Molecular Genetics of Arrhythmogenic Right Ventricular and Dilated Cardiomyopathy in South Africans

Mzwandile Mbele

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Medicine

UNIVERSITY OF CAPE TOWN

February 2014

Supervisor:

Professor Bongani M. Mayosi

Co-supervisors:

Professor Bernard Keavney

Dr Gasnat Shaboodien

ABSTRACT

Molecular Genetics of Arrhythmogenic Right Ventricular and Dilated Cardiomyopathy in South Africans

Mzwandile Mbele

Introduction: Little is known about the molecular genetics of cardiomyopathy in Africans.

Aims: to (1) determine the prevalence of desmosomal gene mutations in arrhythmogenic right ventricular cardiomyopathy (ARVC) and dilated cardiomyopathy (DCM) in desmosomal protein genes (i.e., plakophyllin 2, desmocollin 2, desmoglein 2, and plakoglobin), (2) establish the presence of a founder effect in families with a recurrent mutation in the plakophilin 2 (*PKP2*) gene, (3) investigate whether single nucleotide polymorphisms found in desmosomal genes affect gene expression, and (4) search for new candidate genes for ARVC in families where no causal mutation was found in desmosomal protein genes.

Methods: 177 participants with cardiomyopathy were screened for desmosomal gene mutations which were confirmed by Sanger sequencing. The following methods were used: in the founder effect study we used haplotyping with microsatellite markers; for total gene expression we used real time polymerase chain reaction and allelic expression imbalance and exome sequencing was used for mutation screening in two siblings with severe early onset ARVC. To all novel variants identified prediction tools were used to predict the pathogenicity of the variant in uestion.

Results: 21.5% of ARVC probands had a disease-causing mutation in one of four desmosomal genes; no disease-causing mutation was found in the 112 DCM index cases. A recurrent *PKP2* mutation occurred on a common haplotype background in four white South African probands with cardiomyopathy. Investigation of a common *PKP2* polymorphism had no effect on total gene expression nor was there evidence of allelic expression imbalance. Finally, rare mutations were found in *PARVA* and *HMGXB3* by exome sequencing of two siblings.

Conclusion: Desmosomal gene mutations account for few cases of ARVC in South Africa but are a rare cause of DCM. The founder effect in *PKP2* holds prospects for the creation of a panel of families to examine genotype-phenotype correlations in ARVC. Our work demonstrates the no differential impact of polymorphisms in desmosomal genes on gene expression in peripheral blood. Two rare, possibly deleterious, variants in *PARVA* were identified in ARVC probands. The identification of *PARVA* as a candidate gene for cardiomyopathy holds the prospect of a new diagnostic and therapeutic target for cardiomyopathy.

DECLARATIONS

I, Mzwandile Mbele, hereby:

(a) grant the University of Cape Town free licence to reproduce the above thesis in whole or in part, for the purpose of research;

(b) declare that:

(i) the above thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisor, I have received no assistance.

(ii) neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

I am now presenting the thesis for examination for the degree of PhD.

Candidate's Signature		**Date**	

Acknowledgement

I would like to thank people who have contributed to this research work.

My special thanks must go to my supervisor professor Bongani M. Mayosi for his guidance, help, enthusiasm and expertise, without which this research would not be possible.

Thanks also to the collaborators, Professor Bernard Keavney who made it possible to visit and his supervision when I was learning total gene expression, allelic expression imbalance and exome sequencing in his laboratory (Institute of Human Genetics, Newcastle Upon Tyne, UK). I am very grateful to everyone who offered training and guidance, in particular, Dr Darroch Hall, Dr Mauro Santibanez Koref and Dr Elise Glen.

I am grateful to the South African Medical research council, South African National Research Foundation and department of medicine scholarships for funding the project. I received training in Hatter institute under cardiovascular Genetics wing. I also have a special thank to Dr Gasnat Shaboodien for general laboratory and guidance; Dr David Watkins; Neil Hendricks; Sister Veronica and Sister Unita September for their entirely help to recruiting cardiomyopathy patients that made this study possible. Thanks to everyone to cardiovascular genetics and protection lab group, especially Maryam Fish; Gaurang Deshpande; and Babu Muhamed, for their friendship, and finally to my family and my son Athenkosi Nkume-Mbele whose love and patience I could not do without.

Mr Mzwandile Mbele

PhD candidate

Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 5

Chapter 6

List of abbreviations

µ	micro
A	adenine
ACM	arrhythmogenic right ventricular cardiomyopathy sample code
AEI	allelic expression imbalance
aeQTL	allelic expression quantitative trait locus
AER	allelic expression ratio
ANXA7	annexin A7
ARVC	Arrythmogenic Right Ventricular Cardiomyopathy
ACTB	beta actin
BDGP	Berkeley Drosophila Genome Project
B2M	beta-2-microglobulin
BLAST	basic local alignment search tool
BMI	body mass index
bp	base pairs
BPS	paxillin binding site
ºC	degrees Celsius
C	cytosine
CAD	coronary artery disease
CASSA	Cardiac Arrhythmia Society of South Africa
cDNA	complementary deoxyribonucleic acid
CEPH	Centre d'Etude du Polymorphisme Humain Caucasian cohort
CEU	Caucasian (CEPH) HapMap cohort
CH	calponin homology
CG	cluster generation
CI	confidence interval

Ct	cycle threshold for real-time PCR analysis
CTNNA3	alpha-T-catenin
DCM	Dilated cardiomyopathy
DES	desmin
dH_2O	distilled water
DNA	deoxyribonucleic acid
dHPLC	denaturing high performance liquid chromatography
dNTP	deoxyribonucleotide triphosphate
DSC2	desmocollin-2
DSG2	desmoglein-2
DSP	desmoplakin
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
eQTL	expression quantitative trait locus
EVS	exome variant server
F	forward
g	gram
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
gDNA	genomic DNA
GVUS	genetic variant of unknown significant
GWA	genome-wide association
HCM	Hypertrophic Cardiomyopathy
HF	heart failure
HIV	Human immune deficiency virus
HMGXB3	homo group box containing domain 3
HRM	high resolution melt

Hyb	hybridization
JUP	plakoglobin
kb	kilobase
L	litre
LD	linkage disequilibrium
LMNA	lamin A/C
LR-PCR	long range polymerase chain reaction
LV	left ventricle
M	molar
MAF	minor allele frequency
min	minute
MALDI-TOF	matrix-assisted laser desorption/ionisation time-of-flight
$MgCl_2$	magnesium chloride
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
NE	Caucasian northeast cohort
PARVA	Parvin alpha
PCR	polymerase chain reaction
PERP	PERP, TP53 apoptosis effector
PLN	phospholamban
PKP2	Plakophilin-2
PKP4	Plakophilin-4
R	reverse
RCM	Restrictive Cardiomyopathy
RHD	Rheumatic heart disease
RNA	ribonucleic acid

RN18S1	18S ribosomal RNA
RT	reverse transcription
RYR2	ryanodine receptor 2
SA	South Africa
SA cohort	South African cohort
SAP	shrimp alkaline phosphatase
SBS	sequence by synthesis
SR-PCR	spacer region polymerase chain reaction
sec	second
SNP	single nucleotide polymorphism
TAE	Tris-acetate-EDTA electrophoresis buffer
TB	tuberculosis
TE	Tris-EDTA buffer
TGFβ3	transforming growth factor-β3
TMEM43	transmembrane protein 43
TTN	titin
UK	United Kingdom
USA	United States of America
UTR	untranslated region
UV	ultra violet
V	volts
WPBTS	Western Cape blood transfusion service
WT	wildtype
YRI	Yoruba in Ibadan Nigeria HapMap cohort

Chapter 1

Molecular Genetics of Arrhythmogenic Right Ventricular and Dilated Cardiomyopathy in South Africans

1.1 Background

Overall mortality in the developed world shows the pre-eminence of cardiovascular and neoplastic conditions. By contrast, infective and parasitic conditions remain the dominant causes of death and disability in Africa. Although the causes of heart disease vary within and between African countries, the etiology remains largely non-ischemic, with hypertension, rheumatic heart disease, and cardiomyopathy being the major causes of cardiovascular disease, whereas tuberculous pericarditis and pulmonary heart disease account for the remainder (Sliwa and Mayosi 2013).

The cardiomyopathies pose the greatest challenge of all the cardiovascular diseases in Africa because of their greater prevalence in societies still plagued by diseases of famine and pestilence, the difficulty in diagnosis, which often requires specialized cardiological investigations that are lacking in resource-poor environments, the lack of access to effective interventions, such as heart transplantation and the high mortality associated with these often irreversible disorders of heart muscle (Sliwa, Damasceno *et al.* 2005). Since the first reports of heart failure of unknown cause in Africa appeared approximately 60 years ago (Sliwa, Damasceno *et al.* 2005), it has been recognized that the entity of "cryptogenic heart disease" in Africa represented more than one syndrome, although terms such as "cardiopathy", "primary parietal endomyocarditis", "primary mural endocardial disease", and "cardiomyopathy" were initially used to embrace all forms of idiopathic heart disease.

