge 30 s at 11,000 x g.
    c. Discard flow-through and place column back into collection tube.
    Recommended: Repeat washing step to minimise chaotropic salt carry-over. **(done)**
4. Dry silica membrane:
    a. Centrifuge 1 min at 11,000 x g, to remove residual ethanol.
    b. Place ISOLATE II PCR and Gel Column in a 1.5 ml microcentrifuge tube (not supplied).
5. Elute DNA:
    a. Add 15-30 µl Elution Buffer C directly onto silica membrane. **(20 µl was used)**
    b. Incubate at room temperature for 1 min.
    c. Centrifuge 1 min at 11,000 x g.

Isolate II PCR and gel kit – PCR purification protocol

1. Sample preparation:
    a. For volumes < 30 µl, adjust volume to 50-100 µl with water.
    b. Mix 1 volume of sample with 2 volumes of Binding Buffer CB.
2. Bind DNA:
    a. Place ISOLATE II PCR and Gel Column in a 2 ml Collection Tube and load sample.
    b. Centrifuge 30 s at 11,000 x g and discard flow-through.
    c. Reuse collection tube for step 3.
3. Wash silica membrane:
    a. Add 700 µl Wash Buffer CW to ISOLATE II PCR and Gel Column.
    b. Centrifuge 30 s at 11,000 x g.
    c. Discard flow-through and place column back into collection tube.
    Recommended: Repeat washing step to minimise chaotropic salt carry-over. **(done)**
4. Dry silica membrane:
    a. Centrifuge 1 min at 11,000 x g, to remove residual ethanol.
    b. Place ISOLATE II PCR and Gel Column in a 1.5 ml microcentrifuge tube (not supplied).
5. Elute DNA:
    a. Add 15-30 µl Elution Buffer C directly onto silica membrane. **(20 µl was used)**
    b. Incubate at room temperature for 1 min.
    c. Centrifuge 1 min at 11,000 x g.

Isolate II plasmid mini kit protocol

1. Harvest bacterial cells:
    a. Pellet 1-5 ml of a saturated *E. coli* LB culture for 30 s at 11,000 x g. **(the total volume of *E. coli* in LB was centrifuged – to do this, cultures were separated into aliquots of 5 ml)**
    b. Discard supernatant and remove as much liquid as possible.
2. Lyse cells:
    a. Add 250 µl Resuspension Buffer P1 and resuspend cell pellet by vortexing or pipetting up and down.
    b. Add 250 µl Lysis Buffer P2. Mix gently by inverting tube 6-8 times. Incubate at room temperature for up to 5 min or until lysate appears clear.
    c. Add 300 µl Neutralisation Buffer P3. Mix thoroughly by inverting tube 6-8 times.
3. Clarification of lysate:
    a. Centrifuge 5 min at 11,000 x g at room temperature.
4. Bind DNA:
    a. Place ISOLATE II Plasmid Mini Spin Column in a 2 ml Collection Tube (supplied).
    b. Decant or pipette a maximum of 750 µl of clarified sample supernatant onto column. **(750 µl of each sample was used)**
    c. Centrifuge 1 min at 11,000 x g and discard flow-through.
5. Wash silica membrane:
    a. If plasmid DNA is prepared from host strains containing high levels of nucleases, an extra wash with Wash Buffer PW1 is strongly recommended.
    b. (Optional) Add 500 µl Wash Buffer PW1 preheated to 50°C. Centrifuge 1 min at 11,000 x g. **(done)**
    c. Add 600 µl Wash Buffer PW2 (supplemented with ethanol). Centrifuge 1 min at 11,000 x g.
    d. Discard flow-through and reuse Collection Tube.
6. Dry silica membrane:
    a. Centrifuge 2 min at 11,000 x g, to remove residual ethanol.
    b. Place ISOLATE II Plasmid Mini Spin Column in a 1.5 ml microcentrifuge tube (not supplied).
7. Elute DNA:
    a. Add 50 µl Elution Buffer P directly onto centre of silica membrane. **(50 µl nuclease-free water was used instead of Elution Solution)**
    b. Incubate at room temperature for 1 min.
    c. Centrifuge 1 min at 11,000 x g.

Appendix J Filtered variants in Family 1

Table J.1: Filtered cardiomyopathy panel variants in Family 1

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
MYH7	c.4394C>T	p.S1465L	0	0	0	0	D (0.581)	D (0.67)	D (0)	B (0.322)	25.4	4514.0	Yes	7/7
MYOM1	c.139A>G	p.S47G	0.003	0.005376	0.007187	0.00201	NS	D (0.76)	B (0.08)	B (0.079)	21.9	207.1	NA	5/7
PDE4DIP	c.5599C>T	p.R1867C	0	0.500000	0.500000	0.00008	NS	D (0.94)	D (0.01)	D (0.989)	23.3	47.7	Yes	3/7
PDE4DIP	c.824C>T	p.S275L	0	0.500000	0.500000	0.00008	NS	B (0.99)	B (0.15)	B (0.031)	21.9	47.7	Yes	3/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table J.2: Other filtered variants in Family 1

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
ACO2	c.220C>G	p.L74V	0.002	0.004144	0.003738	0.00415	D (0.175)	D (0.99)	D (0.01)	D (0.485)	22.7	252.3	NA	5/7
ADAT1	c.433G>A	p.A145T	0	0	0	0	D (0.362)	D (0.99)	D (0)	D (1)	32.0	1.8	NA	7/7
APMAP	c.479A>G	p.N160S	0.001	0.001351	0.001202	0.00162	B (0.010)	D (0.99)	D (0.04)	B (0.02)	23.2	19.1	NA	6/7
ARHGAP4	c.1751T>G	p.V584G	0	0	0	0	D (0.252)	D (0.99)	D (0)	D (0.987)	25.2	6.3	NA	5/7
ASB10	c.709C>G	p.R237G	0.003	0.004366	0.004653	0.00508	B (0.017)	B (0.99)	B (0.16)	B (0.262)	22.6	12.4	No	2/7
ATG2A	c.1585_6del	p.R529G*4	0	0	0	0	NS	D (1)	NS	NS	27.2	6.8	NA	NS
ATP7B	c.1518A>G	p.I506M	0	0	0	0	D (0.883)	D (0.99)	D (0)	D (0.999)	20.5	1.6	No	5/7
BCAP31	c.584C>T	p.T195M	0.001	0.001327	0.001356	0.00114	D (0.414)	B (0.00)	D (0.01)	D (0.986)	25.9	42.9	NA	5/7
CCDC92	c.602C>T	p.T201M	0.0004	0.000231	0.000181	0.00031	D (0.075)	D (0.99)	B (0.09)	D (0.996)	23.5	14.6	No	5/7
CEP350	c.1270T>A	p.S424T	0	0.000683	0.000354	0.00046	B (0.019)	D (0.99)	B (0.13)	D (0.955)	23.5	3.3	NA	6/7
CMKLR1	c.311A>G	p.H104R	0	0	0	0	D (0.100)	B (0.52)	D (0)	D (0.980)	23.6	2.4	No	4/7
CPEB2	c.337G>T	p.A113S	0	0	0	0	D (0.619)	B (0.99)	B (0.25)	Unknown	21.4	3.7	Yes	1/7

Continued on next page

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table J.2 continued

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
CSNK1D	c.1_2dup	p.M1?	0	0	0	0	NS	D (0.99)	NS	NS	23.2	23.9	No	NS
DAAM2	c.1664C>T	p.P555L	0.001	0.002073	0.000950	0.00121	D (0.157)	D (0.99)	B (0.18)	D (0.994)	22.0	8.1	No	5/7
DMXL2	c.6046G>A	p.D2016N	0.0004	0.002070	0.001769	0.00270	D (0.027)	D (0.99)	B (0.33)	B (0.012)	20.5	1.3	No	4/7
EIF4ENIF1	c.97G>A	p.E33K	0.003	0.003778	0.003884	0.00315	B (0.016)	D (0.99)	B (0.08)	D (0.991)	32.0	7.6	Yes	7/7
EIF5	c.998T>C	p.I333T	0	0	0	0	D (0.050)	D (0.99)	B (0.19)	D (0.586)	23.8	30.4	NA	5/7
ERCC4	c.1031A>T	p.Y344F	0.001	0.000091	0.000117	0.00054	D (0.026)	D (0.99)	B (0.1)	B (0.026)	22.7	1.3	No	7/7
ESYT1	c.797G>A	p.R266Q	0.0004	0.000223	0.000212	0.00046	B (0.017)	D (0.99)	B (0.13)	B (0.14)	23.6	15.3	NA	6/7
ETV3	c.98G>A	p.R33Q	0	0	0.000032	0	B (0.024)	D (0.99)	D (0)	D (0.991)	31.0	2.9	No	6/7
FAM135A	c.626G>A	p.S209N	0.003	0.006127	0.005981	0.00423	NS	D (0.99)	B (0.66)	D (0.977)	22.9	1.8	NA	6/7
FGD2	c.206G>C	p.S69T	0.001	0.002056	0.001832	0.00277	B (0.015)	B (0.59)	B (0.08)	D (0.971)	23.5	2.0	NA	3/7
FMO2	c.584_5inv	p.S195K	0	0.396000	0.394900	0	NS	D (0.99)	D (0)	B (0.366)	23.7	19.7	No	5/7
GALNS	c.499T>G	p.F167V	0.0004	0.000900	0.001068	0.00115	D (0.887)	D (0.99)	B (0.09)	D (0.529)	24.3	4.6	No	5/7
GORASP1	c.380C>T	p.A127V	0.002	0.005482	0.004923	0.00431	D (0.069)	D (0.99)	D (0)	D (0.971)	24.2	13.2	No	6/7
ITPR2	c.3539G>A	p.R1180Q	0.003	0.004389	0.004538	0.00566	NS	D (0.61)	B (0.99)	B (0)	21.9	1.6	No	4/7
JAG2	c.2459C>T	p.A820V	0	0.002023	0.001231	0.00117	NS	D (0.99)	D (0)	D (0.993)	25.1	12.5	No	5/7
KCNJ12	c.467C>T	p.P156L	0	0.499700	0.499800	0	NS	D (0.99)	D (0.04)	D (1)	29.5	3.8	No	6/7
KIAA1217	c.1523G>A	p.R508H	0.003	0.005745	0.005301	0.00423	NS	D (0.99)	D (0)	D (0.998)	29.2	6.9	No	5/7
KIFC2	c.466del	p.D156M*54	0	0	0	0	NS	D (1)	NS	NS	22.6	5.3	No	NS
LAMA3	c.4643A>G	p.D1548G	0.0002	0.000878	0.000941	0.00138	B (0.012)	D (0.99)	B (0.06)	D (0.795)	25.1	1.2	No	4/7
LEPREL1	c.1694T>A	p.L565Q	0	0	0.000008	0	D (0.025)	D (0.99)	D (0)	D (0.776)	29.9	1.1	No	7/7
LGI4	c.68C>G	p.P23R	0	0	0	0	D (0.025)	D (0.99)	B (0.27)	D (0.659)	23.5	6.8	No	4/7
LRCH4	c.853G>C	p.A285P	0.004	0.005015	0.003946	0.00023	NS	B (0.99)	B (0.28)	B (0.027)	20.2	19.2	NA	5/7
MAP2K3	c.286C>T	p.R96W	0	0.4998	0.490700	0	NS	D (0.99)	D (0)	D (0.992)	33.0	20.7	No	5/7
MAST4	c.3922C>T	p.R1308W	0	0.000124	0.000128	0.00033	D (0.125)	D (0.68)	D (0)	D (1)	24.1	7.6	No	2/7
MDN1	c.14648A>G	p.D4883G	0.006	0.010190	0.009027	0.01084	NS	D (0.99)	D (0)	D (0.494)	25.3	4.2	No	5/7
METTL1	c.137G>C	p.W46S	0	0	0	0	D (0.063)	D (0.99)	D (0)	D (0.988)	31.0	2.9	NA	6/7
MRI1	c.269G>T	p.R90L	0.001	0.002198	0.002628	0.00191	D (0.399)	D (0.99)	D (0.03)	D (1)	26.6	4.7	NA	5/7
MTERF	c.994A>G	p.I332V	0.0002	0.000305	0.000290	0.00008	B (0.004)	D (0.57)	B (0.05)	B (0.015)	21.4	1.5	Yes	4/7
MTIF3	c.413A>G	p.Q138R	0.004	0.008140	0.007611	0.00869	NS	D (0.61)	D (0)	B (0.265)	22.2	61.8	No	7/7
NDEL1	c.100C>G	p.R34G	0	0	0	0	B (0.008)	D (0.99)	D (0)	B (0.258)	25.7	8.2	No	7/7
NDNL2	c.400C>T	p.L134F	0	0	0	0	B (0.008)	D (0.99)	B (0.11)	D (0.914)	24.0	3.3	NA	3/7
NRP1	c.1295C>T	p.S432F	0.0004	0.000348	0.000384	0.00023	D (0.826)	D (0.99)	D (0)	D (0.997)	28.6	16.6	No	6/7
NSL1	c.794C>T	p.P265L	0	0.000734	0.000651	0.00054	B (0.017)	D (0.99)	D (0)	B (0.251)	24.1	4.1	No	4/7
OR51E1	c.896A>G	p.K299R	0.004	0.004533	0.005033	0.00369	NS	B (0.86)	D (0.04)	B (0.025)	23.2	1.6	NA	5/7
PCK2	c.1679G>A	p.R560Q	0.0002	0.000503	0.000446	0.00069	B (0.010)	D (0.99)	B (0.11)	B (0.103)	24.2	1.3	NA	4/7
PCNX	c.3817G>T	p.G1273C	0.0004	0.000973	0.000945	0.00092	D (0.025)	D (0.99)	D (0)	D (0.754)	29.7	5.5	NA	5/7
PHC3	c.1532T>C	p.I511T	0	0.000066	0.000064	0.00008	B (0.005)	B (0.85)	D (0.03)	B (0.035)	22.8	2.4	Yes	5/7

