Analysis of Genetic Variations Associated with Arrhythmogenic Right Ventricular Cardiomyopathy

Maryam Fish

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Medicine

UNIVERSITY OF CAPE TOWN

Supervisor: Professor Bongani Mayosi

Co-supervisor: Dr Gasnat Shaboodien

September 2015

TABLE OF CONTENTS

DECLARATION

I, Maryam Fish, declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that it has not been presented for another degree at this or any other university.

Signature: Signed by candidate

Maryam Fish

Date: 21/09/15

ABSTRACT

Introduction

Cardiomyopathy accounts for 20-30% of acute heart failure cases in adult Africans. Several types of cardiomyopathy have been identified; this study focused primarily on the genetic causes of arrhythmogenic right ventricular cardiomyopathy (ARVC). Many genes are implicated in ARVC pathogenesis, but many remain to be identified.

Methods and Results

We investigated a South African family (ACM2) with autosomal dominant ARVC, for whom the genetic cause of disease was unknown. Extensive genetic analysis was previously performed using genome-wide linkage analysis, but no disease-causing genetic variant was identified. We subsequently performed candidate gene screening of the phospholamban (*PLN*) gene, genome-wide copy number variant (CNV) analysis and whole exome sequencing to identify the causal genetic variant.

The ACM2 family harboured no disease-causing *PLN* variants. However, on screening all cardiomyopathy cases in our registry (ARVC, dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy and peripartum cardiomyopathy), we identified a known pathogenic *PLN* variant (c.25C>T; p.R9C) in a DCM family of European descent. This variant was reported in an American DCM family of European descent. Haplotype analysis revealed independent variant origins in these families.

CNV analysis revealed no disease-causing variants in the ACM2 family. Whole exome sequencing of two affected ACM2 family members revealed 38 variants shared by these individuals. Variants were verified in family members and population controls by high resolution melt analysis and Sanger sequencing, and by bioinformatic analysis to predict variant pathogenicity. A novel N-cadherin (*CDH2*) c.686A>C (p.Q229P) variant segregated with ARVC in the ACM2 family and was bioinformatically predicted to be deleterious. An additional pathogenic *CDH2* variant (c.1219G>A (p.D407N)) was identified in another individual with ARVC after screening 85 cases. These *CDH2* variants were absent in normal population controls. Furthermore, alterations in *Cdh2* are known to cause cardiomyopathy in rodent models. Taken together, these findings support the causal role of N-cadherin gene variants in human cardiomyopathy.

Conclusion

The detection of two different rare pathogenic variants in *CDH2* in individuals with ARVC implicates this gene as a novel cause of the disease and a possible diagnostic and therapeutic target for cardiomyopathy. A known pathogenic variant in *PLN* was confirmed as a cause of DCM.

ACKNOWLEDGEMENTS

- To my supervisor, Professor Bongani Mayosi, for all his invaluable guidance, enthusiasm and assistance throughout the course of this project, without which this research would not have been possible

- To my co-supervisor, Dr Gasnat Shaboodien, for all her invaluable and constant assistance, guidance and support throughout this project

- To Dr Sarah Kraus for her enthusiasm, perseverance and assistance with the clinical investigation of the ACM2 family

- To Professor Peter Schwartz, Dr Lia Crotti and Dr Elisa Mastantuono for their assistance with the SureSelect exome sequencing experiment

- To Dr Ruhan Slabbert, Mr Carel van Heerden and Dr Guillaume Pare for their assistance with the Ion AmpliSeq exome sequencing experiment

- To Dr Jo McBride for her assistance with the Copy Number Variant Analysis experiment

- To all the other members of the Cardiovascular Genetics Team at the Hatter Institute for Cardiovascular Research in Africa, including Dr Gaurang Deshpande, Mrs Lameez Pearce, Mr Stephen Kamuli, Miss Tafadzwa Machipisa and Miss Janine Saaiman, thank you for all your assistance and for making the time spent in the lab enjoyable

- To the other members of the Hatter Institute for Cardiovascular Research in Africa, thank you for all your assistance

- To the South African National Research Foundation, the Medical Research Council of South Africa, the International Centre for Genetic Engineering and Biotechnology, the Italian Government, the Lily and Ernst Hausmann Research Trust and the University of Cape Town for project funding

- Finally, to my parents, siblings and friends for bearing with me, and for providing me with constant support

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

CHAPTER 2 – SCREENING OF THE PHOSPHOLAMBAN (*PLN*) GENE IN THE ACM2 FAMILY AND OTHER CASES OF CARDIOMYOPATHY

CHAPTER 3 – COPY NUMBER VARIATION IN THE ACM2 FAMILY

CHAPTER 4 – SCREENING BY EXOME SEQUENCING IN THE ACM2 FAMILY

CHAPTER 5 – SCREENING OF *CDH2* IN AN ARVC COHORT

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

CHAPTER 2 – SCREENING OF THE PHOSPHOLAMBAN (*PLN*) GENE IN THE ACM2 FAMILY AND OTHER CASES OF CARDIOMYOPATHY

CHAPTER 4 – SCREENING BY EXOME SEQUENCING IN THE ACM2 FAMILY

CHAPTER 5 – SCREENING OF *CDH2* IN AN ARVC COHORT

LIST OF ABBREVIATIONS AND SYMBOLS

µ	Micro
ABI	Applied Biosystems
ACM	Arrhythmogenic Cardiomyopathy
ADAM6	ADAM Metallopeptidase Domain 6 (Pseudogene)
ADD1	α-adducin
AMP	Adenosine Monophosphate
ANKRD62	Ankyrin Repeat Domain 62
ARL17A	Adenosine Diphosphate-Ribosylation Factor-Like 17A
ARL17B	Adenosine Diphosphate-Ribosylation Factor-Like 17A
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
ARVD	Arrhythmogenic Right Ventricular Dysplasia
ASCL1	Achaete-Scute Family BHLH Transcription Factor 1
ATP	Adenosine Triphosphate
B3GNT6	UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase 6
BDH2	3-Hydroxybutyrate Dehydrogenase, Type 2
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BAG3	BCL2-Associated Athanogene 3
BRCA1	Breast Cancer 1, Early Onset
C10orf11	Chromosome 10 Open Reading Frame 11
C15orf62	Chromosome 15 Open Reading Frame 62
C17orf80	Chromosome 17 Open Reading Frame 80
Ca^{2+}	Calcium ion
CAMK4	Calcium/Calmodulin-Dependent Protein Kinase IV

cAMP	Cyclic adenosine monophosphate
CASSA	Cardiac Arrhythmia Society of Southern Africa
CDH2	N-cadherin
CFHR2	Complement Factor H-Related 2
CLTC	Clathrin, Heavy Chain (Human)
CNV	Copy Number Variation
CTNNA3	Catenin (Cadherin-Associated Protein), Alpha 3
CVD	Cardiovascular Disease
dbSNP	National Center for Biotechnology Information Single Nucleotide Polymorphism Database
DCM	Dilated Cardiomyopathy
ddNTPs	Dideoxynucleotide Triphosphates
DDX56	DEAD (Asp-Glu-Ala-Asp) Box Helicase 56
DES	Desmin
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphates
DSC2	Desmocollin 2
dsDNA	Double-stranded deoxyribonucleic acid
DSG1	Desmoglein 1
DSG2	Desmoglein 2
DSP	Desmoplakin
DUSP18	Dual Specificity Phosphatase 18
ECG	Electrocardiography
EDTA	Ethylenediaminetetraacetic acid
EMB	Endomyocardial biopsy
ENPP7	Ectonucleotide Pyrophosphatase/Phosphodiesterase 7

EP400	E1A Binding Protein P400
EPS	Electrophysiological study
ESCO1	Establishment of Sister Chromatid Cohesion N-Acetyltransferase 1
ESE	Exonic Splice Enhancer
EST	Exercise stress test
ExAC	Exome Aggregation Consortium
FSCN3	Fascin Actin-Bundling Protein 3
g	grams
G6PC	Glucose-6-Phosphatase, Catalytic Subunit
GALNT2	Polypeptide N-Acetylgalactosaminyltransferase 2
gDNA	Genomic DNA
GPD1L	Glycerol-3-Phosphate Dehydrogenase 1-Like
G_s	G-protein subunit
HCM	Hypertrophic Cardiomyopathy
HRM	High Resolution Melt
HTR3C	5-Hydroxytryptamine (Serotonin) Receptor 3C
ICD	Implantable cardioverter defibrillator
IDT	Integrated DNA Technologies
JUP	Junction plakoglobin
KANSL1	K (Lysine) Acetyltransferase 8 Regulatory NSL Complex Subunit 1
KANSL1-AS1	KANSL1 Antisense RNA 1
kb	Kilobases
KCNQ1	Potassium Channel, Voltage Gated KQT-Like Subfamily Q, Member 1
KIAA0125	Uncharacterised gene
KIF21B	Kinesin Family Member 21B
KLF11	Kruppel-Like Factor 11

KRT17	Keratin 17, Type I
KRT31	Keratin 31, Type I
KRT34	Keratin 34, Type I
KRTAP9-2	Keratin Associated Protein 9-2
KRTAP16-1	Keratin Associated Protein 16-1
l	Litres
LAMC2	Laminin, Gamma 2
LBBB	Left bundle branch block
LINC00226	Long Intergenic Non-Protein Coding Ribonucleic Acid 226
LMNA	Lamin A/C
LOD	Logarithm of the odds
M	Molar (moles per litre)
MAF	Minor allele frequency
Mb	Megabases
$MgCl_2$	Magnesium Chloride
Min	Minutes
ml	Millilitres
MLPA	Multiplex Ligation-Dependent Probe Amplification
mM	Millimolar
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
ms	Milliseconds
mV	Millivolts
MYBPC3	Myosin Binding Protein C, Cardiac
MYO16	Myosin XVI
Na^+	Sodium ion

NCBI	National Center for Biotechnology Information
NCOA3	Nuclear Receptor Coactivator 3
NEK10	NIMA-Related Kinase 10
NF2	Neurofibromin 2
NHLBI-EVS ESP	National Heart, Lung and Blood Institute Exome Variant Server Exome Sequencing Project
ng	Nanograms
NSF	N-Ethylmaleimide-Sensitive Factor
NSFP1	N-Ethylmaleimide-Sensitive Factor Pseudogene 1
NSVT	Non-sustained ventricular tachycardia
OR4D6	Olfactory Receptor, Family 4, Subfamily D, Member 6
PCR	Polymerase Chain Reaction
PKA	Protein Kinase A
PKP2	Plakophilin 2
PLN	Phospholamban
pM	Picomolar
PolyPhen-2	Polymorphism Phenotyping 2
PPCM	Peripartum Cardiomyopathy
PVC	Premature ventricular contractions
QC	Quality Control
RCM	Restrictive Cardiomyopathy
RNA	Ribonucleic Acid
Rpm	Revolutions per minute
RTL1	Retrotransposon-Like 1
RV	Right ventricle
RVA	Right ventricular akinesia

RVIT	Right ventricular inflow tract
RVOT	Right ventricular outflow tract
RyR2	Cardiac Ryanodine Receptor gene
SAECG	Signal-averaged electrocardiogram
SCN4A	Sodium Channel, Voltage Gated, Type IV Alpha Subunit
SD	Standard Deviation
Sec	Seconds
SERCA2a	Sarcoplasmic reticulum (SR) Ca^{2+} ATPase pump
SIFT	Sorting Intolerant From Tolerant
SNP	Single Nucleotide Polymorphism
SRSF1	Serine/Arginine-Rich Splicing Factor 1
SRSF5	Serine/Arginine-Rich Splicing Factor 5
STL	Six-Twelve Leukemia
STR	Short Tandem Repeat
TAD	Terminal activation duration
TE	Tris- Ethylenediaminetetraacetic acid
TBC1D14	TBC1 Domain Family, Member 14
TGFβ3	Transforming Growth Factor Beta 3
TMEM43	Transmembrane Protein 43
TTC37	Tetratricopeptide Repeat Domain 37
TTN	Titin
TWI	T-wave inversion
U	Units
UCSC	University of California, Santa Cruz
VT	Ventricular tachycardia
WES	Whole Exome Sequencing

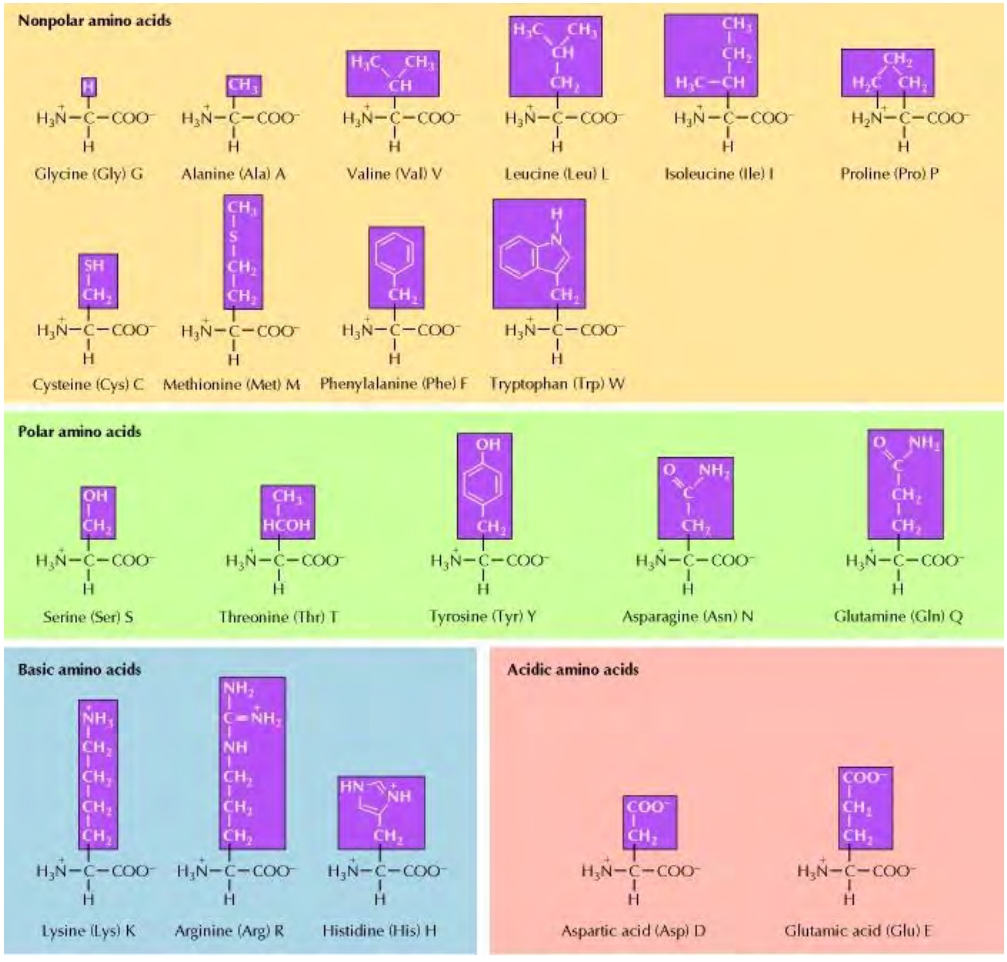
YAP	Yes-associated protein
ZGRF1	Zinc Finger, GRF-Type Containing 1
ZNF860	Zinc Finger Protein 860

Nucleic acid codes

A	Adenine
G	Guanine
C	Cytosine
T	Thymine

Amino acid codes

Indicated below are the names, symbols and properties of the amino acids (from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=cooper&part=A249)



CHAPTER 1
INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of death in all regions of the world except sub-Saharan Africa.[1] CVD has a significant socioeconomic impact on society, causing immense suffering and loss of life as well as posing major economic challenges to developing countries, including costs to the healthcare systems and to the national economy. CVD, which manifests mainly as heart disease and stroke, is attributed to a number of risk factors, including high blood pressure, high plasma cholesterol, tobacco and alcohol use and low fruit and vegetable consumption.[2] In sub-Saharan Africa, the major causes of heart failure (a major form of CVD) are hypertension, rheumatic heart disease and cardiomyopathy.[3] Cardiomyopathy accounts for 20-30% of heart failure in adult African individuals and poses the greatest challenge among the CVD causative factors due to the difficulty of diagnosis, the high prevalence in resource-poor societies, and the associated high disease mortality.[4]

1.1. Cardiomyopathy

Cardiomyopathy is defined as a myocardial disease in which the heart muscle is structurally and functionally abnormal in the absence of coronary artery disease, valvular disease, hypertension, pericardial disease or congenital heart disease sufficient to cause the observed myocardial abnormality.[5] Cardiomyopathies often result in progressive heart failure with significant morbidity and mortality, and may be primary (genetic, acquired or mixed) or secondary (inflammatory, infiltrative or toxic).[6] Cardiomyopathies are classified into morphological and functional phenotypes, which may be familial or non-familial conditions.[5] Various types of cardiomyopathy have been identified, including arrhythmogenic right ventricular cardiomyopathy (ARVC), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM) and left ventricular non-compaction cardiomyopathy.[5] For the purpose of this study I will be focusing on ARVC which is the least understood form of cardiomyopathy in Africans.

1.1.1. Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)

ARVC is a rare inherited muscle disease characterised by replacement of the right ventricular myocardium by fibro-fatty tissue, causing electrical instability including ventricular arrhythmias in the early stages and later reduced contractility and heart failure (Figure 1.1).[7-9] ARVC has been shown to have left-dominant and biventricular subtypes, and the broader term of "arrhythmogenic cardiomyopathy" has therefore been suggested.[7]

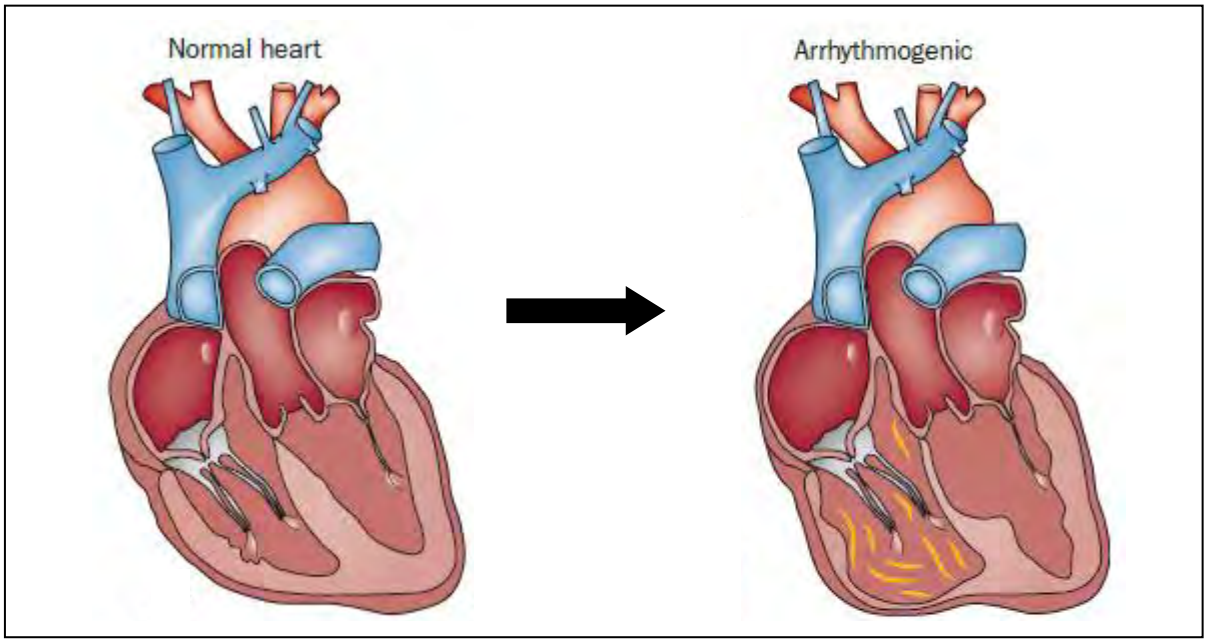


Figure 1.1: Illustration of a heart with arrhythmogenic right ventricular cardiomyopathy. The right ventricular myocytes are replaced with fibro-fatty tissue leading to the arrhythmias characteristic of the disease (adapted from Hershberger *et al*. 2013)[10]

Studies from Italy show that the prevalence of ARVC ranges from 1-2 per 5000 individuals and is responsible for 15-25% of cases of sudden cardiac death in patients <35 years;[11] the prevalence in Africa remains to be determined. The first reported case of ARVC in South Africa was published by Munclinger and colleagues in 2000.[12] The ARVC Registry of South Africa was established in 2004 by the Working Group on Registries of the Cardiac Arrhythmia Society of Southern Africa (CASSA). Preliminary information from this registry suggests that ARVC occurs in all segments of the population and that the clinical features, outcome and frequency of familial disease may be similar to observations in other parts of the world.[13]

The diagnosis of ARVC is made according to a set of Task Force Criteria, originally published in 1994 by McKenna and colleagues[14], and modified in 2010 by Marcus and colleagues.[15] The criteria for the diagnosis of ARVC take into account structural, histological, arrhythmic, electrocardiography (ECG) and familial features of the disease (Table 1.1).[15] Patients are classified as having definite, borderline or possible ARVC dependent on the number of major and minor criteria they fulfil.[15]

Table 1.1: ARVC Task Force Criteria (from Marcus *et al.* 2010)[15]

MAJOR	MINOR
I. Global or regional dysfunction and structural alterations	
By 2D echocardiogram: Regional RV akinesia, dyskinesia, or aneurysm and 1 of the following (end diastole): PLAX RVOT ≥32 mm (corrected for body size [PLAX/BSA] ≥19 mm/m^2) PSAX RVOT ≥36 mm (corrected for body size [PSAX/BSA] ≥21 mm/m^2) or fractional area change ≤33% By MRI: Regional RV akinesia or dyskinesia or dyssynchronous RV contraction and 1 of the following: Ratio of RV end-diastolic volume to BSA ≥110 ml/m^2 (male) or ≥100 ml/m^2 (female) or RV ejection fraction ≤40% By RV angiography: Regional RV akinesia, dyskinesia, or aneurysm	By 2D echocardiogram: Regional RV akinesia or dyskinesia and 1 of the following (end diastole): PLAX RVOT ≥29 to <32 mm (corrected for body size [PLAX/BSA] ≥16 to <19 mm/m^2) PSAX RVOT ≥32 to <36 mm (corrected for body size [PSAX/BSA] ≥18 to <21 mm/m^2) or fractional area change >33% to ≤40% By MRI: Regional RV akinesia or dyskinesia or dyssynchronous RV contraction and 1 of the following: Ratio of RV end-diastolic volume to BSA ≥100 to <110 ml/m^2 (male) or ≥90 to <100 ml/m^2 (female) or RV ejection fraction >40% to ≤45%
II. Tissue characterisation of wall	
Residual myocytes <60% by morphometric analysis (or <50% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy	Residual myocytes 60% to 75% by morphometric analysis (or 50% to 65% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy
III. Repolarisation abnormalities	
Inverted T waves in right precordial leads (V1, V2, and V3) or beyond in individuals >14 years of age (in the absence of complete right bundle-branch block QRS ≥120 ms)	Inverted T waves in leads V1 and V2 in individuals >14 years of age (in the absence of complete right bundle-branch block) or in V4, V5, or V6 Inverted T waves in leads V1, V2, V3, and V4 in individuals >14 years of age in the presence of complete right bundle-branch block
IV. 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 (V1 to V3)	Late potentials by SAECG in ≥1 of 3 parameters in the absence of a QRS duration of ≥110 ms on the standard ECG Filtered QRS duration (fQRS) ≥114 ms Duration of terminal QRS <40 mV (low-amplitude signal duration) ≥38 ms Root-mean-square voltage of terminal 40 ms ≤20 mV Terminal activation duration of QRS ≥55 ms measured from the nadir of the S wave to the end of the QRS, including R0, in V1, V2, or V3, in the absence of complete right bundle-branch block
V. Arrhythmias	
Nonsustained or sustained ventricular tachycardia of left bundle-branch morphology with superior axis (negative or indeterminate QRS in leads II, III, and aVF and positive in lead aVL)	Nonsustained or sustained ventricular tachycardia of RV outflow configuration, left bundle-branch block morphology with inferior axis (positive QRS in leads II, III, and aVF and negative in lead aVL) or of unknown axis >500 ventricular extrasystoles per 24 hours (Holter)
VI. Family history	
ARVC/D confirmed in a first-degree relative who meets current Task Force criteria ARVC/D confirmed pathologically at autopsy or surgery in a first-degree relative Identification of a pathogenic genetic variant categorised as associated or probably associated with ARVC/D in the patient under evaluation	History of ARVC/D in a first-degree relative in whom it is not possible or practical to determine whether the family member meets current Task Force criteria Premature sudden death (<35 years of age) due to suspected ARVC/D in a first-degree relative ARVC/D confirmed pathologically or by current Task Force Criteria in second-degree relative

RV – Right Ventricle; PLAX – Parasternal Long Axis; RVOT – Right Ventricular Outflow Tract; PSAX – Parasternal Short Axis; BSA – body surface area; MRI – Magnetic Resonance Imaging; SAECG – Signal-averaged electrocardiogram; ECG – Electrocardiogram; aVR – Augmented ventricular right; aVL – Augmented ventricular left; ARVC/D – Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia

1.1.1.1. Molecular genetics of ARVC

A familial background of ARVC has been demonstrated in at least 50% of cases, with the disease typically having an autosomal dominant pattern of inheritance.[16,17] The penetrance of ARVC is reduced and frequently age-dependent, while affected relatives often have variable and incomplete disease expression.[7,18] A number of genetic loci have been linked to ARVC to date, and disease-associated genetic variants have been identified at the majority of these loci (Table 1.2).

Among the genes reported to cause ARVC, the main genetic causes reported to date are variants in the plakophilin 2 (*PKP2*), desmoplakin (*DSP*), desmocollin 2 (*DSC2*), desmoglein 2 (*DSG2*) and junction plakoglobin (*JUP*) genes which encode proteins forming part of the desmosome. Desmosomes are cell adhesion junctions particularly abundant in tissues subjected to a high amount of mechanical stress, such as the heart and the skin (Figure 1.2).[19] Disease-associated variants in these genes have been reported in up to 50% of ARVC cases in various patient cohorts worldwide.[20] The main contributor to ARVC is the *PKP2* gene, accounting for 11-51% of disease-associated variants to date.[21] Studies conducted in our laboratory have screened patients with ARVC for variants in the desmosomal genes and have detected a disease-associated variant prevalence of 21.5%.[22] However, variants in desmosomal genes demonstrate incomplete penetrance and variable expressivity, suggesting that genetic modifiers and environmental factors may play a role in the etiology of these disorders.[8] There have also been reports of patients harbouring more than one desmosomal gene variant, either in the same gene or in two different genes (compound heterozygosity), with the presence of multiple variants causing a more severe ARVC phenotype.[23]

In addition to desmosomal genes, other genetic causes of ARVC include *TMEM43,*[24,25] *RyR2,*[26] *TGFβ3,*[27] *PLN,*[28] *CTNNA3,*[29] *DES,*[30] *TTN*[31] and *LMNA.*[32] A number of these genes have also been implicated in DCM (*PLN*, *DES*, *TTN* and *LMNA*) as well as HCM (*DES*, *TTN*, *RyR2*).[10]

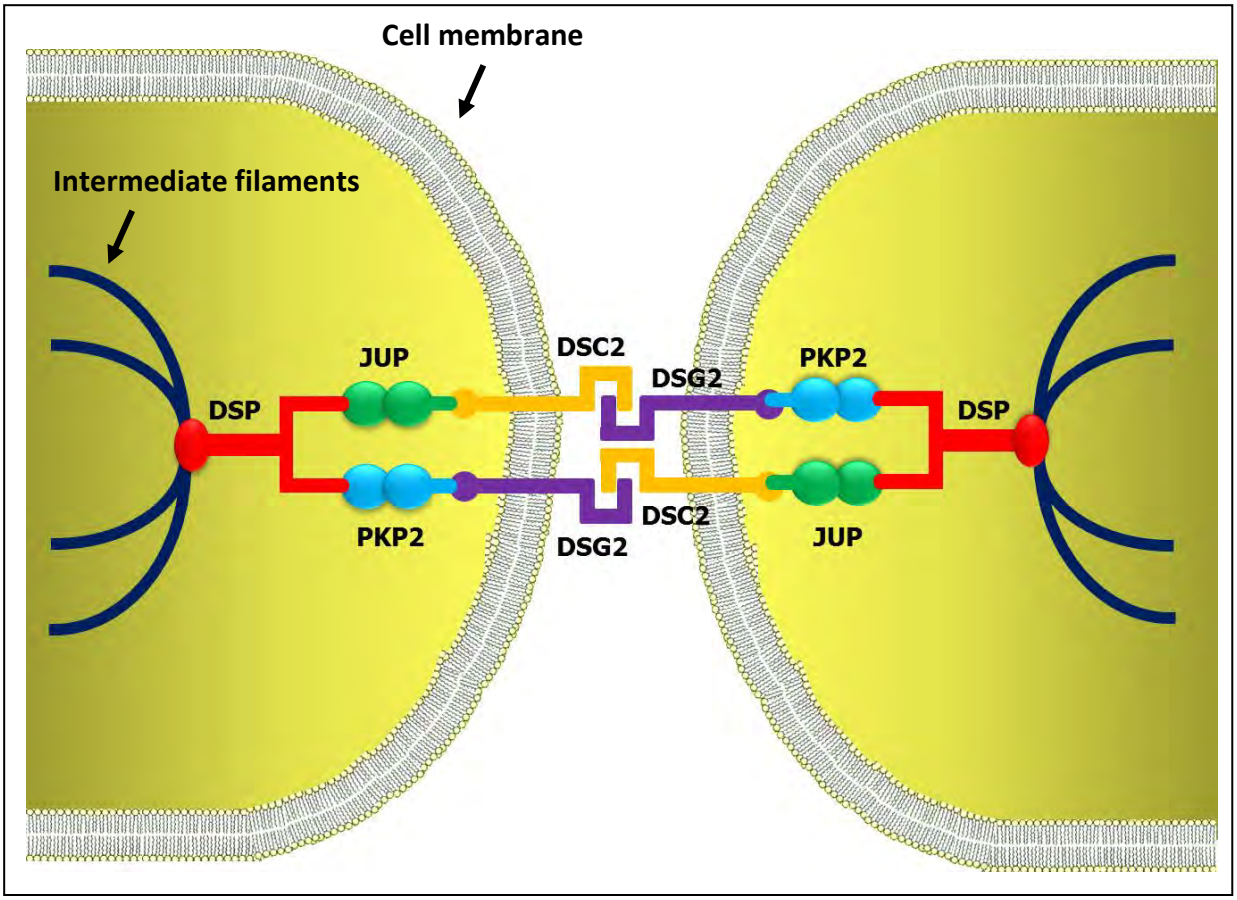


Figure 1.2: Structure of the desmosome. DSP – desmoplakin; JUP – Junction plakoglobin; PKP2 – Plakophilin-2; DSC2 – Desmocollin-2; DSG2 – Desmoglein-2

1.1.1.2. Mechanism of disease

Several mechanisms have been proposed to explain the link between desmosomal gene variants and ARVC, one such mechanism being the degeneration-inflammation model of desmosomal dysfunction. In this model, desmosomal variants resulting in desmosomal dysfunction could impair cell-cell adhesion or intermediate filament function or both, dependent on the effect of the variant on protein structure and function.[7] This would cause a reduction in the ability of the tissue to withstand mechanical stress, causing cellular detachment and necrosis, which may be accompanied by inflammation. When the regenerative capacity of the myocardium is exceeded, the myocytes are instead replaced with fibrous and/or fatty tissue.[7,8]

Another proposed disease model is the transdifferentiation model, involving the Wnt/β-catenin signalling pathway.[7,8] This pathway has been implicated in the regulation of cell fate,

proliferation and apoptosis.[7,8] The Wnt/β-catenin signalling pathway usually promotes gene expression which causes cell differentiation to a cardiomyocyte state and suppression of the expression of genes that promote adipogenesis and fibrogenesis.[7,8] Plakoglobin (*JUP* – one of the desmosomal genes), also known as γ-catenin, is structurally and functionally similar to β-catenin and can shuttle between the cell periphery and the nucleus. In the event of a genetic variant that affects the localisation of plakoglobin (e.g. a variant that suppresses the expression of desmoplakin), plakoglobin translocates to the nucleus where it can take the place of β-catenin and cause upregulation of adipogenic and fibrogenic genes.[7,8]

A recent study has implicated the Hippo/Yes-associated protein (YAP) pathway in the pathogenesis of ARVC.[33] This pathway plays a role in the regulation of cell proliferation and organ growth and has recently been linked to control of cardiomyocyte proliferation and heart size.[34-37] In the event of a desmosomal gene variant causing desmosomal instability, neurofibromin 2 (NF2) (upstream of the Hippo pathway) is activated, resulting in cascade phosphorylation of the Hippo kinases, inactivation of YAP (downstream of the Hippo pathway) and enhanced adipogenesis, one of the hallmarks of ARVC.[33,37]

Other studies have shown evidence of gap junction remodelling and aberrant calcium homeostasis in the presence of desmosomal gene variants.[8] It is not known which of these mechanisms are responsible for the phenotype of ARVC, or whether there is interplay between these mechanisms.[7]

Table 1.2: Genetic loci associated with ARVC to date

DESIGNATION	LOCUS	SYMBOL	CANDIDATE GENE	MODE OF INHERITANCE	LOCUS REFERENCE	VARIANT REFERENCE
ARVD1	14q23-q24	*TGFβ3*	Transforming Growth Factor-β3	AD	[38]	[27]
ARVD2	1q42-q43	*RyR2*	Ryanodine Receptor 2	AD	[39]	[26]
ARVD3	14q12-q22	-	Unknown	AD	[40]	-
ARVD4	2q32.1-q32.3	-	Unknown	AD	[41]	-
ARVD5	3p23	*TMEM43*	Transmembrane protein 43	AD	[42]	[24]
ARVD6	10p12-p14	-	Unknown	AD	[43,44]	-
ARVD7	10q22	-	Unknown	AD	[45]	-
ARVD8	6p24	*DSP*	Desmoplakin	AD/AR	[46]	[46]
ARVD9	12p11	*PKP2*	Plakophilin 2	AD	[47]	[47]
ARVD10	18q12.1- q12.2	*DSG2*	Desmoglein 2	AD	[48]	[48]
ARVD11	18q12.1	*DSC2*	Desmocollin 2	AD	[49]	[49]
ARVD12 / Naxos	17q21	*JUP*	Junction Plakoglobin	AD	[50]	[51]

ARVD – Arrhythmogenic Right Ventricular Dysplasia; AD – autosomal dominant; AR – autosomal recessive

1.1.2. Cardiomyopathy Research in South Africa

To the best of my knowledge, there are only three original scientific studies of ARVC in South Africa that have been reported to date.[12,13] The initial report of ARVC in Africa was published by Munclinger and colleagues in 2000.[12] Mokhobo and Mntla had reported cases of right ventricular cardiomyopathy a few years ahead of this report but it is not known whether these cases met the diagnostic criteria of ARVC or not.[52] The ARVC Registry of South Africa was established in 2004 by CASSA, and the preliminary reports from this registry suggest that ARVC occurs in all population groups and that the frequency of familial disease, clinical features and outcome are similar to observations from other parts of the world.[13]

Molecular genetic studies on ARVC in Africa have largely focused on variants in the desmosomal genes. A study by Mbele and colleagues (available in PhD thesis form only) reported disease-associated variants in desmosomal genes in patients with ARVC.[22] In particular, this study reported the *PKP2* c.1162C>T founder variant in multiple South African ARVC families of European descent.[22] Another unpublished study was from my MSc thesis, which reported variants in the *DSP* desmosomal gene in patients with ARVC.[53] However, the genetic factors responsible for the majority of ARVC cases in South Africa remain unknown; the need to define the genetic factors in South Africans with ARVC serves as the basis for this thesis.