The cardiomyopathies are now defined as myocardial disorders in which the heart muscle is structurally and functionally abnormal, in the absence of coronary artery disease, hypertension, valvular disease and congenital heart disease sufficient to cause the observed myocardial abnormality (Elliott, Andersson *et al.* 2008). They are divided into the following types depending on the structural and functional abnormality delineated by imaging: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), and unclassified types (Fig. 1.1). Each type of cardiomyopathy is classified into familial/genetic cardiomyopathy which defines an occurrence of the disease in more than one family member or due to a *de novo* mutation and non-familial/non-genetic cardiomyopathy (Elliott, Andersson *et al.* 2008).

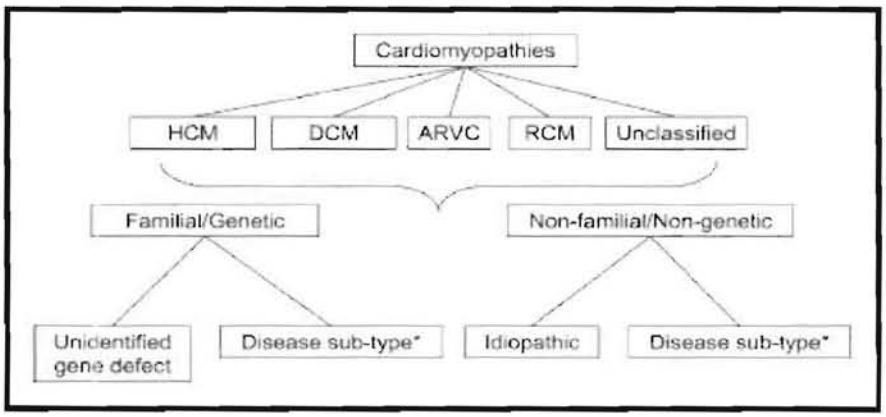


Figure 1.1: The European Society of Cardiology classification of cardiomyopathy (Elliott, Andersson *et al.* 2008).

Familial cardiomyopathies are mostly monogenic disorders where a single gene defect is sufficient to cause the disease. *De novo* mutations are classified as familial/inherited disease because the condition can be transmitted to the next generation. (Elliott, Andersson *et al.* 2008).

The primary objective of this thesis is to delineate the molecular genetics of arrhythmogenic cardiomyopathy and dilated cardiomyopathy in Africa. Other forms of cardiomyopathy that are prevalent in Africa, such as hypertrophic cardiomyopathy, restrictive cardiomyopathy and peripartum cardiomyopathy have been reviewed elsewhere (Sliwa and Mayosi 2013) (Sliwa, Damasceno *et al.* 2005) and will not be considered further in this thesis.

1.2 Arrhythmogenic right ventricular cardiomyopathy in Africans

ARVC is defined by the presence of right ventricular dysfunction (global or regional), with or without left ventricular disease, in the presence of histological evidence for the disease and/or electrocardiographic abnormalities in accordance with published criteria (Marcus FI 2010). The histological hallmark of ARVC is the presence of progressive replacement of right ventricular myocardium with adipose and fibrous tissue often confined to a 'triangle of dysplasia' comprising the right ventricular inflow, outflow and apex. While these pathologic abnormalities can result in functional and morphological right ventricular abnormalities, they also occur in the left ventricle, producing a DCM phenotype, or can be present in the absence of clinically detectable structural changes in either ventricle (Elliott, O'Mahony *et al.* 2010).

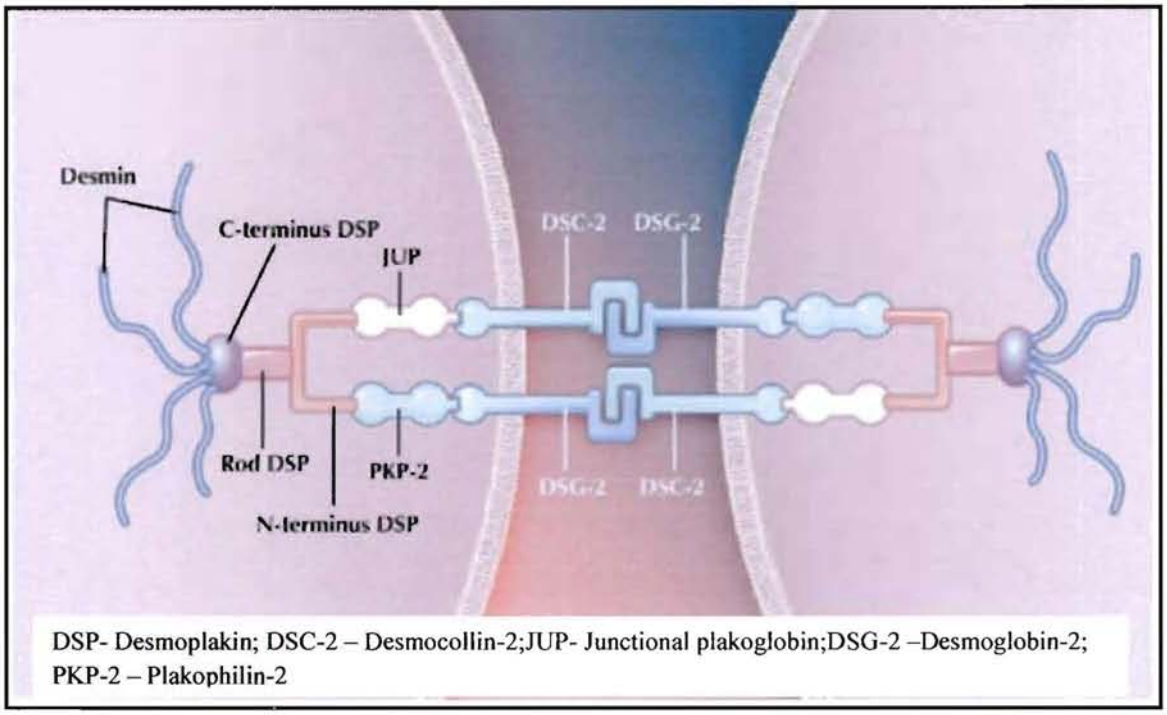
Although uncommon (estimated prevalence 1:5000), ARVC is a frequent cause of sudden death in young people and athletes due to re-entrant tachyarrhythmias of right ventricular origin (La Gerche, Robberecht et al. 2010). To the best of my knowledge, ARVC was reported for the first time in Africa in 2000, approximately 40 years after the first case was described (Munclinger, Patel *et al.* 2000) and the possibility that the disease might have reported before this in South Africa, but under a different name that the disease might have reported before this in South Africa, but under a different name, for example Hereditary dysrhythmic congestive cardiomyopathy (Brink, Torrington et al. 1976). Mokhobo and Mntla

have subsequently reported a series of eight patients with isolated right ventricular cardiomyopathy who presented with symptoms of heart failure and no apparent arrhythmias; no special imaging, electrophysiological, or histological tests were performed to identify features of ARVC in these patients (Mokhobo and Mntla 1997). The prevalence and incidence of the disease in Africa is unknown.

The death reports on ARVC in Africa are probably related to the lack of sophisticated cardiac electrophysiology facilities and expertise required for the diagnosis of the disease (Millar and Mayosi 2003). Initial information from our ARVC Registry of South Africa suggests that ARVC occurs in all segments of the population and that its clinical features, frequency of familial disease, and outcome are similar to experience that has been gathered elsewhere in the world (Watkins, Hendricks *et al.* 2009).

The aetiology of ARVC is not fully understood. The disease is familial in approximately 30-50% of affected individuals and sporadic in the rest of cases (Lopes and Elliott 2013, Te Rijdt, Jongbloed *et al.* 2013). It is thought that a genetic or environmental insult causes cardiac cell apoptosis and replacement by fatty tissue or fibrosis. Inherited forms of ARVC are caused mainly by mutations in desmosomal protein genes such as plakophilin 2 (*PKP2*; locus 12p11.21), desmoglein 2 (*DSG2*; locus 18q12.1), desmocollin 2 (*DSC2*; locus 18q12.1), desmoplakin (*DSP*; locus 6p24.3) and plakoglobin (*JUP*; locus 17q21.2). Figure 1.2 illustrates desmosomal protein interactions that are involved in ARVC. Non-desmosomal gene mutations that have been implicated in ARVC have been found in desmin (*DES*; locus 2q35), titin (*TTN*; locus 2q31.2), lamin A/C (*LMNA*; locus 1q22), phospholamban (*PLN*; locus 6q22.31), ryanodine receptor 2 (*RYR2*; locus 1q43), alpha T catenin (*CTNNA3*, locus 10q21.3), and transmembrane protein 43 (*TMEM 43*; locus 3p25) (Te Rijdt, Jongbloed *et al.* 2013). All these gene mutations are inherited in an autosomal dominant pattern with a few

cases in *JUP* and *DSP* inherited in an autosomal recessive fashion. Table 1.1 illustrates both desmosomal and non-desmosomal genes that were mapped and are involved in ARVC.