Continued on next page

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table J.2 continued

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
PITPNM1	c.2398G>A	p.G800S	0.0002	0.000062	0.000028	0	B (0.021)	D (0.99)	B (0.36)	D (0.649)	23.5	14.4	No	5/7
PLXND1	c.2174T>C	p.F725S	0	0	0	0	B (0.009)	B (0.97)	B (0.39)	B (0.007)	21.0	18.4	Yes	4/7
PMM1	c.588G>C	p.W196C	0	0.000482	0.000439	0.00031	D (0.880)	D (0.99)	D (0.01)	D (1)	24.5	12.8	No	5/7
PPP1R3A	c.785A>G	p.K262R	0	0	0	0	B (0.009)	D (0.99)	D (0.02)	B (0.354)	24.1	17.3	No	6/7
RIC3	c.934G>A	p.D312N	0.003	0.003748	0.003882	0.00570	B (0.013)	D (0.99)	B (0.05)	D (0.998)	24.4	2.9	NA	6/7
RIOK2	c.730A>G	p.M244V	0.004	0.009764	0.008764	0.00754	NS	D (0.99)	B (0.27)	B (0.088)	21.8	3.4	Yes	6/7
RRAGD	c.203T>C	p.M68T	0	0	0	0	D (0.400)	D (0.99)	D (0)	D (0.997)	26.4	24.2	NA	6/7
SBF2	c.2938A>G	p.I980V	0	0.000074	0.000060	0.00023	D (0.036)	D (0.99)	B (0.06)	B (0.109)	21.7	4.2	No	6/7
SEC24A	c.3053C>T	p.P1018L	0.001	0.003018	0.002998	0.00136	NS	D (0.99)	D (0)	D (0.923)	27.2	2.6	No	6/7
SEMA6C	c.824G>A	p.R275H	0.0004	0.000129	0.002628	0	B (0.010)	D (0.99)	D (0)	D (0.954)	28.1	17.7	No	4/7
SIPA1L1	c.166C>A	p.P56T	0.002	0.007484	0.007982	0.00869	NS	D (0.99)	B (0.08)	B (0.193)	20.6	1.1	No	6/7
SLC9A9	c.991G>A	p.G331S	0.0002	0.000083	0.000078	0	B (0.008)	D (0.99)	D (0.02)	B (0.302)	23.7	1.3	NA	5/7
SORL1	c.6194A>T	p.D2065V	0.001	0.002823	0.002537	0.00423	D (0.071)	D (0.99)	D (0)	D (0.997)	28.5	1.0	No	5/7
SPAG4	c.5G>A	p.R2Q	0.002	0.000544	0.000540	0.00178	B (0.011)	B (0.62)	B (0.06)	B (0.003)	21.8	1.4	No	4/7
STAB1	c.5195C>T	p.A1732V	0	0	0	0	D (0.141)	D (0.52)	D (0.03)	D (0.999)	24.0	19.8	NA	5/7
STARD13	c.3022C>T	p.P1008S	0.004	0.000840	0.000870	0.00223	NS	D (0.99)	D (0)	D (0.996)	31.0	5.9	No	7/7
SYTL3	c.1169G>A	p.R390Q	0	0.000161	0.000196	0.00015	B (0.019)	D (0.98)	D (0)	B (0.434)	22.7	3.9	NA	5/7
TBC1D10A	c.997G>A	p.G333S	0	0	0	0	B (0.017)	D (0.99)	D (0)	B (0.223)	26.0	4.8	No	6/7
TJP1	c.1412A>G	p.N471S	0.005	0.009164	0.009063	0.00975	NS	D (0.99)	D (0.02)	B (0.04)	23.0	15.0	Yes	7/7
TMED8	c.361G>A	p.D121N	0.001	0.003839	0.004707	0.00269	NS	D (0.99)	D (0.02)	D (0.996)	28.8	2.0	NA	6/7
TMEM41A	c.586T>C	p.Y196H	0.002	0.005235	0.005435	0.00538	NS	D (0.99)	D (0.03)	D (1)	32.0	5. 5	NA	7/7
TNXB	c.1307C>G	p.P436R	0	0	0	0	B (0.018)	B (0.80)	D (0)	D (0.997)	25.3	10.6	Yes	1/7
TRIM4	c.586C>T	p.R196*	0	0.000041	0.000032	0	NS	D (0.99)	NS	NS	35.0	3.8	NA	1/7
TTC21B	c.3004C>G	p.L1002V	0.001	0.005608	0.005246	0.00654	D (0.027)	D (0.99)	D (0)	D (0.628)	24.3	3.4	No	6/7
TTI1	c.2089T>C	p.Y697H	0.001	0.000264	0.000308	0.00077	D (0.042)	D (0.99)	D (0)	D (0.999)	28.3	4.9	NA	7/7
UGGT2	c.3820T>C	p.F1274L	0.005	0.007618	0.007852	0.00716	NS	D (0.99)	B (0.17)	D (0.791)	25.7	2.3	NA	6/7
VILL	c.838G>C	p.E280Q	0.0004	0.000369	0.000277	0.00046	B (0.012)	B (0.85)	B (0.07)	B (0.296)	20.0	1.9	NA	3/7
VTN	c.1208C>T	p.S403F	0	0.000272	0.000264	0	B (0.006)	B (0.67)	B (0.15)	D (0.473)	20.8	5.5	No	5/7
XRN2	c.1173G>C	p.M391I	0.0002	0.000363	0.000305	0.00054	B (0.009)	D (0.99)	B (0.09)	B (0.003)	22.7	17.4	NA	7/7
ZNF548	c.982A>G	p.K328E	0	0	0	0	B (0.004)	B (0.98)	D (0)	D (0.964)	24.7	2.6	NA	1/7
ZNF623	c.446A>T	p.K149I	0.002	0.004336	0.004117	0.00546	D (0.054)	B (0.80)	D (0)	D (0.887)	23.9	2.7	NA	2/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Appendix K Segregation of candidate variants in Family 1

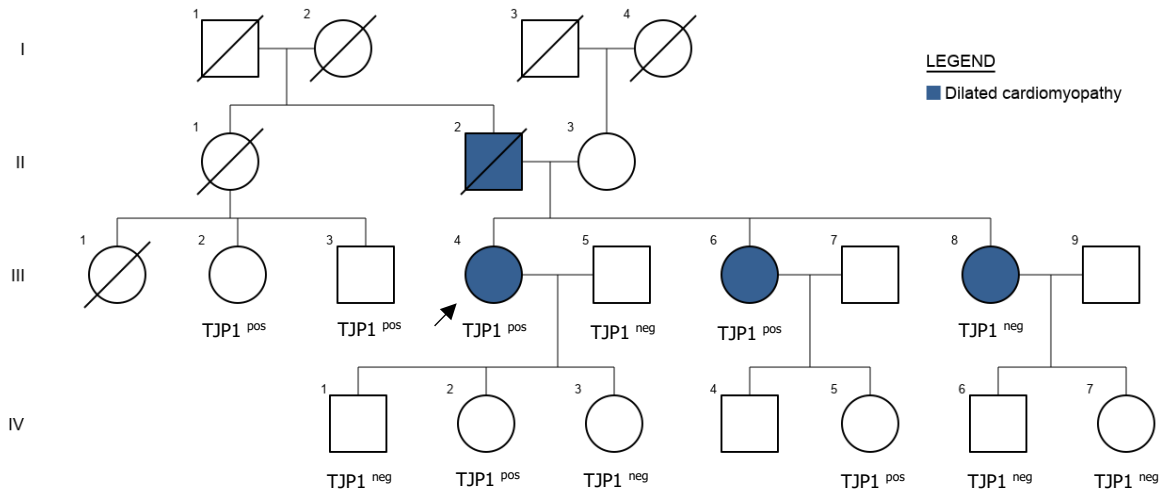


Figure K.1: Pedigree indicating segregation of *TJP1* c.1412A>G in Family 1. Individuals marked 'TJP1 ^{pos}' were heterozygous for the variant and those marked 'TJP1 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates individuals diagnosed with DCM. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

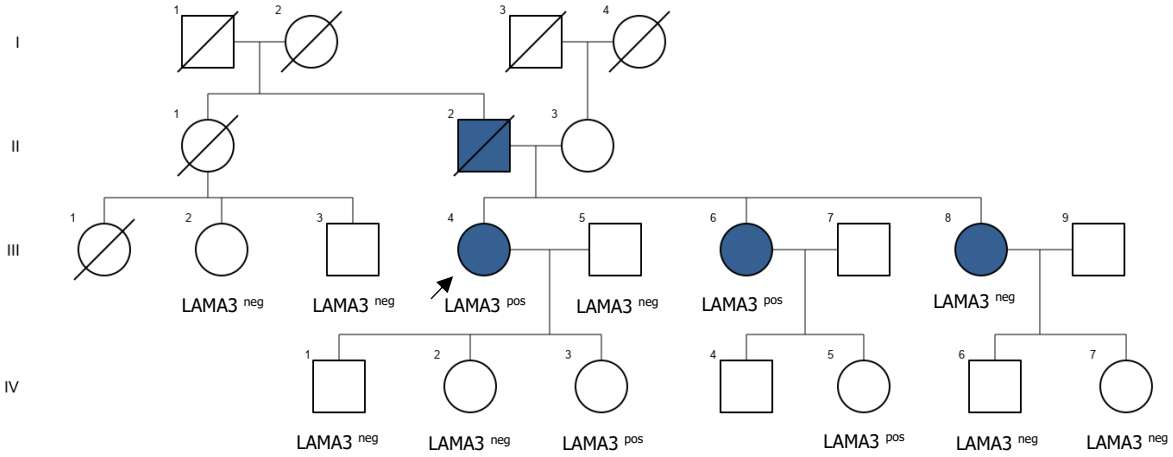


Figure K.2: Pedigree indicating segregation of *LAMA3* c.4643A>G in Family 1. Individuals marked 'LAMA3 ^{pos}' were heterozygous for the variant and those marked 'LAMA3 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates individuals diagnosed with DCM. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

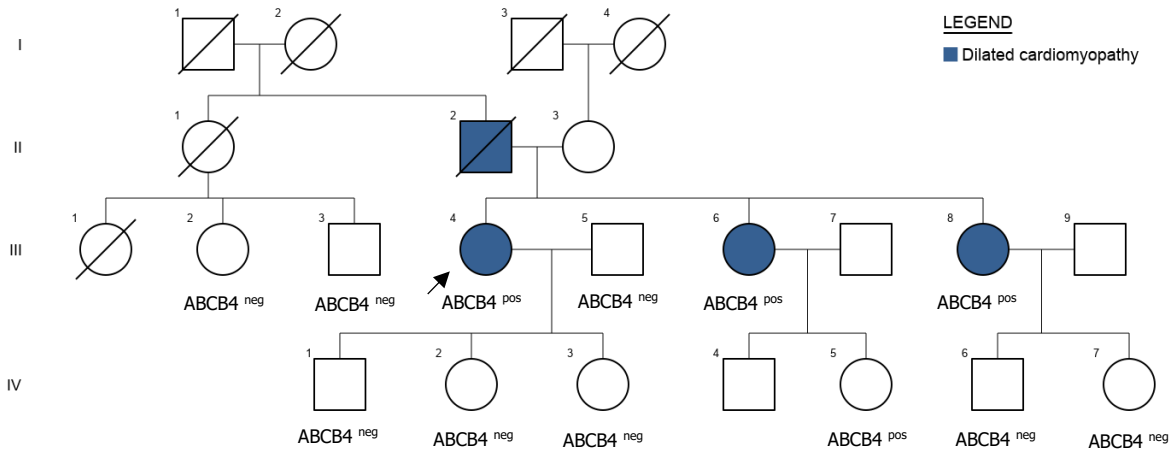


Figure K.3: Pedigree indicating segregation of *ABCB4* c.523A>G in Family 1. Individuals marked 'ABCB4 ^{pos}' were heterozygous for the variant and those marked 'ABCB4 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates individuals diagnosed with DCM. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

Appendix L Filtered variants in Family 2

Table L.1: Filtered cardiomyopathy panel variants in Family 2

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
GBE1	c.664A>G	p.N222D	0	0	0	0	D (0.040)	D (0.99)	B (0.37)	B (0.021)	22.4	24.5	Yes	3/7
GLA	c.774_775delAC	p.P259R*5	0	0	0	0	NS	D (1)	NS	NS	35.0	4.0	Yes	NS
LAMA2	c.7040G>T	p.G2347V	0.004	0.000157	0.0001156	0	D (0.148)	D (0.99)	D (0)	D (0.952)	32.0	33.1	No	5/7
MYBPC3	c.3392T>C	p.I1131T	0	0.001064	0.0008403	0.00057	D (0.184)	D (0.99)	D (0.04)	B (0.400)	24.3	1369.0	Yes	7/7
PDE4DIP	c.1864C>T	p.R622*	0	0.500000	0.500000	0	NS	D (1)	NS	NS	35.0	47.7	Yes	NS
PKP2	c.1592T>G	p.I531S	0.001	0.004722	0.004722	0.00323	NS	D (0.80)	D (0)	D (0.974)	24.8	84.1	Yes	6/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table L.2: Other filtered variants in Family 2

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
ADAM17	c.148A>G	p.I50V	0.002	0.003696	0.003778	0.00300	B (0.004)	D (0.77)	B (0.25)	B (0.354)	20.8	3.5	Yes	5/7
ADCK1	c.386A>G	p.E129G	0	0.000008	0.000048	0	D (0.066)	D (0.99)	D (0.01)	B (0.127)	24.0	5.6	No	6/7
AGRN	c.1660G>A	p.V554M	0.011	0.008835	0.008361	0.00300	NS	D (0.99)	D (0)	D (0.997)	24.9	6.8	No	6/7
ANK3	c.4400A>G	p.K1467R	0.0002	0.000675	0.000587	0.00100	B (0.022)	D (0.99)	D (0.01)	D (0.824)	26.7	5.7	No	5/7
BLOC1S2	c.11C>T	p.A4V	0	0.000228	0.000206	0.00025	B (0.020)	B (0.99)	D (0.03)	B (0)	20.9	6.4	Yes	2/7
C21orf91	c.16C>A	p.Q6K	0	0	0	0	B (0.006)	D (0.99)	D (0.02)	B (0.19)	23.7	1.7	NA	5/7
C2CD3	c.2659G>A	p.V887M	0.003	0.004750	0.004726	0.00524	NS	B (0.74)	B (0.35)	D (0.565)	23.2	2.1	Yes	3/7
CAD	c.1433G>C	p.G478A	0	0	0	0	D (0.620)	D (0.99)	D (0)	D (0.962)	25.2	2.8	No	5/7
CCDC85B	c.451C>G	p.R151G	0	0	0	0	D (0.114)	D (0.99)	B (0.23)	B (0.434)	23.0	33.7	NA	5/7
CELSR1	c.4447G>A	p.G1483S	0	0	0.000004	0	D (0.258)	D (0.99)	D (0)	D (0.998)	28.4	1.2	No	1/7