A relatively new candidate gene for ARVC is the phospholamban (*PLN*) gene. Recent reports by van der Zwaag and colleagues have reported disease-associated variants in the *PLN* gene that encode the phospholamban protein in patients with ARVC.[28,54] They detected a founder variant in *PLN* that was common among individuals from the Dutch population as well as populations known to share common ancestry with this population.[28,54] As many South African Caucasian individuals also share common ancestry with this population, there was the possibility that the *PLN* founder variants may have been associated with ARVC in this population. We aimed to address this question in this study.

1.2. The ACM2 family

We previously reported a four-generation South African family of northern European descent with a dominant form of ARVC (family ACM2).[44] Extensive genetic analysis was previously performed and involved techniques such as exclusion mapping and whole genome single nucleotide polymorphism (SNP) linkage analysis.[44,55] Exclusion mapping resulted in the exclusion of all known ARVC loci with the exception of the ARVC6 locus on chromosome 10p12-p14 (Table 1.2) which showed borderline evidence of linkage (highest LOD score of 2.93).[44] Screening of the coding regions of the 15 genes that occur at this locus revealed no disease-associated variants.[55] Genome-wide linkage studies using SNP arrays revealed suggestive evidence of linkage at chromosomes 1, 4, 10, and 12 (Two-point LOD scores: 2.73, 2.78, 2.72 and 2.56 respectively), with the highest multipoint LOD score being 2.79 for chromosomes 4 and 10.[55] However, no disease-associated variants were identified at these loci.

The design of this current study focused on finding the disease-associated variant/s in the ACM2 family using three different approaches. First, we used candidate gene screening of the phospholamban gene (*PLN*) (chapter 2) then genome-wide copy number variant (CNV) analysis (chapter 3), and finally, whole exome sequencing (chapter 4) to identify the disease-causing genetic variant in this South African kindred with autosomal dominant ARVC.

1.3. Hypothesis

We propose that the disease-causing genetic variant in the ACM2 family is likely to lie within the exome, as more than 85% of disease-associated genetic variants lie within this region, and there is a strong possibility that this is the case in this family, particularly due to the dominant pattern of inheritance of the disease.

Although many genes have been associated with ARVC to date, disease-associated genetic variants remain to be identified in a large proportion of individuals with this disease. In the ACM2 family with autosomal dominant ARVC, no genetic cause has been identified since the family was first identified in 1996. We hypothesise that the genetic cause of disease in this

family may lie in either *PLN* or a novel ARVC genetic locus and may involve a gene that has not been previously associated with this disease.

1.4. Objective of the project

To identify the genetic cause of disease in the ACM2 family with autosomal dominant ARVC

1.5. Specific aims

The specific aims of this project were:

(1) To screen the ACM2 family for disease-causing variants in *PLN* and determine the prevalence of variants in this gene in other cases of cardiomyopathy (ARVC, DCM, PCM and HCM). If no disease-causing variant in *PLN* is found in the ACM2 family, then –

(2) To screen the ACM2 family for the presence of disease-causing copy number variations (CNVs). If no pathogenic CNV is found in the ACM2 family, then –

(3) To screen the ACM2 family for disease-causing genetic variants by whole exome sequencing

CHAPTER 2
SCREENING OF THE PHOSPHOLAMBAN (*PLN*) GENE IN THE ACM2 FAMILY AND OTHER CASES OF CARDIOMYOPATHY

2.1. Introduction

Cardiomyopathy is caused by genetic factors in a large proportion of cases, and our knowledge of these factors has increased monumentally in recent years.[56] Arrhythmogenic right ventricular cardiomyopathy (ARVC) has been associated with genes involved in cell-to-cell adhesion. Dilated cardiomyopathy (DCM) has been associated with genes involved in maintaining the structural integrity of the heart, muscle contraction and mitochondrial function while hypertrophic cardiomyopathy (HCM) is associated mainly with genes coding for proteins of the sarcomere.[57]

The ARVC Registry of South Africa was established in 2004 by the Cardiac Arrhythmia Society of Southern Africa (CASSA).[58] Suspected cases of ARVC are referred to the co-ordinating centre in the Cardiac Clinic at Groote Schuur Hospital, Cape Town. As of September 2015, 85 unrelated ARVC patients have been enrolled in the registry.[13] This registry houses one of our largest ARVC families (ACM2) and includes 20 members – six of whom are affected (Figure 2.1).[44,55]

The design of this study focused on finding the disease-causing variant/s in this family. Their ancestry traces back to that of northern European descent and an autosomal dominant inheritance pattern was noted. Extensive genetic analysis was previously performed and involved techniques such as whole genome SNP linkage analysis; unfortunately, the data was inconclusive and no disease-causing genetic variant was identified in this family.[44,55]

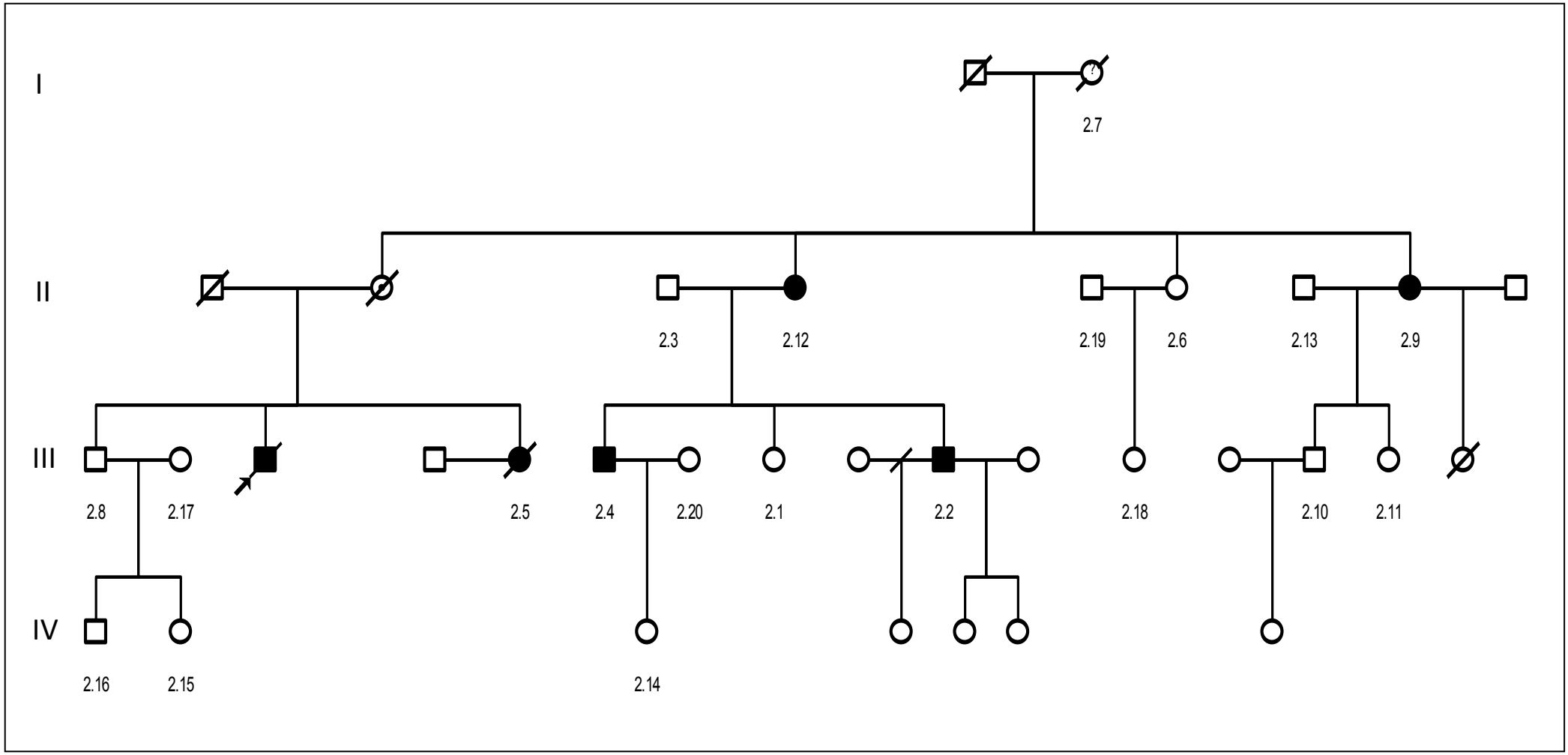


Figure 2.1: Pedigree of the ACM2 family with ARVC. Shaded squares indicate affected males and shaded circles indicate affected females

2.2. The phospholamban (*PLN*) gene

Recently, variants in the phospholamban (*PLN)* gene have been associated with ARVC.[28] Variants in this gene had previously been reported in cases of DCM[28,59-62] or HCM.[63,64] *PLN* is located on the long arm of chromosome 6 (6q22.1) and encodes phospholamban (PLN), a 52 amino acid protein with four domains: domains Ia, Ib and the loop domain located in the cytoplasm region and domain II located in the transmembrane region (Figure 2.2).[65] PLN exists as a monomer or forms a homopentamer in a pinwheel topology.[65] These two forms of PLN exist in equilibrium, a fact that is known to be important for PLN activity.[65] PLN plays an important role in cardiac contractility.

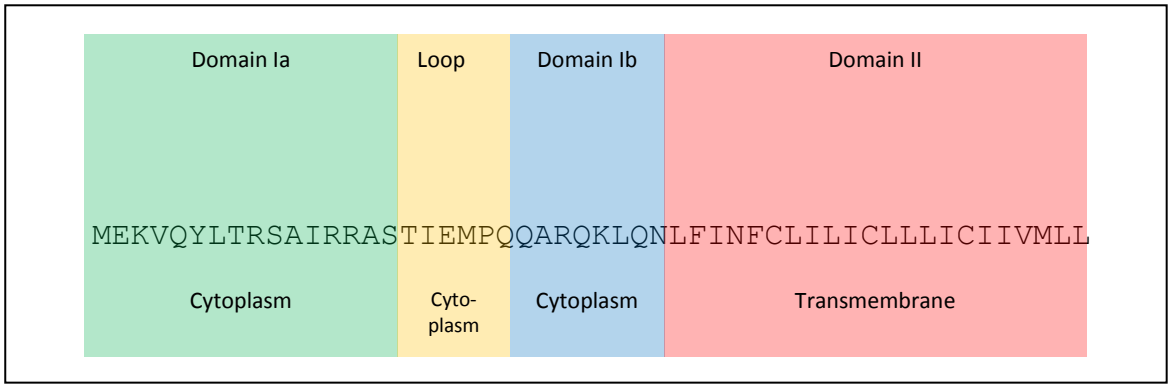


Figure 2.2: Domains of the PLN protein (adapted from Ha *et al.* 2011[65])

2.2.1. The Role of *PLN* in Cardiac Contractility

Adrenaline and other β-agonists bind to and activate β-adrenergic receptors on the cell membrane (Figure 2.3).[66] G_s proteins then stimulate adenylate cyclase, which leads to the formation of cyclic adenosine monophosphate (cAMP).[66] The rise in cAMP concentration activates cAMP-dependent protein kinase A (PKA)[66]. PKA then phosphorylates proteins such as PLN that play vital roles in cardiac function.[66]

PLN is involved in regulation of the sarcoplasmic reticulum (SR) Ca^{2+} ATPase pump (SERCA2a) that regulates Ca^{2+} reuptake into the sarcoplasmic reticulum and thereby allows muscle relaxation (Figure 2.3).[66] In its unphosphorylated state, PLN inhibits the activity of the SERCA2a pump, causing muscle contraction.[66] Phosphorylation by PKA allows PLN to be released from SERCA2a, allowing muscle relaxation.[66] Other proteins involved in Ca^{2+}

regulation include the cardiac ryanodine receptor (RyR), sodium calcium exchangers (NCX), dihydropyridine receptors (DHPR) and plasma membrane Ca^{2+} ATPases (PMCA) (Figure 2.3).[66]

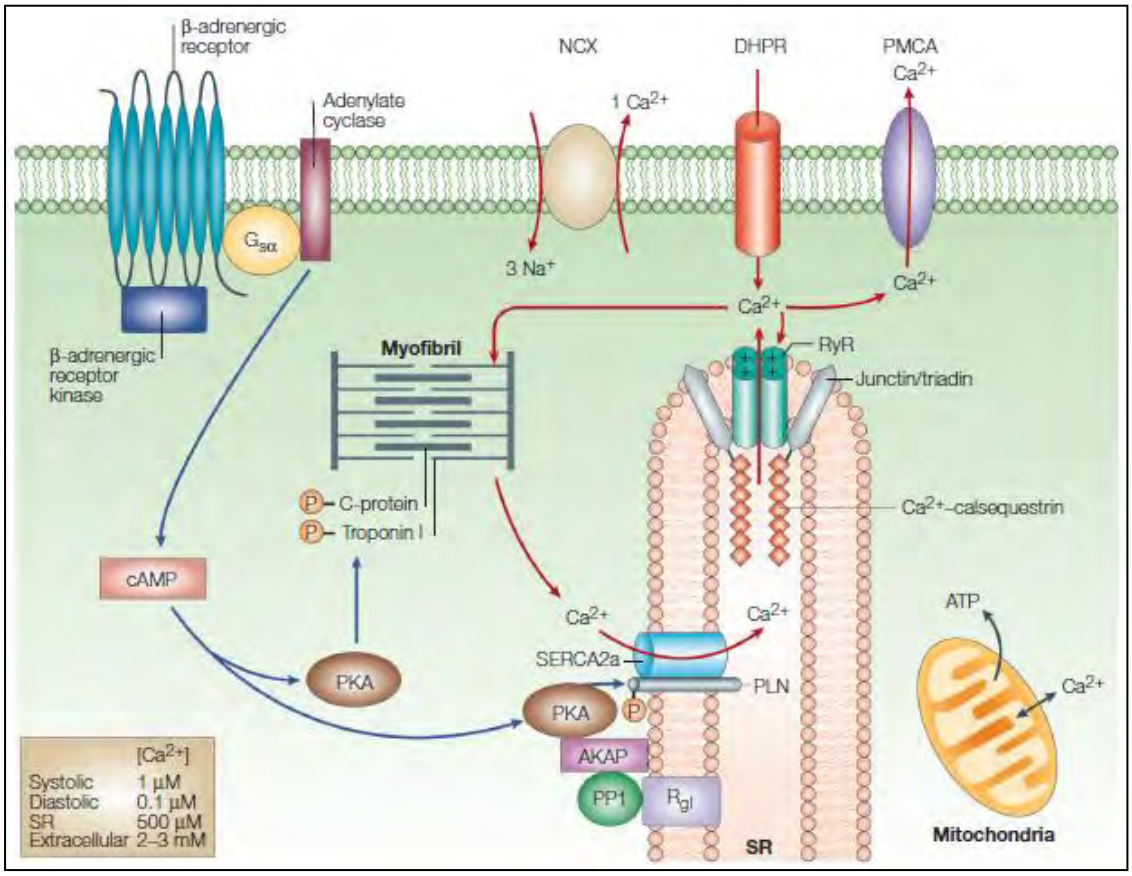


Figure 2.3: The role of PLN in cardiac contractility (from MacLennan and Kranias[66])

$G_{s\alpha}$ – G-protein subunit alpha; NCX – Sodium calcium exchanger; DHPR – Dihydropyridine receptor; ATP – Adenosine triphosphate; PMCA – Plasma membrane Ca^{2+} ATPase; cAMP – Cyclic adenosine monophosphate; RyR – Ryanodine receptor; PKA – Protein kinase A; PLN – Phospholamban; AKAP – A-kinase anchoring protein; PP1 – Protein phosphatase 1; RgI – Regulatory binding subunit A; SR – Sarcoplasmic reticulum; Na^{+} – Sodium ion; Ca^{2+} – Calcium ion

2.2.2. Variants in *PLN* associated with Disease

As mentioned above, variants in *PLN* have previously been associated with various types of cardiomyopathy including DCM, HCM and more recently ARVC. Van der Zwaag and colleagues recently identified the p.R14del variant in the *PLN* gene in patients with ARVC as well as patients with DCM.[28] This variant was determined to be a founder variant in the genetically isolated Newfoundland population, and associated with autosomal dominant

ARVC as well as with DCM.[28] The current Newfoundland population is known to share ancestors with populations from New Zealand, Australia and South Africa, with the founder population having originated in Europe.[54] As variants in *PLN* have been associated with ARVC, DCM and HCM in populations who share common ancestors with the South African Caucasian population, we hypothesise that *PLN* may account for some of the genetic causes of cardiomyopathy in patients with ARVC (such as the ACM2 family) and other patients with cardiomyopathy.

2.3. Objective

The objective of this study was to determine whether *PLN* variants cause cardiomyopathy in the South African population with ARVC, DCM, HCM and peripartum cardiomyopathy (PPCM). By screening large numbers of affected individuals, we aimed to provide a definitive perspective on the contribution of *PLN* variants to the etiology of cardiomyopathy in South Africa.

2.4. Specific aims

(1) To screen for DNA variants in *PLN* in probands with ARVC, DCM, HCM and PPCM, and (2) To determine the prevalence of *PLN* variants in normal controls drawn from the South African population.

2.5. Methods

This study involved screening patients with cardiomyopathy for variants in *PLN* by high resolution melt (HRM) analysis and Sanger DNA sequencing, as well as performing haplotype analysis to determine the presence of possible founder genetic variants.

2.5.1. Cohorts

The patient cohort consisted of ARVC, DCM, PPCM and HCM patients and ethnically matched controls. Informed consent was obtained from all participants (Appendix 1). These patients had been previously screened for variants in the desmosomal genes known to

cause ARVC (*DSP*, *PKP2*, *DSC2*, *DSG2* and *JUP*) and were found not to harbour any disease-causing variants in these genes.[13,22,53]

2.5.1.1. ARVC Cases

ARVC cases were investigated at the Cardiac Clinic, Groote Schuur Hospital, Cape Town, South Africa, and enrolled in the ARVC registry of South Africa.[67] A standardised case report form is completed for all participants and available family members with ARVC, DCM and other cardiomyopathies. Information on presenting symptoms, family history, electrocardiographic findings, cardiac imaging studies, histology, and outcome was collected. Informed consent was obtained from all participants (Appendix 1).

A diagnostic panel consisting of a group of cardiologists determined, by consensus, whether or not patients met the 2010 diagnostic criteria of ARVC.[15] The participants were classified as having definite ARVC (if Task Force criteria were met), probable or possible ARVC (if some criteria were met and no alternative diagnosis was found), or not having ARVC (if there was no evidence of ARVC and/or an alternative diagnosis was present). The diagnosis of ARVC in first-degree relatives of affected individuals was made based on the modified criteria of Hamid and colleagues.[68]

2.5.1.2. DCM, PPCM and HCM Cases

Index cases were recruited from the Cardiac Clinic, Groote Schuur Hospital in Cape Town; the University of Cape Town Private Academic Hospital in Cape Town; Inkosi Albert Luthuli Hospital in Durban; as well as Milpark Hospital, Chris Hani Baragwanath Hospital and Sunninghill Hospital in Johannesburg.

A standardised case report form was completed for all study participants, which included information regarding presenting symptoms, electrocardiographic findings, cardiac imaging studies, family history and outcome. Blood samples were collected for molecular genetic testing. Subjects were reviewed by a panel of cardiologists to determine if the index cases fulfilled the diagnostic criteria for DCM, PPCM or HCM established by the European Society

of Cardiology Working Group on Myocardial and Pericardial Diseases.[5] Patients determined to have secondary causes of cardiomyopathy or left ventricular dysfunction (including hypertension, valvular heart disease, diabetes, coronary artery disease or sepsis) were excluded from the study.

2.5.1.3. Controls

Healthy, anonymous blood donors with no history of heart disease from the Western Province Blood Transfusion Service provided blood samples for DNA isolation. Self-reported ethnicity of these individuals was Black African, Caucasian, Mixed ancestry and Indian. Controls were matched with cardiomyopathy cases according to ethnic origin.

2.5.2. DNA Isolation and Storage

Sample DNA was extracted from 5-10ml blood using the PureGene™ DNA Purification System (Gentra systems) according to the manufacturer's instructions. Written informed consent was obtained from all individuals enrolled in the study (Appendix 1). Extracted DNA samples were stored in 1.5ml Eppendorf tubes at -80°C for long term storage. All individuals were given codes to anonymise the samples and ensure patient confidentiality.

2.5.3. DNA Quality Control

All DNA samples were run on a 1% agarose gel (Appendix 2.2) to verify integrity. Sample concentration was determined with the Qubit® Fluorometer (Life technologies) using the Qubit® dsDNA BR Assay Kit and also with the NanoDrop™ 2000 Spectrophotometer. The NanoDrop™ 2000 Spectrophotometer was also used to verify that samples were free of contaminants.

2.5.4. High Resolution Melt Analysis

HRM analysis with the RotorGene 6000 (Corbett Life Sciences) was used for the screening of *PLN*. Genetic variant detection by HRM analysis involves two steps: amplification of the DNA

sequence of interest using PCR and analysis of the produced amplicon by HRM. With the RotorGene 6000 (Corbett Life Sciences), these two steps can be combined into a single procedure where DNA samples are amplified and then immediately subjected to HRM without the need for human intervention.

2.5.4.1. Primer design

Primer sequences were selected manually by (1) referring to the annotated gene sequences, and (2) analysing the selected sequences using the IDT OligoAnalyzer (https://eu.idtdna.com/calc/analyzer) and NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) web-based tools in order to determine the optimal primer sequences for the amplification of the gene.

The primers were required to have: (1) good sequence diversity with minimal sequence repeats and secondary structure, (2) a length of between 18 and 25 base pairs (bp), (3) a melting temperature (Tm) of between 45°C and 65°C, (4) a GC content of between 45% and 65%, and (5) a G or C clamp at the 3' end. If all criteria could not be met, the best primer option was chosen. The amplicon sequences were also restricted to a 250-350bp length, as increased amplicon size is known to lead to decreased resolution in amplicon screening and genetic variant detection with HRM. Primers were designed to span the exonic regions of the *PLN* gene (Table 2.1). These primers were used for combined real-time PCR and HRM analysis as well as for DNA sequencing.

Table 2.1: Details of primers for the PCR, HRM and sequencing of the *PLN* gene

PRIMER	SEQUENCE	SIZE (bp)	Tm (°C)	GC CONTENT (%)	PRODUCT SIZE (bp)
F	CCA GGC TAC CTA AAA GAA GAC	21	52.6	47.6	332
R	TTC CTG TCT GCA TGG GAT G	19	54.8	52.6	

2.5.4.2. HRM Protocol

Probands with ARVC, DCM, PPCM and HCM were screened for disease-associated genetic variants by comparing sample HRM profiles to those of control samples known to be negative for these diseases. The reagents used for the combined real-time PCR and HRM of the *PLN* gene are shown in Table 2.2 and the reaction conditions are shown in Table 2.3.

Table 2.2: Reagents for the PCR and HRM of the *PLN* gene

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	0.8 µM / 0.5 µl
Reverse Primer (20 µM)	0.8 µM / 0.5 µl
dNTPs (20 µM) (Bioline)	0.8 µM / 1 µl
GoTaq Polymerase (5 U/ ul) (Promega)	0.5 U / 0.1 µl
GoTaq FlexiBuffer (5X) (Promega)	1 X / 5 µl
$MgCl_2$ (25 mM) (Promega)	1.5 mM / 3 µl
EvaGreen dye (Anatech)	1X / 1 µl
DNA	200 ng
FINAL REACTION VOLUME	25 µl

Table 2.3: Optimised temperature cycling conditions for the PCR and HRM of the *PLN* gene

CONDITION	TEMPERATURE (TIME)	
Initial denaturation	95°C – 10 seconds	
Denaturation Primer Annealing Template Elongation	95°C – 5 seconds 55°C – 10 seconds 72°C – 10 seconds	50 cycles
High Resolution Melt	72 – 95°C (0.1°C increments)	

2.5.5. Sequence Analysis

DNA sequencing was performed on all samples that showed a change relative to the controls with HRM analysis.

2.5.5.1. HRM product purification

After analysis by HRM, the samples marked for sequencing were purified using the *Shrimp Alkaline Phosphatase* and *Exonuclease I* enzymes. The reagents used are shown in Table 2.4 and the reaction conditions are shown in Table 2.5.

Table 2.4: Reagents for the purification of the *PLN* HRM Products

REAGENT	QUANTITY/VOLUME PER REACTION
Exonuclease I (New England Biolabs)	1 U / 0.1 µl
Shrimp Alkaline Phosphatase (Promega)	2 U / 1 µl
HRM Product	5 µl
FINAL REACTION VOLUME	20 µl

Table 2.5: Optimised temperature cycling conditions for the purification of the *PLN* HRM Products

CONDITION	TEMPERATURE (TIME)
Incubation	37°C – 1 hour
Deactivation	75°C – 15 min

Sequencing was performed with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The reagents used for the sequencing reactions are described in Table 2.6, and the temperature cycling conditions used are described in Table 2.7.

Table 2.6: Reagents for the sequencing of the *PLN* HRM Products

REAGENT	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	2 mM / 2 µl
HRM product	3 µl
BigDye ® Terminator v3.1. Ready Reaction Mix (Applied Biosystems)	2 µl
5 X Sequencing Buffer (Applied Biosystems)	1 X
FINAL REACTION VOLUME	20 µl

Table 2.7: Optimised temperature cycling conditions for the sequencing of the *PLN* HRM Products

CONDITION	TEMPERATURE (TIME)	
Initial denaturation	96°C – 5 minutes	
Denaturation	96°C – 30 seconds	25 cycles
Primer Annealing	50°C – 15 seconds	
Template Elongation	60°C – 4 minutes	

2.5.5.2. Capillary Electrophoresis

The sequencing products were analysed using capillary electrophoresis with the ABI PRISM® 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA) at the Stellenbosch DNA Sequencing Unit (Department of Genetics, Stellenbosch University, Cape Town).

2.5.5.3. Sequencing Alignment and Analysis

The DNA sequences generated in these experiments were aligned with the wild type sequence obtained from the Ensembl database (http://www.ensembl.org/) and analysed using the BioEdit Sequence Alignment Editor (Ibis Biosciences, Carlsbad, CA, USA).

2.5.6. Population Screening

For variants of interest, 200 ethnically matched control chromosomes were screened by HRM analysis and Sanger sequencing to determine the prevalence of these variants in the general population.

2.5.7. Bioinformatic Analysis

The novelty of detected variants was determined by searching the 1000 Genomes (http://browser.1000genomes.org/index.html), National Heart, Lung and Blood Institute Exome Variant Server Exome Sequencing Project (NHLBI-EVS ESP) (http://evs.gs.washington.edu/EVS/), NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) databases, as well as literature review. Bioinformatic analysis of the *PLN* variants of interest was conducted using various bioinformatic tools to assess the impact of the variants on *PLN* structure and function at the messenger ribonucleic acid (mRNA) and protein levels.

2.5.7.1. Bioinformatic Analysis of the Effect of Nucleic Acid Substitutions on Exon/Intron Splice Sites

In order to determine whether novel sequence variants in *PLN* caused a change in mRNA splicing, the sequences were analysed using the Berkeley Drosophila Genome Project NNSPLICE 0.9 (http://www.fruitfly.org/seq_tools/splice.html)[69], SplicePort (http://spliceport.cbcb.umd.edu/)[70], Human Splicing Finder (http://www.umd.be/HSF/)[71] and NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/)[72] web-based analysis tools.

2.5.7.2. Bioinformatic Analysis of the Effect of Nucleic Acid Substitutions on Exonic Splice Enhancer (ESE) Sites

In order to determine whether novel sequence changes could disrupt existing ESEs, sequences were analysed using the ESEFinder web-based tool (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home).[73]

2.5.7.3. Bioinformatic Analysis of the Effect of Nucleic Acid Substitutions on the Secondary Structure of RNA

Control and mutant sequences were analysed using the Mfold (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form)[74] and RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)[75] web-based tools to predict RNA folding.

2.5.7.4. Bioinformatic Analysis of the Effect of Amino Acid Substitutions on Protein Structure and Function

The Align-GVGD (http://agvgd.iarc.fr/agvgd_input.php)[76], PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/)[77] and SIFT (Sorting Intolerant From Tolerant) (http://sift.jcvi.org/www/SIFT_enst_submit.html)[78] bioinformatic tools were used to predict whether non-synonymous variants had deleterious effects on the PLN protein. The MutationTaster tool (http://www.mutationtaster.org/)[79] was used to predict whether variants would have an overall effect on protein function.

2.5.8. Haplotype analysis

DNA samples for microsatellite analysis of the family described by Schmitt *et al*.[61] were obtained from Professor Christine Seidman (Department of Genetics, Harvard Medical School, Boston, MA 02115, USA). Microsatellite markers were used to construct haplotypes of the *PLN* gene. The markers included were D6S454[54], PLN-200K[54], PLN+200K[54] and D6S412.[54] The primers for the PCR amplification of these markers are listed in Table 2.8, the reagents are listed in Tables 2.9 (D6S454, PLN+200K and D6S412) and 2.10 (PLN-200K) and the reaction conditions in Table 2.11.

PCR products were run on a 2% agarose gel (Appendix 2.3) for verification of reaction success and product size and specificity. These products were then analysed using capillary electrophoresis with the ABI PRISM® 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA) at the Division of Human Genetics, University of Cape Town. Haplotypes were constructed using Cyrillic v2.0 (Cyrillic Software, United Kingdom).