DSP- Desmoplakin; DSC-2 – Desmocollin-2;JUP- Junctional plakoglobin;DSG-2 –Desmoglobin-2; PKP-2 – Plakophilin-2

Figure 1.2: Schematic of desmosomes structure (Sen-Chowdhry, Syrris *et al.* 2007)

The molecular genetics of ARVC in Africans is largely unknown. There have been two reports on the molecular genetics of ARVC from the African continent. The first was by Matolweni and others who reported the linkage of a large South Africa pedigree of European descent to the ARVC6 locus (Matolweni, Bardien *et al.* 2006). The causal genetic mutation has not yet been identified. The second report was based on my Masters thesis study of the first 36 participants in the ARVC Registry of South Africa which showed that *PKP2* mutations occur in about a quarter of patients with ARVC in South Africa (Watkins, Hendricks *et al.* 2009).

The ARVC Registry of South Africa has subsequently been expanded to 65 cases of ARVC, and I report on screening for known and novel genetic causes of ARVC in this work.

Table 1.1: Loci and gene mutations that are associated with arrhythmogenic right ventricular cardiomyopathy

Type of ARVC	*Reference on locus identification*	*MIM*	*Chromosome*	*Mode of inheritance*	*Gene*	*Reference on identification of genetic mutation*
ARVC-1	(Rampazzo, Nava *et al.* 1994)	107970	14q23-q24	AD	Transforming Growth Factor-β3 (*TGF β3*)	(Beffagna, Occhi *et al.* 2005)
ARVC-2	(Rampazzo, Nava *et al.* 1995)	600996	1q42-q43	AD	Cardiac Ryanodine Receptor (*RyR2*)	(Tiso, Stephan *et al.* 2001)
ARVC-3	(Severini, Krajinovic *et al.* 1996)	602086	14q12-q22	AD	Unknown	-
ARVC-4	(Rampazzo, Nava *et al.* 1997)	602087	2q32.1-q32.3	AD	Unknown	-
ARVC-5	(Ahmad, Li *et al.* 1998)	604400	3p23	AD	Transmembrane protein 43 *(TMEM 43)*	(Merner, Hodgkinson *et al.* 2008)
ARVC-6	(Li, Ahmad *et al.* 2000); (Matolweni, Bardien *et al.*)	604401	10p12-p14	AD	Unknown	-
ARVC-7	(Melberg, Oldfors *et al.* 1999)	609160	10q22.3	AD	Unknown	-
ARVC-8	(Rampazzo, Nava *et al.* 2002)	607450	6p24	AD/AR	Desmoplakin (*DSP*)	(Rampazzo, Nava *et al.* 2002)
ARVC-9	(Gerull, Heuser *et al.* 2004)	609040	12p11	AD	Plakophilin-2 (*PKP2*)	(Gerull, Heuser *et al.* 2004)
ARVC-10	(Pilichou, Nava *et al.* 2006)	Nm	18q12.1	AD	Desmoglein-2 (*DSG2*)	(Pilichou, Nava *et al.* 2006)
Naxos	(Coonar, Protonotarios *et al.* 1998)	601214	17q21	AR	Plakoglobin (*JUP*)	(McKoy, Protonotarios *et al.* 2000)
ARVC/APC	(Frances, Rodriguez Benitez *et al.* 1997)	115650	14q24-q terminal	AR	Unknown	-

1.3 Dilated cardiomyopathy in Africans

Dilated cardiomyopathy (DCM) is defined by the presence of left ventricular dilatation and left ventricular systolic dysfunction in the absence of abnormal loading conditions (such as hypertension or valve disease) or coronary artery disease sufficient to cause global systolic impairment. Right ventricular dilation and dysfunction may be present but are not necessary for the diagnosis (Elliott, Andersson *et al.* 2008). Dilated cardiomyopathy continues to be an important cause of heart failure contributing between 20% and 30% of cases of heart failure in adult Africans (Damasceno, Mayosi *et al.* 2012).

The prevalence of DCM in the general population is unknown, but it clearly varies with age and geography. A frequency of 25% in patients from South African populations have evidence of familial disease with predominantly autosomal dominant inheritance (Ntusi, Wonkam *et al.* 2011). Autosomal dominant forms of the disease are caused by mutations in cytoskeletal, sarcomeric protein/ Z-band, nuclear membrane and intercalated disc protein genes. Table 1.2 illustrates reported nuclear genes that are implicated in DCM. X-linked diseases associated with DCM include muscular dystrophies (e.g. Becker and Duchenne) and X-linked DCM (Watkins, Ashrafian *et al.* 2011). DCM may also occur in patients with mitochondrial cytopathies and inherited metabolic disorders (e.g. haemochromatosis). Examples of acquired causes of DCM include nutritional deficiencies, endocrine dysfunction, HIV infection and the administration of cardiotoxic drugs.

Dilated cardiomyopathy can occur at a late stage following cardiac infection and inflammation (Shaboodien, Maske *et al.* 2013). In contrast to active or fulminant myocarditis, which is, by definition, an acute inflammatory disorder of the heart, often with preserved left ventricular size, inflammatory DCM is defined by the presence of chronic inflammatory cells in association with

left ventricular dilatation and reduced ejection fraction; histology and/or immunocytochemistry are, therefore, necessary for the diagnosis. A proportion of individuals with inflammatory DCM have persistence of viral proteins in the myocardium; viral persistence can also be observed in the absence of inflammation (Elliott, Andersson *et al.* 2008).

Peripartum cardiomyopathy (PPCM) is a form of DCM that presents with signs of cardiac failure during the last month of pregnancy or within 5 months of delivery (Sliwa, Hilfiker-Kleiner *et al.* 2010). Suggested aetiological factors in PPCM include myocarditis, autoimmunity caused by chimerism of haematopoetic lineage cells from the foetus to the mother and the haemodynamic stress of pregnancy (Ntusi and Mayosi 2009). PPCM has also been shown to be part of the spectrum of familial DCM in about 7% of cases (Ntusi, Wonkam *et al.* 2011). PPCM can occur at any age but is more common in women older than 30 years. It affects women of all ethnic groups, is almost equally associated with first/second and multiple pregnancies and is strongly associated with twin pregnancy and tocolytic therapy.

Several association studies suggest that genes may play a role in the susceptibility to DCM in African populations. An association with HLA-DR1 and DRw10 antigens has been reported in South African patients, implying that genetically determined immune-response factors play a role in the pathogenesis of some individuals with DCM (Maharaj and Hammond 1990). A common mitochondrial DNA polymorphism (T16189C) has also been found to be a genetic risk factor for DCM in a South African cohort, with a population-attributable risk of 6% (Khogali, Mayosi *et al.* 2001). The association of the mitochondrial T16189C mutation with DCM has been replicated in the British populations, suggesting that it is likely to represent a genuine genetic risk factor for DCM worldwide (Khogali, Mayosi *et al.* 2001). Mutation screening studies in patients with idiopathic and familial DCM have identified a family with early-onset

DCM caused by a known mutation in the troponin T gene (Arg141Trp) (Mayosi BM 2004), but there were no causal genetic mutations for DCM in cardiac and skeletal actin genes (Mayosi, Khogali *et al.* 1999). The investigation of the genetic association of β1- and α2c-adrenoreceptor variants and the G308A polymorphism of the tumor necrosis factor α gene with idiopathic DCM and the mitochondria DNA T16189C polymorphism with HIV associated cardiomyopathy have also proven to be negative in African patients (Badenhorst, van Staden et al. 2008, Brooksbank, Badenhorst et al. 2008, Du Preez, Matolweni et al. 2008, Woodiwiss, Badenhorst et al. 2008, Shaboodien, Engel et al. 2009).Previously, positive associations were noted with regard to the aldosterone synthase gene and improvement in left ventricular ejection fraction in DCM (Tiago, Badenhorst *et al.* 2002), and the mitochondrial DNA T16189C polymorphism and HLA variants in increasing the risk of DCM as stated above (Maharaj and Hammond 1990, Sliwa and Mayosi 2013).