Continued on next page

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table L.2 continued

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
CENPO	c.583G>C	p.D195H	0.0004	0.000066	0.000120	0.00015	B (0.013)	D (0.98)	D (0)	D (1)	28.0	1.6	No	4/7
CEP68	c.1882A>G	p.K628E	0	0	0	0	D (0.041)	D (0.80)	D (0.02)	D (0.453)	24.4	4.9	No	5/7
CES1	c.148A>G	p.I50V	0	0.316800	0.185300	0	NS	B (0.99)	B (1)	B (0)	22.7	32.1	No	5/7
CHST3	c.277C>G	p.R93G	0	0	0	0	D (0.061)	D (0.92)	D (0.02)	B (0.182)	23.7	8.1	No	5/7
CREB3L1	c.599A>T	p.D200V	0.0002	0.003861	0.003617	0.00413	B (0.010)	D (0.99)	D (0)	B (0.091)	23.7	14.1	No	4/7
CREBBP	c.5837del	p.P1946H*30	0	0	0	0	NS	D (1)	NS	NS	33.0	7.3	Yes	NS
CTNNAL1	c.2041A>G	p.K681E	0	0	0	0	D (0.025)	B (0.91)	D (0.01)	B (0.189)	23.8	21.8	No	3/7
DDRGK1	c.916C>T	p.R306W	0	0.000025	0.000028	0.00008	B (0.007)	B (0.99)	D (0.02)	B (0.001)	24.1	18.9	No	4/7
DDX24	c.1805G>A	p.R602H	0.0002	0.001458	0.001280	0.00138	B (0.011)	D (0.99)	D (0)	D (0.997)	34.0	44.4	No	3/7
DDX31	c.1061G>A	p.R354H	0.001	0.002503	0.002378	0.00208	D (0.036)	D (0.99)	D (0)	D (1)	33.0	1.5	No	5/7
DST	c.7678G>A	p.D2560N	0	0.000017	0.000016	0	B (0.003)	D (0.54)	D (0.02)	B (0.236)	24.9	18.8	No	5/7
ERC1	c.659A>C	p.E220A	0	0.000042	0.000080	0.00023	B (0.017)	D (0.99)	B (0.14)	D (0.84)	25.0	5.6	No	5/7
FAM13A	c.1793C>T	p.P598L	0.0002	0.000058	0.000085	0.00008	D (0.069)	D (0.99)	D (0.01)	B (0.397)	23.2	1.9	No	7/7
FAT1	c.7700G>A	p.R2567H	0.004	0.008599	0.008242	0.00996	NS	D (0.73)	B (0.56)	D (0.888)	23.2	5.1	No	3/7
FGD6	c.2255G>T	p.R752L	0.005	0.007959	0.008498	0.00792	NS	B (0.99)	D (0.04)	B (0.103)	22.9	1.6	NA	3/7
GAK	c.2359G>T	p.D787Y	0.004	0.013170	0.008759	0.01054	NS	B (0.99)	D (0)	D (0.73)	21.9	9.5	No	5/7
ICAM1	c.1432C>T	p.R478W	0.003	0.004105	0.003812	0.00461	B (0.010)	B (0.99)	D (0.04)	D (0.797)	21.2	5.5	Yes	1/7
KAT6B	c.5749A>G	p.I1917V	0	0.000190	0.000199	0.00015	D (0.016)	D (0.99)	D (0)	B (0.045)	23.6	3.1	No	5/7
KIAA0100	c.3263C>T	p.S1088L	0	0.000034	0.000024	0.00008	B (0.021)	D (0.99)	B (0.14)	D (0.997)	25.3	17.4	NA	5/7
KIF7	c.2552G>A	p.R851H	0	0.000044	0.000028	0	D (0.084)	D (0.99)	D (0.04)	D (0.974)	31.0	2.2	Yes	6/7
KMT2D	c.2243insC	p.E748D*10	0	0	0	0	NS	D (1)	NS	NS	23.6	6.0	Yes	NS
LGI4	c.68C>G	p.P23R	0	0	0	0	D (0.025)	D (0.99)	B (0.27)	D (0.659)	23.5	6.8	No	4/7
LRP5L	c.118G>C	p.A40P	0.001	0.000948	0.001185	0.00162	D (0.035)	D (0.99)	D (0)	D (0.999)	22.9	1.7	NA	5/7
LUM	c.596T>C	p.L199P	0.001	0.002397	0.002246	0.00331	D (0.153)	D (0.99)	D (0)	D (0.971)	28.9	33.9	Yes	7/7
MKL1	c.1322T>G	p.V441G	0.004	0.003795	0.001992	0	NS	D (0.99)	D (0.01)	B (0.399)	23.8	7.0	Yes	5/7
MOXD1	c.1267G>T	p.E423*	0	0	0.000011	0	NS	D (1)	NS	NS	44.0	1.1	No	5/7
MROH7	c.3091C>G	p.R1031G	0	0.000251	0.000202	0.00024	B (0.011)	B (0.99)	B (0.15)	B (0.43)	22.0	5.8	NA	3/7
MYBBP1A	c.358G>C	p.D120H	0.002	0.007821	0.007401	0.00761	NS	B (0.99)	D (0.01)	D (0.682)	22.4	9.7	No	3/7
MYO1C	c.2574A>C	p.K858N	0.004	0.007805	0.006439	0.00715	NS	D (0.91)	D (0.04)	B (0.024)	21.8	46.8	No	4/7
NAPEPLD	c.718G>A	p.G240R	0	0.000008	0.000014	0	D (0.039)	D (0.99)	B (0.06)	B (0.04)	22.8	2.4	No	6/7
NOC3L	c.971del	p.L324R*6	0	0.000824	0.000962	0	NS	D (1)	NS	NS	35.0	2.5	No	NS
NRDE2	c.89G>A	p.S30N	0.0002	0.001895	0.001960	0.00338	D (0.032)	D (0.99)	D (0)	D (0.979)	25.1	1.8	No	6/7
NSA2	c.767C>T	p.A256V	0	0.000050	0.000046	0	B (0.008)	D (0.99)	B (0.05)	D (0.996)	27.1	27.0	No	7/7
NUP133	c.2561G>A	p.G854D	0.002	0.004386	0.004562	0.004466	B (0.022)	D (0.99)	D (0.03)	D (0.966)	27.9	9.1	Yes	6/7
NUP210	c.5504G>T	p.R1835L	0	0	0	0	D (0.052)	D (0.97)	D (0.04)	B (0.103)	22.3	1.3	No	6/7
ODF2L	c.1689dup	p.K564*	0.001	0.000479	0.000474	0	NS	D (0.99)	NS	NS	35.0	1.9	NA	4/7
OLFML1	c.158C>T	p.T53M	0.001	0.003292	0.003179	0.00354	D (0.074)	D (0.91)	B (0.13)	D (0.893)	23.8	3.0	NA	4/7
PCNX	c.6067A>G	p.R2023G	0	0.000091	0.000078	0.00008	B (0.020)	D (0.99)	B (0.06)	B (0.021)	22.7	5.5	NA	6/7

Continued on next page

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table L.2 continued

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
PDE4C	c.430G>A	p.D144N	0.001	0.000389	0.000432	0.00038	D (0.051)	D (0.99)	D (0.05)	B (0.265)	24.4	1.1	No	2/7
PFKP	c.1577A>G	p.E526G	0.003	0.005307	0.004841	0.00577	D (0.032)	D (0.81)	B (0.06)	B (0.02)	23.3	44.8	NA	6/7
PIP4K2B	c.553G>A	p.G185S	0.001	0.002867	0.002966	0.00292	D (0.065)	D (0.99)	D (0.01)	D (0.956)	29.5	16.0	No	5/7
PIP5K1C	c.1648A>G	p.R550G	0	0.000125	0.000110	0.00031	D (0.069)	D (0.98)	D (0)	B (0.231)	23.3	13.8	No	4/7
PPM1F	c.136G>T	p.G46W	0.0002	0.000025	0.000012	0	D (0.076)	B (0.99)	D (0)	D (0.963)	22.3	6.6	No	5/7
PPP1R21	c.1253C>G	p.T418R	0	0.000017	0.000008	0	B (0.004)	D (0.55)	B (0.2)	B (0.023)	22.3	4.4	No	5/7
PPP6R1	c.1988G>A	p.R663Q	0.003	0.014040	0.005753	0.00401	NS	D (0.70)	D (0.05)	D (0.953)	26.9	19.5	NA	3/7
PRKDC	c.8689C>T	p.L2897F	0	0.000047	0.000054	0	B (0.008)	D (0.86)	D (0.01)	D (0.807)	23.2	7.1	No	6/7
PROS1	c.939A>T	p.L313F	0	0	0	0	D (0.049)	D (0.99)	D (0)	D (0.996)	24.0	16.7	Yes	5/7
PTPN13	c.2258C>T	p.T753I	0	0	0	0	D (0.045)	D (0.99)	D (0)	D (0.988)	26.3	1.4	No	6/7
PTPN2	c.244A>C	p.S82R	0	0	0.000008	0	B (0.014)	B (0.52)	B (0.11)	B (0.072)	22.9	2.6	No	3/7
RAD23B	c.464C>T	p.T155I	0	0.000008	0.000008	0.00008	B (0.009)	D (0.93)	D (0.05)	B (0.034)	23.8	31.7	Yes	3/7
RC3H2	c.2030C>T	p.P677L	0	0.0000911	0.000093	0	B (0.018)	D (0.99)	D (0.02)	D (0.982)	28.2	4.8	No	5/7
REPS1	c.1967C>T	p.P656L	0.003	0.0006591	0.000721	0.00169	NS	D (0.99)	B (0.17)	B (0.023)	23.9	5.3	NA	3/7
RIC3	c.362G>C	p.G121A	0.003	0.007348	0.007448	0.00970	NS	D (0.99)	D (0.04)	D (0.999)	26.1	2.9	NA	6/7
SASH1	c.2459G>A	p.R820Q	0.002	0.001606	0.001584	0.00131	B (0.017)	D (0.99)	D (0)	D (0.994)	32.0	6.7	Yes	4/7
SEC24C	c.2483C>T	p.A828V	0	0.000692	0.000598	0.00062	D (0.089)	D (0.99)	D (0)	D (0.683)	28.0	15.3	No	7/7
SLC20A2	c.295G>A	p.A99T	0	0.000059	0.000054	0.00008	D (0.238)	D (0.99)	D (0)	D (0.844)	24.6	19.5	No	7/7
SLC25A25	c.1456A>G	p.I486V	0	0.000018	0.000019	0	D (0.055)	D (0.99)	B (0.05)	B (0.365)	23.1	2.9	No	7/7
SLC9A8	c.26A>C	p.E9A	0.004	0.008874	0.008854	0.01136	D (0.125)	B (0.92)	B (0.38)	B (0)	23.8	3.9	No	3/7
SMARCD1	c.1444G>A	p.G482R	0	0.000008	0.000004	0	D (0.050)	D (0.99)	B (0.09)	B (0.031)	23.0	3.6	Yes	5/7
SNRPA	c.208C>T	p.R70C	0	0	0.000004	0	D (0.078)	D (0.99)	D (0)	B (0.001)	23.1	19.2	NA	5/7
SORBS2	c.322T>C	p.F108L	0.004	0.008899	0.009501	0.00754	NS	D (0.99)	B (0.15)	D (0.995)	24.7	67.1	Yes	7/7
SPATA13	c.594C>A	p.H198Q	0	0	0	0	D (0.187)	P (0.95)	D (0.01)	D (0.996)	23.8	3.3	No	2/7
STARD3	c.1055G>A	p.R352H	0.0002	0.001709	0.001536	0.00138	D (0.112)	D (0.99)	B (0.34)	D (0.626)	24.0	10.5	No	6/7
TDRD7	c.1760A>G	p.D587G	0	0	0	0	B (0.014)	D (0.99)	D (0.03)	D (0.734)	26.1	3.3	No	6/7
TGS1	c.344C>T	p.S115F	0.003	0.025590	0.009407	0.00808	NS	B (0.97)	D (0)	D (0.936)	24.0	3.3	No	4/7
TNXB	c.1307C>G	p.P436R	0	0	0	0	B (0.018)	P (0.80)	D (0)	D (0.997)	25.3	10.6	Yes	1/7
TRAP1	c.776delinsAC	p.S259L	0	0	0	0	NS	D (0.97)	B (0.28)	B (0.015)	20.9	22.1	No	3/7
TRPM4	c.2674C>T	p.R892C	0	0.000511	0.000577	0.00077	D (0.283)	D (0.99)	D (0)	D (0.993)	26.9	12.1	Yes	5/7
UAP1	c.989A>G	p.N330S	0.001	0.002489	0.002554	0.00238	D (0.032)	D (0.99)	D (0.01)	D (0.994)	26.0	14.3	NA	7/7
VPS13D	c.7294C>T	p.R2432C	0.002	0.002117	0.002288	0.00315	B (0.003)	D (0.99)	D (0.02)	B (0.062)	23.2	9.7	No	3/7
ZNF3	c.679C>T	p.P227S	0	0	0	0	B (0.017)	D (0.99)	D (0.01)	D (0.725)	26.1	5.4	NA	4/7
ZNF35	c.752A>T	p.E251V	0	0	0	0	B (0.006)	B (0.99)	D (0.02)	D (0.475)	24.5	2.5	No	4/7
ZNF527	c.1396G>A	p.G466R	0	0.000008	0.000021	0	B (0.007)	D (0.98)	D (0)	D (0.884)	25.2	1.1	NA	3/7
ZNF580	c.75G>C	p.K25N	0	0	0	0	D (0.041)	B (0.95)	D (0.04)	B (0.234)	23.9	11.2	NA	2/7
ZNF638	c.4856T>C	p.V1619A	0	0	0	0	B (0.018)	D (0.92)	B (0.19)	D (0.675)	24.2	12.0	NA	5/7
ZZEF1	c.764A>G	p.Y255C	0.002	0.007044	0.005582	0.00469	B (0.022)	B (0.87)	B (0.15)	D (0.806)	23.6	7.4	No	5/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Appendix M Filtered variants in Family 3

Table M.1: Filtered cardiomyopathy panel variants in Family 3

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
TTN	c.16529A>G	p.Y5510C	0.007	0.002240	0.002213	0.00025	NS	D (0.99)	NS	D (0.715)	21.0	67.7	Yes	3/7
TTN	c.76739C>T	p.T25580M	0.005	0.001661	0.001618	0.00917	NS	D (0.99)	NS	B (0.191)	22.2	67.7	Yes	5/7
SYNE2	c.12001_12002inv	p.W4001Q	0	0	0	0	NS	B (0.99)	NS	B (0)	23.7	9.1	No	0/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table M.2: Other filtered variants in Family 3

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
ADCK1	c.1456G>A	p.E486K	0	0.000025	0.000014	0.00008	D (0.044)	D (0.99)	B (0.22)	B (0.163)	22.5	5.6	No	7/7
ARFGAP3	c.1210C>T	p.R404C	0.0002	0.000140	0.000135	0.00031	D (0.095)	D (0.99)	D (0.01)	D (0.912)	32.0	12.1	NA	5/7
ARRDC2	c.40G>C	p.D14H	0.001	0.000281	0.000256	0.00131	D (0.091)	D (0.99)	D (0)	D (0.9)	24.9	12.5	No	6/7
C22orf23	c.323G>A	p.R108Q	0	0.000008	0.000012	0	D (0.062)	D (0.99)	D (0)	D (0.999)	31.0	1.9	NA	6/7
DIAPH1	c.2525A>C	p.Q842P	0.005	0.003277	0.002593	0	NS	D (0.99)	B (0.36)	D (0.966)	24.4	17.6	No	3/7
DNAJC28	c.472A>G	p.R158G	0	0	0	0	D (0.032)	D (0.99)	D (0)	D (0.997)	25.0	1.4	No	4/7
FAM104B	c.331C>T	p.R111*	0	0.000461	0.000034	0	NS	B (0.98)	NS	NS	34.0	3.8	NA	NS
FN3KRP	c.259G>T	p.V87L	0	0	0	0	B (0.009)	D (0.99)	D (0.02)	B (0.443)	23.0	7.5	NA	6/7
FN3KRP	c.468+1G>A	-	0	0.000025	0.000008	0	NS	D (1)	NS	NS	34.0	7.5	NA	NS
GADD45GIP1	c.281C>G	p.P94R	0.004	0.000972	0.000951	0.00300	NS	D (0.99)	D (0)	D (0.993)	23.1	47.5	No	4/7
MANBA	c.2356A>G	p.N786D	0	0	0.000004	0	D (0.048)	D (0.99)	D (0.02)	B (0.26)	23.2	5.0	Yes	6/7

Continued on next page

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Table M.2 continued

VARIANT IDENTIFIERS			POPULATION FREQUENCIES				PATHOGENICITY PREDICTION					OTHER DATABASES		
Gene	**Variant**	**Protein**	**1000G**	**ExAC**	**gnomAD**	**EVS**	**M-CAP**	**MT**	**SIFT**	**PP-2**	**CADD**	**GTex**	**MGI***	**Cons.**
TIPARP	c.569A>G	p.Q190R	0	0	0	0	B (0.011)	D (0.99)	D (0.03)	D (0.549)	24.3	6.7	Yes	6/7
TMEM184B	c.1048G>A	p.V350M	0	0.000017	0.000025	0	B (0.016)	D (0.99)	B (0.37)	B (0.324)	22.7	21.5	Yes	6/7
TRPV1	c.860C>T	p.T287M	0.001	0.000504	0.000223	0.00086	D (0.035)	D (0.99)	D (0.03)	D (0.943)	24.4	4.0	Yes	5/7
XRN1	c.5006A>G	p.E1669G	0	0.000008	0.000004	0	B (0.006)	D (0.50)	B (0.2)	B (0.007)	21.9	2.5	NA	4/7
ZNF3	c.679C>T	p.P227S	0	0	0	0	B (0.017)	D (0.99)	D (0.01)	D (0.725)	26.1	5.4	No	4/7
ZNF548	c.980G>A	p.G327E	0	0	0	0	B (0.018)	B (0.52)	D (0)	D (0.991)	24.7	2.6	NA	1/7
ZNF548	c.982A>G	p.K328E	0	0	0	0	B (0.004)	B (0.98)	D (0)	D (0.964)	24.7	2.6	NA	1/7

* Yes indicates mice with cardiovascular phenotypes have been recorded in the MGI database, while No indicates that no cardiovascular phenotype was found for these mice. NA indicates no phenotypes are recorded for the gene, or no mouse orthologue for the gene exists.