Table 2.8: Details for *PLN* microsatellite marker primers

PRIMER	SEQUENCE	TAG	POSITION RELATIVE TO *PLN* GENE (bp)	SIZE (bp)	Tm (°C)	GC CONTENT (%)	PRODUCT SIZE (bp)
D6S454_F	CAATGCTGCCCTGAGATTG	5' HEX	-3 500 000	19	54.1	52.6	136
D6S454_R	TTATTCAGAATTTCTTGCCTATAG	-		24	48.7	29.2	
PLN-200K_F	ACTCCAGCCTGGGTGACAAATCAAG	5' HEX	-200 000	25	61.1	52.0	451
PLN-200K_R	AGTCTCAGCTCCACCATTTG	-		20	54.8	50.0	
PLN+200K_F	TTGTCAGAGGGCTGTGCTTC	5' HEX	+200 000	20	57.5	50.0	332
PLN+200K_R	GTCAGTCAGGAGGATAGTTC	-		20	51.6	50.0	
D6S412_F	GCACAATTTTAGAAGGTTCACAAAG	5' HEX	+1 650 000	25	53.3	36.0	208
D6S412_R	GCCCTACAATCACATGAGCC	-		20	56.0	55.0	

Table 2.9: Reagents for the PCR of the D6S454, PLN+200K and D6S412 *PLN* microsatellite markers

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	0.8 µM
Reverse Primer (20 µM)	0.8 µM
dNTPs (20 µM) (Bioline)	0.8 µM
GoTaq Polymerase (5U/ ul) (Promega)	0.5 U
GoTaq FlexiBuffer (5X) (Promega)	1 X
$MgCl_2$ (25 mM) (Promega)	1.5 mM
DNA	200 ng
FINAL REACTION VOLUME	25 µl

Table 2.10: Reagents for the PCR of the PLN-200K *PLN* microsatellite marker

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	0.8 µM
Reverse Primer (20 µM)	0.8 µM
dNTPs (20 µM) (Bioline)	0.8 µM
GoTaq Polymerase (5U/ ul) (Promega)	0.5 U
GoTaq FlexiBuffer (5X) (Promega)	1 X
$MgCl_2$ (25 mM) (Promega)	0.75 mM
DNA	200 ng
FINAL REACTION VOLUME	25 µl

Table 2.11: Optimised temperature cycling conditions for the PCR of the *PLN* microsatellite markers

CONDITION	TEMPERATURE (TIME)
Initial denaturation	94°C - 4 minutes
Denaturation	94°C - 30 seconds (33 cycles)
Primer Annealing	* - 45 seconds (33 cycles)
Template Elongation	72°C - 50 seconds (33 cycles)
Final Elongation	72°C - 7 minutes

*- D6S454 – 56°C, PLN-200K – 63°C, PLN+200K and D6S412 – 59°C

2.6. Results

The cohorts screened for variants in *PLN* included 85 ARVC cases, 95 DCM cases, 69 PPCM cases and 40 HCM cases. The characteristics of the ARVC, DCM, PPCM and HCM cohorts are shown in Table 2.12.

Table 2.12: Characteristics of cardiomyopathy cohorts screened for variants in *PLN*

CHARACTERISTIC	ARVC PATIENTS	DCM PATIENTS	PPCM PATIENTS	HCM PATIENTS
No. of Patients	85	95	69	40
Gender: N (%)				
Male	56 (66)	60 (63)	0 (0)	25 (62)
Female	29 (34)	32 (34)	69 (100)	15 (38)
Not recorded	0 (0)	3 (3)	0 (0)	0 (0)
Average age in years (±SD)	46±16	54±15	52±15	43±14
Ethnicity: N (%)				
Black African	6 (7)	8 (8)	47 (68)	12 (32)
Caucasian	43 (51)	13 (14)	7 (10)	12 (32)
Mixed Ancestry	19 (22)	55 (58)	7 (10)	10 (26)
Indian	3 (3)	3 (3)	1 (2)	1 (2)
Not recorded	14 (17)	16 (17)	7 (10)	3(8)

Screening of *PLN* identified the previously reported c.25C>T (p.R9C)[61,62] variant in individual DCM320.1. This variant changed an amino acid that was highly conserved among *Pan troglodytes*, *Macaca mulatta*, *Felis catus*, *Mus musculus* and *Gallus gallus* (Figure 2.4). Population screens revealed this variant to be absent in 200 ethnically matched control chromosomes, which is in agreement with what was shown by Schmitt and colleagues, who found this variant to be absent in more than 200 control chromosomes.[61] It was not reported in the Exome Variant Server or ExAC databases, but was reported as rs111033559 in the 1000 Genomes and dbSNP databases.

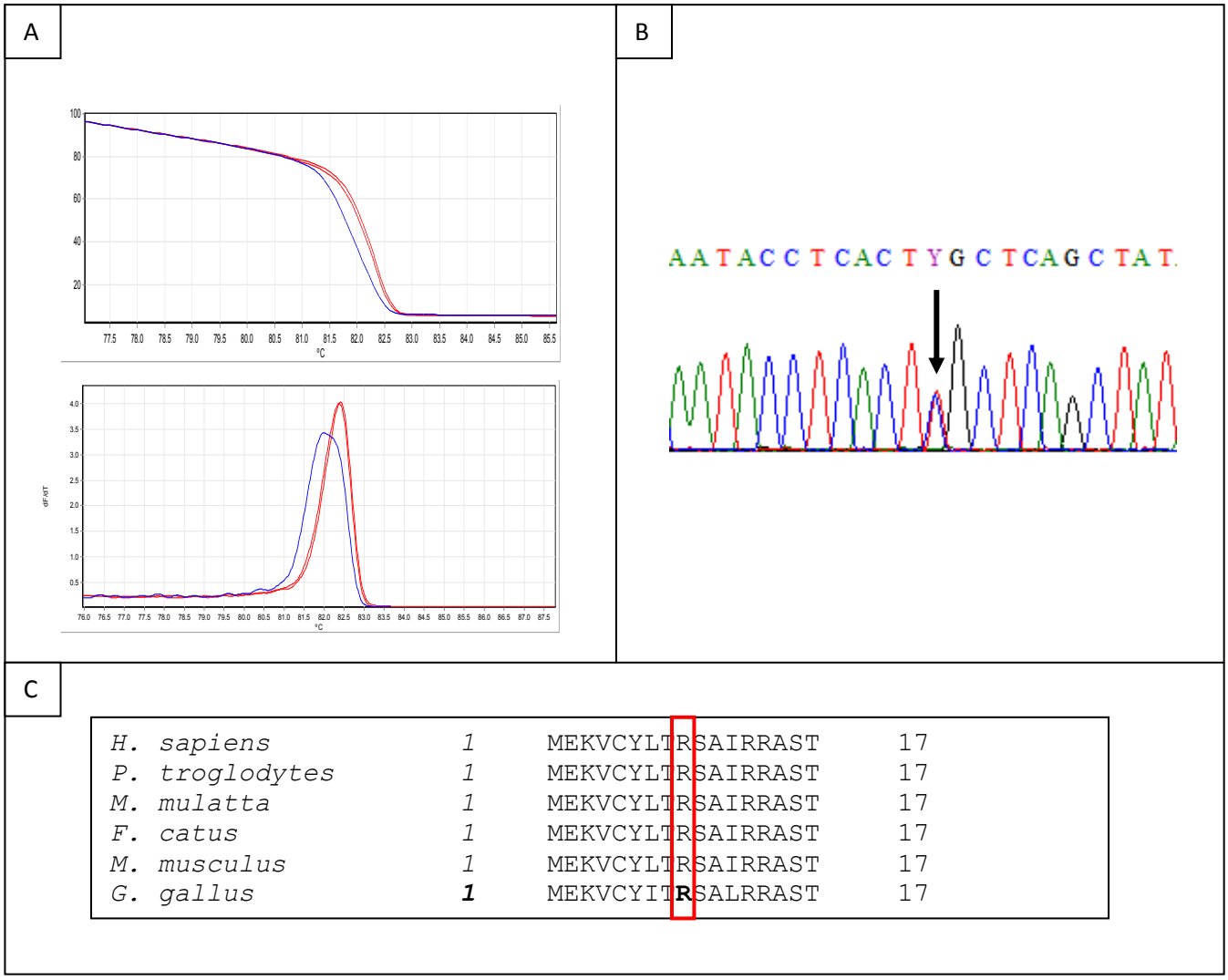


Figure 2.4: Identification of the c.25C>T (p.R9C) variant. (a) HRM and derivative graphs showing control samples (red) and variant sample (blue) amplicon melting (Y-axis for top graph – Normalised fluorescence; Y-axis for bottom graph – dF/dT); (b) Electropherogram showing the c.25C>T sequence change; (c) Multiple species protein alignment of this sequence

Detailed investigation into the family history of this proband pointed to an autosomal dominant inheritance pattern of severe DCM in this family (Figure 2.5). The proband (DCM320.1; II:2) presented with DCM after pregnancy and had a heart transplant at the age of 35 years while the proband's son (DCM320.3; III:1) was similarly affected with DCM and had also undergone a heart transplant at the age of 25 years. The proband's younger sister (II:3) likewise presented with DCM after pregnancy and required a heart transplant at the

age of 39 years. Although the proband's daughter (DCM320.4; III:3) showed early signs of DCM on echocardiography, the diagnosis was not confirmed. Further enquiry also revealed that the proband's mother had died of a heart condition at the age of 36 years. However, we were not able to confirm whether this individual was affected with DCM.

Subsequent variant screening of *PLN* for the available members of this family found the c.25C>T (p.R9C) variant in the proband's affected son (DCM320.3; Individual III:1) and daughter (DCM320.4; Individual III:3) as well as the proband's affected sister (Individual II:3), reflecting segregation of the c.25C>T variant with DCM in this family (Figure 2.5). No other variants in *PLN* were detected.

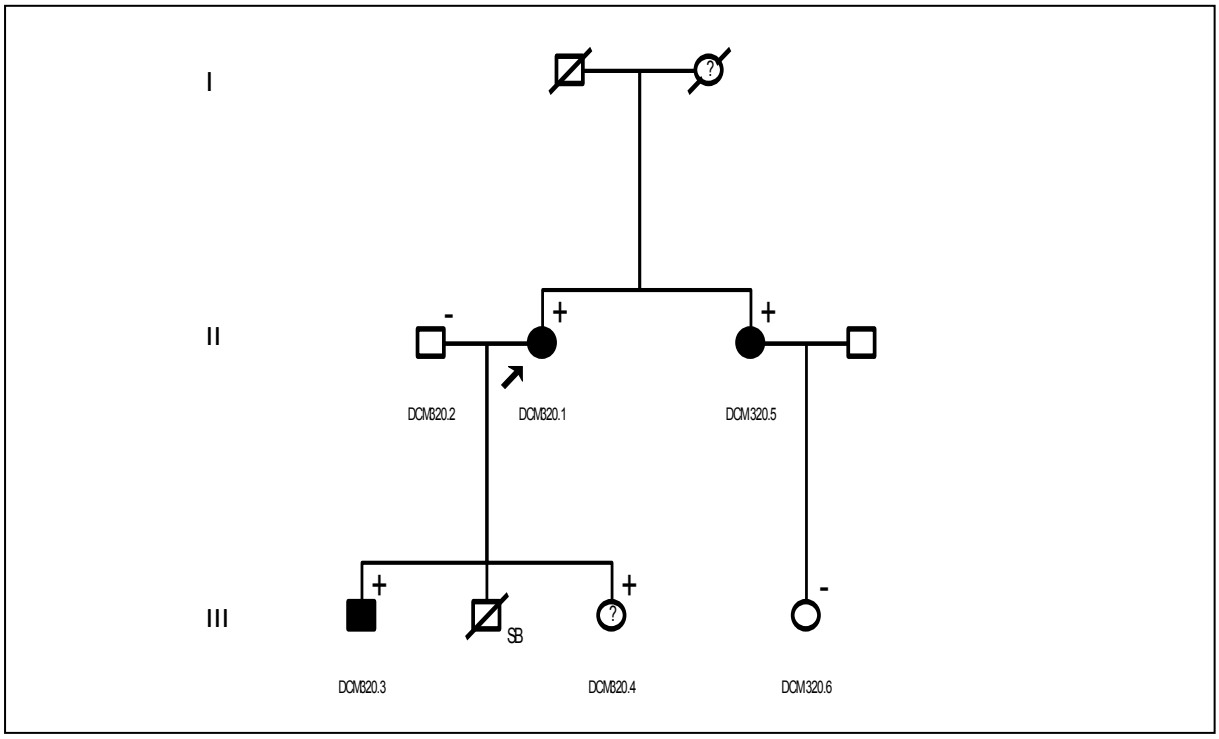


Figure 2.5: Pedigree of family DCM320 showing the c.25C>T variant and individuals with DCM. SB: Stillbirth; $\square^{+}$: c.25C>T positive; $\square^{-}$: c.25C>T negative.

Bioinformatic analysis revealed that this variant was predicted to be disease-causing by the MutationTaster, PolyPhen-2 and Align GVGD bioinformatic tools. This variant was also predicted to cause a change in mRNA secondary structure by the Mfold and RNAfold

bioinformatic tools (Figure 2.6). The ESEfinder tool predicted that this variant would alter SRSF1 and SRSF5 exonic splice enhancer recognition sites (Figure 2.7).

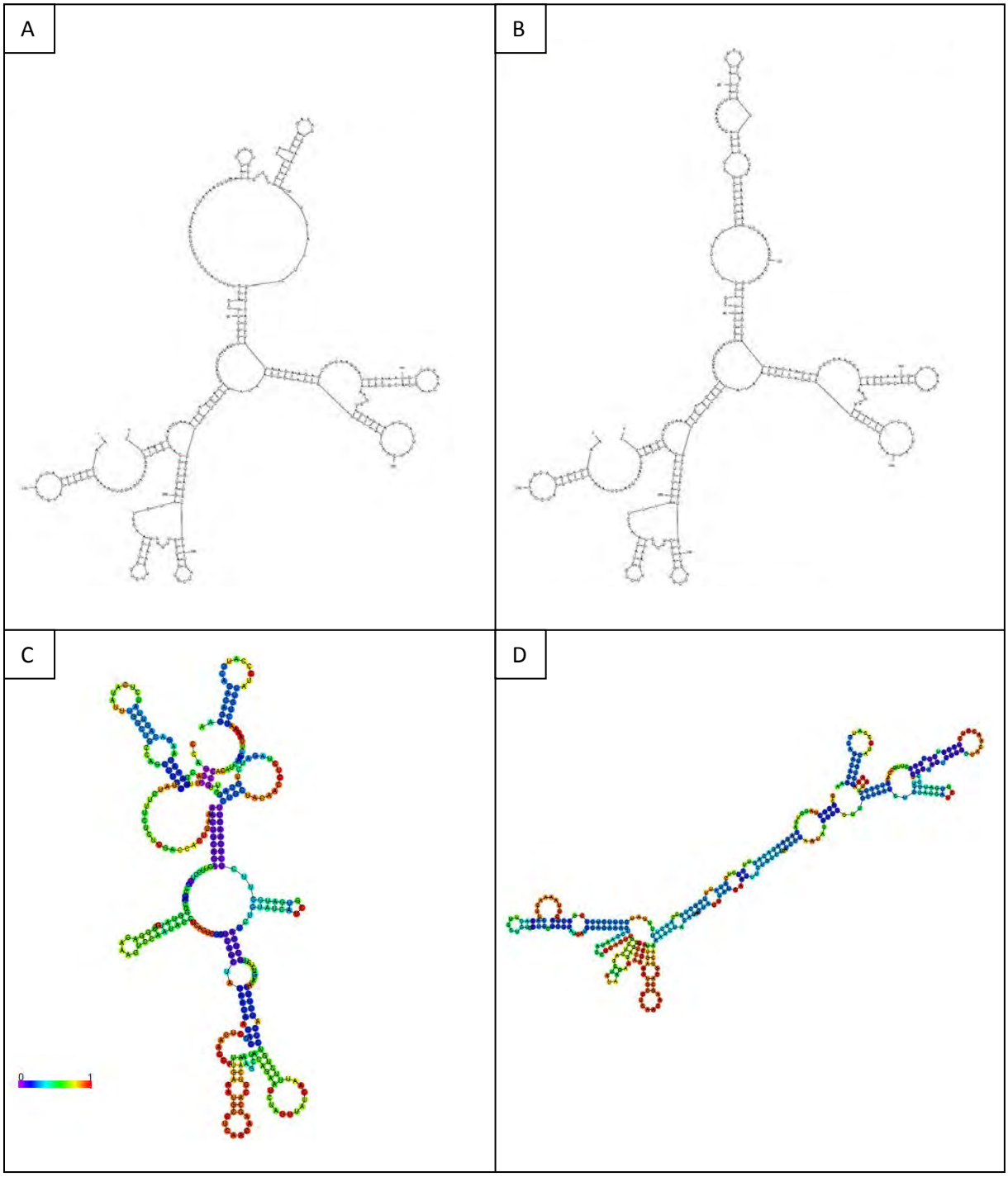


Figure 2.6: MFold and RNAfold predictions for wild type and c.25C>T *PLN*. (a) MFold wild type *PLN* predicted structure; (b) MFold c.25C>T *PLN* predicted structure; (c) RNAfold wild type *PLN* predicted structure; (d) RNAfold c.25C>T *PLN* predicted structure.

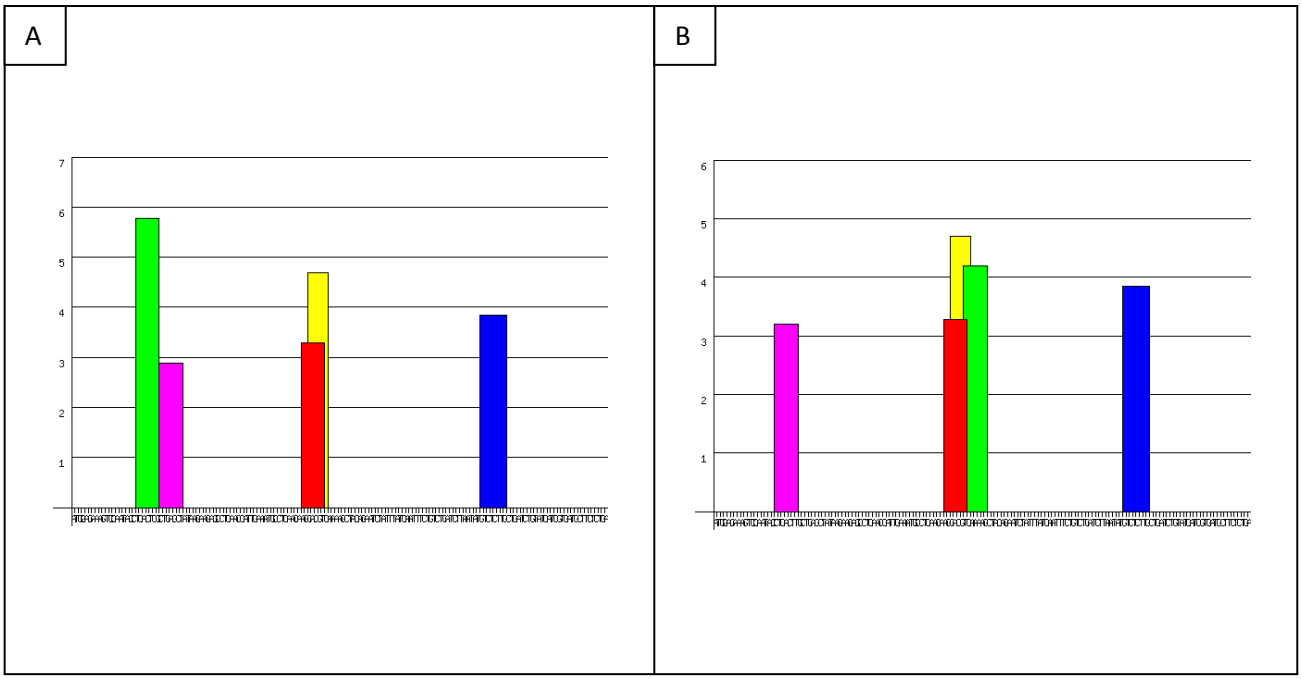


Figure 2.7: ESEfinder results for wild type and c.25C>T *PLN*. (a) ESEfinder predictions for wild type *PLN*. (b) ESEfinder predictions for c.25C>T *PLN* (SRSF – Serine/arginine-rich splicing factors; IgM – Immunoglobulin M; BRCA1 – Breast Cancer 1, Early Onset).

As the *PLN* c.25C>T variant was initially reported in a single family of European descent[61], it was important to determine if these variants arose independently or if they were inherited from a common ancestor, possibly as early as the 17th century when the European colonisation of South Africa began. Haplotypes were constructed using the microsatellite markers mentioned above (D6S454, PLN-200K, PLN+200K and D6S412)[54] spanning 5.15Mb across the *PLN* gene. In the DCM320 family, the *PLN* c.25C>T variant was found to be part of a disease-specific haplotype spanning this 5.15Mb region (Figure 2.8). The *PLN* c.25C>T variant was found to be part of a different haplotype in the family described by Schmitt *et al*. (Figure 2.8).[61] We were unable to determine the possibility of common ancestry between the South African family and the Polish variant carrier, as this individual was only reported recently.[62]

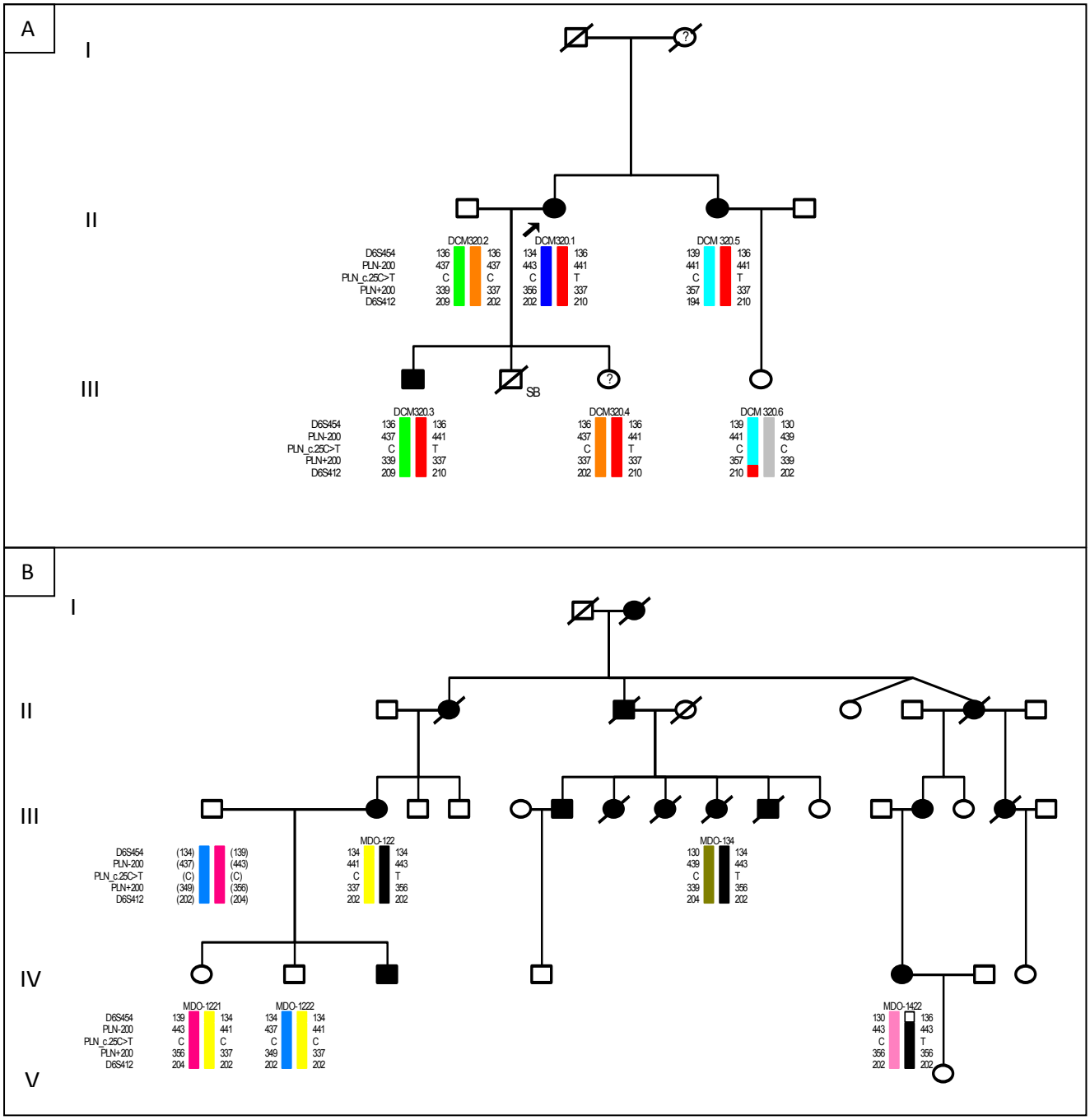


Figure 2.8: Haplotype analysis. (A). Pedigree of the DCM320 family. (B). Pedigree of the MDO DCM family.

In the original report of this *PLN* c.25C>T (p.R9C) variant by Schmitt and colleagues[61], it was reported that this variant was associated with an autosomal dominant inheritance pattern of DCM which was characterised by increased cardiac chamber dimensions, decreased contractile function at 20-30 years of age and progression to heart failure within 5-10 years after symptom onset, a phenotype similar to that observed in the DCM320 family. The

individual harbouring this variant who was reported by Truszkowska and colleagues exhibited onset of severe DCM at the age of 21 and required a heart transplant at the age of 22.[62] There is consequently strong evidence that the *PLN* c.25C>T (p.R9C) variant is disease-causing in the DCM320 family, based on segregation with disease in this family and previous reports of segregation with DCM by Schmitt[61] and Truszkowska.[62] We did not detect any other variants in *PLN*.

2.7. Discussion

We screened patients with ARVC, DCM, HCM or PPCM for variants in *PLN* and detected the previously reported c.25C>T variant in individual DCM320.1. We also detected this variant in three of this individual's family members, two of whom were severely affected with DCM and one who showed early signs of the disease, indicating segregation of this variant with autosomal dominant DCM in this family. This variant was previously reported as disease-causing by Schmitt and colleagues.[61] They found this variant to be associated with autosomal dominant DCM involving decreased contractile function at 20-30 years of age, increased chamber dimensions and progression to heart failure within 5-10 years following symptom onset, a phenotype similar to that found in the DCM320 family. It was also reported in an individual with severe DCM by Truszkowska and colleagues.[62]

A number of PPCM cases are in reality cases of familial DCM occurring around the time of pregnancy. Patients with PPCM often make a full recovery after pregnancy, but others do not.[80] Those who fail to recover after pregnancy are thought to have a genetic cause of disease, and that the phenotype they are displaying is DCM manifesting around the time of pregnancy due to the increased load placed on the heart and the rise of oxidative stress that occurs during this period.[80] It has been shown that oxidative stress worsens the cardiotoxic effect of the *PLN* p.R9C variant,[65] and the manifestation of disease after pregnancy in the DCM320 family appears to be a clear example of this effect.

A transgenic mouse model harbouring the p.R9C variant is characterised by biventricular cardiac dilation which started at the age of 4 months, increase in heart chamber size, decrease in left ventricular contractility and symptoms of terminal heart failure.[61] Histology

revealed myocyte enlargement.[61] The examination of these mice showed that the p.R9C variant is associated with the development of DCM.

An *in vitro* study of the mechanism of disease for this variant revealed that mutant PLN prevented phosphorylation of the wild type PLN by trapping protein kinase A and thus inhibiting phosphorylation of wild type PLN.[61] Mutant cells (HEK293) displayed depressed calcium kinetics and delayed myocyte relaxation.[61] The authors therefore concluded that p.R9C leads to the development of DCM by alteration of myocyte calcium kinetics and therefore muscle contraction and relaxation, causing the development of DCM. A study by Ha and colleagues showed that the p.R9C variant increases the stability of the PLN pentamer, preventing it from binding to Ca^{2+}-ATPase and phosphorylation by protein kinase A.[65] They also showed that these effects are enhanced under conditions of oxidative stress, suggesting that these conditions may worsen the cardiotoxic effects of the PLN p.R9C variant.[65] A study by Yu and Lorigan demonstrated that cytoplasmic domain Ia (amino acids 1-16) of phosphorylated PLN and p.R9C-PLN adopted a similar conformation, shifting away from the membrane surface, that may relate to relief of SERCA2a inhibition.[81] There is therefore strong functional evidence that the *PLN* c.25C>T (p.R9C) variant causes the development of severe autosomal dominant DCM.

Although a total of 289 South African cardiomyopathy patients (ARVC, DCM, PPCM and HCM) were screened in this study, only one disease-causing variant was found in *PLN* (frequency = 0.2%) illustrating that *PLN* is not a common cause of cardiomyopathy in South Africans. This is the first time that *PLN* has been evaluated as a potential cause of disease in a PPCM cohort. An earlier report by van Spaendonck-Zwarts and colleagues had examined *PLN* in a single family with PPCM but no disease-causing variants were found.[82] To date, there are three known *PLN* variants (c.25C>T (p.R9C)[61,62], c.40_42delAGA (p.R14del)[28,60] and c.116T>G (p.L39X)[59]) associated with DCM[28], one variant associated with ACM (c.40_42delAGA (p.R14del)[28]) and two variants associated with HCM (c.1-77A>G[64] and c.1-42C>G[63]) (Table 2.13). The majority of these variants have been associated with severe cardiomyopathy. Recently, a duplication of the entire *PLN* gene has been reported in a

paediatric case of dilated cardiomyopathy.[83] This is an indication of the important function of this gene, as variations in this gene do not appear to be well-tolerated.

Table 2.13: Known *PLN* variants associated with cardiomyopathy

Variant	Disease	No. of probands carrying variant	Frequency in control chromosomes	Reference
c.25C>T (p.R9C)	DCM	1	0/200	[61,62]
c.40_42delAGA (p.R14del)	DCM, ARVC	39 DCM, 13 ACM[28] 1 DCM[60]	1/946[28] 0/582[60]	[28,60]
c.116T>G (p.L39X)	DCM	1	0/500	[59]
c.1-77A>G	HCM	1	0/592	[64]
c.1-42C>G	HCM	1	0/100	[63]

We have established that the *PLN* c.25C>T variant forms part of a disease-specific haplotype spanning 5.15Mb across *PLN* in the DCM320 family that differed from the haplotype observed in the family described by Schmitt *et al.*[61]. It is therefore unlikely that this variant was inherited from a common ancestor.

Evidence obtained from family segregation analysis (DCM320), bioinformatics and previous studies that have focused on the p.R9C variant allows us to conclude that this variant is causative of DCM in family DCM320.

2.8. Conclusion

We detected the known c.25C>T *PLN* variant which leads to a severe form of DCM. We have also shown that *PLN* variants are a rare cause of cardiomyopathy in South Africa. No deleterious variants in *PLN* were detected in the ACM2 family or other cases of ARVC, HCM or PPCM.

CHAPTER 3
COPY NUMBER VARIATION IN THE ACM2 FAMILY

3.1. Introduction

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is known to be caused by various genetic factors, but ~50% of ARVC cases have no identified genetic cause to date.[84] As described in chapter 2, the design of this study focused on finding the disease-causing variant/s in one such family (ACM2) with autosomal dominant ARVC. Extensive genetic analysis was previously performed using techniques such as whole genome single nucleotide polymorphism (SNP) linkage analysis, but no disease-causing variants were identified.[44,55]

The ACM2 family was originally linked to the ARVC6 locus on chromosome 10 based on borderline statistical evidence of linkage (LOD score 2.93) (Fig. 3.1).[44] However, an exhaustive search for a causal gene in the linkage region on chromosome 10 was unsuccessful.[55] Furthermore, whole genome SNP linkage analysis identified new linkage regions on chromosomes 1, 4 and 12, in addition to chromosome 10,[55] but candidate gene analysis in the these regions was negative. As discussed in chapter 2, candidate gene analysis of the *PLN* gene was also negative.

Recent research has shown that studying smaller variants alone is not sufficient to capture the full extent of meaningful genomic variation[85] and larger genomic changes such as copy number variations (CNVs) may assist in the identification of causal variants. CNVs are duplications or deletions of DNA segments 1kb or larger in size and present in variable copy numbers in comparison to reference human genomes.[86] A CNV may have a simple structure, such as tandem duplication, or may have a more complex involvement of gains or losses of homologous sequences at multiple sites in the genome.[86] Thousands of CNVs have been described in the human genome to date, indicating that they contribute to genomic variation as much as single nucleotide polymorphisms (SNPs), short tandem repeats (STRs) and other small genetic changes.[87] CNVs can encompass millions of DNA bases containing entire genes and their regulatory regions.[87] Structural variations in certain genomic regions

may be benign and have no obvious phenotypic consequence or may be a contributing factor to disease development.[87]

CNVs contribute to disease development through a variety of mechanisms. One of the most common mechanisms is alteration of the copy number of a gene (or genes) or regulatory region sensitive to dosage effect.[88] CNVs can also cause a position-dependent effect in the event of a translocation, even if the breakpoints are as far away as 1 megabase (Mb) from the causative gene.[88] Another mechanism of disease for CNVs is the unmasking of recessive variants or functional polymorphisms of the remaining allele in the event of a deletion.[88] CNVs may have a similar effect on protein function and disease development as loss-of-function variants in the case of a deletion, and may therefore play an important role in disease pathogenesis.