Table 1.2: List of DCM reported genes (Arbustini, Narula *et al.* 2013)

Nuclear genes	MIM	Chromosome	Mode of inheritance	Protein
ABCC9	601439	Xq28	AD	ATP-binding cassette, subfamily C, member 9
ABLIM1	602330	10q25	AD	Limatin (actin-binding LIM domain protein)
ACTC1	102540	15q14	AD	Cardiac actin alpha
ACTN2	102573	1q43	AD	Alpha-actinin 2
ALMS1	606844	2p13	AD	ALMS1-C
ANO5	608662	11p14	AD	Anoctamin 5
ANKRD1	609599	10q23	AD	Ankyrin repeat domain-containing protein 1
ARVC-8	607450	6p24	AD/AR	Desmoplakin (*DSP*)
ARVC-9	609040	12p11	AD	Plakophilin-2 (*PKP2*)
ARVC-10	Nm	18q12.1	AD	Desmoglein-2 (*DSG2*)
JUP	601214	17q21	AD/AR	Plakoglobin (*JUP*)
DSC2	125645	18q12.1	AD	Desmocollin 2
EMD	300384	Xq28	X-linked	Emerin
EYA4	603550	6q23	AD	Eyes absent 4
FHL1	300163	Xq26	X-linked	Four-and-a-half LIM domains 1
GATAD1	614518	7q21.2	AD	GATA zinc finger domain containing protein 1
ILK	602366	11p15.4	AD	Integrin-linked kinase
LMNA	150330	1q22	AD	Lamin A/C
LAMA4	600133	6q21		Laminin alpha 4

Table 1.2: list of reported DCM genes (Arbustini, Narula *et al.* 2013) (continue)

Nuclear genes	MIM	Chromosome	Mode of inheritance	Protein
LDB3	605906	10q22	AD	LIM domain- binding 3
MYBPC3	600958	11p11.2	AD	Myosin-binding protein C
MYH6	160710	14q12	AD	Alpha-myosin heavy chain 6
MYH7	160760	14q12	AD	Beta-myosin heavy chain 7
MYOZ1	605603	10q22.1	AD	Myozenin 1
MYPN	608517	10q21.3	AD	Myopalladin
NEBL	605491	10p12	AD	Nebulette
NKX2-5	600584	5q34	AD	NK2 homeobox 5; cardiac-specific homeobox 1
PDLIM3	605889	4q35	AD	PDZ LIM domain protein 3
PLN	172405	6q22.1	AR	Phospholamban
PSEN2	600759	1q31-q42	AD	Presenilin 2
RBM20	613171	10q25.2	AD	RNA-binding protein 20
SCN5A	600163	3p21	AD	Sodium channel, voltage gated, type V, alpha subunit
SGCD	601411	5q33.2	AD	Delta-sarcoglycan
SYNE1	608441	6p25.2	AD	Nesprin 1, synaptic nuclear envelop protein 1
TCAP	604488	17q12	AD	Titin-cap; telethonin
TCF21	603306	6q23.2	AD	Transcription factor 21, epicardin
TGFB3	190230	14q24.3	AD	Transforming growth factor beta-3
TMEM43	612048	3p25.1	AD	Transmembrane domain 43
TMPO	188380	12q22	AD	Thymopoietin

1.4 Overlap between the phenotype and genetic causes of ARVC and DCM

It has already been mentioned above that ARVC may also be associated with left ventricular involvement in a significant proportion of cases. Indeed, there are instances when arrhythmogenic cardiomyopathy is predominantly left sided, in the entity that has been called arrhythmogenic left ventricular cardiomyopathy (Coats, Quarta *et al.* 2009). Furthermore, the primary presentation of ARVC may be that of heart failure rather than an arrhythmia. Therefore the phenotype and presentation of ARVC may be difficult to distinguish from that of DCM.

The two cardiomyopathies have recently been shown to be caused by mutations in desmosomal protein genes (Elliott, O'Mahony *et al.* 2010). In 2002, Itoh-Satoh and his colleagues reported a genetic mutation in titin that causes DCM; mutations in the same gene have recently been implicated in ARVC by Taylor and his colleagues (Itoh-Satoh, Hayashi *et al.* 2002, Taylor, Graw *et al.* 2011). There is therefore an overlap between the phenotype and molecular genetics of ARVC and DCM (Posch, Posch *et al.* 2008, Elliott, O'Mahony *et al.* 2010). On the basis of this observation, I sought to determine in this work the prevalence of desmosomal gene mutations in South African patients with ARVC and DCM.

1.5 Additional considerations in the study of the molecular genetics of cardiomyopathy in South Africa

There are three factors that may play a role in the molecular genetics of cardiomyopathy in South Africa that have been studied in this thesis. The first is the possibility of the presence of founder genetic effects based on the population history of South African sub-populations that are over-represented in patients with ARVC. The second factor relates to the role of common polymorphisms in modifying the expression of genes that cause ARVC and DCM.

Finally, I sought to identify novel genetic mutations that may play a role in the pathogenesis of cardiomyopathy in South Africans.

1.5.1 Founder effects in African heart diseases

Previous reports have defined founder populations as the families with a disease allele descended from the common ancestor (Brink and Schwartz 2009). In South Africa, genetic founder effects are common among the Afrikaner population which is predominantly of German, Dutch, and French origin. The founder effects in South Africa likely resulted from the fact that a small group of Dutch people harbouring a variety of rare, single gene diseases originally settled in the Western Cape Province in the 17th century (Moolman-Smook, De Lange *et al.* 1999) and later migrated to the rest of South Africa. Those single gene diseases have expanded in the Afrikaner population because of relative population isolation due to perceptions of common origin, culture and religion (Brink and Schwartz 2009). Genetic founder effects have been observed among Afrikaners and South Africans of mixed ancestry in several heart diseases such as long QT syndrome (LQTS) (Brink and Schwartz 2009), progressive familial heart block type I and II (Fernandez, Moolman-Smook *et al.* 2005, Kruse, Schulze-Bahr *et al.* 2009), familial hypercholesterolemia (Kotze, Langenhoven *et al.* 1991, Kotze, De Villiers *et al.* 1993) and HCM (Moolman-Smook, De Lange *et al.* 1999). Founder populations are not unique to Africa; they are also common in Finland where genetic founder effects have been identified in many diseases (Norio 2003) and also to many regions of the world.

1.5.2 Role of single nucleotide polymorphisms in gene expression

Polymorphisms are defined as common sequence variations, which occur in coding; non-coding regions of genes and can also occur in non genetic regions, with a frequency of ≥ 1% in the general population (Matkovich, Van Booven *et al.* 2010). There are different types of polymorphisms including single nucleotide polymorphisms (SNPs) and microsatellite polymorphisms. Both SNPs and microsatellites could regulate promoter activity and gene expression (Hassel, Dahme *et al.* 2009). They are also associated with a variety of diseases such as cancer and Crohn's disease (Tuch, Laborde *et al.* 2010, Hu and Peter 2013), and are used as markers in founder effect studies (Hassel, Dahme et al. 2009, Sheu, Zhai et al. 2009). SNPs may be homozygous or heterozygous, synonymous or non-synonymous or deletions or insertions.

Understanding of the cellular mechanisms that modulate gene expression is required in order to define their genetic contribution to human phenotypic variations (Campino, Forton *et al.* 2008, Cunnington, Kay *et al.* 2009). Cis regulatory elements are cis-acting factors which they regulate sequences to which they bind to are on the same DNA or RNA as the gene or RNA transcript that is being regulated (Strachan and Read 1999). Trans-regulatory elements might be located on different chromosomes and regulate both copies of the gene of which their expression levels can be affected by environment variables (Campino, Forton *et al.* 2008, Pham, Bonello *et al.* 2012). The effect of cis-regulatory elements on gene expression is investigated in this work through quantification of the allelic expression effects by the SNP of interest (Cunnington, Kay *et al.* 2009).