1000G, 1000 Genomes Project; B, benign; CADD, Combined Annotation Dependent Depletion; Cons.: conservation; D, deleterious; ExAC, Exome Aggregation Consortium; EVS, Exome Variant Server; gnomAD, Genome Aggregation Database; GTex, Genotype-Tissue Expression; M-CAP, Mendelian Clinically Applicable Pathogenicity; MGI, Mouse Genome Informatics; MT, MutationTaster; NS, not scored; PP-2, PolyPhen-2; SIFT, Sorting Intolerant From Tolerant

Appendix N Analysis of *KCNK10* c.1052A>G in Family 3

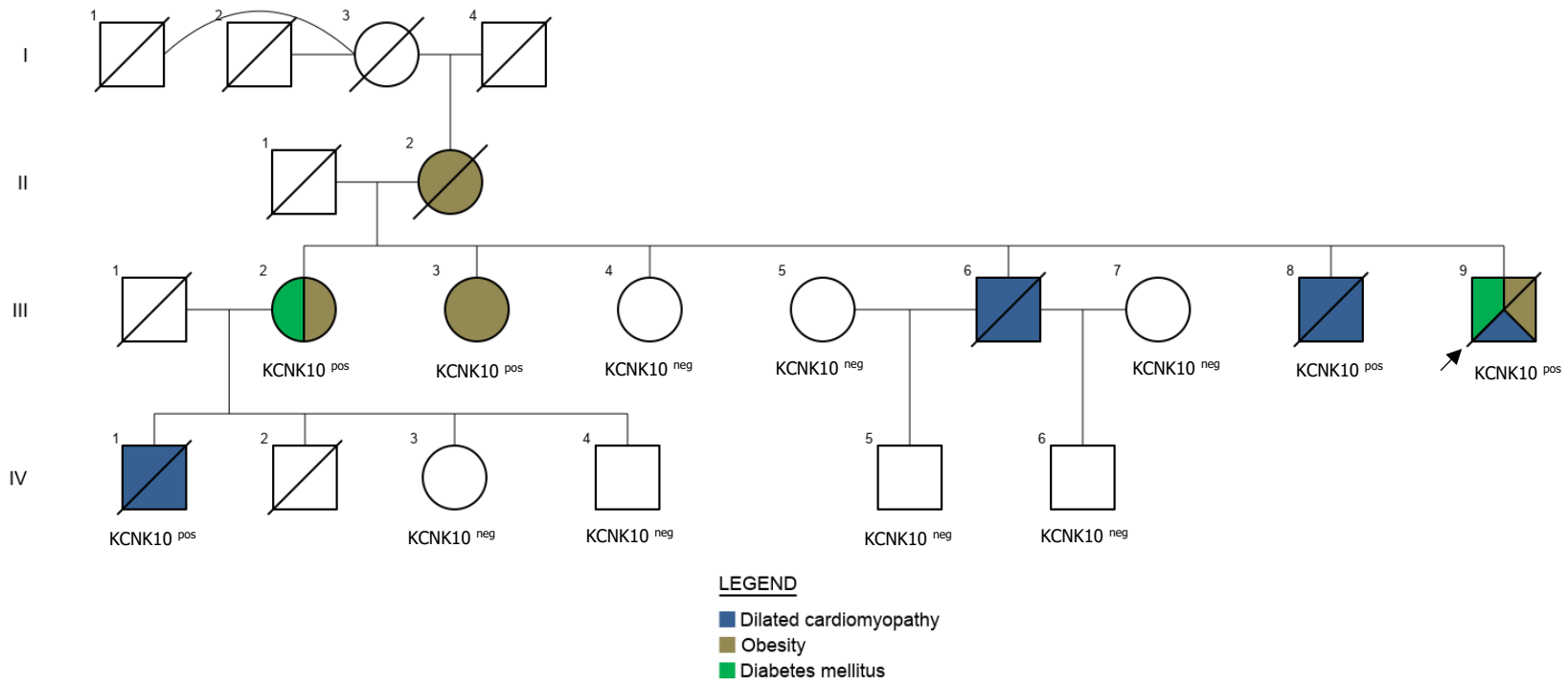
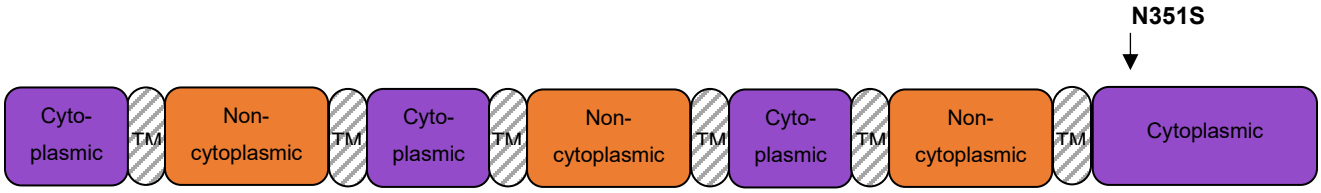


Figure N.1: Pedigree indicating segregation of *KCNK10* c.1052A>G in Family 3. Individuals marked 'KCNK10 ^{pos}' were heterozygous for the variant and those marked 'KCNK10 ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Blue shading indicates DCM, green shading indicates diabetes mellitus and brown shading indicates obesity. No shading indicates that the individual has no reported conditions. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

A



B

H. sapiens	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
M. mulatta	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
P. troglodites	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
F. catus	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
M. musculus	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
G. gallus	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R
D. rerio	E	I	K	A	H	A	A	E	W	K	A	N	V	R	A	E	L	R	E	T	R	R	R
X. tropicalis	E	I	K	A	H	A	A	E	W	K	A	N	V	T	A	E	F	R	E	T	R	R	R

Figure N.2: *KCNK10* c.1052A>G identified in Family 3. (A) Protein structure of KCNK10 with position of the p.N351S mutation marked with an arrow, (B) Multiple species conservation alignment of the p.N351 residue of KCNK10 (shaded in blue). *TM, transmembrane*

Appendix O Segregation of candidate variants in Family 4

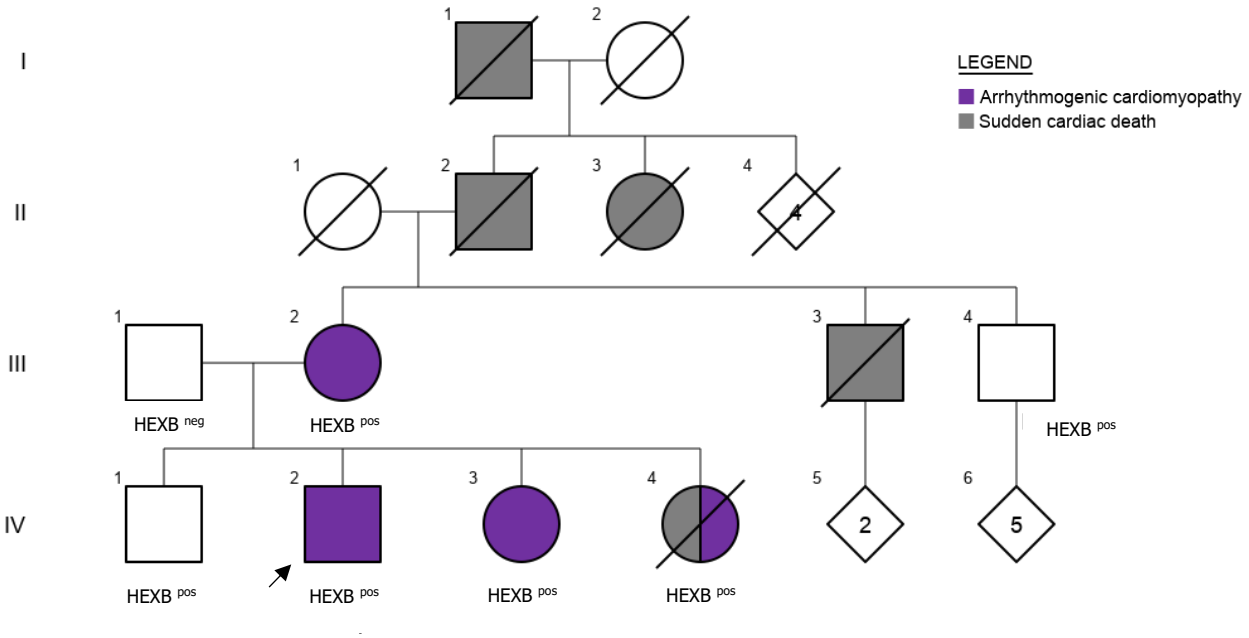


Figure O.1: Pedigree indicating segregation of *HEXB* c.1250C>T in Family 4. Individuals marked 'HEXB ^{pos}' were heterozygous for the variant and those marked 'HEXB ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

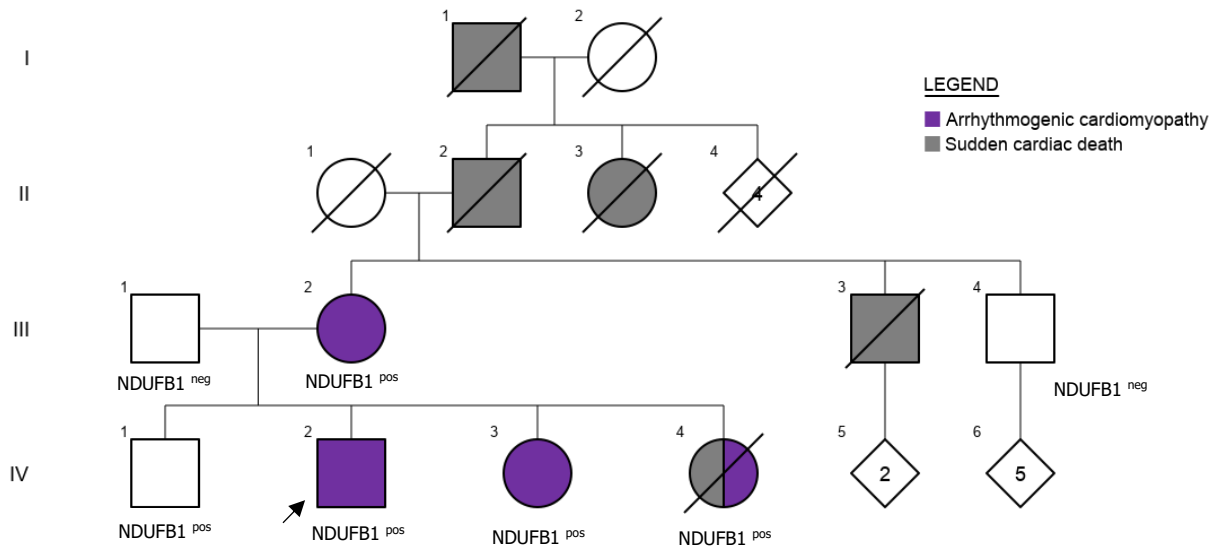


Figure O.2: Pedigree indicating segregation of *NDUFB1* c.257G>C in Family 4. Individuals marked 'NDUFB1 ^{pos}' were heterozygous for the variant and those marked 'NDUFB1 ^{neg}' (B) tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

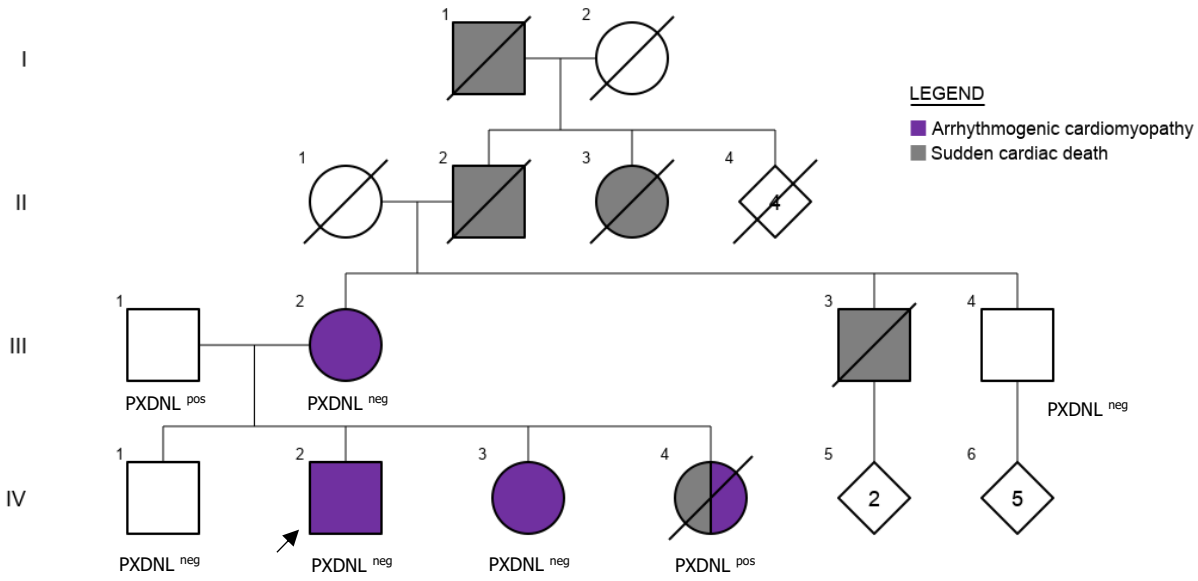


Figure O.3: Pedigree indicating segregation of *PXDNL* c.2218C>T in Family 4. Individuals marked 'PXDNL ^{pos}' were heterozygous for the variant and those marked 'PXDNL ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