CNVs have predominantly been associated with *de novo* complex disorders (e.g. autism and schizophrenia) and disease susceptibility (e.g. psoriasis, emphysema).[88] There have also been reported cases of CNVs associated with Mendelian diseases, although they account for a minority of these cases in relation to other variant classes in most disease types.[89-91] However, CNVs are the major class of variants responsible for Mendelian conditions such as Duchenne muscular dystrophy, Charcot-Marie tooth disease, α-thalassemia and Prader-Willi syndrome.[88] There are limited reports of involvement of CNVs in cardiovascular disease to date, with the majority of these reports focused on congenital heart diseases. Many reported CNVs overlap genes implicated in cardiovascular disease, such as *RyR2* (ARVC), *TTN* (DCM) and *KCNQ1* (Long QT syndrome).[85] Losses in copy number in the desmosomal genes *PKP2* and *DSP*, which are most commonly associated with ARVC, have been reported to cause this disease.[92-94] Bearing this in mind, we elected to perform genome-wide CNV analysis to determine if there were any insertions, deletions or chromosomal rearrangements within the DNA segments of the ACM2 family that could account for the ARVC phenotype. Previous studies have identified CNVs in the desmosomal genes by multiplex ligation-dependent probe amplification (MLPA) or SNP genotyping arrays[92-94], but this has a limited capacity for detection of CNVs as compared to the newer CytoScan® HD Array (Affymetrix) technology.

The CytoScan® HD Array is reported to have the widest genomic coverage and best performance for the detection of human chromosomal aberrations.[95] It can reliably detect 25-50kb copy number variations throughout the genome at high specificity and >99% sensitivity.[95] This array has >2.6 million copy number markers including 750 000 SNPs to enable accurate detection of changes in genomic copy number for accurate detection of breakpoint estimation, loss of heterozygosity (including copy-neutral loss of heterozygosity due to segment uniparental disomy), regions identical-by-descent, maternal contamination and low-level mosaicism.[95] The simultaneous use of both probe sets allows validation of the results obtained by either of these probe sets.

3.2. Objective

The objective of this study was to determine if CNVs play a causal role in ARVC in the ACM2 family.

3.3. Specific aim

- To screen for pathogenic CNVs in three members of the ACM2 family (2 affected individuals and 1 unaffected individual).

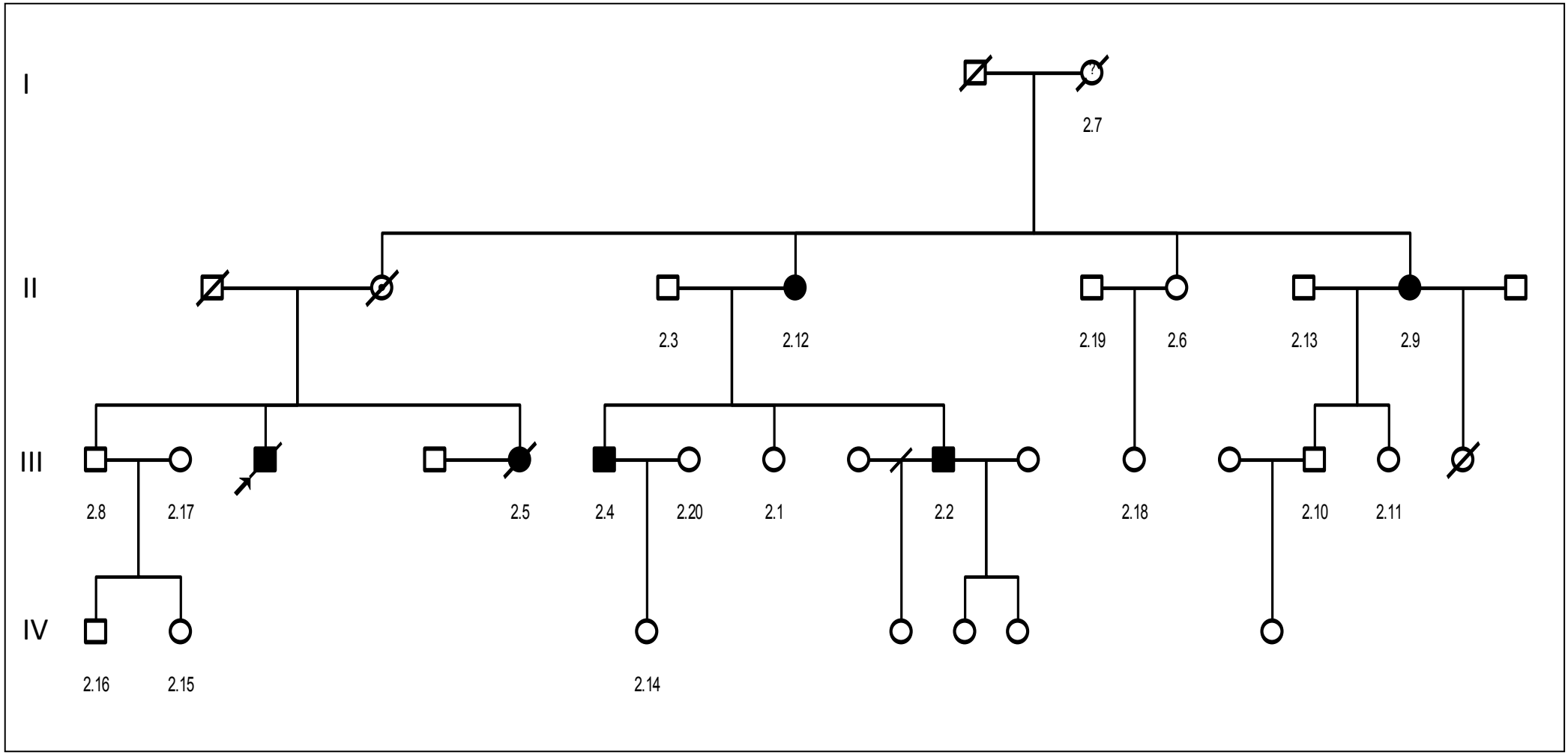


Figure 3.1: Pedigree of the ACM2 family with ARVC. Shaded squares indicate affected males and shaded circles indicate affected females

3.4. Methods

Individuals 2.4, 2.5 and 2.11 of the ACM2 family were screened for CNVs with the CytoScan® HD Array (Affymetrix). ACM2.4 and 2.5 are affected patients with ARVC while individual 2.11 was unaffected, and was therefore used for exclusion of CNVs unlikely to play a role in the development of ARVC in the ACM2 family.

3.4.1. ACM2 Family Evaluation

ACM2 family members were evaluated as described in section 2.5.1.1. and classified as having definite, borderline or possible ARVC or being unaffected with this disease.

3.4.2. DNA Isolation and Storage

Sample DNA was isolated from blood samples and stored as described in section 2.5.2.

3.4.3. Sample Quality Control – Concentration and Quality

DNA sample concentration was determined using a PicoGreen® quantitation assay (Invitrogen). All DNA samples had the required concentration of >50ng/µl.

3.4.4. Sample Quality Control – DNA Integrity

All DNA samples were run on a 2% agarose gel to determine molecular weight and integrity. All samples were of sufficient molecular weight and free of degradation.

3.4.5. Genomic DNA Preparation

All samples were normalised to a concentration of 50ng/µl using Low EDTA TE buffer and 250ng (5µl) of each sample was used in the CytoScan® assay. The Genomic DNA Control supplied in the CytoScan® Reagent Kit served as a positive control (concentration = 50ng/µl) to monitor assay performance, while low EDTA TE buffer provided with the CytoScan® Reagent Kit was used as a negative control.

3.4.6. Target Preparation

3.4.6.1. Amplification

Each sample was processed in accordance with the Affymetrix CytoScan® protocol. Samples were digested, ligated and subsequently amplified by PCR. Aliquots of the PCR product samples were run on a 2% gel to determine the success of the PCR as well as whether the PCR product generated was of the correct size (150-2000 bp)**.** The PCR products were then purified and quantified**.** The required PCR product concentration is >300ng/µl for hybridisation to the array, and all samples met these quality control (QC) criteria.

3.4.6.2. Fragmentation and labelling

All samples were fragmented in accordance with the modified protocol provided by Affymetrix. An aliquot of each sample was run on a 4% gel to determine successful fragmentation of PCR products (distribution of fragmented samples should be 25-125 bp). All samples were successfully fragmented. The fragmented samples were subsequently labelled with a proprietary labelling reagent.

3.4.7. Hybridisation

Proprietary hybridisation buffers and an oligonucleotide control reagent were added to each sample to create a hybridisation cocktail. 200 µl of sample hybridization cocktail was injected into each array and the arrays were hybridised in an Affymetrix Hybridization Oven 645 set at 50°C, 60 rpm for 17 hours.

3.4.8. Washing, staining and scanning

At the end of hybridisation, the arrays were washed and stained using the GeneChip® Fluidics Station 450 (Affymetrix) and scanned using the GeneChip® Scanner 3000 7G (Affymetrix).

3.4.9. Analysis

Following scanning, the CEL files generated by the Affymetrix Command Console® were analysed with the Affymetrix® Chromosome Analysis Suite (ChAS) (version 2.1.0.16). This software includes a quality control (QC) metric which identifies sample data that meet the required quality gates. The default QC threshold was used where:

- SNP QC > 15
- Median absolute pairwise difference (MAPD) <0.25
- Waviness standard deviation (SD) <0.12

Analysis was based on the reference genome sequence of the University of California, Santa Cruz (UCSC) Genome Browser hg19, February 2009 (GRCh37/hg19). Copy number variant analysis was performed using the ChAS software default filter settings which allowed detection of CNVs ≥400kb in size (marker count ≥ 50). Subsequent CNV analysis was conducted using the high resolution ChAS setting, which allowed detection of variants ≥100kb in size (marker count ≥ 50). Final custom analysis settings allowed detection of all copy number variants ≥1kb in size (marker count ≥ 50).

Variants were filtered according to presence in the Database of Genomic Variants[96] (http://dgv.tcag.ca/dgv/app/home), UCSC Genome Browser[97] (https://genome.ucsc.edu/), DECIPHER database[98] (https://decipher.sanger.ac.uk/) and the International Standards for Cytogenomic Arrays (ISCA) Consortium database (https://www.clinicalgenome.org/).

3.5. Results

The DNA samples of the three members (ACM2.4, ACM2.5 and ACM2.11) of the ACM2 family passed all quality control measures.

3.5.1. Analysis of CNV data using standard settings (CNVs ≥400kb)

Analysis using the standard ChAS settings identified no variants of interest (common between ACM2.4 and 2.5) (Figure 3.2).

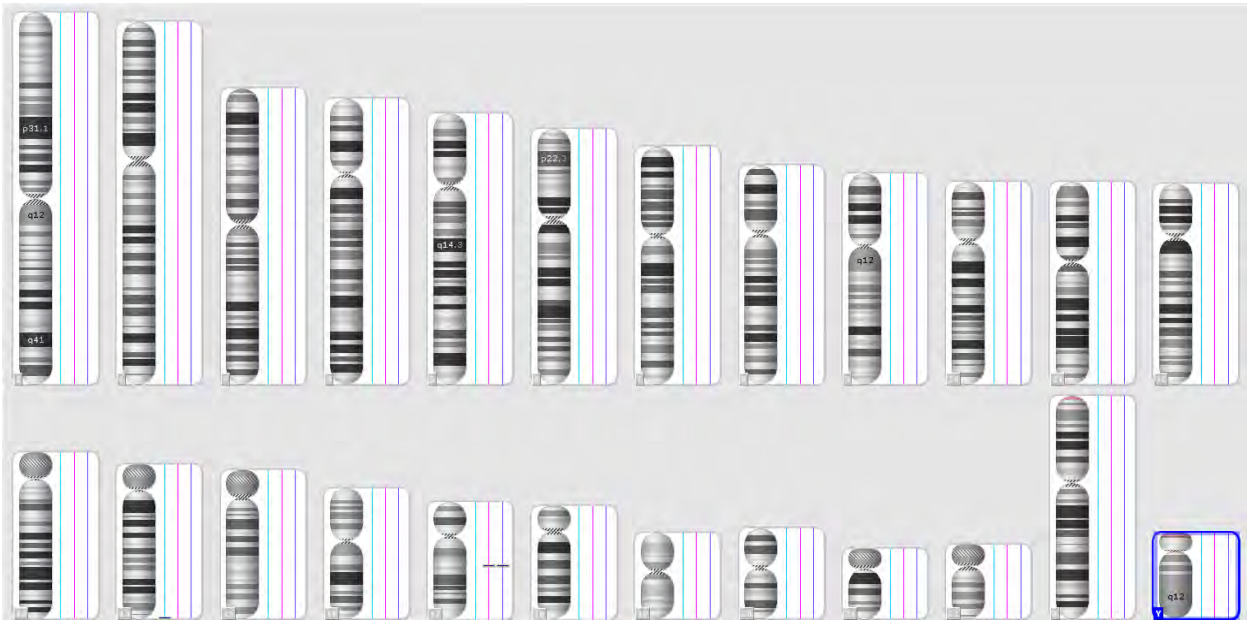


Figure 3.2: Genome-wide occurrence of regions with changes in copy number (≥400kb) in samples ACM2.4, 2.5 and 2.11. Chromosomes are listed from 1-12 (row 1) and 13-22, X and Y (row 2). Vertical lines (light blue, pink and purple) indicate characteristics of the different samples (green – 2.5, red – 2.4; purple – 2.11). Horizontal lines indicate regions showing losses (blue) and gains (red) in copy number.

3.5.2. Analysis of CNV data using high resolution settings (CNVs ≥100kb)

Analysis using this high resolution ChAS setting identified the ~201 kb CNV on chromosome 14 (position 106,329,183 - 106,530,460), which was present in 3 copies in ACM2.4 and 2.5 (Figure 3.3). This contains the *KIAA0125* (RNA-coding) and *ADAM6* (pseudogene) genes (Figure 3.4). This region was reported as highly variable in the Database of Genomic Variants[96], as well as in the local population (local unpublished data).

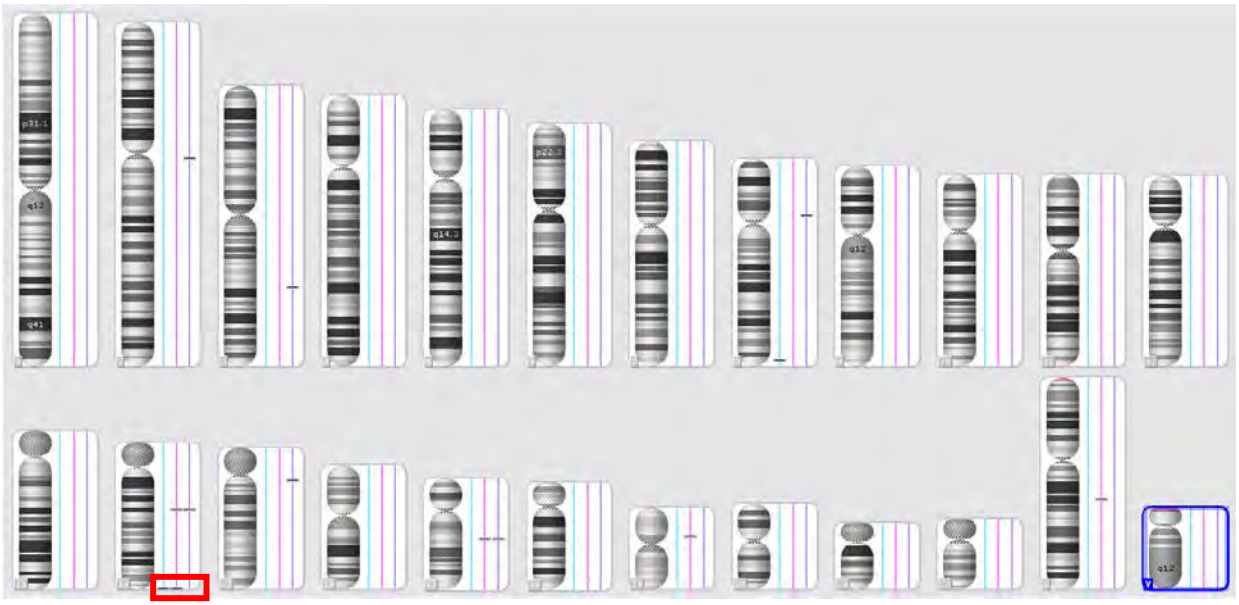


Figure 3.3: Genome-wide occurrence of regions with changes in copy number (≥100kb) in samples ACM2.4, 2.5 and 2.11. Chromosomes are listed from 1-12 (row 1) and 13-22, X and Y (row 2). Vertical lines (light blue, pink and purple) indicate characteristics of the different samples (light blue – 2.5, pink – 2.4; purple – 2.11). Horizontal lines indicate regions showing losses (blue) and gains (red) in copy number.

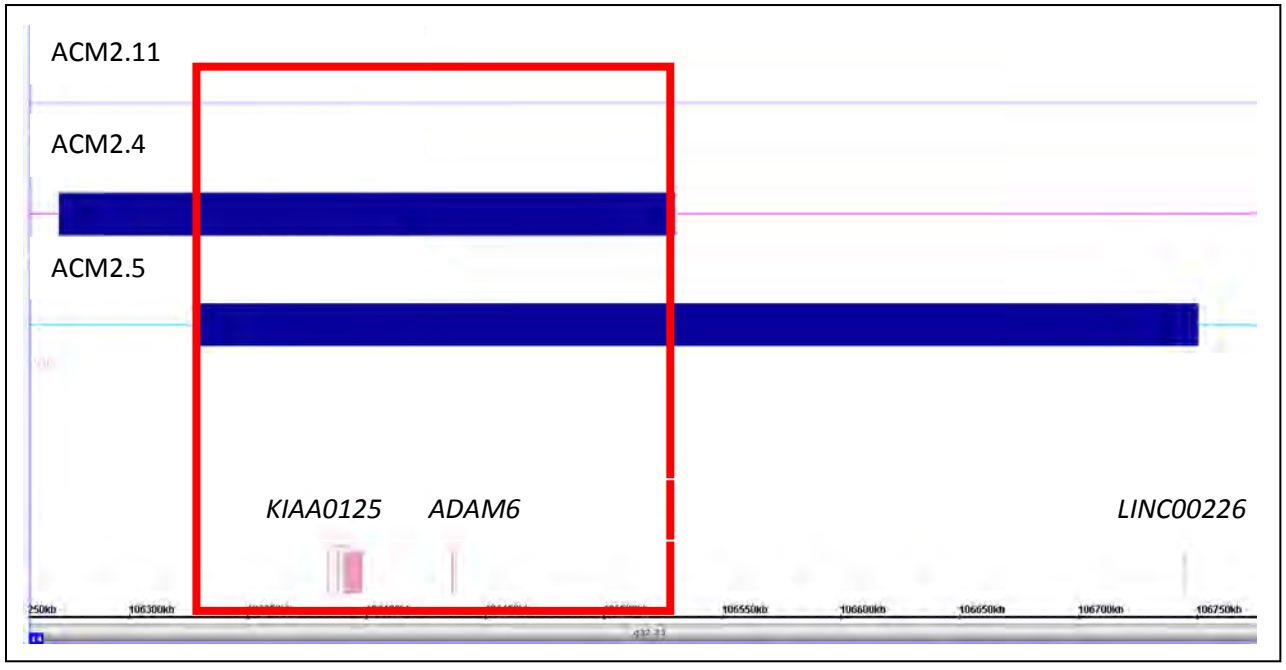


Figure 3.4: Region on chromosome 14 with CNV gain common between ACM2.4 and 2.5. Horizontal lines (light blue, pink and purple) indicate characteristics of the different samples (light blue – 2.5, pink – 2.4; purple – 2.11). Blue blocks indicate regions with a gain of copy number. Genes in the vicinity are indicated below the copy number losses.

3.5.3. Analysis of CNV data using custom high resolution settings (CNVs ≥1kb)

Analysis using the custom ChAS settings to detect CNVs ≥1kb identified one region on chromosome 17 which displayed a gain in copy number that was common to ACM2.4, 2.5 and 2.11 (Figure 3.5); this was in addition to the region already reported in section 3.5.2. This region on chromosome 17 was ~62.9kb in size (position 44,225,439 - 44,288,296) and displayed a three copy gain in CNV across the *KANSL1* and *KANSL1-AS1* genes (Figure 3.6) in all three individuals. This region has been reported in the Database of Genomic Variants as it was found in multiple control individuals.

Figure 3.5: Genome-wide occurrence of regions with changes in copy number (≥1kb) in samples ACM2.4, 2.5 and 2.11. Chromosomes are listed from 1-12 (row 1) and 13-22, X and Y (row 2). Vertical lines (light blue, pink and purple) indicate characteristics of the different samples (light blue – 2.5, pink – 2.4; purple – 2.11). Horizontal lines indicate regions showing losses (blue) and gains (red) in copy number.

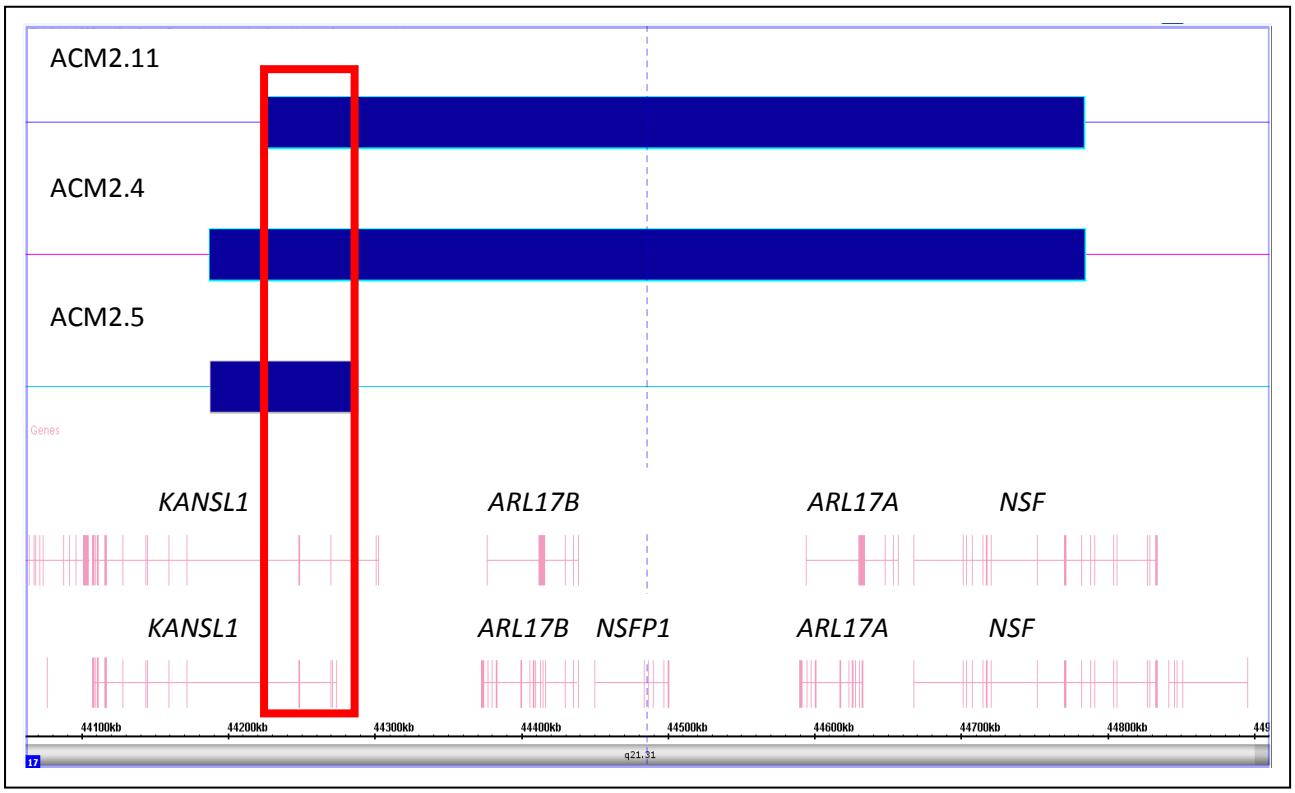


Figure 3.6: Region on chromosome 17 with CNV gain common between ACM2.4, 2.5 and 2.11. Horizontal lines (light blue, pink and purple) indicate characteristics of the different samples (light blue – 2.5, pink – 2.4; purple – 2.11). Navy blocks indicate regions with a gain of copy number. Genes in the vicinity are indicated below the copy number losses.

All variants were reported as part of larger pathogenic CNVs in the DECIPHER and ISCA databases. However, no pathogenic CNVs encompassing only regions defined by our CNV analysis were detected. CNVs affecting genes implicated in cardiomyopathy were not detected.

3.6. Discussion

Analysis of data for ACM2.4, 2.5 and 2.11 obtained with the CytoScan® HD Array by standard settings revealed no CNVs of interest ≥400kb in size.

Analysis with high resolution settings to detect CNVs ≥100kb in size identified a region on chromosome 14 that was present in 3 copies in ACM2.4 and 2.5. This region harbours the *KIAA0125* RNA-coding gene and the *ADAM6* pseudogene. However, this region was

reported in the Database of Genomic Variants as highly variable in many control individuals, and is therefore most likely a copy number polymorphism.

Analysis with custom setting to detect all CNVs ≥1kb in size identified a region on chromosome 17 that displayed a copy gain across the *KANSL1* and *KANSL1-AS1* genes (Figure 3.6) in all three individuals (ACM2.4, 2.5 and 2.11). The *KANSL1* gene has previously been implicated in 17q21.31 microdeletion syndrome (Koolen-de Vries syndrome)[99,100] which is characterised by hypotonia, highly distinctive facial features and moderate-to-severe intellectual disability.[99,100] Congenital cardiac defects have been reported in ~27% of cases, but this is not consistent with the phenotype observed in ARVC patients.[101] Due to the role of this gene in histone acetylation and transcriptional regulation, and its previous association with syndromes which share no common phenotypical characteristics with ARVC, its role in the development of ARVC in the ACM2 family is unknown. The *KANSL-AS1* gene has no reported disease association to date. This region has been reported in the Database of Genomic Variants due to its presence in multiple control individuals, and is therefore very likely a copy number polymorphism.

3.7. Conclusion

We did not detect any chromosomal rearrangements or large CNVs that would account for the ARVC within this ACM2 family. It is therefore likely that the genetic cause of disease is either a point genetic variant or a small DNA deletion or insertion (most likely less <1kb in size). With the suspicion that we were looking for a small genetic variant, we elected to perform whole exome sequencing.

CHAPTER 4
SCREENING BY EXOME SEQUENCING IN THE ACM2 FAMILY

4.1. Introduction

Arrhythmogenic right ventricular cardiomyopathy (ARVC) has been attributed to variants in various genes, but the genetic factors responsible for the remaining ~50% of ARVC cases remain to be identified. As described in chapters 2 and 3, the main objective of this study was to identify the disease-causing genetic variant in the South African family (ACM2) with autosomal dominant ARVC (Figure 4.1). Extensive genetic analysis was previously performed using techniques such as whole genome single nucleotide polymorphism (SNP) linkage analysis, but no disease-causing variants were identified.[44,55] As discussed in chapter 2, we performed candidate gene screening of the phospholamban gene *(PLN)*, a newly identified cause of ARVC, but no disease-causing variants were identified in this family. We subsequently performed genome-wide copy number variant analysis but no significant change in copy number was seen which could account for the disease development in this family (chapter 3). In previous work (PhD thesis of Mzwandile Mbele and MSc thesis of Maryam Fish) we had also screened this family for the five known desmosomal genes[22,53] but all screens were negative. We hypothesised that the disease-causing variant in this family was likely to reside in a gene that had not previously been associated with ARVC.

Due to the dominant pattern of inheritance of this disease and the fact that 85% of disease-causing variants lie within the exome,[102] the possibility existed that the variant that we were looking for would also be found in the exome. The advent of the whole exome sequencing (WES) technique in 2009 subsequently allowed us to use this approach to identify the causal variant in this family.[102]

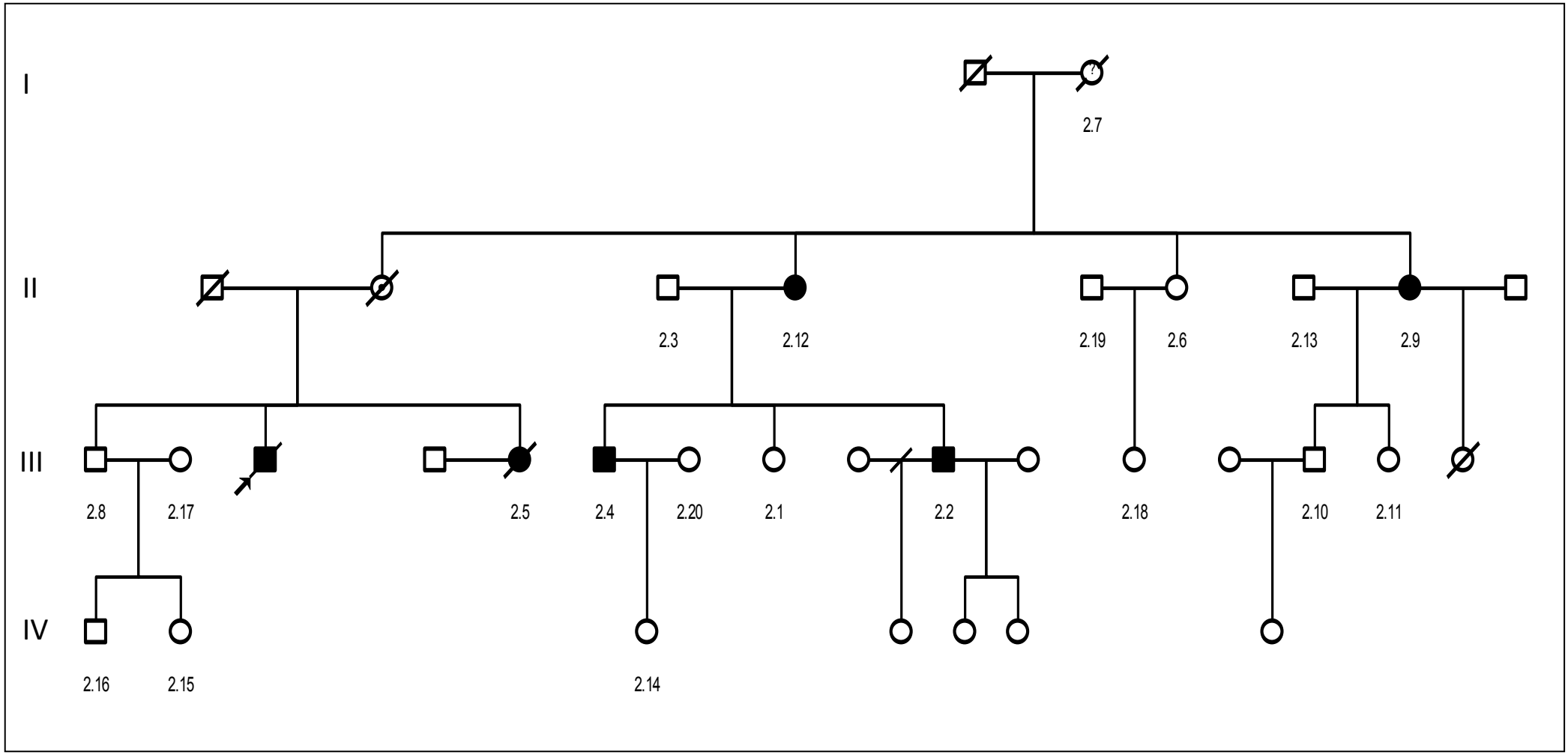


Figure 4.1: Pedigree of the ACM2 family with ARVC. Shaded squares indicate affected males and shaded circles indicate affected females

4.2. Objective

The objective of this study was to determine if a pathogenic exonic variant could be identified as the cause of ARVC in the ACM2 family.

4.3. Specific aim

- To perform whole exome sequencing of two members of the ACM2 family and screen for a disease-causing genetic variant

4.4. Methods

Exome sequencing was performed in individuals ACM2.4 and 2.5 (Fig. 4.1) using two different methods – the SureSelect Human All Exon 50Mb kit (Agilent Technologies, Inc.) and the Ion AmpliSeq™ Exome Kit (Life Technologies).

4.4.1. ACM2 Family Evaluation

ACM2 family members were investigated at the Cardiac Clinic, Groote Schuur Hospital, Cape Town, and enrolled in the ARVC registry, and diagnosed according to the 2010 diagnostic criteria of ARVC[15] as described in section 2.5.1.1.

4.4.2. Control selection

Controls were chosen from a South African population comprising anonymous blood donors. Self-reported ethnicity for these individuals was Caucasian. Informed consent was obtained from all participants (Appendix 1).

4.4.3. DNA Isolation and Storage

DNA was isolated and stored as described in section 2.5.2.

4.4.4. Exome sequencing

Two affected individuals from the ACM2 family (ACM2.4 and ACM2.5) were selected for exome sequencing using two different methods. For the first method, the SureSelect Human

All Exon 50Mb kit (Agilent Technologies, Inc.) was used and samples were sequenced with the Illumina HiSeq 2000. For the second method, the Ion AmpliSeq™ Exome Kit (Life Technologies) was used and samples were sequenced with the Ion Proton™ Sequencer (Life Technologies). Sequence analysis for this method was conducted using the Ion Reporter™ Software.

We chose to perform two exome sequencing experiments for the following reasons: (1) the use of the newer Ion AmpliSeq™ Exome Kit (kit 2) would allow for detection of any genetic variation that was out of the range of the older SureSelect Human All Exon 50Mb kit (kit 1), as the newer kit provided more comprehensive exome coverage, (2) the results obtained with kit 2 would allow validation of the results obtained with kit 1 and (3) the experiment using kit 2 was run on a new, untested next generation sequencing platform at Stellenbosch University, Cape Town. As the samples had previously been successfully sequenced with kit 1 at another facility in Germany, we felt that the sequencing of these samples with the new system would validate the accuracy and efficiency of this system here in Cape Town, and would help with the establishment of next generation sequencing capacity in South Africa.