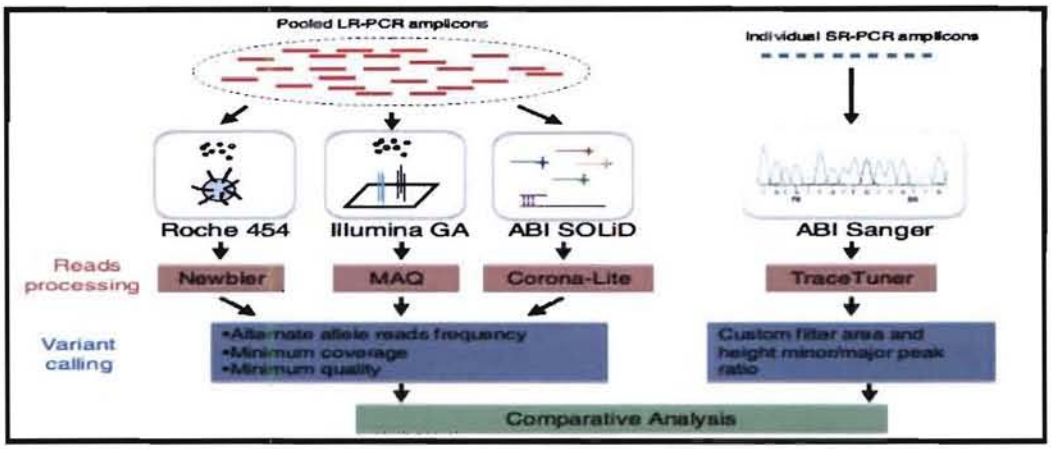
A powerful approach to investigate allelic expression imbalance (AEI) is to perform a quantification of the expression of transcripts obtained from each individual with heterozygous alleles (Beltelshees, Aquilante *et al.* 2012, Pham, Bonello *et al.* 2012). The approach is based on the fact that in the absence of allelic expression variation there will be

an equal amount of expression from both the paternal and maternal alleles as they are exposed to same environment. In comparison, individuals heterozygous for cis-acting polymorphisms that affect gene expression or messenger ribonucleic acid (mRNA) processing will show an alteration in the level of mRNA expression from one allele compared with its partner allele and this variation is termed allelic expression imbalance (AEI) (Pham, Bonello *et al.* 2012).

1.5.3 Novel genes that contribute to the disease phenotype

The known genetic mutations that cause ARVC (Table 1.1) explain about 50% of cases of ARVC (Basso, Corrado *et al.* 2012). It is likely therefore that there remain a number of novel genetic causes of ARVC that have not been discovered. The advent of new methods of sequencing, such as exome sequencing, has improved the chances of detection of causal genes in small families or individuals with inherited conditions (Yang, Muzny *et al.* 2013). Sanger sequencing and fluorescence-based electrophoresis technologies have been used to determine variants in somatic and germline genetic studies. To increase the throughput of Sanger sequencing, the use of massive parallel sequencing was developed (Reis-Filho 2009, Ng, Buckingham *et al.* 2010). (Reis-Filho 2009, Ng, Buckingham *et al.* 2010). Massive parallel resequencing technology renders resequencing of the whole genome practical but cost remains an important consideration. Its approach involves the targeted sequencing of protein-coding sequences (exome sequencing) which was reported to require approximately 5% as much sequencing as a whole human genome (Ng, Turner et al. 2009, Ng, Buckingham et al. 2010) Exome sequencing is reported to be a more efficient strategy to investigate for alleles underlying rare Mendelian disorders (Ng, Buckingham *et al.* 2010, Yang, Muzny *et al.* 2013). Exome sequencing is part of the next generation sequencing technologies recently developed, which include Roche 454, Illumina GA and ABI SOLID; these are considered to be more productive and less expensive as compared to Sanger sequencing (Harismendy, Ng

et al. 2009). An overview of exome sequencing technologies compared with Sanger sequencing is summarized in figure 1.3 below (Harismendy, Ng *et al.* 2009). Exome sequencing for novel disease-causing mutations was carried out in this study because there are many cardiomyopathy families who did not show any gene mutations in known genes.



SR-PCR: short range PCR and LR-PCR: Long range PCR

Figure 1.3: Comparison of exome sequencing with Sanger sequencing (Harismendy, Ng *et al.* 2009)

1.6 Aims of this study

The aims of this study were:

1) To determine the prevalence of desmosomal gene mutations in South African patients with ARVC and DCM;

2) To investigate the existence of founder effects in the event that recurrent mutations were found;

3) To determine whether polymorphisms in desmosomal genes influence total gene expression and allele-specific expression;

4) To perform exome sequencing to identify known or novel gene mutations that causes ARVC in families with no desmosomal gene mutation.

Chapter 2

MATERIALS AND METHODS

This chapter provides a detailed outline of the materials and methods that have been used throughout this thesis. More details are provided in the appendices: appendix 1 is an approval letter from the ethics committee; appendix 2 is the desmosomal gene (Tables 1-4) and *PARVA* (Table 5) designed primers; appendix 3 is solutions and buffers and appendix 4 is the GeneRuler^{TM} 100bp DNA Ladder Plus. The appendices are presented at the back of the thesis.

2.1 STUDY PARTICIPANTS

The cases recruited for this study are patients with arrhythmogenic right ventricular cardiomyopathy (ARVC) and dilated cardiomyopathy (DCM).

2.1.1 ARVC Cases

The ARVC cases were enrolled in the ARVC Registry of South Africa which was established by the Working Group on Registries of the Cardiac Arrhythmia Society of Southern Africa (CASSA), as reported previously (Latib, Michaels *et al.* 2004). The registry is approved by the Human Research Ethics Committee of the Faculty of Health Sciences of the University of Cape Town (Appendix 1). Physicians refer suspected cases of ARVC and their first-degree relatives to the Coordinating Centre in the Cardiac Clinic, Groote Schuur Hospital, Cape Town, for consideration for enrollment in the registry.

A group of cardiologists from the Cardiac Clinic at Groote Schuur Hospital established a diagnostic panel that reviews the diagnosis of each of the potential ARVC patients and determines whether they meet the diagnostic criteria for ARVC or not. A definite diagnosis of ARVC was made if the patient met the diagnostic criteria set by the Task Force of the Working Group on Myocardial and Pericardial Diseases of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology (McKenna, Thiene *et al.* 1994). If some of the criteria were met and no alternative diagnosis was found, a diagnosis of possible or probable ARVC was made. If there was no evidence of ARVC and/or an alternative diagnosis was present, the case was considered not to have ARVC. The diagnosis of the first degree relatives of affected individuals was made based on the modified criteria of Hamid and colleagues (Hamid, Norman *et al.* 2002).

Table 2.1: The diagnostic criteria of ARVC

I Global and/or regional dysfunction and structural alterations	**IV Depolarization/ conduction abnormalities**
Major	Major
Severe dilatation and reduction of right ventricular ejection fraction with no (or only mild) LV imparement	Epsilon waves or localized prolongation (>110ms) of the QR complex in right precordial leads (V1- V3)
Minor	Minor
Mild global right ventricular dilatation and/or ejection fraction reduction with normal left ventricle Mild segmental dilatation of the right ventricle Regional right ventricle Hypokinesia	Late potentials (signal averaged ECG)
II Tissue characterization of the walls	**V Arrhythmias**
Major	Minor
Fibrofatty replacement of myocardium on endomyocardial biopsy	Left bundle branch block type ventricular tachycardia (sustained and non-sustained) (ECG, Holter and exercise). Frequent ventricular extrasystoles(more than 1000/24h)(Holter)
III Repolarization abnormalities	**VI Family history**
Minor	Major
Inverted T waves in right precordial leads (V2 andV3)(people aged more than 12 yrs, in absence of right branch bundle block)	Familial disease confirmed at necropsy or surgery
	Minor
	Family history of premature sudden death (<35yrs) due to suspected right ventricular dysplasia

2.1.2 DCM cases

The DCM cases were recruited from Groote Schuur Hospital and UCT Private Hospital in Cape Town; Chris Hani Baragwanath Hospital, Milpark Hospital and Sunninghill Hospital in Johannesburg; Dr George Mukhari Hospital in Pretoria; as well as Inkosi Albert Luthuli Hospital in Durban. as reported previously (Du Preez, Matolweni *et al.* 2008).

The diagnosis of DCM was based on the criteria from the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases (Elliott, Andersson *et al.* 2008).

2.1.3 Controls

Controls for genetic screening of gene variants comprised genomic deoxyribonucleic acid (gDNA) only and controls for total gene expression and allelic expression included gDNA and RNA. gDNA was used to genotype the SNPs of interest and RNA was used to study the expression of the gene harboring the SNP. Three sets of population controls were used in this study: (1) the Cape Town population controls were used for the screening of genetic mutations that cause cardiomyopathy (Chapters 3 and 6) and (2) the South African population controls and the North East of England population controls were used in the study of gene expression and allele expression imbalance (AEI) (Chapter 5).

2.1.3.1 Cape Town controls

The Cape Town population controls were used to determine the population frequencies of novel variants identified in desmosomal and other genes (Chapters 3 and 6). Two hundred and thirty two anonymous blood donors from the Western Province Blood Transfusion Service (WPBTS) provided blood samples for DNA extraction. The blood samples were obtained from 99 people of mixed ancestry, 62 black Africans, 11 people of Indian origin and 60 white South Africans.