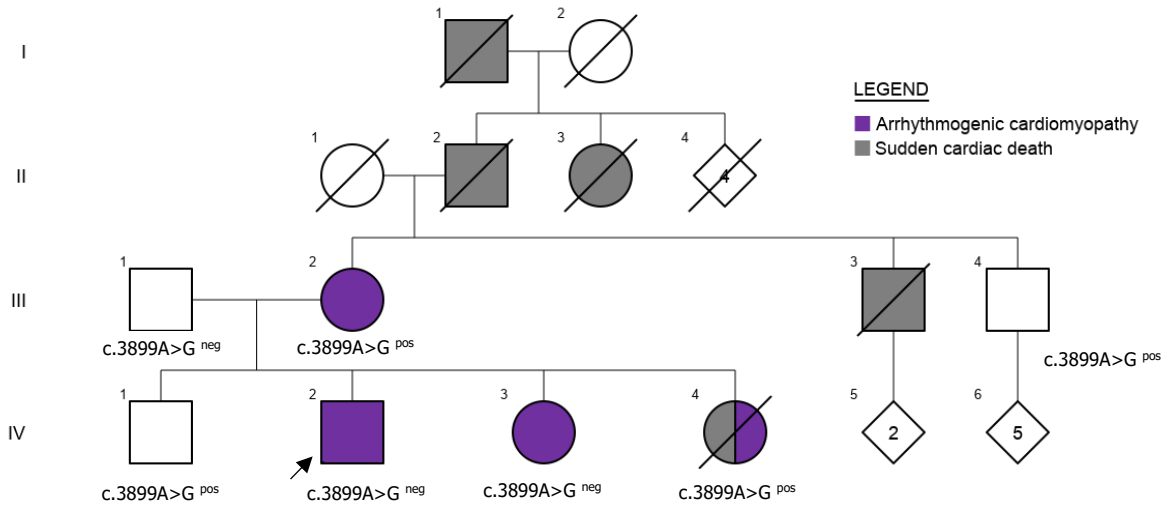


Figure O.4: pedigree indicating segregation of *TTN* c.3899A>G in Family 4. Individuals marked 'c.3899A>G ^{pos}' were heterozygous for the variant and those marked 'c.3899A>G ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

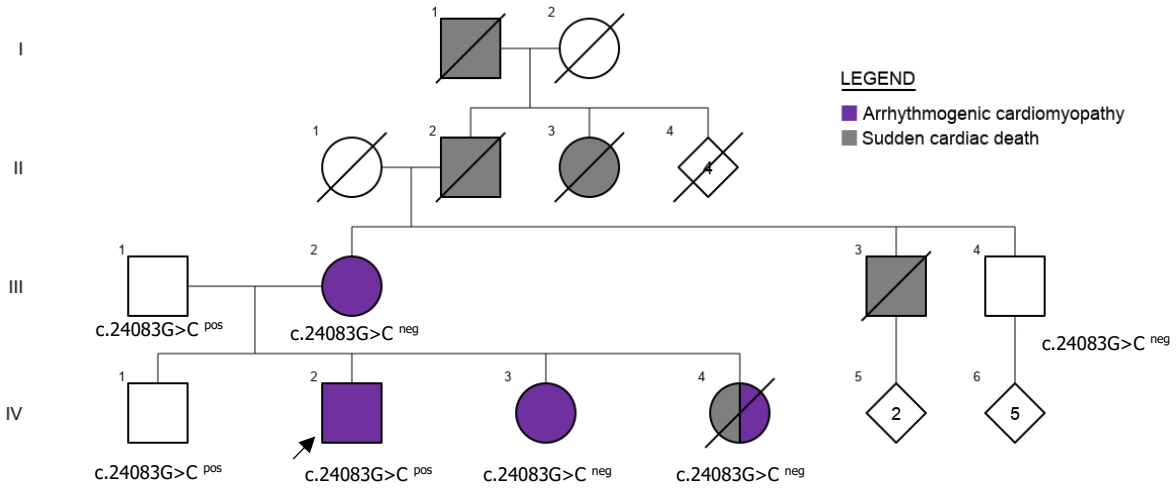
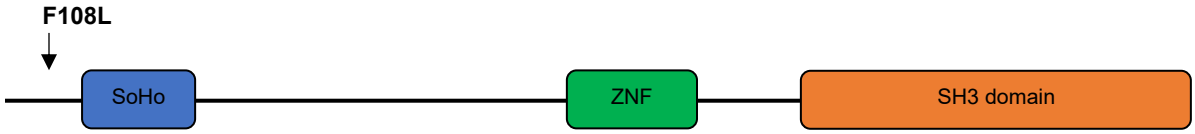


Figure O.5: Pedigree indicating segregation of *TTN* c.24083G>C in Family 4. Individuals marked 'c.24083G>C ^{pos}' were heterozygous for the variant and those marked 'c.24083G>C ^{neg}' tested negative for it. DNA was not available for unmarked individuals. Squares represent males and circles represent females. Symbols that are crossed out indicate deceased individuals. Shaded symbols indicate affected individuals. Purple shading indicates individuals diagnosed with ACM, while grey shading indicates sudden cardiac death. No shading indicates that the individual has no reported cardiomyopathy. Numbers in Roman numerals are the generation number while Arabic numerals denote individuals. The index case is indicated with an arrow.

Appendix P Analysis of *SORBS2* c.322T>C in Family 4

A



B

H. sapiens	A	V	S	P	M	S	Y	Y	Q	R	P	F	S	P	S	A	Y	S	L	P	A	S	L
M. mulatta	A	V	S	P	M	S	Y	Y	Q	R	P	F	S	P	S	A	Y	S	L	P	A	S	L
P. troglodites	A	V	S	P	M	S	Y	Y	Q	R	P	F	S	P	S	A	Y	S	L	P	A	S	L
F. catus	A	V	S	P	M	S	Y	Y	Q	R	P	F	S	P	S	A	Y	S	L	P	G	S	L
M. musculus	A	V	S	P	V	S	Y	Y	Q	R	P	F	S	P	S	A	Y	S	L	P	A	S	L
G. gallus	A	A	S	P	V	S	H	F	Q	R	P	F	S	P	SS	A	Y	S	P	P	A	S	L
D. rerio	P	A	S	P	R	S	H	I	Q	R	P	F	S	P	S	T	Y	P	P	P	P	S	L
X. tropicalis	P	A	S	P	M	S	H	Y	Q	R	T	F	S	P	SP	T	Y	P	S	S	R	S	L

Figure P.1: *SORBS2* c.322T>C identified in Family 4. (A) Protein structure of SORBS2 with position of the p.F108L mutation marked with an arrow, (B) Multiple species conservation alignment of the p.F108 residue of SORBS2 (shaded in blue). *SH3, Src Homology-3; SoHo, Sorbin-homology domain; ZNF, zinc finger C2H2-type domain*

Appendix Q Coding region variation in well-characterised DCM genes in Family 5

Table Q.1: Coding region variation in well-characterised DCM genes in the proband of Family 5

Gene*	Variant	Protein change	Carriers	gnomAD MAF	MT	CADD
Sarcomeric genes						
ACTN2	c.351T>C	p.I117=	I:1, I:2, II:3	0.997800	B (1.4^{-21})	10.1
ACTN2	c.378C>T	p.N126=	I:1, II:3	0.958800	B (1.5^{-17})	8.3
ACTN2	c.1296G>A	p.A432=	I:1, II:3	0.018660	B (5.2^{-8})	0.1
LDB3	c.302C>T	p.P101L	I:1, II:3	0.003162	D (0.99)†	24.8
LDB3	c.609G>A	p.S203=	I:2, II:3	0.004768	D (1)†	9.0
MYBPC3	c.833G>A	p.G278E	I:2, II:3	0.001332	D (0.97)†	21.3
MYH7	c.1128C>T	p.D376=	I:1, II:3	0.096820	B (1.6^{-17})	8.1
MYH7	c.1605A>G	p.E535=	I:1, I:2, II:3	0.010740	B (7.7^{-22})	10.1
TNNT2	c.207G>A	p.S69=	I;1, II:3	0.063130	B (0.02)	5.4
TNNT2	c.318C>T	p.I106=	I:1, II:3	0.719400	B (1.2^{-22})	10.4
TTN	c.1002C>T	p.T334=	I:2, II:3	0.000407	B (0.99)	0.3
TTN	c.2432A>T	p.T811I	I:1, II:3	0.168700	B (0.90)	22.4
TTN	c.2949C>T	p.I983=	I:1, II:3	0.004338	D (1)†	16.2
TTN	c.3601A>G	p.K1201E	I:1, II:3	0.688700	B (0.99)	22.4
TTN	c.3884C>T	p.S1295L	I:1, I:2, II:3	0.960100	B (0.99)	22.3
TTN	c.4177A>G	p.I1393V	I:2, II:3	0.012490	B (0.99)	19.1
TTN	c.9781G>A	p.V3261M	I:2, II:3	0.844500	B (0.99)	19.9
TTN	c.9879A>G	p.E3293=	I:2, II:3	0.991700	B (4.8^{-13})	17.9
TTN	c.10104T>G	p.V3368=	I:2, II:3	0.000403	D (1)†	14.0
TTN	c.10256G>A	p.S3419N	I:1, I:2, II:3	0.845900	B (0.99)	11.6
TTN	c.10726G>A	p.A3576T	I:1, I:2, II:3	0.999700	B (0.99)	20.1
TTN	c.11252G>A	p.G3751D	I:2, II:3	0.843400	B (0.72)	21.9
TTN	c.11422C>T	p.P3808S	I:1, I:2, II:3	0.182900	B (0.99)	21.9
TTN	c.14525G>A	p.R4842K	I:2, II:3	0.100500	B (0.99)	13.4
TTN	c.14610C>T	p.S4870=	I:1, I:2, II:3	0.994400	B (2.6^{-22})	14.5
TTN	c.21106G>A	p.D7036N	I:2, II:3	0.005536	D (0.99)†	22.8
TTN	c.23223G>A	p.Q7741=	I:1, I:2, II:3	0.994500	B (1.7^{-18})	15.9
TTN	c.25064C>A	p.A8355E	II:3	0.306300	B (0.99)	22.7
TTN	c.25398T>A	p.D8466E	I:2, II:3	0.011050	B (0.99)	19.9
TTN	c.26091A>T	p.L8697=	II:3	0.306800	B (0.18)	11.6
TTN	c.26289A>G	p.E8763=	I:1, II:3	0.318800	B (1.3^{-26})	18.1
TTN	c.26655C>T	p.S8885=	II:3	0.306000	B (3.4^{-16})	14.9
TTN	c.26818G>A	p.G8940S	I:2, II:3	0.000311	D (0.97)†	22.2
TTN	c.28662G>A	p.R9554=	I:2, II:3	0.047670	B (1.6^{-29})	15.5
TTN	c.29763T>C	p.I9921=	I:2, II:3	0.054550	B (1.6^{-27})	18.7
TTN	c.29799G>A	p.S9933=	I:2, II:3	0.049550	B (7.2^{-23})	15.7
TTN	c.31564A>G	p.I10522V	I:1, II:3	0.337300	B (0.99)	21.4
TTN	c.32593G>C	p.V10865L	I:2, II:3	0.000421	B (0.99)	18.1
TTN	c.33834G>A	p.E11278=	I:2, II:3	0.186000	B (1.2^{-9})	14.9
TTN	c.56101A>G	p.N18701D	I:1, II:3	0.354700	B (0.94)	23.3
TTN	c.58436G>A	p.R19479H	I:2, II:3	0.174300	B (4.8^{-10})	24.7
TTN	c.59585C>T	p.P19862L	I:2, II:3	0.173400	B (0.0001)	23.0
TTN	c.61029T>C	p.F20343=	I:2, II:3	0.000440	D (1)†	17.6
TTN	c.61245A>G	p.T20415=	I:1, II:3	0.345100	B (0.99)	8.8
TTN	c.62058T>C	p.Y20686=	I:1, II:3	0.354400	B (1.4^{-16})	16.1
TTN	c.64208C>T	p.T21403I	I:1, II:3	0.350400	B (0.99)	23.2
TTN	c.65682A>G	p.T21894=	I:1, II:3	0.354500	B (0.88)	11.2
TTN	c.67075G>A	p.V22359I	II:3	0.240200	B (0.01)	16.9
TTN	c.67246G>C	p.A22416P	I:1, I:2, II:3	0.998500	B (0.99)	12.5
TTN	c.74839C>T	p.R24947C	I:2, II:3	0.171500	B (3.0^{-6})	23.8
TTN	c.79862C>T	p.T26621M	II:3	0.233200	B (0.04)	22.5

Continued on next page

* List of paediatric DCM genes adapted from Lee TM et al. 2017. *Circ Res*, **121**(7):33-35.

† Although predicted deleterious by MT, this variant was benign by other pathogenicity prediction tools and/or ClinVar

B, benign; CADD, Combined Annotation Dependent Depletion; D, deleterious; gnomAD, Genome Aggregation Database; MT, MutationTaster

Table Q.1 continued

Gene*	Variant	Protein change	Carriers	gnomAD MAF	MT	CADD
TTN	c.83323A>G	p.I27775V	I:1, II:3	0.352000	B (0.99)	22.0
TTN	c.83673T>C	p.G27891=	I:1, II:3	0.351000	B (3.4^{-23})	16.3
TTN	c.88187T>C	p.I29396T	I:1, II:3	0.350500	B (0.99)	12.6
TTN	c.93243C>T	p.A31081=	I:2, II:3	0.180900	B (0.0003)	10.9
TTN	c.97613G>A	p.R32538H	I:2, II:3	0.171900	B (1.1^{-8})	25.5
TTN	c.98267C>T	p.T32756I	I:2, II:3	0.000453	D (0.98)†	23.6
TTN	c.102519C>T	p.G34173=	I:2, II:3	0.179700	B (2.0^{-14})	4.8
TTN	c.103781G>A	p.R34594H	I:2, II:3	0.172100	B (1.2^{-7})	25.2
TTN	c.104988C>T	p.V34996=	I:2, II:3	0.179600	B (8.2^{-12})	12.8
TTN	c.105384A>G	p.A35128=	I:1, II:3	0.352000	B (1.3^{-15})	11.2
Nuclear membrane genes						
LMNA	c.861T>C	p.A287=	I:1, I:2, II:3	0.109600	B (3.7^{-14})	9.4
LMNA	c.1338T>C	p.D446=	I:1, I:2, II:3	0.128500	B (2.5^{-17})	10.7
SYNE1	c.1794A>C	p.S598=	I:1, II:3	0.001240	D (0.99)†	0.6
SYNE1	c.2653T>G	p.L885V	I:2, II:3	0.055500	B (0.99)	14.7
SYNE1	c.3104T>C	p.V1035A	I:1, I:2, II:3	0.475400	B (0.91)	20.1
SYNE1	c.3306C>T	p.H1102=	I:2, II:3	0.057260	B (0.53)	0.1
SYNE1	c.5190T>A	p.D1730E	I:2, II:3	0.043340	B (0.99)	10.9
SYNE1	c.6470A>G	p.K2157R	I:1, II:3	0.007910	B (0.76)	21.3
SYNE1	c.8737G>T	p.A2913S	I:2, II:3	0.000054	B (0.61)	22.9
SYNE1	c.9495A>G	p.E3165=	I:1, II:3	0.504100	B (1.6^{-12})	8.5
SYNE1	c.10191C>A	p.G3397=	I:1, II:3	0.658200	B (3.9^{-20})	0.1
SYNE1	c.10866T>C	p.S3622=	I:1, II:3	0.584700	B (0.99)	0.3
SYNE1	c.12180G>T	p.E4060D	I:1, I:2, II:3	0.580500	B (0.99)	6.4
SYNE1	c.12362A>G	p.K4121R	I:1, I:2, II:3	0.757200	B (0.99)	20.4
SYNE1	c.13786T>A	p.S4596T	I:1, I:2, II:3	0.784200	B (0.99)	0.001
SYNE1	c.15043T>A	p.L5015M	I:1, I:2, II:3	0.787900	B (0.99)	21.0
SYNE1	c.21904T>G	p.F7302V	I:1, I:2, II:3	0.991900	B (0.99)	16.8
SYNE1	c.25038T>C	p.R8346=	I:1, II:3	0.611200	B (5.5^{-8})	1.1
SYNE2	c.399G>C	p.L133=	I:1, II:3	0.087370	B (7.6^{-25})	11.8
SYNE2	c.8404A>G	p.S2802G	I:1, II:3	0.909400	B (0.99)	0.8
SYNE2	c.8597C>T	p.T2866M	I:1, II:3	0.004167	B (0.99)	23.4
SYNE2	c.9023T>C	p.I3008T	I:1, II:3	0.003793	B (0.99)	12.4
SYNE2	c.9389A>G	p.N3130S	I:2, II:3	0.043650	B (0.99)	11.7
SYNE2	c.9757G>C	p.D3253H	II:3	0.803800	B (0.99)	21.0
SYNE2	c.9926A>G	p.H3309R	II:3	0.804400	B (0.99)	14.3
SYNE2	c.10567A>C	p.K3523Q	I:1, II:3	0.025200	B (0.99)	8.8
SYNE2	c.11385G>A	p.K3795=	I:1, II:3	0.006816	B (2.1^{-17})	18.2
SYNE2	c.11613A>G	p.V3871=	I:2, II:3	0.085970	B (0.29)	11.5
SYNE2	c.11944A>C	p.N3982H	I:2, II:3	0.066610	B (0.98)	22.6
SYNE2	c.14734C>G	p.P4912A	I:2, II:3	0.033450	B (0.23)	22.9
SYNE2	c.15543C>T	p.I5181=	I:2, II:3	0.371300	B (0.0001)	13.3
SYNE2	c.15556C>A	p.L5186M	I:1, I:2, II:3	0.504900	B (0.99)	12.1
SYNE2	c.16722A>G	p.Q5574=	I:2, II:3	0.036290	B (2.8^{-13})	15.5
SYNE2	c.17202C>A	p.L5734=	I:2, II:3	0.331300	B (0.006)	15.9
Other genes						
DES	c.669T>C	p.I223=	I:2, II:3	0.009637	B (2.9^{-22})	19.9
DES	c.1026C>T	p.N342=	I:1, II:3	0.006725	B (4.5^{-18})	12.3
RBM20	c.90G>A	p.R30=	I:1, II:3	0.155100	B (0.99)	14.7
RBM20	c.2303G>C	p.W768S	I:1, I:2, II:3	0.997100	B (0.99)	0.02
VCL	c.945C>A	p.G315=	I:1, I:2, II:3	0.008791	B (7.1^{-21})	12.5
VCL	c.2814C>G	p.G938=	I:1, I:2, II:3	0.676000	B (3.1^{-20})	11.4