4.4.4.1. SureSelect Human All Exon 50Mb kit

For each of the two samples, 3µg of genomic DNA was analysed using the SureSelect Human All Exon 50Mb kit, which contains the exonic sequences for ~24 000 genes and covers 50Mb of genomic sequence.

The exome sequencing process involves 3 steps: (1) DNA library preparation (section 4.4.4.1.1.), (2) Cluster generation (library amplification) (section 4.4.4.1.2.) and (3) Sequence by synthesis (sequencing of the amplified library fragments) (section 4.4.4.1.3.). This experiment was performed at the Institute of Human Genetics, Helmholtz Zentrum München, Germany, with data analysis performed with the assistance of Dr Lia Crotti and Dr Elisa Mastantuono.

4.4.4.1.1. DNA library preparation

DNA samples were prepared using the Illumina Paired End Sample Preparation Kit (Illumina) and 50Mb SureSelect Human All Exon Capture kit (Agilent Technologies) according to instructions of the Illumina Paired End Sample Preparation Guide (Illumina).

4.4.4.1.1.1. DNA quantification and fragmentation

DNA sample concentrations were measured with the Qubit® fluorometer (Life Technologies) and the Qubit® dsDNA BR Assay Kit (Life Technologies) and NanoDrop 2000 spectrophotometer (Thermo Scientific). DNA was then fragmented by the Covaris shearing method. Samples were purified with the QIAquick PCR Purification Kit (Qiagen), and samples were then concentrated to 30 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific). Concentrated DNA samples were then assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies) and the Agilent DNA 1000 Kit (Agilent Technologies).

4.4.4.1.1.2. DNA library preparation

4.4.4.1.1.2.1. End repair

After DNA fragmentation by the Covaris method, the DNA fragments have 3' and 5' overhangs and blunt ends. End repair is conducted to convert all the overhanging fragments to blunt ends. End repair was performed with the Illumina Paired End Sample Preparation kit (Illumina). Samples were purified with the QIAquick PCR Purification Kit (Qiagen), and concentrated to 32 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific).

4.4.4.1.1.2.2. Adenylation of 3' ends

Adenylation of fragments involved the addition of a single adenine nucleotide to the 3' ends of the blunt fragments to prevent them from ligating to each other. A corresponding thymine nucleotide was added to the 3' end of the adapter to provide a complimentary overhang that enables the adapter to ligate to the fragment. This was achieved with the addition of the Klenow fragment (3'→5' exo-) (New England Biolabs) according to the manufacturer's instructions. Samples were then purified with the Qiagen MinElute PCR

Purification columns (Qiagen) and then concentrated to 10 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific).

4.4.4.1.1.2.3. Paired end adapter ligation

Adapters were ligated to all adenylated fragments with the Illumina Paired End Sample Preparation Kit (Illumina). Ligation reactions were purified with Agencourt AMPure XP beads (Beckman Coulter) and concentrated to 30 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific).

4.4.4.1.1.3. PCR enrichment of paired end adapter ligated DNA fragments

PCR was performed with the Illumina Paired End Sample Preparation Kit (Illumina). PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter) and concentrated to 30 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific). After PCR purification and concentration, DNA fragment sizes were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies) and the Agilent DNA 1000 Kit (Agilent Technologies), in order to determine whether sizes differed from starting sizes after initial fragmentation. 500 ng of DNA is required for the exome enrichment step.

4.4.4.1.1.4. Exome enrichment – hybridisation step

Exome enrichment was performed with the 50Mb SureSelect Human All Exon Capture kit (Agilent Technologies) according to the manufacturer's instructions. This procedure has three parts: 1) Preparation of the DNA capture plate, 2) Hybridisation buffer preparation and 3) Final library preparation. The end products of these three parts were combined to give a mixture of prepared library, baits and paired end blockers, which was then incubated in a thermal cycler for 24 hours at 65°C. After the 24 hour incubation, fragments that had not hybridised to the baits, i.e. non-exonic regions, were removed. RNA baits were then digested from the exonic region library fragments.

4.4.4.1.1.5. Post-hybridisation amplification

After exome enrichment, library samples were amplified by PCR with the Illumina Paired End Sample Preparation Kit (Illumina). PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter) and concentrated to 30 µl with the Savant™ DNA SpeedVac™ Concentrator (Thermo Scientific).

4.4.4.1.1.6. Paired end library quantification

The libraries were quantified using the 2100 Bioanalyzer (Agilent Technologies) and Agilent High Sensitivity DNA Kit (Agilent Technologies) before sequencing.

4.4.4.1.2. Cluster generation

Cluster generation involved the amplification of DNA libraries in an Illumina Flow Cell. The flow cell has eight lanes with each having a dense lawn of oligonucleotides attached to the inside surface which are complimentary to the paired end adapters flanking the gDNA libraries. The first step was denaturation of the libraries to single-stranded fragments at a set concentration, and then loading the denatured library in each lane to facilitate template hybridisation to the flow cell via the lawn of oligonucleotides. “Bridge PCR” was then performed, allowing fragments to be clonally amplified and several million clusters per lane to be generated. The reverse strands were then cleaved and washed away. The ends were blocked and the sequencing primer was hybridised to the DNA template, and the flow cell was then ready to load on the Illumina HiSeq 2000 sequencer for sequencing.

4.4.4.1.3. Sequencing by synthesis

After cluster generation, the flow cell was sequenced on the Illumina HiSeq 2000 sequencer with a 75 cycle Paired End protocol, which included 75 cycles of sequencing in the forward direction (termed Read 1) followed by 75 cycles in the reverse direction (termed Read 2). This allowed for the generation of 55 to 82 million reads per lane.

4.4.4.1.4. Data analysis

4.4.4.1.4.1. Alignment and variant calling

Reads were aligned against the human assembly hg19 (GRCh37) using Burrows-Wheeler Aligner (BWA v 0.7.5).

Variant calling was performed using SAMtools (v0.1.19) and PINDEL (v0.2.4t). Variant quality was determined using the SAMtools varFilter script, using the default parameters (except for the setting of the minimum P-value for base quality bias to 1e-400). In addition, a custom script was applied to mark all variants having adjacent bases of low median base quality. All variants were then annotated using custom Perl scripts, including information about known transcripts (UCSC Known Genes and RefSeq genes), known variants (dbSNP v135), type of variant and, if applicable, amino acid change in the corresponding protein.

4.4.4.1.4.2. Variant filtering

The annotated variants were then inserted into the Helmholtz in-house database. All variants identified by exome sequencing were filtered according to a set of criteria. Genetic variants were verified in the 1000 Genomes project browser (www.1000genomes.org/), the Exome Aggregation Consortium (ExAC) database (http://exac.broadinstitute.org) and internal Helmholtz database (5978 exomes).

Predicted functional effect of non-synonymous variants was surveyed using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph) and SIFT (http://sift.jcvi.org/).

To identify putative disease-causing variants, the Helmholtz database was queried to show only variants identified in both affected subjects. To reduce false positives, variants that had a variant quality of less than 30 were excluded. As an internal frequency filter, only variants present in less than 4 internal controls were considered. All synonymous and intronic (other than canonical splice site) variants were excluded.

The raw read data of the variants identified was then manually investigated using the Integrative Genomics Viewer (IGV).

4.4.4.2. Exome Sequencing with the Ion AmpliSeq™ Exome Kit

The Ion AmpliSeq™ Exome Kit includes 294,000 primer pairs that amplify the whole exome in 12 primer pools using as little as 50 ng of DNA. Sample DNA was amplified using the Ion AmpliSeq™ Exome Kit and sequencing was performed with the Ion Proton™ Sequencer and the Ion PI™ Chip Kit v2 (Life Technologies) according to the manufacturer's instructions. This experiment was performed at the Central Analytical Facilities Sequencing Lab at Stellenbosch University. Data analysis was performed with the assistance of Dr Guillaume Pare at McMaster University in Toronto, Canada.

4.4.4.2.1. Exome sequencing method

A total of 100 ng of genomic DNA per sample was amplified using the Ion Ampliseq™ Exome Primer Panel (Life Technologies) (part of the Ion AmpliSeq™ Exome Kit (Life Technologies)). DNA quantification and quality assessment was performed with the Qubit™ 2.0 Fluorometer (Life Technologies) and the Qubit® dsDNA HS Assay Kit (Life Technologies). Amplification was performed using a Veriti® Thermal Cycler (Life Technologies) using the following cycling conditions: an activation step at 99°C for 2 min followed by 10 cycles of 99°C for 15 sec and 60°C for 16 min.

The library was then prepared with the Ion AmpliSeq™ Library Kit V2 (Life Technologies) and Ion Xpress™ Barcode Adapters (Life Technologies). In short: the 12 individual reaction products were pooled and the primers were digested using the supplied FuPa reagent (Life Technologies). This was followed by adapter ligation during which the barcodes were added. The post-ligation products were purified using Agencourt AMPure XP reagent (Beckman Coulter) after which the products were amplified by limited PCR. Amplification was done with a Veriti® Thermal Cycler (Life Technologies) as follows: an activation step at 98°C for 2 min was followed by 5 cycles of 98°C for 15 sec and 64°C for 1 min. The amplified library was then purified with Agencourt AMPure XP reagent (Beckman Coulter). The libraries were quantified using real time PCR with the Ion Library Quantitation Kit (Life Technologies) using a StepOnePlus™ Real-Time PCR System (Life Technologies). The libraries were normalised to 100 pM and pooled in equal volumes prior to template preparation on the Ion OneTouch™ 2 System (Life Technologies) with the Ion PI™ Template OT2 200 Kit v2 (Life Technologies). Enrichment and purification of the Ion Sphere Particles (ISPs) was performed on the Ion

OneTouch™ ES (Life Technologies) using the Ion PI™ Template OT2 200 Kit v2 (Life Technologies). The enriched ISPs were analysed on the Ion Proton™ System (Life Technologies) using the Ion PI™ Sequencing 200 Kit v2 (Life Technologies) and the Ion PI™ Chip v2 Kit (Life Technologies). Data was collected and analysed using the Torrent Server with Torrent Suite Version 4.0.2 (Life Technologies).

4.4.4.3. Validation of potentially disease-causing variants

The shortlisted variants were then validated with high resolution melt analysis and Sanger sequencing. All available ACM2 family members (affected and unaffected) were screened for the shortlisted variants in order to determine segregation of these variants with disease. Variants were then assessed using various bioinformatic tools in order to determine if they were likely to affect mRNA splicing, mRNA secondary structure or protein function as described in section 2.5.7.

4.4.5. HRM Analysis

HRM analysis was used to screen for the variants of interest in the entire ACM2 family to determine whether or not these variants segregated with disease in the family. This technique was also used to screen 100 Caucasian controls (200 chromosomes) for the variants of interest identified by exome sequencing. The samples of interest were compared to samples known to be positive for the variants of interest, namely ACM2.4 and ACM2.5.

Primers were designed according to criteria described in section 2.5.4.1. to amplify the regions surrounding the variants identified by exome sequencing. The primers used for the combined real-time PCR and HRM of the genes of interest are shown in Table 4.1, while the reagents are shown in Table 4.2 and the reaction conditions in Table 4.3.

Table 4.1: Primers for the PCR and HRM of the variants of interest

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE	VARIANT
ADD1_F	CCA CAG CTC CTA CCA TAT CCT TC	23	52.2	56.6	273	c.1810C>G; p.Q604E
ADD1_R	GAA AGG TGT CAC GTC CTG GC	20	60	58.5		
ANKRD62_F	ATG ACT AAT ATG ACA GAC TGA G	22	36.4	49.1	285	c.1071-1073delATA
ANKRD62_R	ACC TAT CAA ACT CGC CAT TC	20	45	52.5		
ASCL1_F	TTT TTA ACT TCC GTC AGG GCT C	22	45.5	55.3	451	c.86C>T; p.A29V
ASCL1_R	GTT TGC AGC GCA TCA GTT CGG	21	57.1	59.7		
B3GNT6_F	CGC ACA GCA GTC CTC CTT C	19	63.2	58.4	194	c.1073G>A; p.R358H
B3GNT6_R	TTC TCA GCA TGG ACA TGG TTG	21	47.6	55.4		
BDH2_F	ATC ATG GAA CTG TCC TGG	18	50	51.5	135	c.326T>C; p.M109T
BDH2_R	CAG TAC AGA GAC CTA ATG TGG TG	23	47.8	54.6		
BRCA1_F	TTG GAA CCA GGT TTT TGT G	19	42.1	51	154	c.4132T>C; p.I1378V
BRCA1_R	TCT ATG AAA AGC ACC TTA GGA G	22	40.9	51.9		
C15orf62_F	ACC ACG AGG AGT ACA GCA AC	20	55	56.9	225	c.217_delCG; p.R73fs
C15orf62_R	ACA GAG AGA GCG CCG ATG	18	61.1	57.2		
C17orf80_F	CGT GTG GGA GCA AAG GAA ATG	21	52.4	56.8	184	c.1007A>C; p.N336T
C17orf80_R	GAA TGA AAC TGG TAA GTC GTC	21	42.9	51.1		
CAMK4_F	TGT TTC CAA TCA GGC AGC GG	20	55	58.4	230	c.1110A>C; p.K370N
CAMK4_R	AGG CTT GCA CCT TCA TCA GCT C	22	54.5	59.9		
CDH2_F	GAT AAA AAC CTT TCA CTG CGG	21	42.9	52.2	241	c.686A>C; p.Q229P
CDH2_R	TCA GTA TTG AGG CTT TCA GAC	21	42.9	52.1		
CFHR2_F	GTA TAT GCT CCA GGT TCA TC	20	45	50.1	173	c.595G>T; p.E199X
CFHR2_R	AAT ATC AGA CTC ATC ACA CTG	21	28.1	49.5		
CLTC_F	GAC AGA ACT GAT ATT CGC AC	20	45	51	145	c.3568A>G; p.I1190V
CLTC_R	CAA GCT AAT CAA GTA GCC C	19	47.4	50.5		

Table 4.1: Primers for the PCR and HRM of the variants of interest (continued)

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE	VARIANT
DDX56_F	GGC CAG ACC AGT TAC AGC AG	20	60	58.1	250	c.824G>A; p.R275H
DDX56_R	CTT CAA CTT ATT TCA ACC AAG ATG	24	33.3	50.5		
DSG1_F	GTT AGC CAA TGC CCA CAA TG	20	50	54.7	169	c.2914G>A; p.G972S
DSG1_R	CAC TTA TGC CAG CTC CAC TC	20	55	55.6		
DUSP18_F	GCC CAT CAT CCG ACC CAA C	19	63.2	58.8	195	c.490_delTG; p.V164fs
DUSP18_R	ATC TGT ACC TCT GAC TCC AAT G	22	45.5	53.7		
ENPP7_F	ACA GGT ACG GCC CCG AGT C	19	68.4	61.8	167	c.712C>A; p.R238S
ENPP7_R	GTC GCC AGC CCG TTT GTC	18	66.7	59.7		
EP400_F	CCT CTC CAG GCT TCC AGT TC	20	60	57.4	255	c.442C>A; p.Q148K
EP400_R	GAG GGG CTC AAG CTG ATC TG	20	60	57.7		
ESCO1_F	TCA AAA ATG GCT ACT TCA GTG G	22	40.9	53.3	160	c.826C>G; p.L276V
ESCO1_R	TAC CTC GTT TGC TCT TTC CTG	21	47.6	54.7		
FSCN3_F	CCA TCC AGA TAA CTC CTT CC	20	50	52.1	180	c.1181G>A; p.R394H
FSCN3_R	CGG CAG GGT AGT AGA TGA ATG	21	52.4	54.9		
G6PC_F	GTT GGG ATT CTG GGC TGT G	19	57.9	56.2	172	c.558G>T; p.L186
G6PC_R	CTG GGA TTT TGT CCT GAT TAG	21	42.9	50.8		
GALNT2_F	GAG CTG GTG GTC ATG TGC AG	20	60	58.6	159	c.1333A>G; p.I445V
GALNT2_R	TGA CAT TCA TAA ACT CCA ACC	21	38.1	50.4		
HTR3C_F	CAT GGA TGT GGA TCA GAC GCC	21	57.1	58.1	241	c.556A>G; p.T186A
HTR3C_R	TCT AAT GAT ACC TAC ACC ACT C	22	40.9	50.9		
KIF21B_F	GTC GGC CAG CAC TAC CTC ATC	21	61.9	59.9	143	c.2641T>A; p.L881M
KIF21B_R	CTT ACC GGG CAG GAC GGG	18	72.2	60.9		
KLF11_F	AGG GGA GAA GAA GTT TGT GTG	21	47.6	54.7	224	c.1430T>G; p.M477R
KLF11_R	AAA TCC CAT GAG TGA TGT CC	20	45	52.3		

Table 4.1: Primers for the PCR and HRM of the variants of interest (continued)

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE	VARIANT
KRT17_F	TGG AGG CCG ACA TCA ATG G	19	57.9	57.6	223	c.667C>T; p.E223K
KRT17_R	CCT GCT CTG CTC TCT CCC AC	20	65	57.9		
KRT31_F	CTT GAC CCA GGC ATT CTT AC	20	50	53.3	249	c.1151C>T; p.P384L
KRT31_R	TAG CGC ACG AAG GAA TTG	18	50	52.9		
KRT34_F	ATT CCA ACT AAC TCC AGC C	19	47.4	52	327	c.1256C>T; p.A419V
KRT34_R	AAA TAA CAT AGA GGC AAG ATG AG	23	34.8	50.5		
KRTAP9-2_F	CTG GAA GCC CAC CAC TGT G	19	63.2	58.6	300	c.250C>A; p.P84T
KRTAP9-2_R	GG TAG TAG CAG GTT CTT C	19	52.6	52		
KRTAP16-1_F	AGC CAC CTC TCT CTG CTC CAC	21	61.9	61	329	c.217A>G; p.C73R
KRTAP16-1_R	AGC AAG AAG GCT CAC AGA TG	20	50	55		
LAMC2_F	GAC AAG GGA GGC CAC TCA AG	20	60	57.9	218	c.2570G>C; p.R857P
LAMC2_R	CAA CAT GGA GAA AAA GCT GAA TG	23	33.1	52.9		
MYO16_F	TCC TCT CTG TTG TTT TCA GTG	21	42.9	52.9	240	c.334G>A; p.V112I
MYO16_R	TGC AGC TTA TCA TCC AAG CCT G	22	50	57.6		
NCOA3_F	GAG AGC TGC TAA GTC ATC AC	20	50	52.7	241	c.T3729del-GCAGCAGCA; p.M1243delQQ?
NCOA3_R	TTG GTT GAT ATG GAA ACT GTT G	22	36.4	51.2		
NEK10_F	GTT ATG GAG CTG ATA GAA GGA G	22	45.5	52.2	189	c.1889G>A; p.R630K
NEK10_R	CCA AAG ATG ACT ACT GCA CGT C	22	50	55.7		
OR4D6_F	TGA TCT CTA ACA ACG GAC TG	20	45	51.5	173	c.665T>C; p.L222P
OR4D6_R	GTA AAC ACA AGG CAC GAA G	19	47.4	51.6		
RTL1_F	GAA CCT GAC CTG GAT CGA CTC	21	57.1	57	124	c.1195G>A; p.E399K
RTL1_R	GTT CAC TCT CAC CAT GAG	18	50	49.7		
SCN4A_F	GAG TCC CTC ATC AAG ATA CTG	21	47.6	51.9	274	c.606G>C; p.M202I
SCN4A_R	GAG AGT GCT GAG CCA GAC	18	61.1	55.2		

Table 4.1: Primers for the PCR and HRM of the variants of interest (continued)

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE	VARIANT
TTC37_F	GAG TCA CAA AAG CCA CTC CCA G	22	54.5	58.1	244	c.4187A>G; p.N1396S
TTC37_R	TAG TTC CTC TAA CCA CAA TAC C	22	40.9	51.3		
ZGRF1_F	CCC AGG TTC CAC TTA TTA CTT TG	23	43.5	53.5	157	c.3256A>G; p.M1086V
ZGRF1_R	AAG AAC TGC TGA CTT TTC ACA C	22	40.9	53.6		

Table 4.2: Reagents for the PCR and HRM of the variants of interest

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	0.8 µM / 0.5 µl
Reverse Primer (20 µM)	0.8 µM /0.5 µl
dNTPs (20 µM) (Bioline)	0.8 µM / 1 µl
GoTaq Polymerase (5U/ ul) (Promega)	0.5 U / 0.1 µl
GoTaq FlexiBuffer (5X) (Promega)	1 X / 5 µl
$MgCl_2$ (25 mM) (Promega)	1.5 mM / 3 µl
EvaGreen dye (Anatech)	1X / 1 µl
DNA	200 ng
FINAL REACTION VOLUME	**25 µl**

Table 4.3: Optimised temperature cycling conditions for the PCR and HRM of the variants of interest

CONDITION	TEMPERATURE (TIME)
Initial denaturation	95°C – 10 seconds
Denaturation Primer Annealing Template Elongation	95°C – 5 seconds 55°C – 10 seconds 72°C – 10 seconds 50 cycles
High Resolution Melt	72 – 95°C (0.1°C increments)

4.4.6. Sequencing

For the screening of the ACM2 family, the HRM amplicons for all family members were subjected to Sanger sequencing to confirm the presence of the variants of interest. For the population control screening, any sample that displayed a similar HRM profile to the positive control from the ACM2 family was subjected to Sanger sequencing to confirm the presence of the variant of interest. Sequencing was performed as described in section 2.5.5.

4.5. Results

4.5.1. Sample characteristics

The ACM2 samples (ACM2.4 and ACM2.5) are described in Table 4.4. Five individuals (the proband (Individual III:3 for whom we did not have DNA), ACM2.2 (Figure 4.1, III:10), ACM2.4 (Figure 4.1, III:6), ACM2.5 (Figure 4.1, III:5) and ACM2.12 (Figure 4.1, II:4)) fulfilled the criteria for definite ARVC while one individual (ACM2.9 – Figure 4.1, II:8) fulfilled criteria for borderline ARVC.[15] The control population consisted of 100 healthy Caucasian individuals (200 chromosomes).

4.5.2. Exome sequencing output

Exome sequencing of ACM2.4 and ACM2.5 was performed with the SureSelect Human All Exon 50Mb kit (Agilent Technologies, Inc.) and run on the Illumina HiSeq 2000. For ACM2.4, the average coverage depth was 199.8 and 20x coverage for 92.16% of the target bases, while for ACM2.5, the average coverage depth was 204.8 and 20x coverage for 92.06% of the target bases (Table 4.5).

Exome sequencing of these samples was also performed with the Ion AmpliSeq™ Exome Kit (Life Technologies) and samples were run on the Ion Proton™ Sequencer (Life Technologies). For ACM2.4, the average coverage depth was 129.9 and 5x coverage for 93.21% of the target bases, while for ACM2.5, the average coverage depth was 121.4 and 5x coverage for 93.91% of the target bases (Table 4.5).

After bioinformatic analysis of data from both exome sequencing experiments, a number of variants common between ACM2.4 and ACM2.5 were highlighted. A search was then conducted in various genetic variation databases to determine the novelty and prevalence of these variants. These include the 1000 Genomes database (http://browser.1000genomes.org/index.html), the Exome Aggregation Consortium (ExAC) database as well as a database of in-house control exomes.

4.5.2.1. Variants identified with the SureSelect Human All Exon 50Mb kit

These variants included 21 missense variants and 1 synonymous variant (22 variants in total) (Table 4.6 and Figure 4.2a). Of these variants, 9 were detected on chromosome 17 with many other variants found on chromosomes 18 (n = 4) and 11 (n = 2) (Figure 4.3a).

Table 4.4: Clinical characteristics of ACM2 family members

ID	Position in pedigree	Sex	Age at presentation	Presenting symptoms	Cardiac imaging	Tissue characteristics	Repolarization abnormalities	Depolarization/ conduction abnormalities	Arrhythmias	Diagnostic criteria	Outcome
Proband	IV: 5	M	15	Palpitations with sweating	RVA: dilated RV with regional akinesia (outflow tract)	No information	TWI V1-V5	Epsilon wave TAD >55ms	VT (LBBB with superior axis)	DEFINITE ARVC 4 major, 1 minor	Died (age 21)
ACM2.1	IV:10	F	22	Asymptomatic	Normal MRI	No information	No TWI	No epsilon wave No late potentials	PVC	UNAFFECTED 1 major	No information
ACM2.2	IV:12	M	20	Palpitations, associated with chest pain	RVA: anterior hypokinesia	Fibro-fatty replacement	No TWI	No epsilon wave No late potentials	PVC 2035/24h	DEFINITE ARVC 2 major, 1 minor	No information
ACM2.3	III:14	M	45	Asymptomatic	Normal echocardiogram	No information	No TWI	No information	No information	UNAFFECTED	No information
ACM2.4	IV:8	M	23	Chest pain, occasional palpitations	RVA: bulging hypokinetic RV inferior wall and sub-tricuspid dyskinesia	Fatty infiltration, scattered chronic inflammatory cells	TWI V1-V5	No epsilon wave No late potentials	EPS: NSVT LBBB morphology superior axis PVC 9341/24h (no meds 1995) PVC 12620/24h (Sotalol 1997)	DEFINITE ARVC 3 major, 1 minor	No information
ACM2.5	IV:7	F	No information	Symptomatic	No information	No information	No information	No information	VT on amiodarone	DEFINITE ARVC No data available	Died (age<35y)

Table 4.4: Clinical characteristics of ACM2 family members (continued)

ID	Position in pedigree	Sex	Age at presentation	Presenting symptoms	Cardiac imaging	Tissue characteristics	Repolarization abnormalities	Depolarization/ conduction abnormalities	Arrhythmias	Diagnostic criteria	Outcome
ACM2.6	III:17	F	No information	No information	No information	No information	No information	No information	No information	No information	No information
ACM2.7	II:17	F	75	No information	No information	No information	No information	No information	No information	No information	No information
ACM2.8	IV:3	M	44	Asymptomatic	MRI: Focal area of dyskinesia RVOT but did not fulfil criteria for ARVC Echocardiogram normal	EMB: Non-diagnostic, mild endocardial and myocardial fibrosis, mildly hypertrophic myocytes, no fatty infiltration	No TWI	No epsilon wave No late potentials	Holter normal	UNAFFECTED	No information
ACM2.9	III:19	F	44	Palpitations, chest pain	Normal echocardiogram MRI reported as normal	No information	TWI V1 and V2	No epsilon wave No late potentials	No holter done EST normal	BORDERLINE ARVC 1 majot, 1 minor	No information
ACM2.10	IV:16	M	27	Asymptomatic	MRI: Normal RV function and volumes Echocardiogram: mild RVIT dilatation, otherwise normal	Non-diagnostic	No TWI	No epsilon wave No late potentials	NSVT on Holter but does not fulfil criteria PVC 40/24h EST normal	UNAFFECTED	No information

Table 4.4: Clinical characteristics of ACM2 family members (continued)

ID	Position in pedigree	Sex	Age at presentation	Presenting symptoms	Cardiac imaging	Tissue characteristics	Repolarization abnormalities	Depolarization/ conduction abnormalities	Arrhythmias	Diagnostic criteria	Outcome
ACM2.11	IV:17	F	21	Palpitations, chest pain, episode of near syncope	MRI reported as normal Echocardiogram normal	No information	No TWI	No epsilon wave Positive for late potentials (Filtered QRS 119)	EPS negative 2009 and 2014 EST normal No holter done ICD 2009 – no VT	UNAFFECTED 1 minor	No information
ACM2.12	III:15	F	57	Palpitations, chest pain, presyncope	Echocardiogram normal	No information	TWI V1-V3	No epsilon wave, No SAECG TAD normal	No holter done No history of VT	DEFINITE ARVC 2 major	No information
ACM2.13	III:18	M	43	Asymptomatic	No information	No information	No information	No information	No information	No information	No information
ACM2.14	V:3	F	11	Chest pain	No information	No information	No information	No information	No reported VT	No information	No information
ACM2.15	V:2	F	2	No information	No information	No information	No information	No information	No information	No information	No information
ACM2.16	V:1	M	7	No information	No information	No information	No information	No information	No information	No information	No information

Table 4.4: Clinical characteristics of ACM2 family members (continued)

ID	Position in pedigree	Sex	Age at presentation	Presenting symptoms	Cardiac imaging	Tissue characteristics	Repolarization abnormalities	Depolarization/ conduction abnormalities	Arrhythmias	Diagnostic criteria	Outcome
ACM2.17	IV:4	F	34	No information	No information	No information	No information	No information	No information	UNAFFECTED	No information
ACM2.18	IV:14	F	19	No information	No information	No information	No information	No information	No information	No information	No information
ACM2.19	III:16	M	53	No information	No information	No information	No information	No information	No information	UNAFFECTED	No information
ACM2.20	IV:9	F	23	No information	No information	No information	No information	No information	No information	UNAFFECTED	No information

M – Male; F – Female; RV – Right ventricle; RVA – Right ventricular akinesia; TWI – T-wave inversion; TAD – Terminal activation duration; VT – Ventricular tachycardia; LBBB – Left bundle branch block; PVC – Premature ventricular contractions; EPS – Electrophysiological study; NSVT – Non-sustained ventricular tachycardia; MRI – Magnetic resonance imaging; EMB – Endomyocardial biopsy; EST – Exercise stress test; RVIT – Right ventricular inflow tract; ICD – Implantable cardioverter defibrillator; RVOT – Right ventricular outflow tract; SAECG – Signal-averaged electrocardiogram

Table 4.5: Exome sequencing results overview

	SureSelect Human All Exon 50 Mb kit		Ion AmpliSeq™ Exome Kit	
	ACM 2.4	ACM 2.5	ACM 2.4	ACM 2.5
Total variants called	Off target variants are not called 76.5% reads on target	Off target variants are not called 76.0% reads on target	49996	49590
Total variants on target	76400	72102	49006	48658
Coverage	199.8X; 92.16% @ 20X	204.5X; 92.06% @20X	129.9X; 93.21% @ 5X	121.4X; 93.9% @ 5X
Transition/transversion ratio	2.47	2.52	3.12	3.16
Variants present in the 1000Genomes database	58433	57707	45234	44843
Variants present in the dbSNP database	60354	59603	47471	47158
Novel variants	7132	7616	1303	1272
Variant consequence				
Missense	9482	9391	9282	9245
Nonsense	69	81	95	103
Synonymous	10606	10564	10387	10408
Frameshift	285	271	156	133
Non-frameshift Insertion/deletion	209	210	133	136
Splice	133	134	61	66
5’ UTR	2927	2879	946	894
3’ UTR	3442	3252	1565	1544
Intronic	35869	35731	26381	26129

4.5.2.2. Variants identified with the Ion AmpliSeq™ Exome Kit

These variants included 23 missense variants, 1 synonymous variant, 2 frameshift variants, 1 non-frameshift deletion and 1 nonsense variant (Table 4.7 and Figure 4.2b). Six of these variants were found on chromosome 17 and many other variants were also detected on chromosomes 1 (n = 4) and 4 (n = 3) (Figure 4.3b).