2.1.3.2 North East England population controls

The UK population controls were recruited in hospitals based on the fact that they do not have heart diseases. They were included to the study for the comparison with South African cohort.These controls were used in the total gene expression and allele expression imbalance experiments (Chapter 5). A total of 192 gDNA and RNA samples were obtained from anonymous individuals who participated in this study. They were all white persons who were born in the United Kingdom (UK). These anonymous individuals were recruited from two hospitals: the Freeman Hospital in Newcastle-upon-Tyne and the James Cook University Hospital in Middlesborough.

2.1.3.3 South African controls

These controls were used to study total gene expression and allelic expression imbalance of variants in chapter 5. A different set of 307 blood donors from the WPBTS contributed blood samples from which gDNA and RNA were obtained. The participants included people of mixed ancestry (n = 200), black Africans (n = 67), Indian South Africans (n = 30) and white South Africans (n = 10). The South African participants were the cohort of choice for allelic expression studies as they were expected to exhibit greater genetic diversity and higher allelic heterozygosity compared to European Caucasians who demonstrated low levels of heterozygosity in previous studies (Salisbury, Pungliya *et al.* 2003). The South African population controls were used to perform allelic expression imbalance (AEI) and the North East of England population controls were used to investigate total gene expression.

2.2 DNA EXTRACTION AND QUALITY CONTROL

2.2.1 DNA extraction

Peripheral blood samples were collected from participants in 5 ml ethylenediaminetetraacetic acid (EDTA) tubes. The samples were stored at -80°C until they were extracted. All blood samples were kept at room temperature for at least 2 hours prior to extraction. The gDNA was extracted from peripheral blood using the PureGene™ DNA isolation kit (Gentra system, USA) and was extracted according to the manufacturer's instructions. The gDNA samples were given disease code numbers (ACM to refer to the disease ARVC) to anonymise them and were filed in the database of the Cardiovascular Genetics Laboratory at the Hatter Institute for Cardiovascular Research in Africa, Department of Medicine, University of Cape Town.

2.2.2 Quality Control

The purity and concentration of gDNA were measured using a NanoDrop™ ND-1000 spectrophotometer (Thermo Scientific). All extracted gDNA samples were quantified at an absorbance of 260/280nm using a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, UK) in 1µl solution. All the gDNA samples were vortexed and spun down prior to measurement and each sample was measured twice for confirmation of its concentration. The average of two measurements was taken as the concentration of each sample.

All gDNA samples were divided into long term stock solutions which were stored in a -80°C freezer; working stocks of DNA were made and used for experiments. For experiments, working stocks of DNA were made into aliquots of 100ul which were standardized to 25ng/µl and stored at 4°C.

All primary DNA extractions were also validated using the Qubit™ Fluorometer (Invitrogen, USA) which is considered an accurate determinant of nucleic acid concentrations. All samples were mixed by vortexing before concentration measurements were taken; the manufacturer's instructions for measurement of DNA concentrations were followed. Samples were also run on 1.5% agarose gels to check the integrity of the DNA (appendix 2).

2.2.3 Agarose gel electrophoresis

Agarose (Seakem® LE, Whitehead Scientific (Pty) Ltd, SA) was used to visualize the products of the polymerase chain reaction (PCR). It is a standard method for separation of PCR products. Agarose is cross-linked to form pores through which DNA particles migrate at different speeds depending on the size of the fragment. The PCR products migrate towards the cathode (positive terminal) in an electric field due to the negative charge of DNA. The velocity at which molecules move through the agarose pores depends on the size of the pore, which is related to the percentage of the agarose in the gel. To visualize the PCR products in

the agarose gel, GelRed™, a fluorescent nucleic acid stain, was used to stain the PCR products. GelRed™ intercalates between the base pairs of the DNA strands and is visualized under ultraviolet (UV) light. The loading of the DNA and the monitoring of electrophoretic progression were assisted with the electrophoresis buffer, 1µl of 1X loading dye with a DNA ladder (GeneRuler™ 100bp DNA ladder Plus) (appendix 3) which was run alongside PCR products to give the sizes of the PCR products, which were electrophoresed through a 2% agarose gel in 1X TAE buffer.

2.3. POLYMERASE CHAIN REACTION (PCR) amplification

2.3.1 Principle of PCR

The principle of the PCR is based on the amplification of a target piece of DNA using two primers mediated by enzymes to generate thousands to millions copies of that particular DNA sequence (Mohini Joshi 2010). The technique of PCR is widely used in molecular biology to amplify a section of gDNA. The technique uses oligonucleotide primers to anneal to specific gDNA template sequences which flank the genomic region of interest. A gDNA polymerase enzyme is used to extend the sequence from the primers across the target region in many cycles and each cycle consists of denaturation, annealing and elongation in order to produce numerous copies of the target sequence. With PCR it is possible to generate millions of gDNA copies of interest for biological use (Santos, Sakai *et al.* 2004).

2.3.2 Optimization of the PCR conditions

All the gDNA samples from patients under investigation were diluted to a concentration of 25ng/µl. PCR optimization was first performed on one gDNA control sample in a PCR

gradient. The temperature producing the correct amplicon size on an agarose gel was then selected. The desmosomal primers used are presented in appendix 4 (desmosomal gene and *PARVA* primers).

PCR amplification was performed on a pre- PCR bench with pre-PCR pipettes to a total volume of 25µl. The reaction mixture contained: 25ng of gDNA, 1X GoTaq ™ Flexi Buffer, 1.5mM $MgCl_2$, 0.8mM dNTPs, 0.5 µM of each of forward (F) and reverse (R) primers each and 0.5 Units (U) of GoTaq Flexi polymerase. The cycling conditions used during PCR are presented in Table 2.1.

Table 2.2: PCR cycling conditions used for DNA expression in peripheral blood

Temperature (°C)	Time	Number of cycles
94	15 minutes	1
94	30 seconds	35
Annealing temperature*	30 seconds	
72	90 seconds	
72	5 minutes	1

*Depends on annealing temperature of the particular amplicon

PCR was run on the Peltier thermal cycler (Tetrad2). A negative control (whereby deionised H_2O was used instead of DNA) was included in all PCR experiments. This was done to identify possible contamination that could interfere within the reactants.

2.4 DESMOSOMAL SCREEN

The mutation screening of four desmosomal genes (i.e. plakophilin 2 (*PKP2*), desmoglein 2 (*DSG2*), desmocollin 2 (*DSC2*) and plakoglobin (*JUP*)) was performed using either the denaturing high performance liquid chromatography (dHPLC)-based WAVE® system (Transgenomic, Inc., Omaha, NE) or High Resolution Melt (HRM) analysis with a Rotor

gene 6000 (Corbett Life Science) for mutation screening, followed by direct Sanger sequencing of abnormal amplimers on the screening test. The methods that were used were based on the screening technique that was in place in the laboratory. As the screening technology changed, so did our screening methods. The dHPLC technique was replaced by HRM technology about half-way through this study. All amplicons larger than 300 base pairs were deemed unsuitable for HRM analysis and were directly sequenced.

2.4.1 Denaturing High-performance Liquid Chromatography

The principle underpinning the dHPLC-based WAVE® system (Transgenomic, Inc., Omaha, NE) is to separate heteroduplex gDNA fragments from homoduplex gDNA fragments by ion-pair reverse-phase liquid chromatography (Xiao, Stern *et al.* 2001, Frueh and Noyer-Weidner 2003). A heteroduplex is a double-stranded gDNA complex in which the two strands do not have perfect base complementarity and a homoduplex is a double stranded gDNA with perfect base complementarity (figure 2.1)

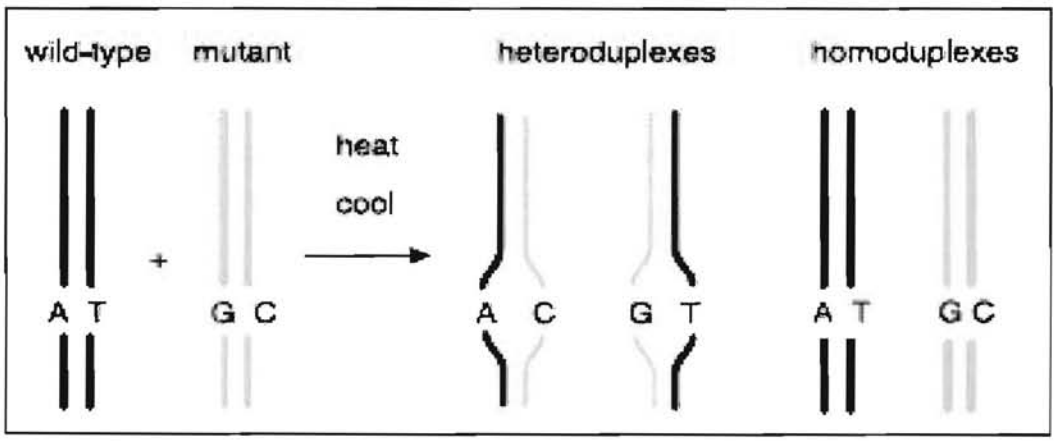


Figure 2.1: Illustration of the formation of both the heteroduplexes and the homoduplexes in the process of denaturing high-performance liquid chromatography (Frueh and Noyer-Weidner 2003).