* List of paediatric DCM genes adapted from Lee TM et al. 2017. *Circ Res*, **121**(7):33-35.

† Although predicted deleterious by MT, this variant was benign by other pathogenicity prediction tools and/or ClinVar

B, benign; CADD, Combined Annotation Dependent Depletion; D, deleterious; gnomAD, Genome Aggregation Database; MT, MutationTaster

Appendix R Power analyses of zebrafish experiments

Table R.1: Power analysis of ANOVA tests of CRISPR zebrafish ventricle measurements

CRISPR knockout group	Mean measurement	Standard deviation	Calculated power*
Comparison of ventricle length			
myh7	41.128	9.400	0.753
pkp2	28.140		
polg	46.848		
Uninjected	33.976		
Comparison of ventricle thickness			
myh7	5.254	1.262	0.784
pkp2	7.374		
polg	4.574		
Uninjected	5.910		
Comparison of ventricle area			
myh7	731.324	330.782	**0.850**
pkp2	273.560		
polg	1054.222		
Uninjected	462.428		

* Sufficiently powered tests (> 0.80) are indicated in bold print
ANOVA, analysis of variance; CRISPR, clustered regularly-interspaced short palindromic repeats

Table R.2: Power analysis of pairwise comparisons of CRISPR zebrafish ventricle measurements

CRISPR knockout group	Mean measurement	Standard deviation	Calculated power*
Comparison of ventricle length			
Uninjected (control)	33.976	7.625	NA
myh7	41.128	6.667	0.243
pkp2	28.140	6.164	0.187
polg	46.848	5.328	0.687
Comparison of ventricle thickness			
Uninjected (control)	5.910	0.409	NA
myh7	5.254	0.810	0.252
pkp2	7.374	0.891	0.751
polg	4.574	0.768	0.774
Comparison of ventricle area			
Uninjected (control)	462.428	136.014	NA
myh7	731.324	162.078	0.615
pkp2	273.560	108.032	0.489
polg	1054.222	177.820	**0.999**

* Sufficiently powered tests (> 0.80) are indicated in bold print
CRISPR, clustered regularly-interspaced short palindromic repeats; NA, not applicable

Table R.3: Power analysis of pairwise comparisons of zebrafish heart rates

mRNA injection group	Mean measurement	Standard deviation	Calculated power*
Mutant *POLG* (experimental)	142.332	34.132	NA
Uninjected (control)	183.000	33.028	**> 0.999**
Water (control)	182.647	32.213	**> 0.999**
Wild type *POLG* (control)	180.400	26.582	**> 0.999**

* Sufficiently powered tests (> 0.80) are indicated in bold print
mRNA, messenger ribonucleic acid; NA, not applicable

Table R.4: Power analysis of pairwise comparisons of *POLG* zebrafish ventricle measurements

mRNA injection group	Mean measurement	Standard deviation	Calculated power*
Comparison of ventricle length			
Wild type *POLG*	50.456	5.997	**0.987**
Mutant *POLG*	42.538	4.870	
Comparison of ventricle width			
Wild type *POLG*	28.598	3.418	0.400
Mutant *POLG*	26.185	4.681	
Comparison of ventricle area			
Wild type *POLG*	1236.714	232.319	**0.896**
Mutant *POLG*	957.796	268.057	
Comparison of blood flow			
Wild type *POLG*	9.603	4.227	**0.889**
Mutant *POLG*	4.485	3.600	

* Sufficiently powered tests (> 0.80) are indicated in bold print
mRNA, messenger ribonucleic acid

Appendix S Additional correlation analyses of CRISPR knockout zebrafish

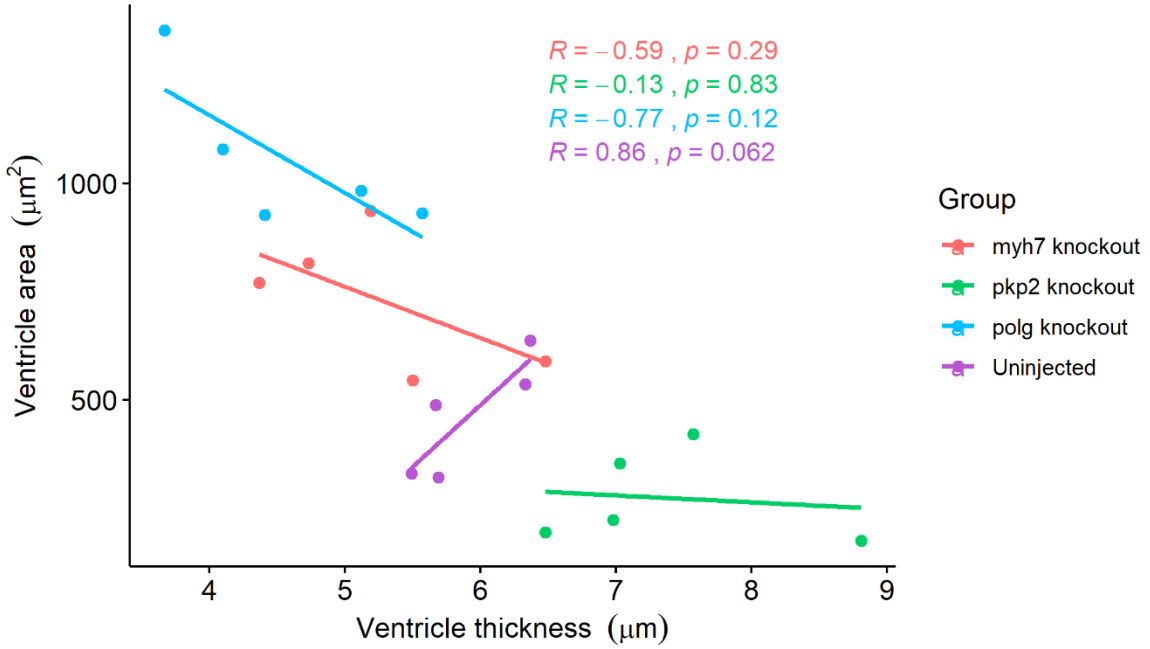


Figure S.1: Correlation of ventricular thickness and area amongst the experimental groups. The thickness is represented on the x-axis, while area is on the y-axis. The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

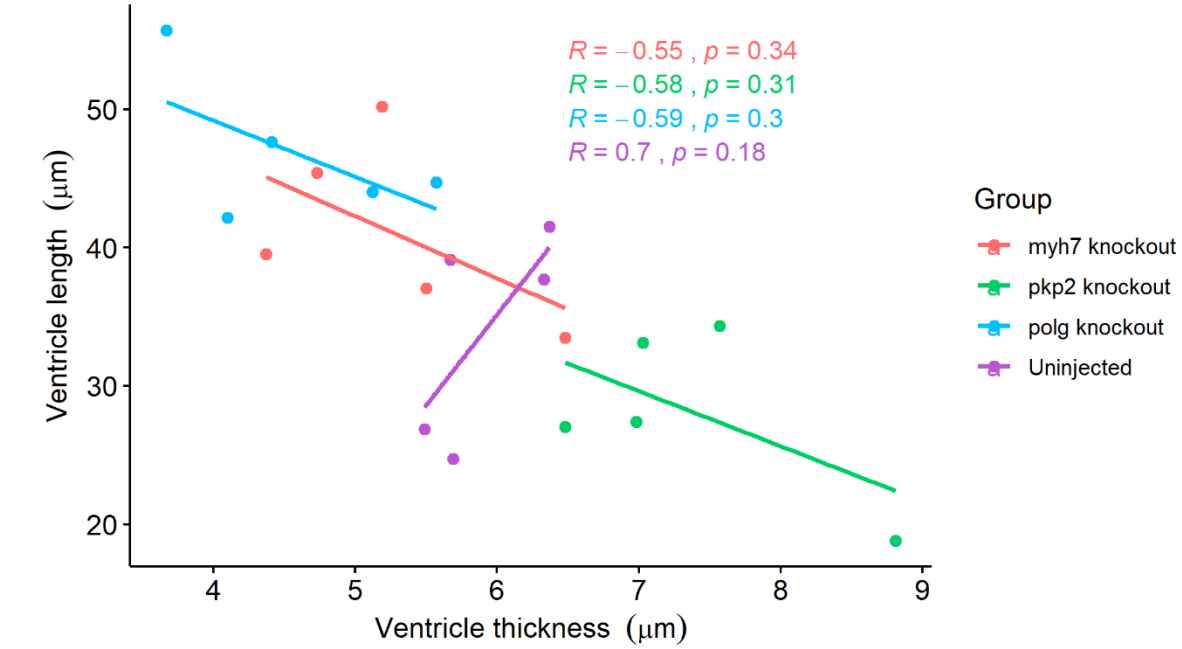


Figure S.2: Correlation of ventricular thickness and length amongst the experimental groups. The thickness is represented on the x-axis, while length is on the y-axis. The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

Appendix T Genotyping of CRISPR mutant zebrafish larvae

Table T.1: Mutation detection rates in CRISPR gene edited zebrafish larvae

Target gene	Fish no.	Site of greatest indel rate	Proportion of frameshift reads*	Proportion of non-frameshift reads*	Mean indel rate*
myh7	1	584-594	16.77%	39.77%	48.43%
	2	583-593	11.89%	39.88%	
	3	584-594	4.97%	50.45%	
	4	584-594	15.25%	38.17%	
	5	584-594	6.81%	18.16%	
pkp2	1	302-312	34.52%	61.30%	78.59%
	2	302-312	18.72%	42.79%	
	3	299-309	7.91%	72.41%	
	4	299-309	7.57%	66.27%	
	5	299-309	0.82%	80.65%	
polg	1	430-440	75.97%	0.65%	73.19%
	2	430-440	64.69%	5.67%	
	3	430-440	72.95%	2.73%	
	4	430-440	71.97%	1.27%	
	5	430-440	62.85%	7.23%	

***** Calculated at site of greatest indel rate
Indel, insertion/deletion mutation; no., number

Table T.2: Genotyping results of *myh7* knockout zebrafish larvae

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
1	589	ATGAAG/-	-6	1217	6502
	588	GAT/-	-3	717	6590
	589	ATGA/-	-4 (frameshift)	590	6541
	589	ATGAAGAG/-	-8 (frameshift)	326	6500
	591	CCA/-	3	256	6604
	583	ATCCAGATGAAG/-	-12	247	6469
	587	AGATG/-	-5	183	6561
	591	GAA/-	-3	159	6580
	591	GAAGAG/-	-6	132	6539
	582	GATCCAGAT/-	-9	125	6559
2	588	GAT/-	-3	1101	5602
	589	ATGAAG/-	-6	667	5478
	592; 603	AA/-; -/A	-2 ; 1 (frameshift)	402	5399
	591	GAAGAGTAT/-	-9	269	5468
	591	-/CCA	3	197	5620
	590	-/TGCC	4 (frameshift)	159	5629
	590	TGAAGAGTA/-	-9	101	5432
	592	AA/-	-2 (frameshift)	97	5664
	582	GATCCAGAT/-	-9	71	5549
	591	GAAGAG/-	-6	70	5532
3	589	ATGAAG/-	-6	599	4389
	588	GAT/-	-3	556	4515
	591	-/CCA	3	349	4505
	591	GAA/-	-3	283	4428
	591	GAAGAGTAT/-	-9	212	4386

Continued on next page

* Entries marked 'frameshift' were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Table T.2 continued

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
3	591	G/-	-1 (frameshift)	150	4505
	589	ATGAAGAGTATGTGA/-	-15	126	4282
	592	-/TGA	3	97	4486
	583	ATCCAGATGA/-	-10 (frameshift)	77	4377
	591	GAAGAGTATGTGAAAGCGTC/-	-20 (frameshift)	69	4142
4	589	ATGAAG/-	-6	1144	4827
	590	TGAAGAG/-	-7 (frameshift)	436	4816
	588	GAT/-	-3	394	4958
	591	GA/-	-2 (frameshift)	139	4915
	591	-/CCA	3	127	4975
	591	GAAGAG/-	-6	113	4830
	589	ATGAAGAG/-	-8 (frameshift)	90	4794
	591	GAAGAGTAT/-	-9	75	4659
	591	GAA/-	-3 (frameshift)	75	4911
5	589	ATGAAG/-	-6	528	6785
	588	GAT/-	-3	491	7318
	591	GAAGAGTAT/-	-9	234	6279
	590	-/TTATG	5 (frameshift)	182	7396
	582	GATCCAGAT/-	-9	131	6899
	591	GAAGAGTATGTG/-	-12 (frameshift)	69	6116
	591	G/-	-1 (frameshift)	83	7366
	589	ATGAAGAG/-	-8 (frameshift)	71	6653
	576	GTGCCTGATCCAGAT/-	-15	68	6616
	587	AGATG/-	-5 (frameshift)	65	7269