Table 4.6: Variants identified with the SureSelect Human All Exon 50Mb kit

CHR	GENE	VARIANT	TYPE
1	*GALNT2*	c.1333A>G; p.I445V	Missense
2	*KLF11*	c.1430T>G; p.M477R	Missense
3	*NEK10*	c.1889G>A; p.R630K	Missense
4	*ADD1*	c.1810C>G; p.Q604E	Missense
7	*FSCN3*	c.1181G>A; p.R394H	Missense
11	*OR4D6*	c.665T>C; p.L222P	Missense
11	*B3GNT6*	c.1073G>A; p.R358H	Missense
12	*EP400*	c.442C>A; p.Q148K	Missense
13	*MYO16*	c.268G>A; p.V90I	Missense
17	*KRTAP9-2*	c.250C>A; p.P84T	Missense
17	*KRTAP16-1*	c.217A>G; p.C73R	Missense
17	*KRT34*	c.1256C>T; p.A419V	Missense
17	*KRT31*	c.1151C>T; p.P384L	Missense
17	*KRT17*	c.667C>T; p.E223K	Missense
17	*G6PC*	c.558G>T; p.L186	Synonymous
17	*BRCA1*	c.4132T>C; p.I1378V	Missense
17	*CLTC*	c.3568A>G; p.I1190V	Missense
17	*SCN4A*	c.606G>C; p.M202I	Missense
18	*ANKRD62*	c.1071_1073delATA	Missense
18	*ESCO1*	c.826G>C; p.L276V	Missense
18	*CDH2*	c.686A>C; p.Q229P	Missense
18	*DSG1*	c.2914G>A; p.G972S	Missense

Table 4.7: Variants identified with the Ion AmpliSeq™ Exome Kit

CHR	GENE	VARIANT	TYPE
1	*LAMC2*	c.2570G>C; p.R857P	Missense
1	*CFHR2*	c.595G>T; p.E199X	Stop
1	*KIF21B*	c.2641T>A; p.L881M	Missense
1	*GALNT2*	c.1333A>G; p.I445V	Missense
2	*KLF11*	c.1430T>G; p.M477R	Missense
3	*NEK10*	c.1889G>A; p.R630K	Missense
3	*HTR3C*	c.556A>G; p.T186A	Missense
4	*ADD1*	c.1810C>G; p.Q604E	Missense
4	*BDH2*	c.326T>C; p.M109T	Missense
4	*ZGRF1*	c.3256A>G; p.M1086V	Missense
5	*TTC37*	c.4187A>G; p.N1396S	Missense
5	*CAMK4*	c.1110A>C; p.K370N	Missense
7	*DDX56*	c.824G>A; p.R275H	Missense
7	*FSCN3*	c.1181G>A; p.R394H	Missense
12	*ASCL1*	c.86C>T; p.A29V	Missense
13	*MYO16*	c.334G>A; p.V112I	Missense
14	*RTL1*	c.1195G>A; p.E399K	Missense
15	*C15orf62*	c.217_delCG; p.R73fs	Frameshift
17	*KRT34*	c.1256C>T; p.A419V	Missense
17	*KRT31*	c.1151C>T; p.P384L	Missense
17	*G6PC*	c.558G>T; p.L186	Synonymous
17	*CLTC*	c.3568A>G; p.I1190V	Missense
17	*C17orf80*	c.1007A>C; p.N336T	Missense
17	*ENPP7*	c.712C>A; p.R238S	Missense
18	*ESCO1*	c.826C>G; p.L276V	Missense
18	*CDH2*	c.686A>C; p.Q229P	Missense
20	*NCOA3*	c.T3729del-GCAGCAGCA; p.M1243delQQ?	Indel
22	*DUSP18*	c.490_delTG; p.V164fs	Frameshift

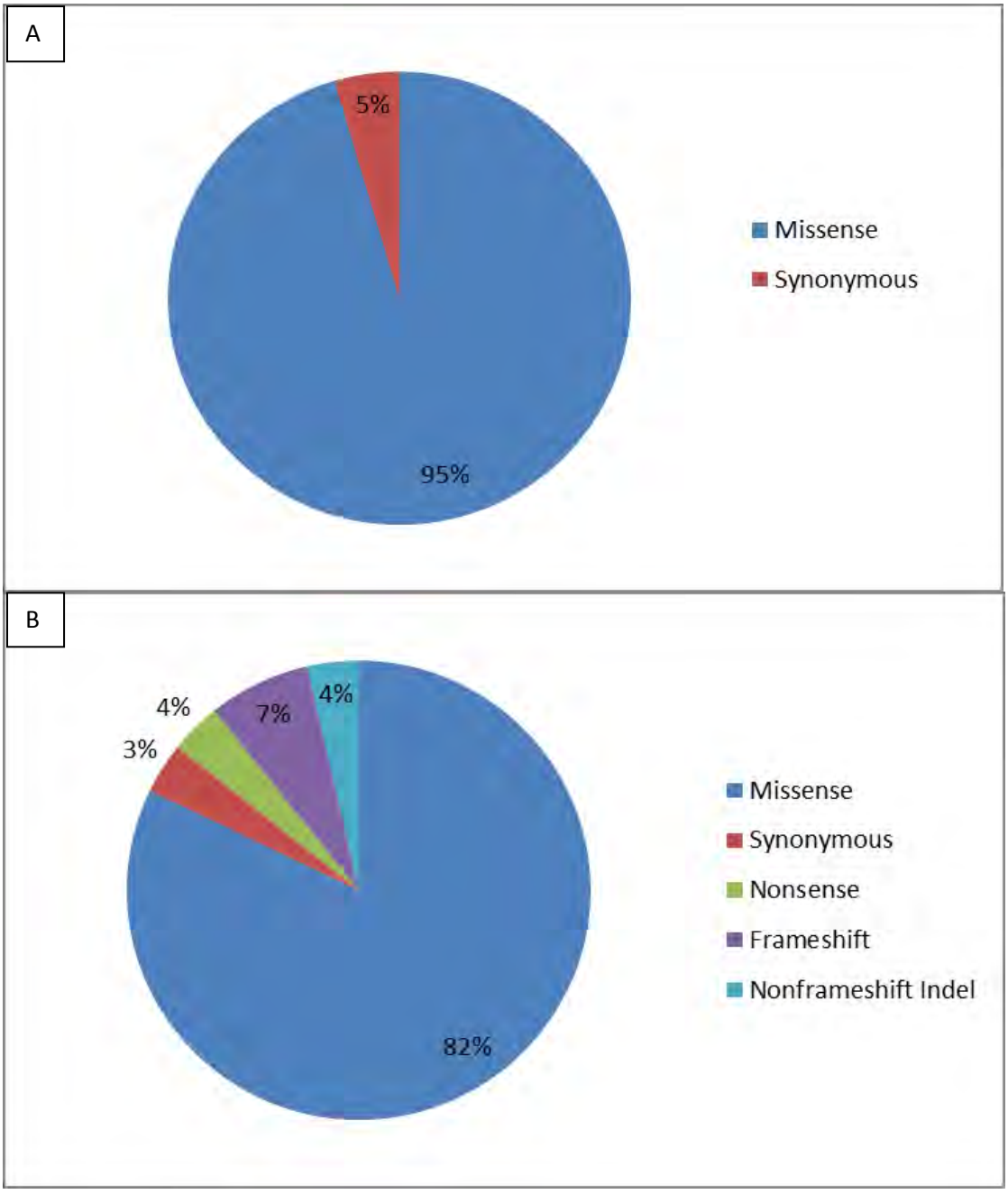


Figure 4.2: Variant distribution by type – SureSelect (A) and Ion AmpliSeq™ (B) Exome Sequencing Experiments

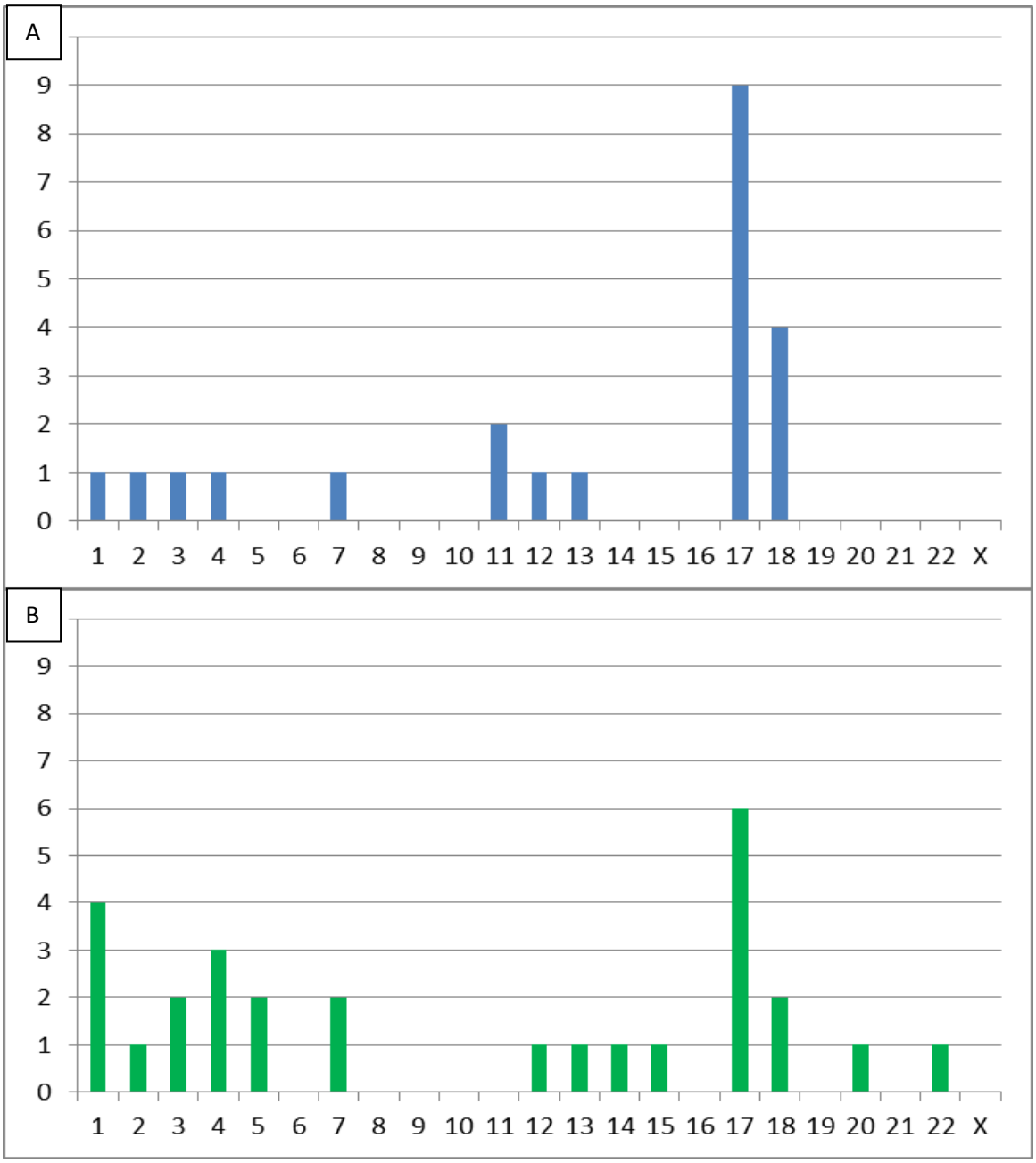


Figure 4.3: Variant distribution by chromosome – SureSelect (A) and Ion AmpliSeq™ (B) Exome Sequencing Experiments

The combination of the variant data from both exome sequencing experiments revealed a list of 12 variants common to the two datasets (Table 4.8). A significant amount of variant overlap was expected, as the same two samples were sequenced in both experiments. The differences in variant lists obtained with these experiments could be due to differences in the chemistry of the whole exome sequencing chips, sensitivity and specificity of the exome

sequencing kits, as well as variations in the variant calling and filtering criteria, as has been reported previously.[103] Follow-up experiments were performed for all variants, irrespective of whether the variant was detected by one or both kits.

Table 4.8: Variants common between the SureSelect and Ion AmpliSeq™ Exome Sequencing Experiments

CHR	GENE	VARIANT	TYPE
1	*GALNT2*	c.1333A>G; p.I445V	Missense
2	*KLF11*	c.1430T>G; p.M477R	Missense
3	*NEK10*	c.1889G>A; p.R630K	Missense
4	*ADD1*	c.1810C>G; p.Q604E	Missense
7	*FSCN3*	c.1181G>A; p.R394H	Missense
13	*MYO16*	c.334G>A; p.V112I	Missense
17	*KRT34*	c.1256C>T; p.A419V	Missense
17	*KRT31*	c.1151C>T; p.P384L	Missense
17	*G6PC*	c.558G>T; p.L186	Synonymous
17	*CLTC*	c.3568A>G; p.I1190V	Missense
18	*ESCO1*	c.826G>C; p.L276V	Missense
18	*CDH2*	c.686A>C; p.Q229P	Missense

The entire ACM2 family (affected and unaffected individuals) was then screened for these variants to determine their segregation with disease. The variants were also subjected to population screens in 200 ethnically matched (Caucasian) control chromosomes. The RNAfold and Mfold bioinformatic analysis tools were used to predict whether these three variants caused structural mRNA changes, while the ESEFinder tool was used to predict whether variants affected mRNA splicing, by determining whether variants affected exonic splice enhancers (ESEs). The MutationTaster, Mutation Assessor, Polyphen-2, SIFT and Align GVGD bioinformatic tools were used to determine whether these variants had a deleterious effect on the protein.

Variants were filtered according to the following criteria (Figure 4.4):

(1) Presence in all affected individuals. We did not exclude variants detected in individuals classified as unaffected due to possible incomplete penetrance.

(2) Absence in the control population. Variants found to be present in ≥1% of the control population were excluded

(3) Amino Acids need to be conserved across multiple species

(4) Prediction of pathogenicity by bioinformatic tools

(5) The affected gene is a functionally plausible ARVC candidate

After filtering according to criteria 1 to 4 listed above, the list of candidate variants included the *KIF21B* c.2641T>A (p.L881M), *TTC37* c.4187A>G (p.N1396S) and *CDH2* c.686A>C (p.Q229P) variants. *KIF21B* encodes the Kinesin Family Member 21B protein which is known to play a role in neuronal component transport.[104] The *TTC37* gene encodes the Tetratricopeptide Repeat Domain 37 protein which is known to be involved in exosomal RNA degradation.[105] The *CDH2* gene encodes the N-cadherin protein which plays an important role in cell adhesion and is known to interact with proteins previously implicated in ARVC,[106] making it the most functionally plausible candidate gene for this disease. Based on all criteria, the most likely pathogenic variant in the ACM2 family is the *CDH2* c.686A>C (p.Q229P) variant.

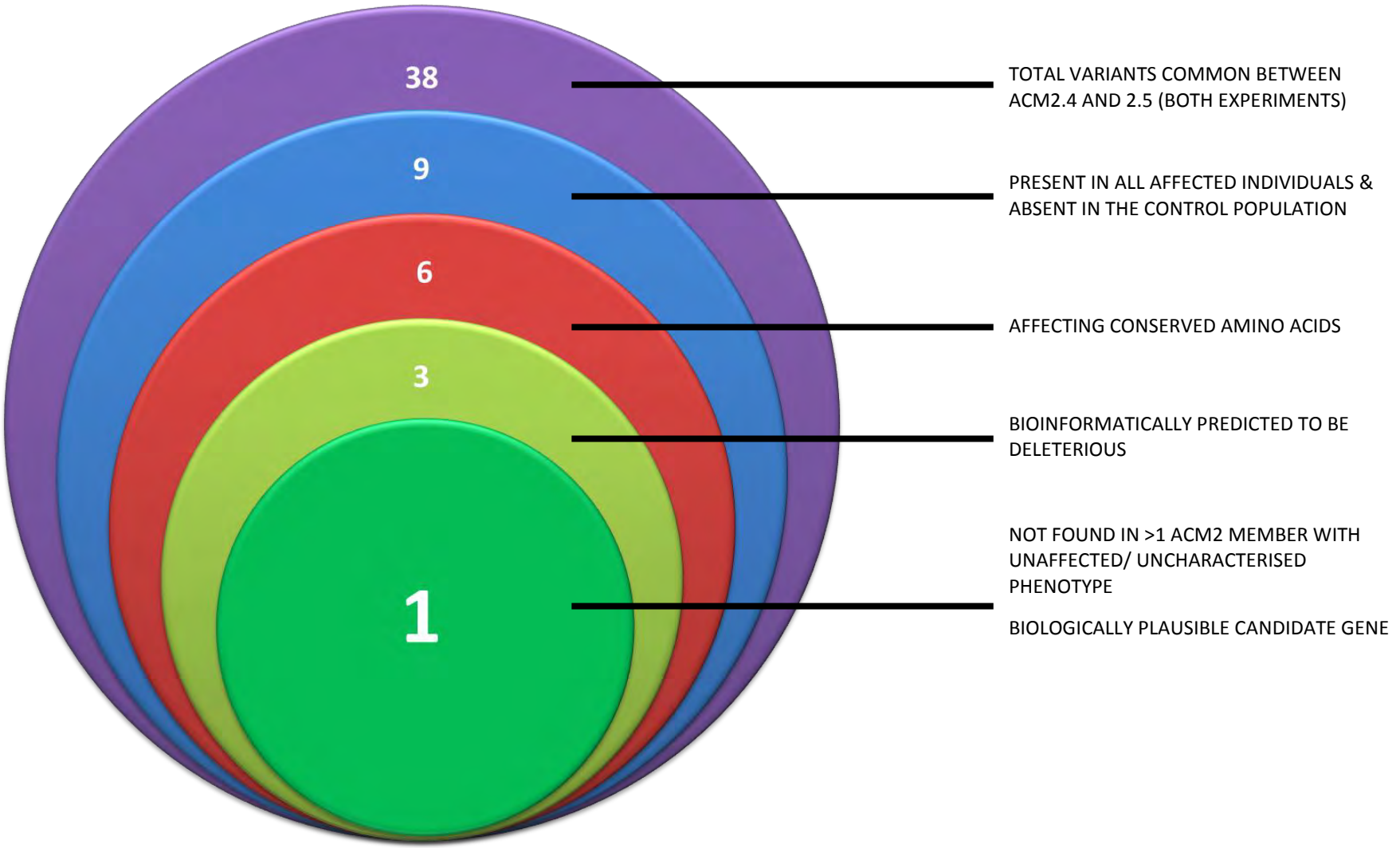


Figure 4.4: Filtering of exome sequencing variants

4.5.2.3. *CDH2* c.686A>C (p.Q229P)

This variant was found in all affected individuals (definite and borderline ARVC) (Figure 4.6), and was not detected in the South African Caucasian control population. This variant is conserved across multiple species (Figure 4.5), and was not reported in the 1000 Genomes, NHLBI Exome Variant Server, NCBI dbSNP and ExAC databases or the Helmholtz in-house database. No medical conditions have been associated with this genetic variant to date.

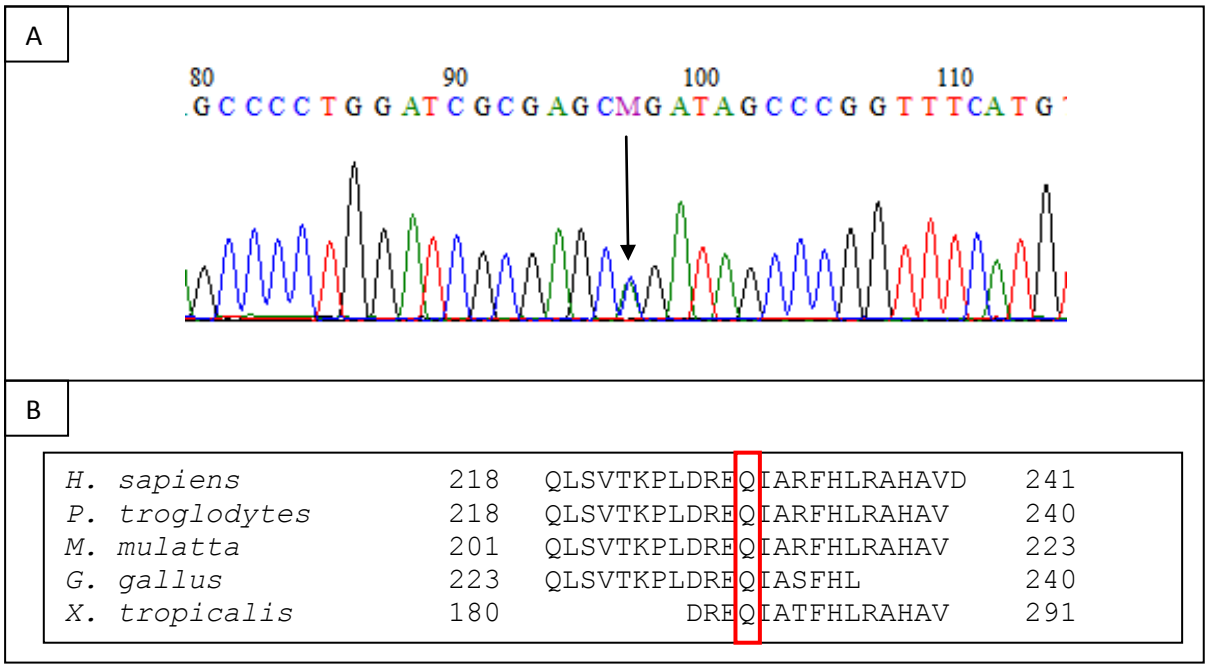


Figure 4.5: Identification and conservation of the *CDH2* c.686A>C (p.Q229P) variant. (A) Electropherogram indicating nucleotide change; (B) Multiple species protein sequence conservation

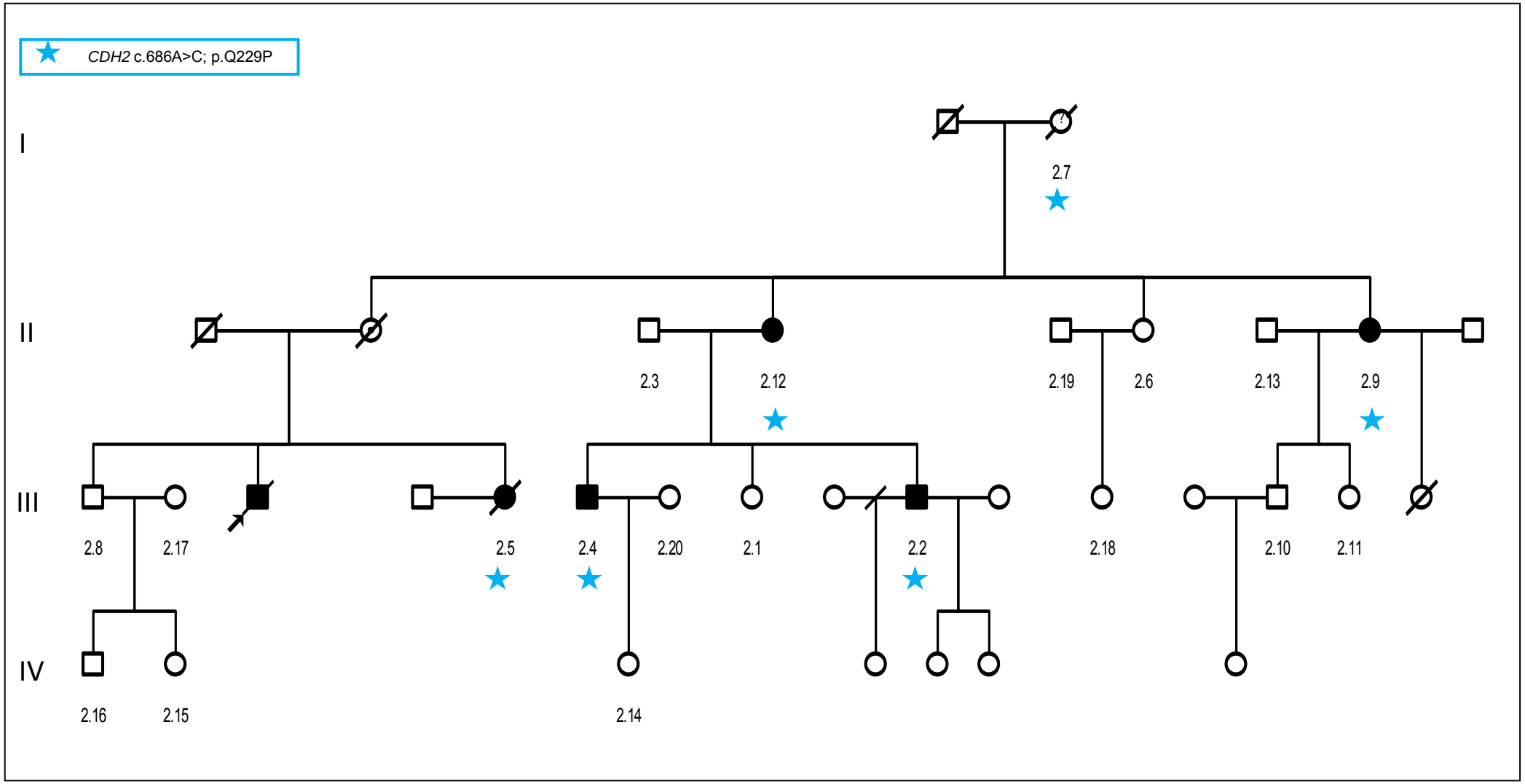


Figure 4.6: ACM2 family pedigree indicating segregation of the *CDH2* c.686A>C variant

Bioinformatic tools (Mfold and RNAFold) predicted that the *CDH2* c.686A>C variant would alter mRNA secondary structure (Figure 4.7), and would cause a change in SRSF1 recognition sites (Figure 4.8). This variant was also predicted to have a deleterious effect on protein function by the MutationTaster, Polyphen-2, SIFT and Align GVGD bioinformatic tools. There are therefore many lines of bioinformatic evidence that suggest the deleterious effect of this variant.

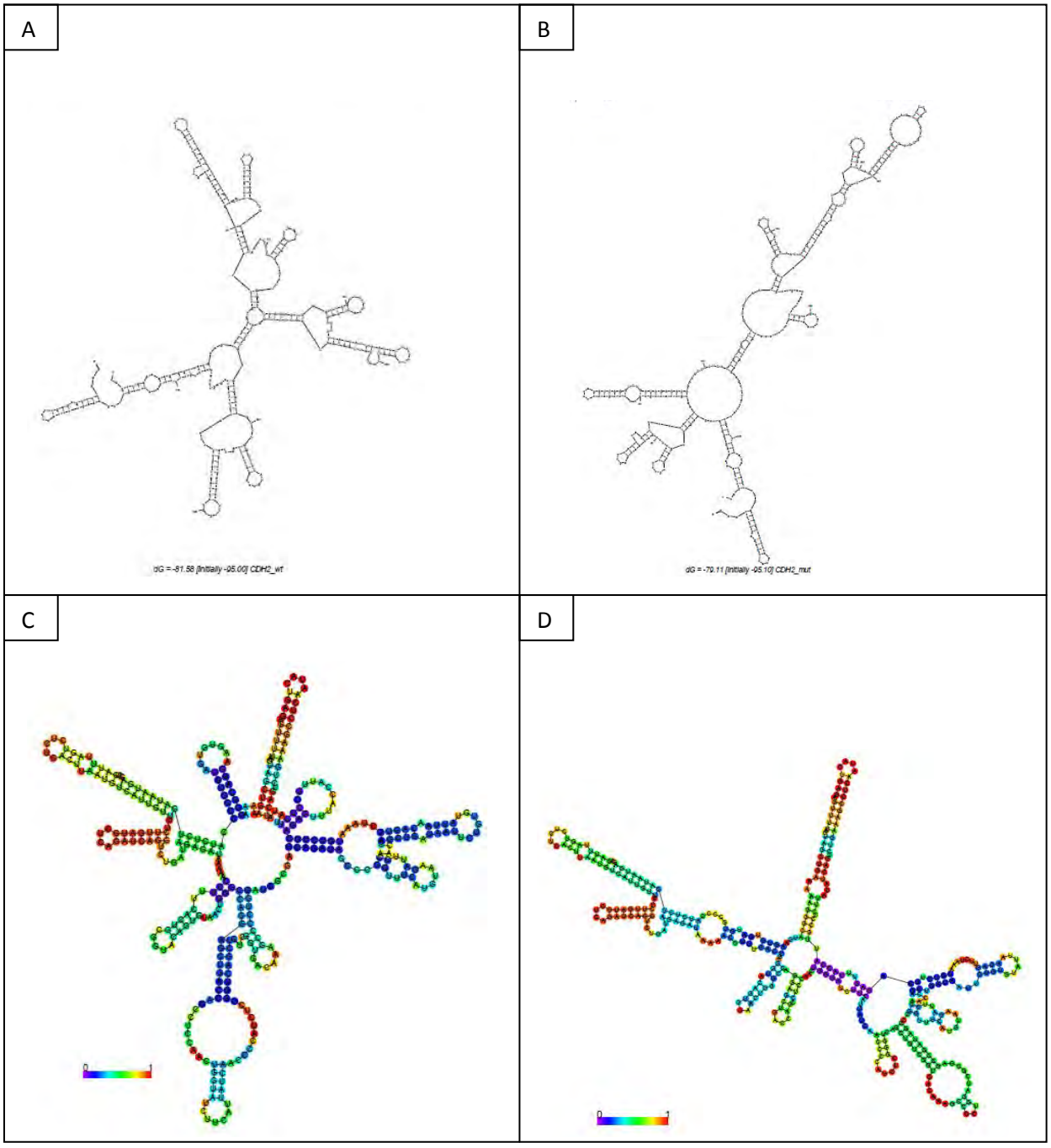


Figure 4.7: Mfold and RNAfold predictions for the *CDH2* c.686A>C variant. (A) Mfold wild type structure; (B) Mfold mutant structure; (C) RNAfold wild type structure; (D) RNAfold mutant structure

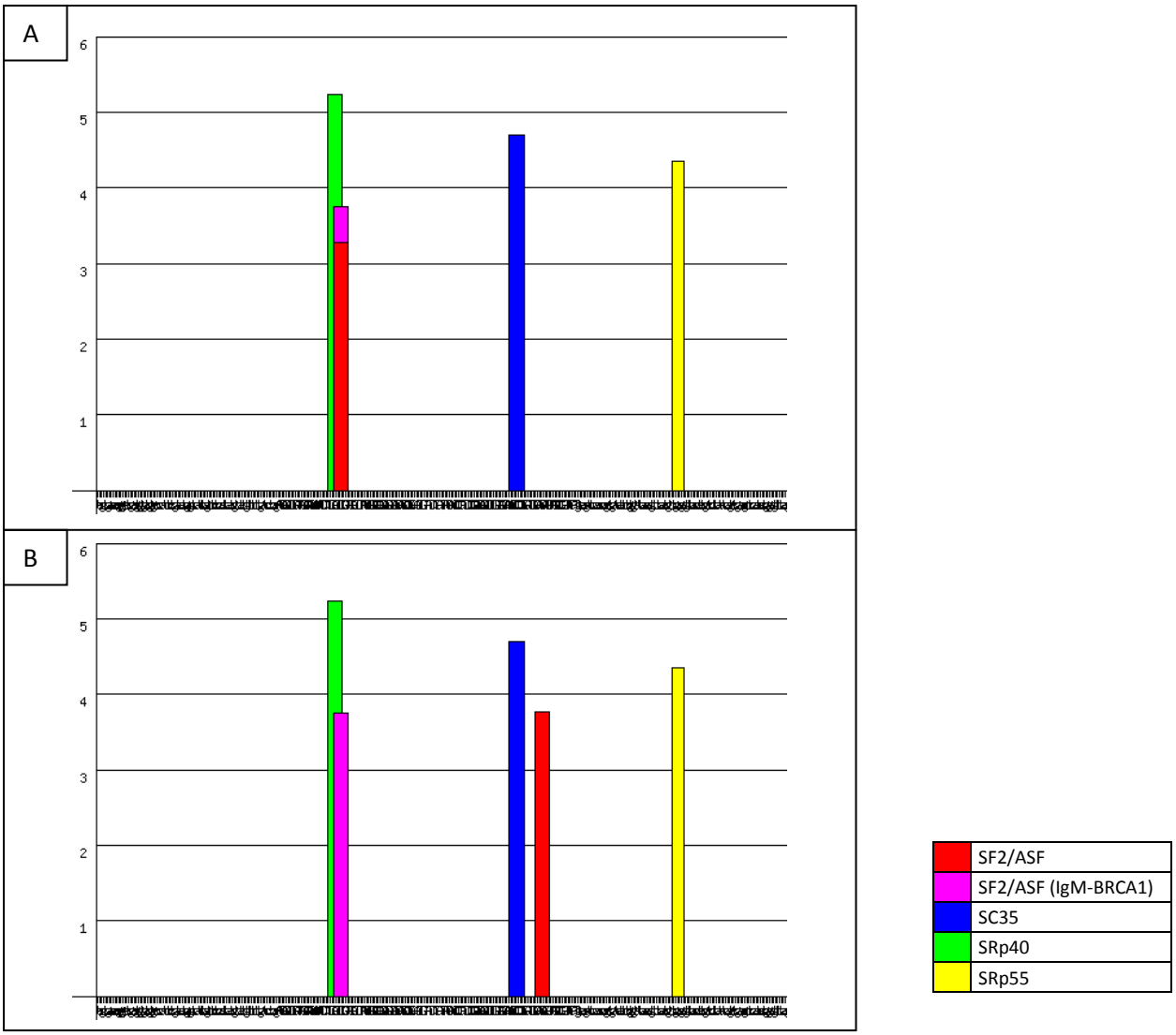


Figure 4.8: ESEFinder predictions for the *CDH2* c.686A>C variant. (A) ESEFinder prediction for wild type sequence; (B) ESEfinder prediction for mutant sequence

Based on the fact that the *CDH2* c.686A>C (p.Q229P) variant (1) is detected in all affected individuals in the ACM2 family, (2) alters a conserved amino acid, (3) is absent in 200 ethnically matched control chromosomes, (4) was not reported in the 1000 Genomes, Exome Variant Server, NCBI dbSNP or ExAC databases, (5) is bioinformatically predicted to be deleterious by various predictive tools and (6) is a biologically plausible candidate (based on the role of N-cadherin in the cell), we classify this variant as a possibly disease-associated variant. This evidence suggests that this variant may play a causal role in the development of ARVC in the ACM2 family.

4.6. Discussion

By combining the information gathered from two whole exome sequencing experiments we believe we have not only identified *CDH2* as the causative gene within this family but we have also identified *CDH2* as a new candidate gene for ARVC and cardiomyopathy in general.

Based on genetic and bioinformatic evidence, the most likely disease-causing variant in the ACM2 family is the *CDH2* c.686A>C (p.Q229P) variant. This variant segregates with disease within this family, is absent in the control population and control databases and is bioinformatically predicted to have a deleterious effect on the function of *CDH2*.