The method detects all types of nucleotide mismatches regardless of the position of the mismatch in the fragment. The WAVE® system of dHPLC operates in three different modes:

partial denaturation, full denaturation and non-denaturation (Xiao, Stern *et al.* 2001, Frueh and Noyer-Weidner 2003). This study utilised the partial denaturation mode, as it is appropriate for the sequence, size and the melting temperature (Tm) of the PCR products. Once the PCR products are loaded in the autosampler, they are passed through the separation cartridge during the mobile phase. For dHPLC two steps are necessary: PCR amplification and heteroduplexing.

2.4.1.1 PCR amplification

A total of 25 µL of PCR reactions were prepared for certain *DSC2* and *PKP2* amplicons following the PCR methods mentioned above. gDNA from both cardiomyopathy patients and population controls was amplified and visualized on agarose gel as mentioned before or otherwise stated.

2.4.1.2 Heteroduplexing of the PCR products

The PCR products of gDNA samples of the individuals affected with cardiomyopathy were mixed in equal volumes (1:1 ratio) with PCR products from the gDNA samples of control individuals. The mixture was denatured at 95°C for 5 minutes.

2.4.1.3 Detection of desmosomal variants using a WAVE system

After denaturation of the heteroduplex samples they were subjected to a WAVE system for the detection of sequence variants. DHPLC-based WAVE® system uses the Wavemaker software programme; the programme was used to create methods for mutation detection for each PCR amplicon.

2.4.2 High resolution melt (HRM)

The principle of HRM is based on the characterization of double stranded DNA according to its annealing behavior. The process is done by monitoring the annealing from double stranded

DNA to single stranded DNA with increasing temperature as explained below. HRM performs the following two steps simultaneously: PCR amplification and analysis of PCR products produced. This technique uses PCR to amplify the DNA regions of interest. HRM analysis is then performed on all amplicons to distinguish nucleotide sequence differences between samples. During HRM the gDNA amplicons are heated from approximately 50°C to 95°C. At a particular point during this process the amplicons melt which leads to the separation of the two strands, (Figure 2.2).

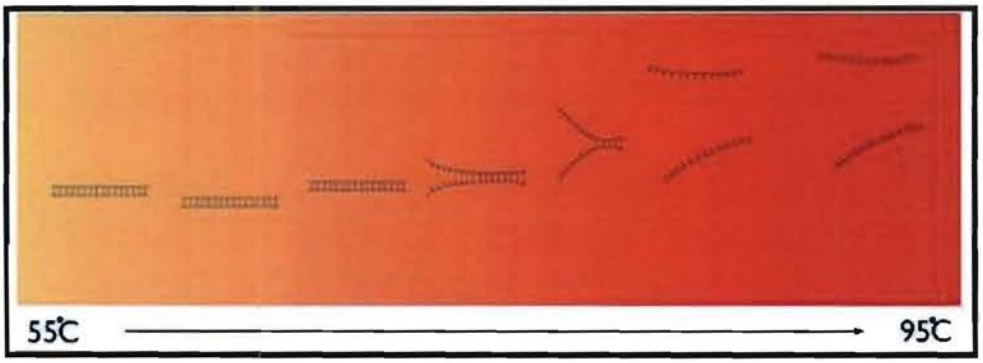


Figure 2.2: Double stranded gDNA process under different temperatures **(http://en.wikipedia.org/wiki/High_Resolution_Melt)**

HRM uses third generation fluorescent dyes that intercalate and bind specifically to double stranded gDNA. High levels of fluorescence in gDNA amplicons are present because of large copy numbers of the amplicon, but as the DNA amplicon is subjected to different temperatures, the fluorescence levels decrease as the double stranded gDNA separates (figure 2.3). The HRM machine monitors this process by measuring the fluorescence of the solution.

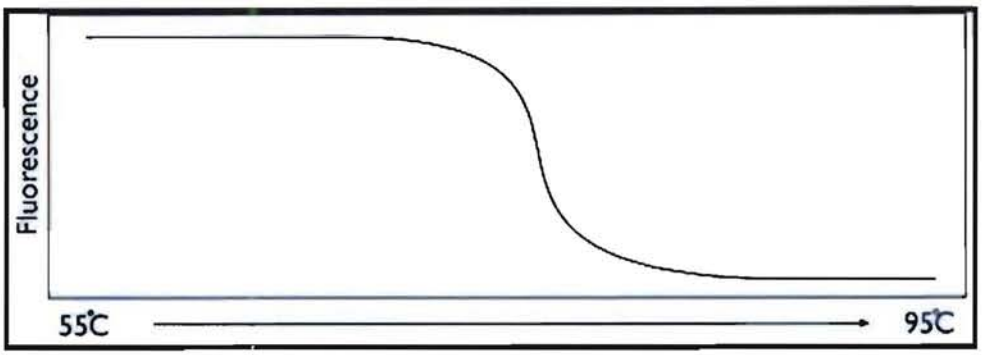


Figure 2.3: A melt curve, showing levels of fluorescence vs temperature **(http://en.wikipedia.org/wiki/High_Resolution_Melt)**

Using this method, two samples from different individuals without a mutation are expected to produce exactly the same melt curve shape, but if there is a mutation or any alterations on the gDNA amplicon sequence, the temperature at which the gDNA strands melt and separate changes, resulting in different melting curves (figure 2.4).

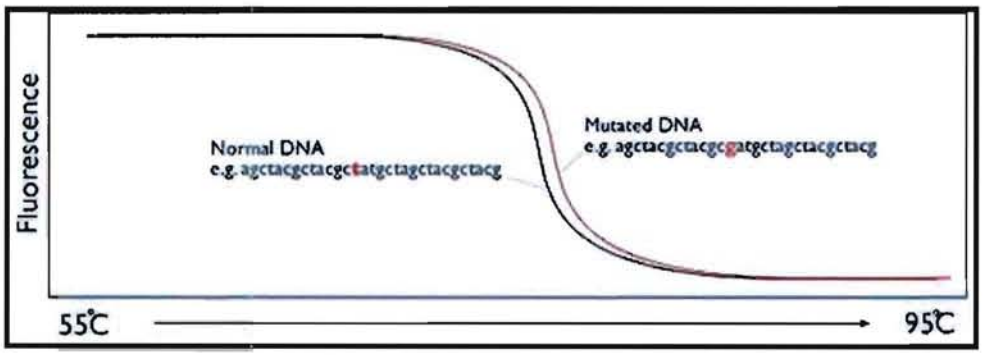


Figure 2.4: Two different curves showing normal gDNA and mutated gDNA **(http://en.wikipedia.org/wiki/High_Resolution_Melt)**

A gene has two alleles. When a gene is amplified by PCR both alleles are amplified and when looking for mutations, three outcomes can be expected: no variations in both alleles (WT), both alleles present variations (homozygote) and only one allele shows a variation (heterozygote). Each of these outcomes produces a different melting curve (figure 2.5).

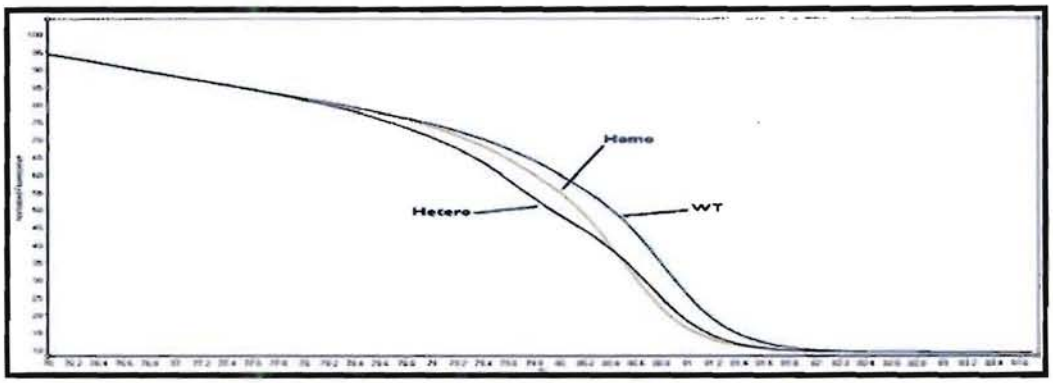


Figure 2.5: A graph showing three possible results of gene alleles **(http://en.wikipedia.org/wiki/High_Resolution_Melt)**

PCR amplification was performed in a total volume of 25µl, which contained 50ng of gDNA, 1.5µM (1X) Go Taq ^{TM} reaction Buffer that contained 1.5µM $MgCl_2$, 200µM dNTPs, 20 µM of each of forward (F) and reversed (R) primers, 1X EvaGreen dye (AnaTech, USA) and 0.5 Units (U) of GoTaq polymerase. The cycling conditions used during PCR are presented in Table 2.2.