* Entries marked ‘frameshift’ were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Table T.3: Genotyping results of *pkp2* knockout zebrafish larvae

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
1	307	CTGAACCGA/-	-9	1324	4181
	309	-/A	1 (frameshift)	1338	4261
	304	CGGCTGAAC/-	-9	1274	4252
	309	GAACC/-	-5 (frameshift)	105	4202
	300	GCACCGGCT/-	-9	41	4146
	308	T/-	-1 (frameshift)	21	4293
	349	A/-	-1 (frameshift)	21	4563
	294	CGACGCGCACCGGCTGAAC/-	-19 (frameshift)	18	4080
2	307	CTGAACCGA/-	-9	1227	6772
	287	CGCTCGCCGACGCGCACCGGCTGAAC/-	-26 (frameshift)	1205	6660
	300	GCACCGGCT/-	-9	1227	6807
	308	T/-	-1 (frameshift)	1150	6819
	304	CGGCTGAAC/-	-9	915	6785
	309	GAA/-	-3	538	6805
	309	G/-	-1	230	6805
	309	-/C	1 (frameshift)	69	6805
	308	-/G	1 (frameshift)	54	6819
3	304	CGGCTGAAC/-	-9	2464	9083
	307	CTGAACCGA/-	-9	2088	9063
	302	ACCGGCTG/-	-8 (frameshift)	719	9072
	309	-/ACA	3	536	9121
	308	T/-	-1	477	9164
	300	GCACCGGCT/-	-9	455	9111
	292	GCCGACGCGCACCGGCTGA/-	-19 (frameshift)	423	8995
	307	CTGAAC/-	-6	237	9089

Continued on next page

* Entries marked 'frameshift' were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Table T.3 continued

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
3	309	-/ACAGCAACAAGTCAA	15	200	9121
	301	CACCGGCTGAACCGACTACAG/-	-21	125	8987
4	304	CGGCTGAAC/-	-9	3421	8811
	307	CTGAACCGA/-	-9	1408	8824
	300	GCACCGGCT/-	-9	1027	8814
	302	ACCGGCTG/-	-8 (frameshift)	196	8799
	291	CGCCGACGCGCACCGGCTGAACCG/-	-24	182	8661
	306	GCTGAA/-	-6 (frameshift)	167	8856
	294	CGACGCGCACCGGCTG/-	-16 (frameshift)	116	8748
	309	GAACC/-	-5 (frameshift)	112	8859
	308	TG/-	-2 (frameshift)	109	8863
	309	G/-	-1 (frameshift)	85	8868
5	304	CGGCTGAAC/-	-9	2641	4805
	307	CTGAACCGA/-	-9	632	4805
	300	GCACCGGCT/-	-9	370	4791
	304; 349	CGGCTGAAC/-; A/-	-9; -1 (frameshift, latter)	16	3528
	297; 309	-/CGCGCACCGG; GA/-	10; -2 (frameshift)	21	4754

* Entries marked ‘frameshift’ were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Table T.4: Genotyping results of *polg* knockout zebrafish larvae

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
1	435	GCCAAAA/-	-7 (frameshift)	2652	5290
	436	C/-	-1 (frameshift)	1066	5410
	436	-/C	1 (frameshift)	207	5410
	436	C/-	-1 (frameshift)	21	1165
	417	ATGGGGCGGATGGAGAGAGCCAAA/-	-24	56	5093
	436	CC/-	-2 (frameshift)	41	5409
	437	CAAAAG/-	-6	35	5295
	431	GAGAGCCAAAA/-	-11 (frameshift)	34	5218
	434	AGCCA/-	-5 (frameshift)	34	5393
	437	C/-	-1 (frameshift)	34	5417
2	435	GCCAAAA/-	-7 (frameshift)	2563	5758
	436	CC/-	-2 (frameshift)	388	5850
	435	GCCAAA/-	-6	330	5778
	436	C/-	-1 (frameshift)	289	5858
	417	ATGGGGCGGATGGAGAGAGCCAA/-	-23 (frameshift)	147	5518
	434	AGCC/-	-4 (frameshift)	147	5847
	436	-/C	1 (frameshift)	147	5858
	437	C/-	-1 (frameshift)	85	5851
	436	CC/-	-2 (frameshift)	77	5850
	435	GCCAA/-	-5 (frameshift)	72	5779
3	435	GCCAAAA/-	-7 (frameshift)	3892	8020
	436	CC/-	-2 (frameshift)	1341	7968
	437	CAAAAG/-	-6	214	7654
	436	-/C	1 (frameshift)	214	7812
	437	C/-	-1 (frameshift)	137	7796

Continued on next page

* Entries marked 'frameshift' were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Table T.4 continued

Fish no.	Sequence position	Genotype (reference/alternate allele)	Indel size (bp)*	Indel count	Total read count
3	417	ATGGGGCGGATGGAGAGAGCCAAA/-	-24	45	7883
	437	C/-	-1 (frameshift)	36	7409
	436	CC/-	-2 (frameshift)	35	8123
	434	AGCC/-	-4 (frameshift)	34	7894
	436	-/C	1 (frameshift)	34	7895
	435	GCCAAAA/-	-7 (frameshift)	3892	8020
4	435	GCCAAAA/-	-7 (frameshift)	2794	5768
	436	C/-	-1 (frameshift)	894	5412
	436	-/C	1 (frameshift)	235	5909
	436	C/-	-1 (frameshift)	102	5904
	417	ATGGGGCGGATGGAGAGAGCCAAA/-	-24	74	6058
	436	CC/-	-2 (frameshift)	74	5872
	435	GCCAAA/-	-6	74	5825
	437	C/-	-1 (frameshift)	43	5778
	431	GAGAGCCAAAA/-	-11 (frameshift)	34	6032
	434	AGCCA/-	-5 (frameshift)	21	5762
	435	GCCAAAA/-	-7 (frameshift)	2794	5768
5	435	GCCAAAA/-	-7 (frameshift)	1892	5840
	436	C/-	-1 (frameshift)	838	5902
	436	-/C	1 (frameshift)	401	5722
	437	CAAAAG/-	-6	405	5662
	431	GAGAGCCAAAA/-	-11 (frameshift)	224	4928
	417	ATGGGGCGGATGGAGAGAGCCAA/-	-23 (frameshift)	122	5421
	437	C/-	-1 (frameshift)	87	5620
	431	GAGAGCCAAAA/-	-11 (frameshift)	55	5872
	434	AGCCA/-	-5 (frameshift)	23	5444

***** Entries marked 'frameshift' were frameshift mutations
Bp, base pairs; indel, insertion/deletion mutation

Appendix U Individual *POLG* overexpression experiment results

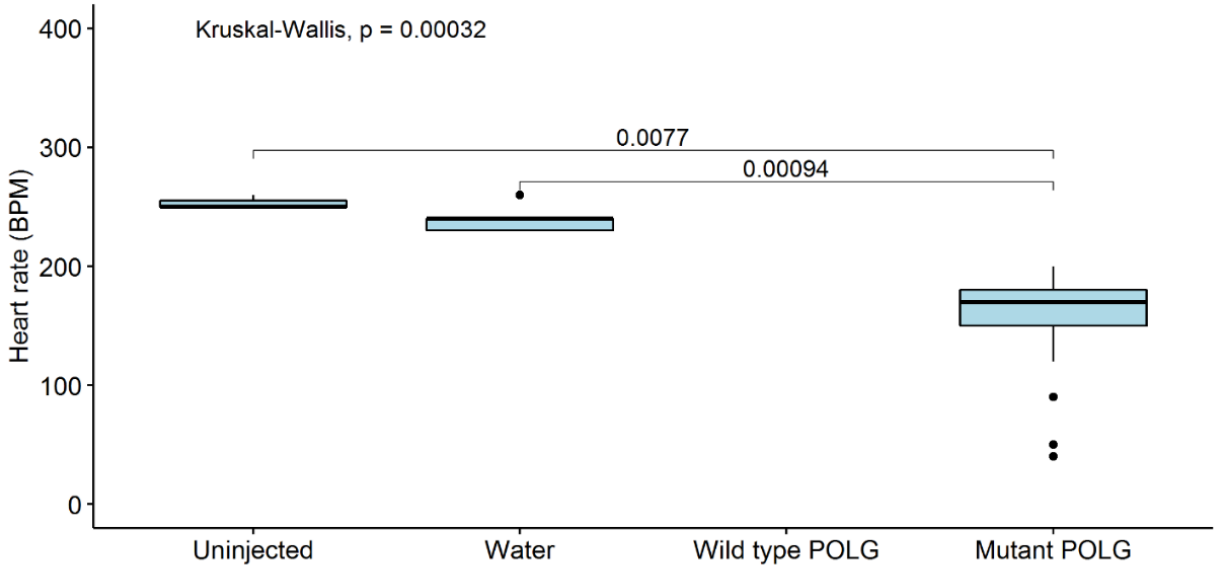


Figure U.1: Average heart rates in *POLG* overexpression experiment 1. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by box-and-whisker plots in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test). Note that no data for wild type *POLG* injection was available in this experiment.

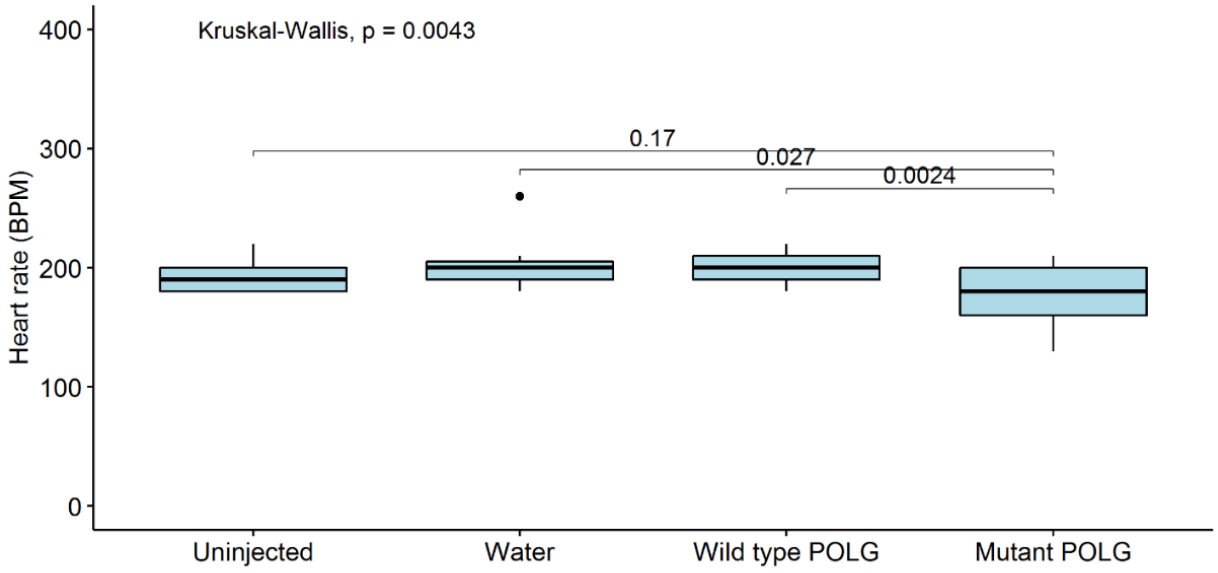


Figure U.2: Average heart rates in *POLG* overexpression experiment 2. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by box-and-whisker plots in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test).

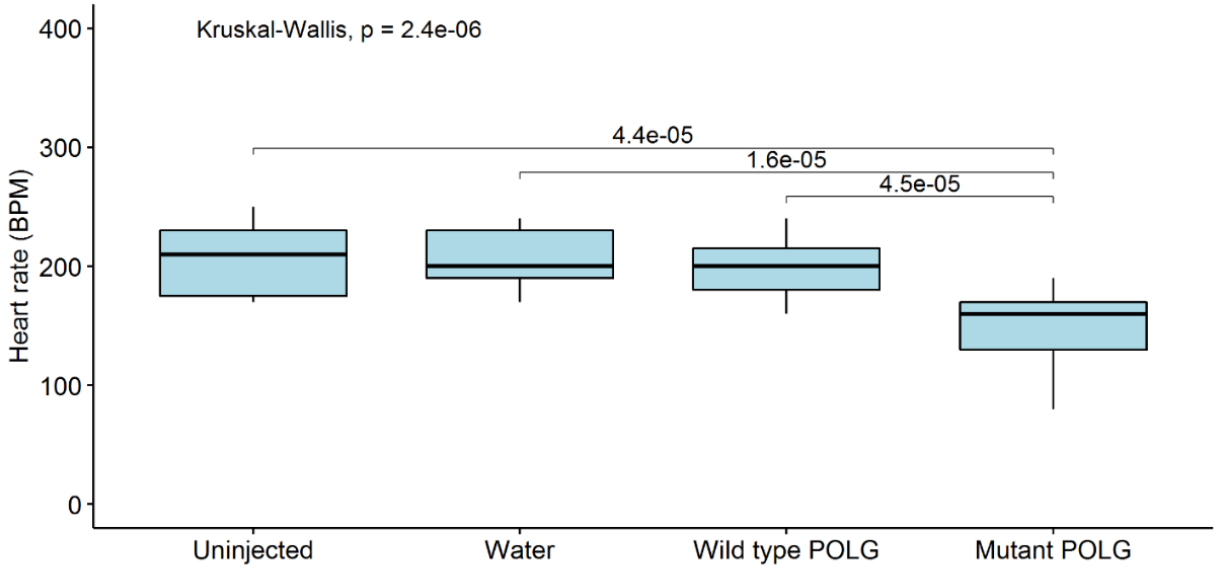


Figure U.3: Average heart rates in *POLG* overexpression experiment 3. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by box-and-whisker plots in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test).

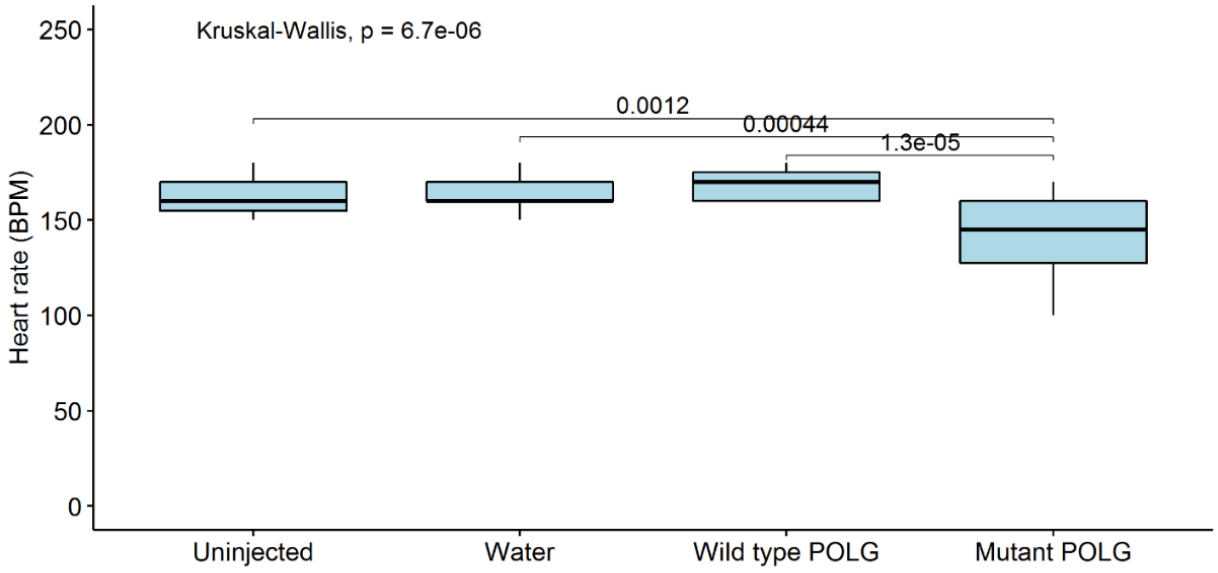


Figure U.4: Average heart rates in *POLG* overexpression experiment 4. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by box-and-whisker plots in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test).