Functionally, the *CDH2* gene encodes the N-cadherin protein, which is a member of the cadherin superfamily of predominantly calcium-dependent cell adhesion proteins.[107] N-cadherin is comprised of five extracellular cadherin repeats, a transmembrane region and a highly conserved cytoplasmic tail (Figure 4.9).[108] The cadherin repeats are vital for the adhesive function of N-cadherin. Calcium ions bind to each cadherin repeat to ensure correct protein folding and confer rigidity to the extracellular domain. Multiple cadherin domains form Ca^{2+}-dependent rod-like structures with conserved Ca^{2+}-binding pockets at the inter-domain region.[109] The p.Q229P variant is located in cadherin repeat 1, and it is therefore plausible that this variant would affect the adhesive function of the N-cadherin protein.

Members of the cadherin superfamily have distinct spatial and temporal expression patterns during embryonic development and adulthood.[110] Changes in cadherin expression are often associated with changes in tissue architecture and cell morphology.[110] In the heart, N-cadherin is located at the intercalated disc and forms part of the adherens junction where it co-localises with the desmosomal proteins known to play a role in the development of ARVC.[106] N-cadherin has been shown to associate directly with plakoglobin, one of the proteins found in the desmosome.[111] Although the main role of N-cadherin is cell adhesion, it is also involved in establishment of embryonic left-right asymmetry[112] and angiogenesis.[113]

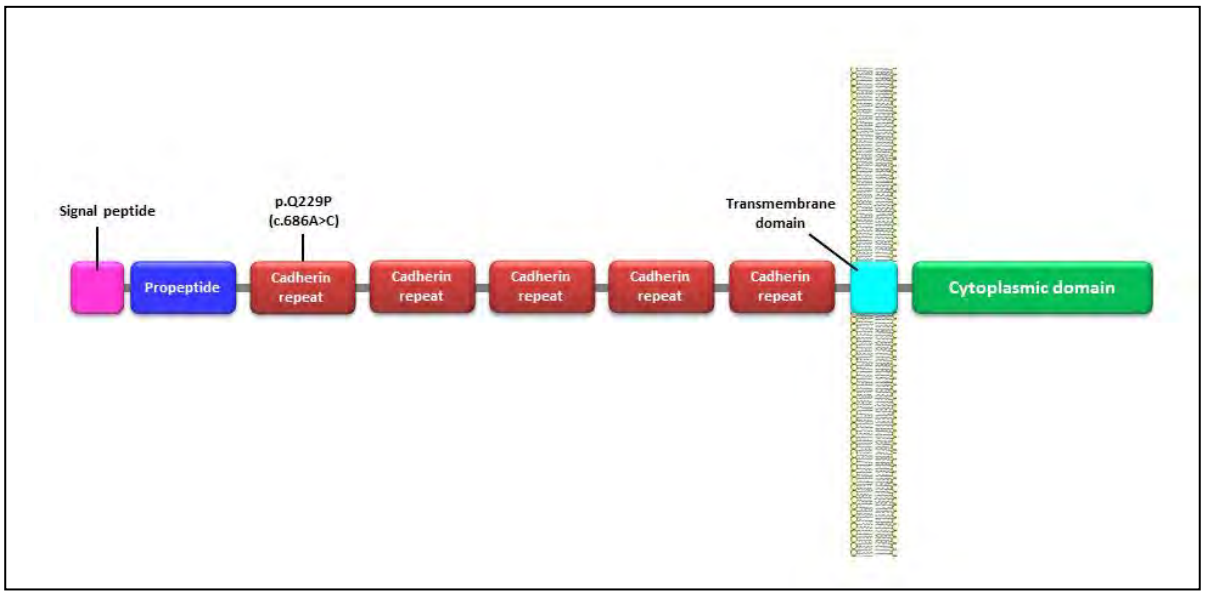


Figure 4.9: Position of the c.686A>C (p.Q229P) variant in the CDH2 protein

Several *CDH2* mouse models have been developed. *CDH2* knockout mice died in the embryonic stage due to a severe cardiac defect.[110] Induced cardiac-specific deletion of N-cadherin in the mouse heart caused dissolution of the intercalated disc structure including desmosomes and adherens junctions[106], an observation which is similar to what is seen in individuals with ARVC. These mice also exhibited dilated cardiomyopathy and the abrupt onset of ventricular tachycardia, providing evidence that the deaths were arrhythmic in nature.[106] Mice with N-cadherin haploinsufficiency exhibited a reduction in Connexin-43 plaque size in the myocardium as well as increased susceptibility to arrhythmias, indicating that N-cadherin influences gap junction function.[114]

N-cadherin is also known to interact directly with components of the Wnt/β-catenin and Hippo signalling pathways known to be involved in ARVC.[33] It is therefore plausible that variants in *CDH2* could lead to the development of ARVC by disruption of the intercalated disc structure and modulation of the Wnt/β-catenin signalling pathway, causing myocyte instability and death and replacement of the myocardium by fibrous fatty tissue.

This is, to our knowledge, the first report of a variation in *CDH2* associated with cardiomyopathy in humans. Although the frequency of variants in this gene is small (one family), the phenotype of individuals with *CDH2* variants appears to be severe, highlighting

the importance of including this gene in routine genetic screening of ARVC patients. Although *CDH2* was included in a candidate gene screen of 13 ARVC patients by Campuzano and colleagues, no disease-causing variants were detected.[115] As the evidence implicating this gene in the development of ARVC is limited to one family, we performed further experiments to determine if this gene plays a role in disease pathogenesis in other ARVC families (Chapter 5).

4.7. Conclusion

We identified the novel *CDH2* c.686A>C (p.Q229P) variant as a disease-associated variant in the ACM2 family, which is likely to play a causal role in the development of ARVC. Our research suggests that *CDH2* may be a novel candidate gene not only for ARVC but for cardiomyopathies in general.

CHAPTER 5
SCREENING OF *CDH2* IN AN ARVC COHORT

5.1. Introduction

Exome sequencing of the ACM2 family with autosomal dominant ARVC identified a putative disease-causing variant in a possibly new ARVC gene i.e. *CDH2* c.686A>C (p.Q229P) (Chapter 4). *CDH2* encodes the N-cadherin protein which is known to play a role in cell adhesion[106], similar to the desmosomal genes which have been widely implicated in ARVC. Mouse models have demonstrated that deletion of *CDH2* causes embryonic lethality due to a cardiac defect[110], while a cardiac-specific knockout model displays disorganisation of the intercalated disc structure[106], indicating the importance of this gene in cardiac structure and function. As the *CDH2* variant was detected in the ACM2 family, it is possible that this variant may cause ARVC in other individuals. We therefore screened the rest of our ARVC cohort for possible disease-associated variants in this gene.

5.2. Objective

The objective of this study was to screen the ARVC cohort for possible disease-associated variants in the *CDH2* gene

5.3. Specific aim

- To screen our ARVC cohort for variants in *CDH2* by high resolution melt analysis and DNA sequencing.

5.4. Methods

Patient and control samples were obtained as described previously (section 2.5.1.). DNA isolation was performed as previously described (section 2.5.2.). HRM analysis was performed as described in section 2.5.4. The primers used for the combined real-time PCR and HRM of *CDH2* are shown in Table 5.1; the reagents are shown in Table 5.2 and the reaction conditions in Table 5.3.

Purification and sequencing of HRM products was performed as described in section 2.5.5., and population screens were performed as described in section 2.5.6. The methods used for bioinformatic analysis of variants of interest have been described in section 2.5.7.

Table 5.1: Primers for the PCR and HRM of the *CDH2* gene

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE
CDH2_Exon1_F	ATC AGC TCG CTC TCC ATT G	19	52.6	54.8	481
CDH2_Exon1_R	AGG AAA GGC GCG AGA GAC CTA C	22	59.1	60.8	
CDH2_Exon2_F	GCT TTA TGC CAA GTG GAA TG	20	45	52.1	323
CDH2_Exon2_R	AGT CTC ATC ATA GTG ATT GTG	21	38.1	49.5	
CDH2_Exon3_F	ACT TCT GTG TTT CAT GCT AC	20	40	50.3	416
CDH2_Exon3_R	TTA GTT CAA ACT GTA TAA AGC TG	23	30.4	48.8	
CDH2_Exon4_F	ACT GTG ATT CCT ATG CTT TCA G	22	40.9	52.6	402
CDH2_Exon4_R	GAG GCT TTC TAC AAC ACT AC	20	45	50.3	
CDH2_Exon5_F	AGG GCA TGA AGT GTG ATG TG	20	50	55.3	322
CDH2_Exon5_R	CAG ACT GAT ATA AGA GGC AAT G	22	40.9	50.6	
CDH2_Exon6_F	ATG ATG TAT TTA TCT TCT CAC AGA C	25	32	50.6	506
CDH2_Exon6_R	CAA CTA AAG CAG AAC GAA AGG	21	42.9	51.9	
CDH2_Exon7_F	CTA TTT TGG GGT CCC TGG AAT G	22	50	55.6	358
CDH2_Exon7_R	GAA AAC GAT TAG CAT TTG AAT AAC	24	29.2	49.2	
CDH2_Exon8_F	GAA TGA TGC CAT GCT CTC CTT G	22	50	56.2	384
CDH2_Exon8_R	GAA ATG TCC GAG TTA GCA GCC	21	52.4	56	
CDH2_Exon9_F	GGT GTG GTC ATC AAG ACA AG	20	50	53.4	421
CDH2_Exon9_R	AAC AGT AAG ATA TAA CCT CCC AG	23	39.1	51.5	
CDH2_Exon10.1_F	TTT CCA GAA ACT TGA GAG GAA C	22	40.9	52.7	415
CDH2_Exon10.1_R	CAC TAC AAA CAG CAT TTC ATA CC	23	39.1	52.2	
CDH2_Exon10.2_F	TGT CTG TTA CAG TTA TTG AC	20	35	47.2	455
CDH2_Exon10.2_R	TGT TTG GCA TAA AGG CAC	18	44.4	50.9	

Table 5.1: Primers for the PCR and HRM of the *CDH2* gene (continued)

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)	PRODUCT SIZE
CDH2_Exon11_F	TAT GTC CAG CCA TCA TCT ACA G	22	45.5	53.8	445
CDH2_Exon11_R	TGC AAA GAC AGT GAA GGA TGT TC	23	43.5	55.6	
CDH2_Exon12_F	TTT CAC CAT AAT TGA GGC TC	20	40	50.2	363
CDH2_Exon12_R	TCA TGC CAG GCT TCA AAA TC	20	45	53.6	
CDH2_Exon13_F	ACA TCC TGA ACA ATG CTA TGT G	22	40.9	53.1	433
CDH2_Exon13_R	GGA AAT GCT GAA AGG TCA AAG	21	42.9	52.2	
CDH2_Exon14_F	ATG TGA GGC CAG AAA TAG ATA G	22	40.9	51.8	366
CDH2_Exon14_R	TGA TGT ATT AAC ACT TCC CAC	21	38.1	49.9	
CDH2_Exon15_F	TTC CTC ATT TCT ATG TGT TCT G	22	36.4	50.2	354
CDH2_Exon15_R	TAG GCC ACA AAT TGC TTG GAA TAC	24	41.7	55.8	
CDH2_Exon16_F	TCA TTA GGA TCT GCT TGT GG	20	45	52	411
CDH2_Exon16_R	TAT GCC AAA GCC TCC AGC AAG	21	52.4	57.9	

Table 5.2: Reagents for the PCR and HRM of the variants of interest

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME PER REACTION
Forward Primer (20 µM)	0.8 µM / 0.5 µl
Reverse Primer (20 µM)	0.8 µM / 0.5 µl
dNTPs (20 µM) (Bioline)	0.8 µM / 1 µl
GoTaq Polymerase (5U/ ul) (Promega)	0.5 U / 0.1 µl
GoTaq FlexiBuffer (5X) (Promega)	1 X / 5 µl
$MgCl_2$ (25 mM) (Promega)	1.5 mM / 3 µl
EvaGreen dye (Anatech)	1X / 1 µl
DNA	200 ng
FINAL REACTION VOLUME	25 µl

Table 5.3: Optimised temperature cycling conditions for the PCR and HRM of the variants of interest

CONDITION	TEMPERATURE (TIME)	
Initial denaturation	95°C – 10 seconds	
Denaturation Primer Annealing Template Elongation	95°C – 5 seconds *55°C – 10 seconds 72°C – 10 seconds	50 cycles
High Resolution Melt	72 – 95°C (0.1°C increments)	

**CDH2* Exons 2, 4 and 7 were amplified at an annealing temperature of 50°C while all other exons were amplified at an annealing temperature of 55°C.

5.5. Results

5.5.1. Sample characteristics

Between January 2004 and September 2015, 85 unrelated individuals diagnosed with definite (45), borderline (19) and possible (21) ARVC were recruited for molecular genetic analysis (Table 5.4). The majority of these patients were male (65.9%), while the average age of disease onset for this cohort was 37 (±15.2) years. The ethnic origin of these 85 index cases was predominantly Caucasian (50.6%). Palpitations were the most frequent presenting symptom (72.9%), followed by dizziness (58.8%), chest pain/discomfort (36.5%) and syncope (32.9%).

Table 5.4: Clinical characteristics of the ARVC cohort

CHARACTERISTIC	VALUE	%
NO. OF PATIENTS	85	100
GENDER		
Male	56	65.9
Female	29	34.1
AVERAGE AGE IN YEARS (±SD)	46 (±15.7)	
AVERAGE AGE OF ONSET IN YEARS (±SD))	37 (±15.2)	
ETHNICITY		
Black	6	7.1
Mixed ancestry	19	22.4
Caucasian	43	50.6
Indian	3	3.5
Not recorded	14	16.5
PRESENTING SYMPTOMS		
Palpitations	62	72.9
Dizziness	50	58.8
Chest pain/discomfort	31	36.5
Syncope	28	32.9
Dyspnoea	25	29.4
Sudden cardiac death	6	7.1
Asymptomatic	3	3.5
HISTORY OF FAMILIAL DISEASE	17	20.0
DIAGNOSIS		
Definite	45	52.9
Borderline	19	22.4
Possible	21	24.7

SD – Standard deviation

5.5.2. *CDH2* screening

We screened a cohort of 85 ARVC patients for variants in the *CDH2* gene by high resolution melt analysis and Sanger sequencing. A cohort of 200 ethnically matched control chromosomes was screened for all variants of interest. We detected a novel *CDH2* c.1219G>A (p.D407N) variant in ACM11.2, an individual with severe ARVC. No other variants of interest were detected.

5.5.2.1. CDH2 c.1219G>A (p.D407N)

We detected the c.1219G>A (p.D407N) variant in patient ACM11.2 (Figure 5.1). This variant was absent in the South African Caucasian background population (as determined by the population screen); it was also absent in the Exome Variant Server (http://evs.gs.washington.edu/EVS/), NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and Exome Aggregation Consortium (ExAC) browser (http://exac.broadinstitute.org/) databases, but was reported in the 1000Genomes database (http://browser.1000genomes.org/index.html) in reference to its presence in the COSMIC database of cancerous tissue variants (minor allele frequency (MAF) < 0.01%), but has not directly been associated with any disease to date. It was predicted to be disease-causing by MutationTaster (Disease-causing – 0.999), Polyphen-2 (Probably damaging – 1.00) and MutationAssessor (Disease-causing – 4.225) tools while it was predicted to be tolerated by the Align GVGD (Class C15) and SIFT (0.12) tools. Bioinformatic tools Mfold and RNAFold predicted that this variant would alter *CDH2* mRNA secondary structure (Figure 5.3).

This variant was not found in individual ACM11.1, the unaffected mother of ACM11.2, indicating that the variant was either (1) inherited from the proband's father or (2) a *de novo* variant which arose in ACM11.2 (Figure 5.2). ACM11.2 was diagnosed with severe ARVC at 15 years of age. According to the 2010 ARVC Task Force criteria, he fulfils major criteria for depolarisation abnormalities (TWI V1-V6), an epsilon wave and structural cardiac changes. He also fulfils minor criteria for late potentials and arrhythmia, and also displayed fibro-fatty tissue infiltration of the myocardium on biopsy. He also has an implantable cardioverter defibrillator.

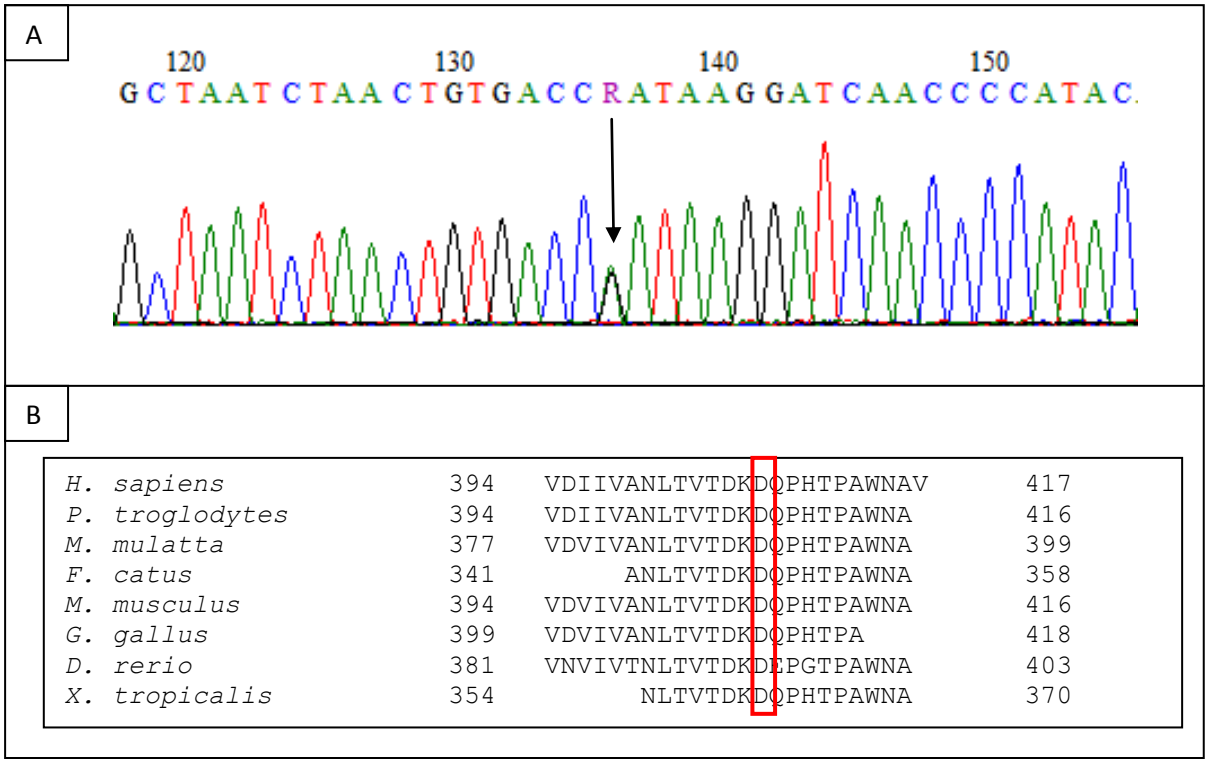


Figure 5.1: Identification and conservation of the *CDH2* c.1219G>A (p.D407N) variant. (A) Electropherogram indicating nucleotide change; (B) Multiple species protein sequence conservation

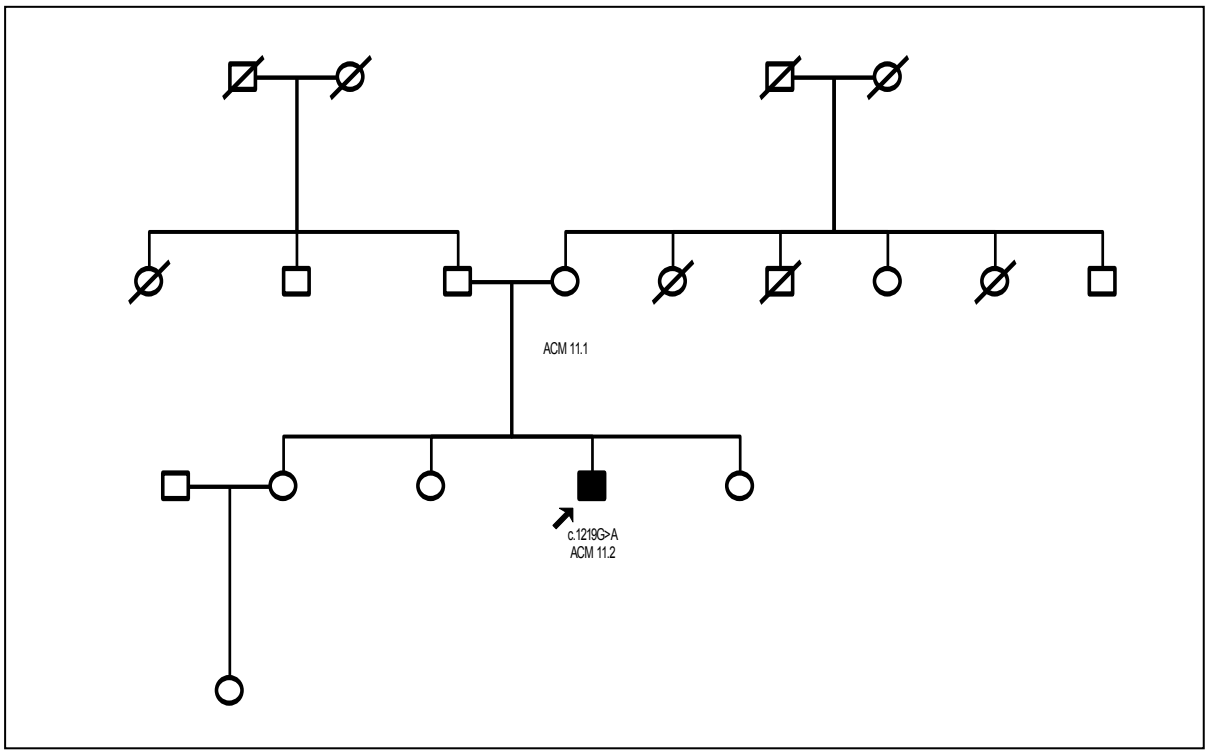


Figure 5.2: ACM11 family pedigree indicating segregation of the *CDH2* c.1219G>A variant

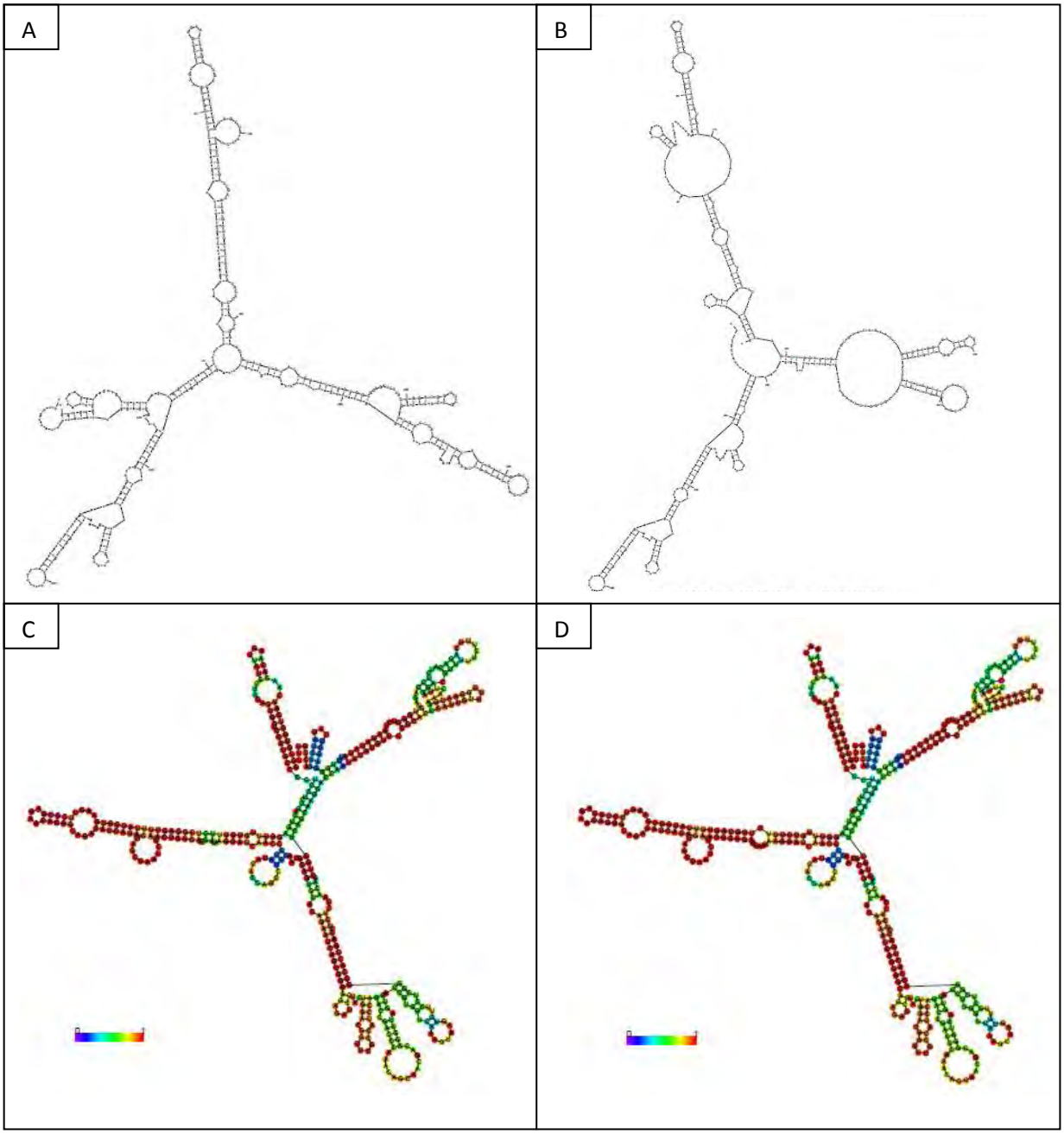


Figure 5.3: Mfold and RNAfold predictions for the *CDH2* c.1219G>A variant. (A) Mfold wild type structure; (B) Mfold mutant structure; (C) RNAfold wild type structure; (D) RNAfold wild type structure

Based on the fact that this variant (1) was absent from all control populations, (2) alters a conserved amino acid and (3) is bioinformatically predicted to be deleterious, this variant is classified as a disease-associated genetic variant.

5.5.2.2. Other CDH2 variants

We have detected many common variations in *CDH2* in our ARVC cohort which had previously been found in individuals from the 1000Genomes project. We did not detect any other possibly pathogenic variants or genetic variants of unknown significance in the ARVC cohort.

5.6. Discussion

Based on genetic and bioinformatic evidence to date, the most likely disease-causing variant in the ACM2 family appears to be the *CDH2* c.686A>C (p.Q229P) variant (Figure 5.4). This variant segregates with disease, is absent in the control population and control databases and is bioinformatically predicted to have a deleterious effect on the function of CDH2. We therefore screened a cohort of 85 ARVC patients for variants in the *CDH2* gene and detected the possibly pathogenic c.1219G>A (p.D407N) variant in ACM11.2 (Figure 5.4).

The *CDH2* gene encodes the N-cadherin protein, which is a member of the cadherin superfamily of predominantly calcium-dependent cell adhesion proteins.[107] As discussed in chapter 4, multiple studies have shown the importance of CDH2 in cell adhesion. N-cadherin is comprised of five extracellular cadherin repeats, a transmembrane region and a highly conserved cytoplasmic tail.[108] The cadherin repeats are vital for the adhesive function of N-cadherin. Calcium ions bind to each cadherin repeat to ensure correct protein folding and confer rigidity to the extracellular domain. Multiple cadherin domains form Ca^{2+}-dependent rod-like structures with conserved Ca^{2+}-binding pockets at the inter-domain region.[109] The p.Q229P variant is located in cadherin repeat 1 while the p.D407N variant is located in cadherin repeat 3 (Figure 5.4). It is therefore plausible that these variants in the cadherin repeat domains could affect the adhesive function of the N-cadherin protein.

CDH2 is known to play a vital role in the maintenance of intercalated disc structure[106], and is also known to be directly involved with the Wnt/β-catenin and hippo signalling pathways.[33] Desmosomal proteins that have the same biological functions have been widely implicated in ARVC, making *CDH2* a strong candidate for playing a causal role in this disease. It is therefore plausible that variants in *CDH2* could lead to the development of ARVC by disruption of the intercalated disc structure and modulation of the Wnt/β-catenin signalling

pathway, causing myocyte instability, myocyte death and replacement of the myocardium by fibrous fatty tissue.

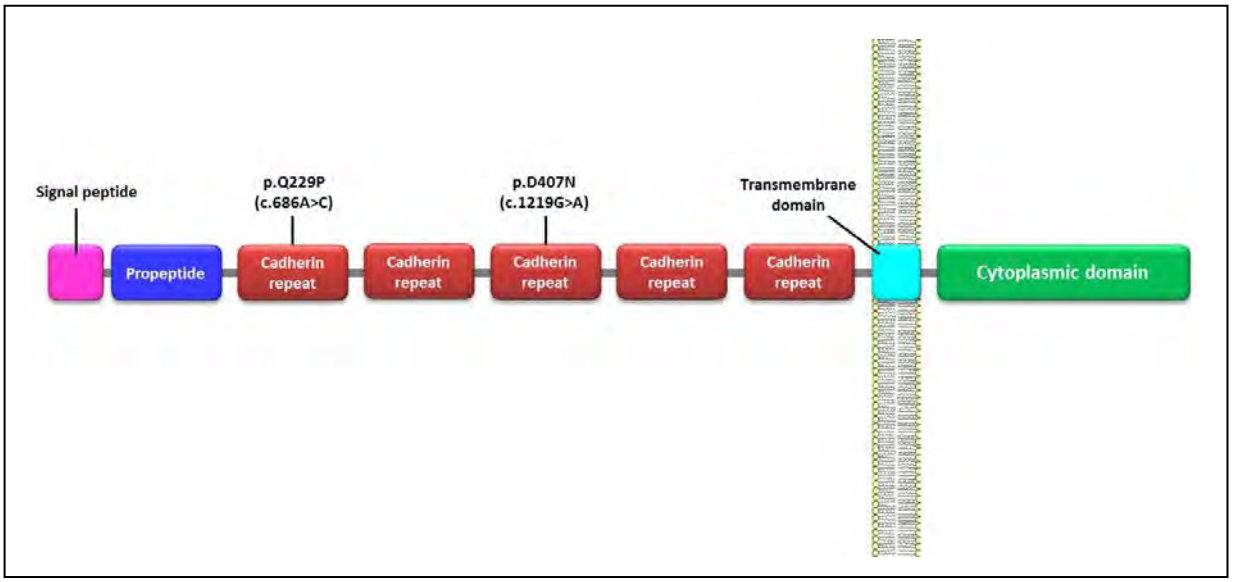


Figure 5.4: Position of detected pathogenic variants in the CDH2 protein

As discussed in chapter 4, several *CDH2* knockout mouse models have been described. The first is a homozygous *Cdh2* knockout mouse model; the mice died in the embryonic stage due to a severe cardiac defect, indicating the importance of *Cdh2* in cardiac development and function.[110] The second is a heterozygous *Cdh2* knockout mouse model which exhibited a reduction in Connexin-43 plaque size in the myocardium as well as increased susceptibility to arrhythmias, showing that N-cadherin influences the function of gap junctions.[114] This suggests that disruption of one copy of *Cdh2* (as in the case of a heterozygous variant or knockout), could lead to an arrhythmic phenotype, which is characteristic of ARVC. The third model is mice with induced cardiac-specific deletion of *Cdh2* displayed dissolution of the intercalated disc structure including desmosomes and adherens junctions[106], a similar observation to what is observed in patients with ARVC. These mouse models also had dilated cardiomyopathy and the abrupt onset of ventricular tachycardia, indicating that the deaths were arrhythmic in nature.[106]

There is consequently overwhelming evidence that disruption of *Cdh2* in rodent models leads to a cardiac phenotype – specifically, an ARVC phenotype. However, mouse models examining the effects of specific *CDH2* variants have not been described, as disease-causing variants in this gene have not previously been reported.

As discussed in chapter 4, this is, to our knowledge, the first report of variants in *CDH2* associated with human disease. The frequency of *CDH2* variants in our ARVC cohort is 2/85 (2.4%). Individuals with *CDH2* variants appear to have a severe ARVC phenotype, highlighting the importance of including this gene in routine genetic screening of ARVC patients.

There is strong evidence that *CDH2* may be a novel causal gene for ARVC based on (1) segregation of *CDH2* variants with disease in the affected members of the ACM2 family, (2) absence of these variants in normal control populations, (3) alteration of conserved amino acids by these variants, (4) biological plausibility based on gene function, (5) bioinformatic predictions of pathogenicity of the variants, and (6) evidence from animal models that disruption of *CDH2* leads to an ARVC phenotype.

5.7. Conclusion

We detected pathogenic *CDH2* variants in two unrelated probands with ARVC (ACM2 and ACM11), implicating *CDH2* in the pathogenesis of ARVC. This is the first report of *CDH2* variants implicated in heart disease, and emphasises the need to include *CDH2* in genetic screening of ARVC patients.

CHAPTER 6
SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

6.1. Summary of findings

The main objective of this study was to identify the genetic cause of arrhythmogenic right ventricular cardiomyopathy (ARVC) in the ACM2 family with autosomal dominant disease.

6.1.1. Candidate gene screening: *PLN*

We used the candidate gene screening approach and screened the ACM2 family for variants in *PLN*, but did not detect any disease-causing variants. Simultaneously, we screened a cohort of 85 other ARVC patients as well as patients with DCM, HCM or PPCM for variants in *PLN* and detected the reported pathogenic *PLN* c.25C>T (p.R9C) variant in the DCM320 family. No other disease-causing *PLN* variants were detected.

6.1.2. CNV analysis

We then performed whole genome copy number variation (CNV) analysis to search for large changes in DNA copy number which could be associated with ARVC, but detected no deleterious CNVs. There have been limited reports of CNVs in inherited cardiomyopathies to date. These CNVs reported as associated with cardiomyopathy are limited to deletions in *BAG3*[116] and *LMNA*[117] in individuals with DCM, deletions in *PKP2*[92,93] and *DSP*[94] in individuals with ARVC and a duplication in *MYBPC3*[118] in individuals with HCM. Given the rarity of pathogenic CNVs in cardiomyopathy, it is not surprising that we failed to detect disease-causing CNVs in the ACM2 family.

6.1.3. Exome sequencing

We performed exome sequencing for two affected members of the ACM2 family. Family screening, population screening and bioinformatic analysis pinpointed the novel deleterious *CDH2* c.686A>C (p.Q229P) variant as the disease-associated variant in the ACM2 family.

6.1.4. Candidate gene screening: *CDH2*

We screened 85 ARVC patients for variants in *CDH2*, and identified the deleterious *CDH2* c.1219G>A (p.D407N) in the ACM11 family. This provides additional evidence for the role of *CDH2* in the pathogenesis of ARVC.

6.2. Conclusions

There have been many reports of genetic factors associated with cardiomyopathy to date. In ARVC, the focus was initially predominantly on the desmosomal genes, which are now known to be responsible for 30-50% of ARVC cases. More recently, other genes with diverse functions have been implicated in this disease. However, the amount of evidence for the involvement of these genes in ARVC is varied, and it has become increasingly challenging to prove the pathogenicity of variants thought to be deleterious.

6.2.1. New sequencing methods facilitate faster variant identification

Early genetic studies, including those involving cardiomyopathy patients, adopted a candidate gene approach for variant identification. Genes of interest were selected based on linkage information or current knowledge about gene function, and the sequencing of these genes was performed by the Sanger sequencing method. Newer methods allow the simultaneous targeting of multiple genes, facilitating a more efficient and less time-consuming approach to pinpointing disease-causing variants. These methods include targeted sequencing of multiple genes of interest, exome sequencing and whole genome sequencing with various next generation sequencing platforms.

6.2.2. Importance of stringent variant classification criteria

In recent years, the number of studies reporting genetic variants in patients with cardiomyopathy has increased dramatically. The advent of next generation sequencing has led to faster identification of variants in patients with various diseases, including cardiomyopathy. This increase in the identification of variants against the background of large amounts of other genetic information from next generation sequencing experiments

has made it more challenging to separate truly causal variants from variants that are associated with disease but are not necessarily pathogenic.

The key to ensuring correct variant classification is a stringent set of criteria which will allow unbiased assessment of any variant and provides the most accurate determination of variant pathogenicity. Using more recent classification criteria, many variants previously deemed pathogenic have been reclassified as variants of unknown significance, as the original evidence provided for causality is no longer substantial according to more recent criteria. In our study, we have used a very stringent set of criteria for variant classification, and variants which we have classified as pathogenic have strong evidence for causality or association with disease.

6.2.3. Importance of accurate phenotyping information

During the course of this project, the importance of current and comprehensive phenotype information was highlighted. Correct classification of patients according to the most recent ARVC Task Force Criteria and based on thorough examination was of great importance in pinpointing the causal variant in the ACM2 family. Uncertainty about the classification of patients due to incomplete patient information causes difficulty when attempting to establish genotype-phenotype relationships. Much care should be taken to ensure rigorous and unbiased phenotyping of patients before inclusion in a genetic study.

6.2.4. The ACM2 family

There has been much work performed on the ACM2 family to date, spanning multiple experiments over a number of years, but no disease-causing variant had been found. In this study, we identified the pathogenic *CDH2* c.686A>C (p.Q229P) variant by exome sequencing. *CDH2* screening of a cohort of 85 other ARVC probands revealed the c.1219G>A (p.D407N) variant, providing further evidence of the contribution of *CDH2* variants to the pathogenesis of ARVC. This is the first report of *CDH2* variants associated with human disease, and establishes this gene as a novel cause of ARVC.

Previous studies on African cardiomyopathy patients implicated the *PKP2* gene in ARVC.[13,22] Variants in *PKP2* included the c.1162C>T (p.R388W) founder variant present in many South African individuals of European descent.[13,22] The identification of *CDH2* as a new ARVC-causing gene in the South Africans with cardiomyopathy sets the stage for inclusion of *CDH2* in routine cardiomyopathy screening both locally and internationally.

6.2.5. Limitations

The different techniques utilised in this study have a number of limitations. The candidate gene screening approach used in chapter 2 is limited in scope as only the *PLN* gene is evaluated. The copy number variant screening approach used in chapter 3 evaluates large genetic variations (>1kb) and is unable to detect smaller genetic changes. The performance of an exome sequencing experiment has addressed the limitations of these techniques. However, exome sequencing does not cover the intergenic and the majority of intronic regions of the genome, but as the inheritance pattern of the ACM2 family is autosomal dominant, it is most likely that the disease-causing variant would be located in the exonic region as has been found in this study.

Another limitation of this study arises with the use of bioinformatic tools to predict the effect of variants on gene function and their consequent likelihood of playing a causal role in disease development. These tools predict the effect of genetic variants based on existing information about the structure, function and conservation of the affected genes, transcripts and proteins.[119] However, as these tools determine variant effect severity in different ways, use of more than one tool is recommended to obtain an overall predictive result and ensure optimal sensitivity and specificity, as we have done in this study. Results obtained with these tools should however be interpreted with caution and require validation with functional experiments.

6.2.6. Future directions

There are several lines of investigation that are required to establish the significance of variants in the *CDH2* gene in cardiomyopathy. First, a variant screening study of the *CDH2* gene is required in a large series of cases and families of ARVC and other forms of cardiomyopathy. These studies will serve to confirm the causal role of *CDH2* variants in cardiomyopathy and determine the prevalence of variants in this gene in patients with cardiomyopathy. Second, studies of the expression of the desmosomal proteins in heart tissues of affected and unaffected individuals will be informative in establishing the role of variants in *CDH2* in human disease.[120] In addition to studies of the expression of normal and mutant protein in human tissues, *in vitro* studies of the interaction of CDH2 with other proteins using systems such as the yeast-2-hybrid system can be conducted. Other experiments to assess the functional impact of the *CDH2* c.686A>C and c.1219G>A variants may involve the use of animal models and induced pluripotent stem cells.[121]

Although mouse models with deletions of one or both copies of *CDH2*, as well as cardiac-specific knockouts, have been established[106,110,114] and have demonstrated that a decrease in *CDH2* levels leads to an ARVC-like phenotype, there are no reported studies assessing the functional impact of *CDH2* point variants. Studies that will establish the mechanism of disease caused by *CDH2* variants will elucidate the molecular pathways involved in the development of ARVC and other related disorders.

REFERENCES

1 Gaziano, T. A. Reducing the growing burden of cardiovascular disease in the developing world. *Health affairs* **26**, 13-24, doi:10.1377/hlthaff.26.1.13 (2007).

2 Yusuf, S., Hawken, S. & Ounpuu, S. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study (vol 364, pg 937m 2004). *Lancet* **364**, 2020-2020 (2004).

3 Damasceno, A. *et al.* The causes, treatment, and outcome of acute heart failure in 1006 Africans from 9 countries. *Archives of internal medicine* **172**, 1386-1394, doi:10.1001/archinternmed.2012.3310 (2012).

4 Sliwa, K., Damasceno, A. & Mayosi, B. M. Epidemiology and etiology of cardiomyopathy in Africa. *Circulation* **112**, 3577-3583, doi:10.1161/CIRCULATIONAHA.105.542894 (2005).

5 Elliott, P. *et al.* Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. *European heart journal* **29**, 270-276, doi:10.1093/eurheartj/ehm342 (2008).

6 Wexler, R. K., Elton, T., Pleister, A. & Feldman, D. Cardiomyopathy: an overview. *American family physician* **79**, 778-784 (2009).

7 Sen-Chowdhry, S., Morgan, R. D., Chambers, J. C. & McKenna, W. J. Arrhythmogenic cardiomyopathy: etiology, diagnosis, and treatment. *Annual review of medicine* **61**, 233-253, doi:10.1146/annurev.med.052208.130419 (2010).

8 Awad, M. M., Calkins, H. & Judge, D. P. Mechanisms of disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Nature clinical practice. Cardiovascular medicine* **5**, 258-267, doi:10.1038/ncpcardio1182 (2008).

9 Kies, P., Bootsma, M., Bax, J., Schalij, M. J. & van der Wall, E. E. Arrhythmogenic right ventricular dysplasia/cardiomyopathy: screening, diagnosis, and treatment. *Heart rhythm : the official journal of the Heart Rhythm Society* **3**, 225-234, doi:10.1016/j.hrthm.2005.10.018 (2006).

10 Hershberger, R. E., Hedges, D. J. & Morales, A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nature reviews. Cardiology* **10**, 531-547, doi:10.1038/nrcardio.2013.105 (2013).

11 Paul, M. *et al.* Arrhythmogenic right ventricular cardiomyopathy: an update on pathophysiology, genetics, diagnosis, and risk stratification. *Herzschrittmachertherapie & Elektrophysiologie* **23**, 186-195, doi:10.1007/s00399-012-0233-7 (2012).

12 Munclinger, M. J., Patel, J. J. & Mitha, A. S. Follow-up of patients with arrhythmogenic right ventricular cardiomyopathy dysplasia. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* **90**, 61-68 (2000).

13 Watkins, D. A. *et al.* 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 : the official journal of the Heart Rhythm Society* **6**, S10-17, doi:10.1016/j.hrthm.2009.08.018 (2009).

14 McKenna, W. J. *et al.* Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. *British heart journal* **71**, 215-218 (1994).

15 Marcus, F. I. *et al.* Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the Task Force Criteria. *European heart journal* **31**, 806-814, doi:10.1093/eurheartj/ehq025 (2010).

16 Corrado, D., Basso, C. & Thiene, G. Arrhythmogenic right ventricular cardiomyopathy: an update. *Heart* **95**, 766-773, doi:10.1136/hrt.2008.149823 (2009).

17 Basso, C., Corrado, D., Marcus, F. I., Nava, A. & Thiene, G. Arrhythmogenic right ventricular cardiomyopathy. *Lancet* **373**, 1289-1300, doi:10.1016/S0140-6736(09)60256-7 (2009).

18 Rigato, I. *et al.* Compound and digenic heterozygosity predicts lifetime arrhythmic outcome and sudden cardiac death in desmosomal gene-related arrhythmogenic right ventricular cardiomyopathy. *Circulation. Cardiovascular genetics* **6**, 533-542, doi:10.1161/CIRCGENETICS.113.000288 (2013).

19 Saffitz, J. E. Arrhythmogenic Cardiomyopathy Advances in Diagnosis and Disease Pathogenesis. *Circulation* **124**, E390-E392, doi:10.1161/Circulationaha.111.064022 (2011).

20 Basso, C., Corrado, D., Bauce, B. & Thiene, G. Arrhythmogenic right ventricular cardiomyopathy. *Circulation. Arrhythmia and electrophysiology* **5**, 1233-1246, doi:10.1161/CIRCEP.111.962035 (2012).

21 Te Rijdt, W. P. *et al.* Clinical utility gene card for: arrhythmogenic right ventricular cardiomyopathy (ARVC). *European journal of human genetics : EJHG* **22**, doi:10.1038/ejhg.2013.124 (2014).

22 Mbele, M. *Molecular Genetics of Arrhythmogenic Right Ventricular and Dilated Cardiomyopathy in South Africans* PhD thesis, University of Cape Town, (2014).

23 Xu, T. *et al.* Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy. *Journal of the American College of Cardiology* **55**, 587-597, doi:10.1016/j.jacc.2009.11.020 (2010).

24 Merner, N. D. *et al.* Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *American journal of human genetics* **82**, 809-821, doi:10.1016/j.ajhg.2008.01.010 (2008).

25 Baskin, B. *et al.* TMEM43 mutations associated with arrhythmogenic right ventricular cardiomyopathy in non-Newfoundland populations. *Human genetics* **132**, 1245-1252, doi:10.1007/s00439-013-1323-2 (2013).

26 Tiso, N. *et al.* Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Human molecular genetics* **10**, 189-194 (2001).

27 Beffagna, G. *et al.* Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovascular research* **65**, 366-373, doi:10.1016/j.cardiores.2004.10.005 (2005).

28 van der Zwaag, P. A. *et al.* Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. *European journal of heart failure* **14**, 1199-1207, doi:10.1093/eurjhf/hfs119 (2012).

29 van Hengel, J. *et al.* Mutations in the area composita protein alphaT-catenin are associated with arrhythmogenic right ventricular cardiomyopathy. *European heart journal* **34**, 201-210, doi:10.1093/eurheartj/ehs373 (2013).

30 Klauke, B. *et al.* De novo desmin-mutation N116S is associated with arrhythmogenic right ventricular cardiomyopathy. *Human molecular genetics* **19**, 4595-4607, doi:10.1093/hmg/ddq387 (2010).

31 Taylor, M. *et al.* Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation* **124**, 876-885, doi:10.1161/CIRCULATIONAHA.110.005405 (2011).

32 Quarta, G. *et al.* Mutations in the Lamin A/C gene mimic arrhythmogenic right ventricular cardiomyopathy. *European heart journal* **33**, 1128-1136, doi:10.1093/eurheartj/ehr451 (2012).

33 Chen, S. N. *et al.* The hippo pathway is activated and is a causal mechanism for adipogenesis in arrhythmogenic cardiomyopathy. *Circulation research* **114**, 454-468, doi:10.1161/CIRCRESAHA.114.302810 (2014).

34 von Gise, A. *et al.* YAP1, the nuclear target of Hippo signaling, stimulates heart growth through cardiomyocyte proliferation but not hypertrophy. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2394-2399, doi:10.1073/pnas.1116136109 (2012).

35 Xin, M. *et al.* Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size. *Science signaling* **4**, ra70, doi:10.1126/scisignal.2002278 (2011).

36 Heallen, T. *et al.* Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science* **332**, 458-461, doi:10.1126/science.1199010 (2011).

37 Hu, Y. & Pu, W. T. Hippo activation in arrhythmogenic cardiomyopathy. *Circulation research* **114**, 402-405, doi:10.1161/CIRCRESAHA.113.303114 (2014).

38 Rampazzo, A. *et al.* The gene for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 14q23-q24. *Human molecular genetics* **3**, 959-962 (1994).

39 Rampazzo, A. *et al.* A new locus for arrhythmogenic right ventricular cardiomyopathy (ARVD2) maps to chromosome 1q42-q43. *Human molecular genetics* **4**, 2151-2154 (1995).

40 Severini, G. M. *et al.* A new locus for arrhythmogenic right ventricular dysplasia on the long arm of chromosome 14. *Genomics* **31**, 193-200 (1996).

41 Rampazzo, A. *et al.* ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy, maps to chromosome 2 long arm. *Genomics* **45**, 259-263, doi:10.1006/geno.1997.4927 (1997).

42 Ahmad, F. *et al.* Localization of a gene responsible for arrhythmogenic right ventricular dysplasia to chromosome 3p23. *Circulation* **98**, 2791-2795 (1998).

43 Li, D. *et al.* The locus of a novel gene responsible for arrhythmogenic right-ventricular dysplasia characterized by early onset and high penetrance maps to chromosome 10p12-p14. *American journal of human genetics* **66**, 148-156, doi:10.1086/302713 (2000).

44 Matolweni, L. O. *et al.* Arrhythmogenic right ventricular cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes. *BMC medical genetics* **7**, 29, doi:10.1186/1471-2350-7-29 (2006).

45 Melberg, A. *et al.* Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy linked to chromosome 10q. *Annals of neurology* **46**, 684-692 (1999).

46 Rampazzo, A. *et al.* Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. *American journal of human genetics* **71**, 1200-1206, doi:10.1086/344208 (2002).

47 Gerull, B. *et al.* Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet* **36**, 1162-1164, doi:10.1038/ng1461 (2004).

48 Pilichou, K. *et al.* Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy. *Circulation* **113**, 1171-1179, doi:10.1161/CIRCULATIONAHA.105.583674 (2006).

49 Syrris, P. *et al.* Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. *American journal of human genetics* **79**, 978-984, doi:10.1086/509122 (2006).

50 Coonar, A. S. *et al.* Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21. *Circulation* **97**, 2049-2058 (1998).

51 McKoy, G. *et al.* Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet* **355**, 2119-2124, doi:10.1016/S0140-6736(00)02379-5 (2000).

52 Mokhobo, K. P. & Mntla, P. S. Lone ventricular cardiomyopathy, 1993-1996. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* **87**, 892-896 (1997).

53 Fish, M. *Analysis of Desmoplakin in Arrhythmogenic Right Ventricular Cardiomyopathy* MSc thesis, University of Cape Town, (2010).

54 van der Zwaag, P. A. *et al.* Recurrent and founder mutations in the Netherlands-Phospholamban p.Arg14del mutation causes arrhythmogenic cardiomyopathy. *Netherlands heart journal : monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation* **21**, 286-293, doi:10.1007/s12471-013-0401-3 (2013).

55 Blanckenberg, J. *Molecular genetics of arrhythmogenic right ventricular cardiomyopathy in South Africa* PhD thesis, University of Cape Town, (2011).

56 Sliwa, K. & Mayosi, B. M. Recent advances in the epidemiology, pathogenesis and prognosis of acute heart failure and cardiomyopathy in Africa. *Heart*, doi:10.1136/heartjnl-2013-303592 (2013).

57 Cahill, T. J., Ashrafian, H. & Watkins, H. Genetic cardiomyopathies causing heart failure. *Circulation research* **113**, 660-675, doi:10.1161/CIRCRESAHA.113.300282 (2013).

58 Latib, M., Michael, KA, Mayosi, BM. Initial Report of the Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa: 2004 [Abstract]. *Cardiovascular Journal of South Africa* **15**, 237 & 238 (2004).

59 Haghighi, K. *et al.* Human phospholamban null results in lethal dilated cardiomyopathy revealing a critical difference between mouse and human. *The Journal of clinical investigation* **111**, 869-876, doi:10.1172/JCI17892 (2003).

60 Haghighi, K. *et al.* A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1388-1393, doi:10.1073/pnas.0510519103 (2006).

61 Schmitt, J. P. *et al.* Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science* **299**, 1410-1413, doi:10.1126/science.1081578 (2003).

62 Truszkowska, G. T. *et al.* A study in Polish patients with cardiomyopathy emphasizes pathogenicity of phospholamban (PLN) mutations at amino acid position 9 and low penetrance of heterozygous null PLN mutations. *BMC medical genetics* **16**, 21, doi:10.1186/s12881-015-0167-0 (2015).

63 Medin, M. *et al.* Mutational screening of phospholamban gene in hypertrophic and idiopathic dilated cardiomyopathy and functional study of the PLN -42 C>G mutation. *European journal of heart failure* **9**, 37-43, doi:10.1016/j.ejheart.2006.04.007 (2007).

64 Minamisawa, S. *et al.* Mutation of the phospholamban promoter associated with hypertrophic cardiomyopathy. *Biochemical and biophysical research communications* **304**, 1-4 (2003).

65 Ha, K. N. *et al.* Lethal Arg9Cys phospholamban mutation hinders Ca2+-ATPase regulation and phosphorylation by protein kinase A. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2735-2740, doi:10.1073/pnas.1013987108 (2011).

66 MacLennan, D. H. & Kranias, E. G. Phospholamban: a crucial regulator of cardiac contractility. *Nature reviews. Molecular cell biology* **4**, 566-577, doi:10.1038/nrm1151 (2003).

67 Hendricks, N., Watkins, D. A. & Mayosi, B. M. Lessons from the first report of the Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa. *Cardiovascular journal of Africa* **21**, 129-130 (2010).

68 Hamid, M. S. *et al.* Prospective evaluation of relatives for familial arrhythmogenic right ventricular cardiomyopathy/dysplasia reveals a need to broaden diagnostic criteria. *Journal of the American College of Cardiology* **40**, 1445-1450 (2002).

69 Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *Journal of computational biology : a journal of computational molecular cell biology* **4**, 311-323 (1997).

70 Dogan, R. I., Getoor, L., Wilbur, W. J. & Mount, S. M. SplicePort--an interactive splice-site analysis tool. *Nucleic acids research* **35**, W285-291, doi:10.1093/nar/gkm407 (2007).

71 Desmet, F. O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* **37**, e67, doi:10.1093/nar/gkp215 (2009).

72 Hebsgaard, S. M. *et al.* Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic acids research* **24**, 3439-3452 (1996).

73 Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research* **31**, 3568-3571 (2003).

74 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**, 3406-3415 (2003).

75 Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic acids research* **31**, 3429-3431 (2003).

76 Tavtigian, S. V. *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of medical genetics* **43**, 295-305, doi:10.1136/jmg.2005.033878 (2006).

77 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

78 Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome research* **11**, 863-874, doi:10.1101/gr.176601 (2001).

79 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**, 575-576, doi:10.1038/nmeth0810-575 (2010).

80 van Spaendonck-Zwarts, K. Y. *et al.* Peripartum cardiomyopathy as a part of familial dilated cardiomyopathy. *Circulation* **121**, 2169-2175, doi:10.1161/CIRCULATIONAHA.109.929646 (2010).

81 Yu, X. & Lorigan, G. A. Secondary structure, backbone dynamics, and structural topology of phospholamban and its phosphorylated and Arg9Cys-mutated forms in phospholipid bilayers utilizing 13C and 15N solid-state NMR spectroscopy. *The journal of physical chemistry. B* **118**, 2124-2133, doi:10.1021/jp500316s (2014).

82 van Spaendonck-Zwarts, K. Y. *et al.* Titin gene mutations are common in families with both peripartum cardiomyopathy and dilated cardiomyopathy. *European heart journal* **35**, 2165-2173, doi:10.1093/eurheartj/ehu050 (2014).

83 Lee, T. M., Addonizio, L. J. & Chung, W. K. Dilated cardiomyopathy due to a phospholamban duplication. *Cardiology in the young* **24**, 953-954, doi:10.1017/S1047951113002035 (2014).

84 Dunn, K. E. & Ashley, E. A. Arrhythmogenic Right Ventricular Cardiomyopathy: Toward a Modern Clinical and Genomic Understanding. *Circulation. Cardiovascular genetics* **8**, 421-424, doi:10.1161/CIRCGENETICS.115.001119 (2015).

85 Pollex, R. L. & Hegele, R. A. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation* **115**, 3130-3138, doi:10.1161/CIRCULATIONAHA.106.677591 (2007).

86 Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444-454, doi:10.1038/nature05329 (2006).

87 Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature reviews. Genetics* **7**, 85-97, doi:10.1038/nrg1767 (2006).

88 Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annual review of medicine* **61**, 437-455, doi:10.1146/annurev-med-100708-204735 (2010).

89 Padiath, Q. S. *et al.* Lamin B1 duplications cause autosomal dominant leukodystrophy. *Nat Genet* **38**, 1114-1123, doi:10.1038/ng1872 (2006).

90 Le Marechal, C. *et al.* Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* **38**, 1372-1374, doi:10.1038/ng1904 (2006).

91 Lee, J. A. & Lupski, J. R. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52**, 103-121, doi:10.1016/j.neuron.2006.09.027 (2006).

92 Li Mura, I. E. *et al.* Identification of a PKP2 gene deletion in a family with arrhythmogenic right ventricular cardiomyopathy. *European journal of human genetics : EJHG* **21**, 1226-1231, doi:10.1038/ejhg.2013.39 (2013).

93 Roberts, J. D. *et al.* Detection of genomic deletions of PKP2 in arrhythmogenic right ventricular cardiomyopathy. *Clinical genetics* **83**, 452-456, doi:10.1111/j.1399-0004.2012.01950.x (2013).

94 Murray, B. *et al.* A family with a complex clinical presentation characterized by arrhythmogenic right ventricular dysplasia/cardiomyopathy and features of branchio-oculo-facial syndrome. *American journal of medical genetics. Part A* **161A**, 371-376, doi:10.1002/ajmg.a.35733 (2013).

95 Affymetrix. CytoScan® Cytogenetics Solution Information Booklet. (2012).

96 MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research* **42**, D986-992, doi:10.1093/nar/gkt958 (2014).

97 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102. Article published online before print in May 2002 (2002).

98 Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American journal of human genetics* **84**, 524-533, doi:10.1016/j.ajhg.2009.03.010 (2009).

99 Koolen, D. A. *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat Genet* **44**, 639-641, doi:10.1038/ng.2262 (2012).

100 Zollino, M. *et al.* Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet* **44**, 636-638, doi:10.1038/ng.2257 (2012).

101 Tan, T. Y. *et al.* Phenotypic expansion and further characterisation of the 17q21.31 microdeletion syndrome. *Journal of medical genetics* **46**, 480-489, doi:10.1136/jmg.2008.065391 (2009).

102 Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19096-19101, doi:10.1073/pnas.0910672106 (2009).

103 Boland, J. F. *et al.* The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Human genetics* **132**, 1153-1163, doi:10.1007/s00439-013-1321-4 (2013).

104 Labonte, D. *et al.* TRIM3 regulates the motility of the kinesin motor protein KIF21B. *PloS one* **8**, e75603, doi:10.1371/journal.pone.0075603 (2013).

105 Fabre, A. *et al.* SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. *American journal of human genetics* **90**, 689-692, doi:10.1016/j.ajhg.2012.02.009 (2012).

106 Kostetskii, I. *et al.* Induced deletion of the N-cadherin gene in the heart leads to dissolution of the intercalated disc structure. *Circulation research* **96**, 346-354, doi:10.1161/01.RES.0000156274.72390.2c (2005).

107 Knudsen, K. A., Soler, A. P., Johnson, K. R. & Wheelock, M. J. Interaction of alpha-actinin with the cadherin/catenin cell-cell adhesion complex via alpha-catenin. *The Journal of cell biology* **130**, 67-77 (1995).

108 Moya, P. R. *et al.* Rare missense neuronal cadherin gene (CDH2) variants in specific obsessive-compulsive disorder and Tourette disorder phenotypes. *European journal of human genetics : EJHG* **21**, 850-854, doi:10.1038/ejhg.2012.245 (2013).

109 Ivanov, D. B., Philippova, M. P. & Tkachuk, V. A. Structure and functions of classical cadherins. *Biochemistry. Biokhimiia* **66**, 1174-1186 (2001).

110 Luo, Y. *et al.* Rescuing the N-cadherin knockout by cardiac-specific expression of N- or E-cadherin. *Development* **128**, 459-469 (2001).

111 Sacco, P. A., McGranahan, T. M., Wheelock, M. J. & Johnson, K. R. Identification of plakoglobin domains required for association with N-cadherin and alpha-catenin. *The Journal of biological chemistry* **270**, 20201-20206 (1995).

112 Garcia-Castro, M. I., Vielmetter, E. & Bronner-Fraser, M. N-Cadherin, a cell adhesion molecule involved in establishment of embryonic left-right asymmetry. *Science* **288**, 1047-1051 (2000).

113 Resink, T. J., Philippova, M., Joshi, M. B., Kyriakakis, E. & Erne, P. Cadherins and cardiovascular disease. *Swiss Med Wkly* **139**, 122-134, doi:smw-12429 (2009).

114 Li, J. *et al.* N-cadherin haploinsufficiency affects cardiac gap junctions and arrhythmic susceptibility. *Journal of molecular and cellular cardiology* **44**, 597-606, doi:10.1016/j.yjmcc.2007.11.013 (2008).

115 Campuzano, O. *et al.* Genetic testing of candidate genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia. *European journal of medical genetics* **55**, 225-234, doi:10.1016/j.ejmg.2012.02.007 (2012).

116 Norton, N. *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *American journal of human genetics* **88**, 273-282, doi:10.1016/j.ajhg.2011.01.016 (2011).

117 Gupta, P. *et al.* Genetic and ultrastructural studies in dilated cardiomyopathy patients: a large deletion in the lamin A/C gene is associated with cardiomyocyte nuclear envelope disruption. *Basic Res Cardiol* **105**, 365-377, doi:10.1007/s00395-010-0085-4 (2010).

118 Meyer, T., Pankuweit, S., Richter, A., Maisch, B. & Ruppert, V. Detection of a large duplication mutation in the myosin-binding protein C3 gene in a case of hypertrophic cardiomyopathy. *Gene* **527**, 416-420, doi:10.1016/j.gene.2013.06.025 (2013).

119 Frousios, K., Iliopoulos, C. S., Schlitt, T. & Simpson, M. A. Predicting the functional consequences of non-synonymous DNA sequence variants--evaluation of bioinformatics tools and development of a consensus strategy. *Genomics* **102**, 223-228, doi:10.1016/j.ygeno.2013.06.005 (2013).

120 Asimaki, A. *et al.* A novel dominant mutation in plakoglobin causes arrhythmogenic right ventricular cardiomyopathy. *American journal of human genetics* **81**, 964-973, doi:10.1086/521633 (2007).

121 Fatima, A. *et al.* Derivation of induced pluripotent stem (iPS) cells from a patient with an arrhythmogenic right ventricular cardiomyopathy (ARVC). *Journal of stem cells & regenerative medicine* **6**, 97 (2010).

APPENDICES

Appendix 1: ARVC study patient consent form

UNIVERSITY OF CAPE TOWN

Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa
Department of Medicine
Cardiology Research Office
Room H46/47 OMB
Groote Schuur Hospital
OBSERVATORY. 7925
Tel: +27-21-447 2777
Fax: +27-21-447 2765 **Email: veronica.francis@uct.ac.za**

Date___/___/_____
DD MM YEAR

Informed Consent Form

I agree to participate in the study of Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC), a heart muscle disease (cardiomyopathy) that predominantly affects the right ventricle and may lead to ventricular arrhythmias and heart failure. I understand that I will be interviewed about my medical history, family history and medications, and that I may undergo further cardiac investigations. However, I will only undergo those investigations necessary to confirm or exclude the diagnosis of ARVC. In addition, I will have a blood sample drawn consisting of 5 tubes for a total of 25 ml of blood. This blood will be used to test for & study the genetic factors that cause ARVC. After my initial consultation, I agree to be followed up annually to assess if there have been any changes in my medical history and medications. As ARVC is a familial disease in 30-50% of people, I agree to have my first-degree relatives contacted for clinical screening & participation in this study. However, each of them will be counselled about ARVC and will also have to sign a consent form before they will be included in this study.

I understand that my participation in this study is entirely voluntary. All information gathered in this study is strictly confidential, and will only be used for research relating to the study of ARVC. This information (including genetic material) will not be used to generate a profit. As well, genetic material will not be used for the purpose of gene alteration, and, prior to blood sampling, I will sign a separate DNA consent form that governs the use of genetic material under the rules of the University of Cape Town Research Ethics Committee. I will not be identified in any published report. I am free to refuse to participate or withdraw from the study at any time, without jeopardising my future care. If I have any questions, I may contact _____at_____.

I agree to participate in the study and I have been given a copy of this form.

Subject name	Subject signature	Date
Witness name	Witness signature	Date
Investigator name	Investigator signature	Date

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory
Division of Human Genetics
1st Floor, Anatomy Building
UCT Medical School, Observatory 7925

Tel: (021) 406 6425 Fax: (021) 448-0906

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel. Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname: _____ First Name(s): _____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____ Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year: _____ Month: _____ Day: _____

Number of children: _____

Ethnic Origin : (please indicate ancestry of both your mother and father) _____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

Arrhythmogenic Right Ventricular Cardiomyopathy ☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental retardation ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

For Laboratory use only:

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

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:* _____

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I might have inherited a disease-causing mutation in the gene for arrhythmogenic right ventricular cardiomyopathy.

2. I understand that the genetic material for analysis is to be obtained from: blood cells/other (specify) (DELETE WHERE NOT APPLICABLE) :

3. I request that **no** portion of the sample be stored for later use. ☐ (MARK IF APPLICABLE)
   Or
   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):
   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____

5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:
   (a) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) the genetics laboratory is under an obligation to respect medical confidentiality .
   (d) genetic analysis may not be informative for some families or family members.
   (e) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM

Appendix 2: Solutions

2.1. 1X TBE Buffer

10.8 g Tris (Promega)

5.5 g Boric Acid (Riedel-de Haën)

7.4 g EDTA (Promega)

Made up to 1 litre with dH_2O

2.2. 1% Agarose Gel

1 g Agarose (Roche)

Volume made up by 100 ml 1X TBE Buffer

2.3. 2% Agarose Gel

2 g Agarose (Roche)

Volume made up by 100 ml 1X TBE Buffer