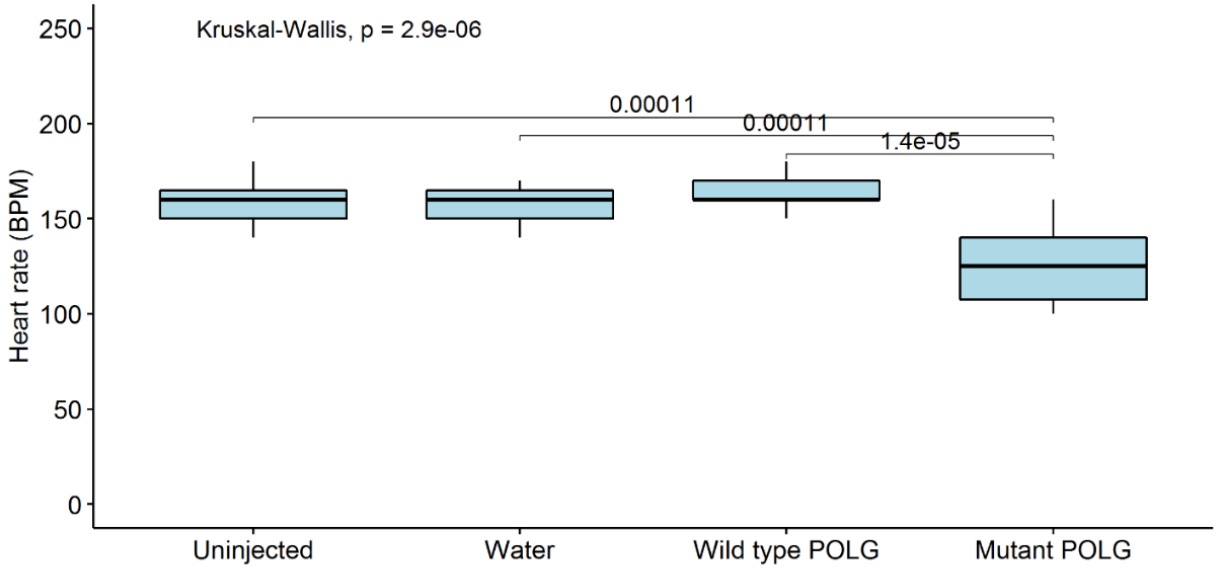


Figure U.5: Average heart rates in *POLG* overexpression experiment 5. Experimental groups are on the x-axis while heart rate is on the y-axis. The distributions of heart rate are represented by box-and-whisker plots in which the boxes delineate the upper and lower quartiles, and the whiskers indicate the minimum and maximum values. Lines within the boxes are the median values. Dots are used to show the distribution of variables in each group. Significance bars show the p-value when comparing each dose group to the uninjected control group (two-sided Mann-Whitney U test).

Appendix V Additional correlation analyses of mRNA injected zebrafish

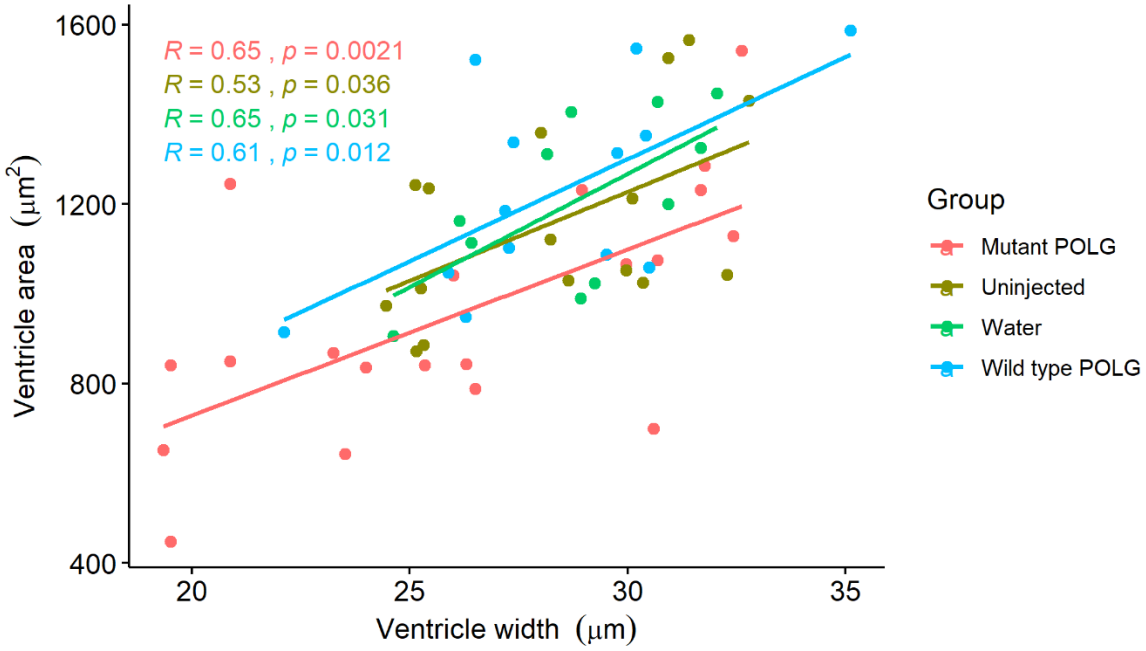


Figure V.1: Correlation of ventricular width and area amongst the mRNA experimental groups. The length is represented on the x-axis, while area is on the y-axis. The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

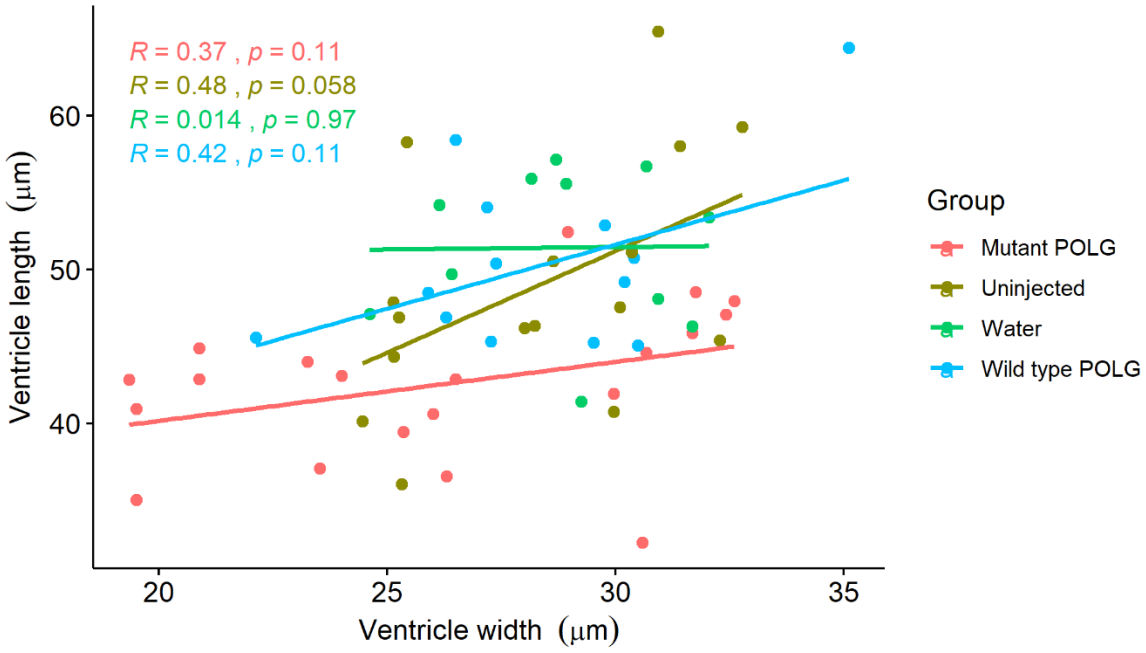


Figure V.2: Correlation of ventricular width and length amongst the mRNA experimental groups. The length is represented on the x-axis, while area is on the y-axis. The Pearson's correlation coefficient is represented for each group (R), as is the significance of the relationship.

Appendix W Analysis of zebrafish larval heart dimensions per mRNA injection group

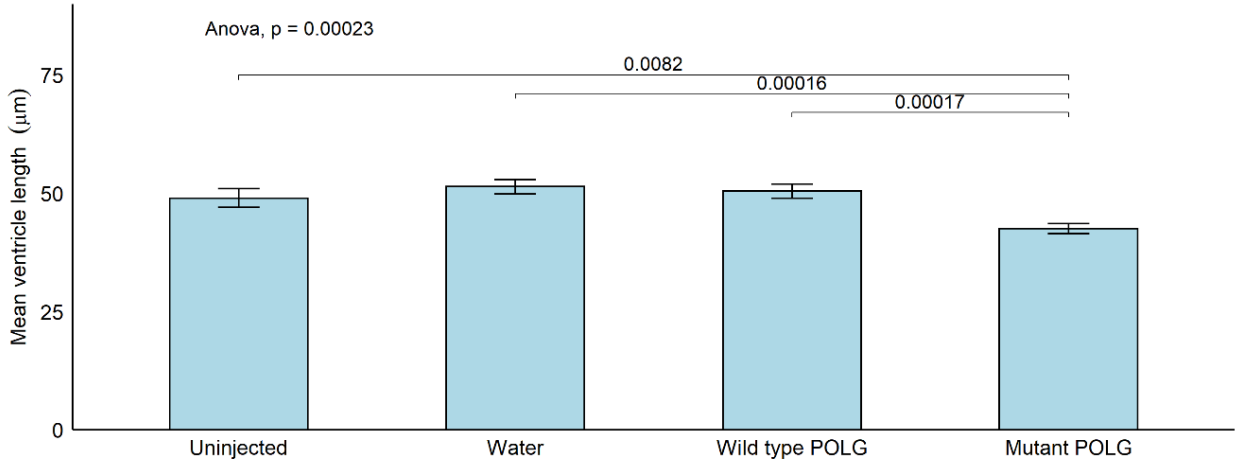


Figure W.1: Comparison of ventricular length amongst the experimental groups. The length is represented on the y-axis, while experimental groups are on the x-axis. Error bars indicate the standard error in each group. Significance bars show the p-value when comparing each experimental group to the uninjected control group (two-sided t-test).

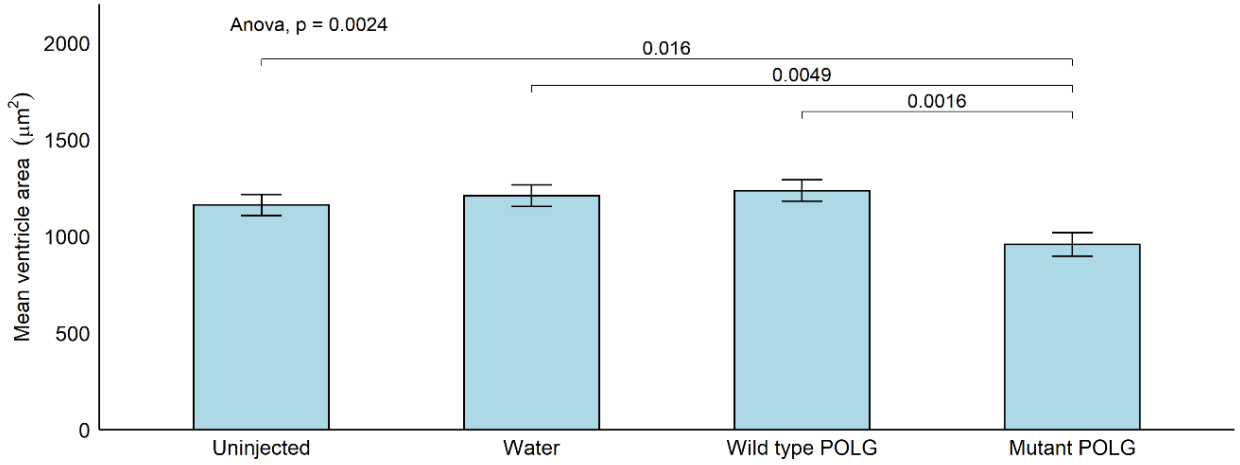


Figure W.2: Comparison of ventricular area amongst the experimental groups. The area is represented on the y-axis, while experimental groups are on the x-axis. Error bars indicate the standard error in each group. Significance bars show the p-value when comparing each experimental group to the uninjected control group (two-sided t-test).

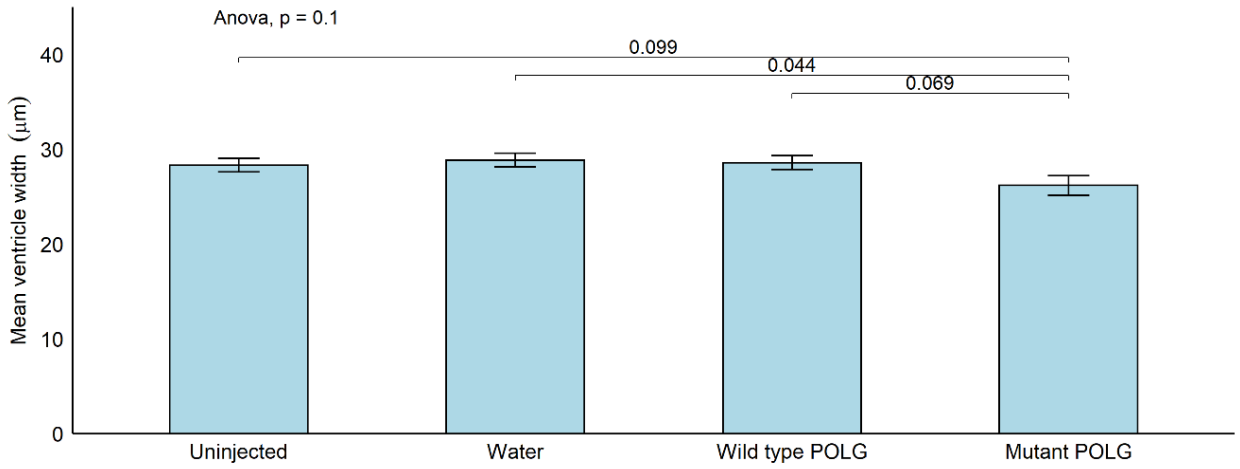


Figure W.3: Comparison of ventricular width amongst the experimental groups. The width is represented on the y-axis, while experimental groups are on the x-axis. Error bars indicate the standard error in each group. Significance bars show the p-value when comparing each experimental group to the uninjected control group (two-sided t-test).