Table 2.3: PCR cycling conditions

Condition	Temperature/ Time	Number of Cycles
Initial denaturation	90 ℃ 5 for 10 seconds	1
Denaturation	95 ℃ for 5 seconds	50
Annealing temperature*	55 ℃ for 10 seconds	
Elongation	72 ℃ for10 seconds	
High Resolution Melt	72 °- 95 ℃ (0.1 increments)	

*Depends on annealing temperature of the particular amplicon

A negative control (whereby dH_2O was used instead of gDNA) was included in all PCR experiments. This was done to identify possible contamination that could have interfered with the reactants.

2.4.3 Direct sequencing

For the mutation screening of amplicon sequences that were larger than 300bps, we performed PCR followed by direct sequencing of the amplicons. Amplicons were then run on an agarose gel to establish whether the PCR reaction worked. Amplicons were then purified and sequenced on an ABI Prism ® 3100 Genetic Analyzer (Applied Biosystems).

2.4.3.1 Exonuclease treatment

All the samples that showed variations on dHPLC or HRM analysis were sequenced. Prior to the sequencing reactions, samples were treated with *Exonuclease I* to remove unincorporated dideoxy-nucleotide triphosphate (ddNTP) terminators. The following protocol was used per sample: 1U of *Exonuclease I* (New England Biolabs), 2U *Shrimp Alkaline Phosphatase* (Promega) to get rid of single stranded products, 5µl PCR product and dH_20 was added to a final reaction volume of 20µl. The reactions were incubated at 37°C for an hour (or overnight depending on the manufacturer's instructions) and the enzyme deactivated at 75°C for 15 minutes.

2.4.3.2 Sequencing reactions

The products were sequenced in both forward (F) and reverse (R) directions to confirm the mutations that were observed. The parameters of half reactions were as follows: a minimum of 20µM of each F and R primers, 5X Dilution Buffer, 5X Termination Mix, and a quantity of gDNA product and dH_2O to make up the total volume of 20µl. The conditions that applied in this sequencing are presented in Table 2.3.

Table 2.4: Dideoxy-Cycle Sequencing conditions

Temperature (°C)	*Time*	*Number of Cycles*
96	5 minutes	**1**
96	30 seconds	**25**
50	15 minutes	
60	**4 minutes**	

2.4.3.3 Sequence analysis

The sequencing electropherograms from the ABI Prism ® 3100 were analyzed using the BioEdit programme. The reverse and the forward electropherograms were aligned with the wildtype sequences from NCBI and analysed for changes. All rare variants were aligned with amino acid sequences of modern species that were obtained from Mutation Taster to determine the conservation of the amino acids, they compare the human sequence with the modern animals' sequence to look if the amino acid in question is shared among the list of modern animals of which if it is shared there the amino acid is predicted to be important and it alteration may cause disease. Splice site prediction by the neural network programme on the Berkeley Drosophila Genome Project (BDGP) website (www.fruitfly.org/seq_tools/splice.html) was used to predict whether intronic variants could be disease-causing or not by altering mRNA splicing.

2.4.4 Screening of population controls for novel variants

To determine the frequency of a novel variant in the population, healthy controls with no history of cardiomyopathy were screened using either direct sequencing (described in section 2.4.3) or the HRM method (described in section 2.4.2) in conjunction with direct sequencing.

2.4.5 Bioinformatics: prediction tools

The pathogenicity of all non-synonymous rare novel variants was assessed using the Polyphen-2. PolyPhen predicts the effect of nsSNPs by combining a number of properties relating to the structure and function of the encoded protein. The user can provide a UniProt ID or copy and paste the query sequence in FASTA format. The substitution (one per run) is also required. (http://genetics.bwh.harvard.edu), SIFT (Sorting IntolerantFrom Tolerant) is a program that predicts whether an amino acid substitution affects protein function so that users can prioritize substitutions for further study. We have shown that SIFT can distinguish between functionally neutral and deleterious amino acid changes in mutagenesis studies and on human polymorphisms. To predict whether an amino acid substitution in a protein will affect protein function, SIFT considers the position at which the change occurred and the type of amino acid change (http://sift.bii.a-star.edu.sg/), and MutationTaster (http://www.mutationtaster.org/) prediction programmes. Only those variants predicted to be disease-causing in a protein of interest by at least two of the prediction tools were considered for further analysis. The pathogenicity of synonymous rare variants was assessed using the RNAfold webserver (http://rna.tbi.univie.ac.at/) and MutationTaster tool. MutationTaster is a fast web-based application to evaluate DNA sequence variants using information from various sources combined and evaluated in a naive Bayes classifier. A prediction is given as either disease-causing or polymorphism along with a P value indicating the security of the prediction (with 1 being most secure). A breakdown of all of the components that contribute to the prediction is also provided (http://www.mutationtaster.org/). In the case of non-coding variants, the Berkeley splice site prediction by neural network tool (http://www.fruitfly.org/seq_tools/splice.html) and alternate splice site predictor (http://wangcomputing.com/assp/index.html) were used to predict splice site mutations.

MutationTaster is a fast web-based application to evaluate DNA sequence variants using information from various sources combined and evaluated in a naive Bayes classifier. A prediction is given as either disease-causing or polymorphism along with a P value indicating the security of the prediction (with 1 being most secure). A breakdown of all of the components that contribute to the prediction is also provided(http://www.mutationtaster.org/).

2.5 FOUNDER EFFECTS

2.5.1 Haplotyping

Three single tandem repeats (STR) were used in this experiment: *PKP2*_3'_CA1 marker at position 2415474 to 241529, a published D12S1692 marker at position 7275538- 727786 and *PKP2*_5'_TG1 marker at position 825316 to 825348, which were selected from the genomic sequence of *PKP2*. Primers for the D12S1692 marker were published in NCBI (rs11341966).

Table 2.4 illustrates primer information. These primers were selected to span 5', intragenic and 3' regions of *PKP2* (Figure 2.6).

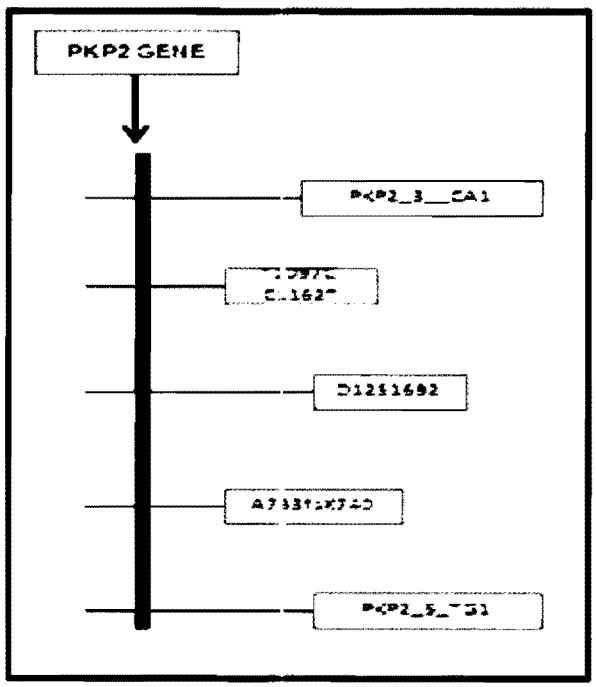


Figure 2.6: Illustration of two manually selected STR markers (*PKP2*_3'_CA1; *PKP2*_5'_TG1) flanking *PKP2*; one reported intragenic marker (D12S1692) and variants that were detected in family probands (red boxes)

Table 2.5: Primer sequences for *PKP2* microsatellite markers

Marker	Primer sequence	Annealing temperature	Amplicon size (bps)
PKP2_3'_CA1	F_GAGGAACAGGTGGGCTGAG	55°C	218bps
	R_GAAACGAACCCCAGTAATGTC TC		
D12S1692	F_CTTTGATTCCATACCCTCCT	52°C	249bps
	R_GCAGCAATTTCAGACTTCTC		
PKP2_5'_TG1	F_TGGGTCTGTATGTGTTTGCT		

	R_CATACACTCTCAAATAGAAATAGG	51	104bps

2.5.2 Variants and SNI's used as markers

Two markers were typed for haplotype construction, in addition to the 5', intragenic and 3' STR markers. These two markers were the common *PKP2* SNP rs1046116 (c.T1097C, p.L366P) in exon 4 and the *PKP2* InDel (A733fsX740) mutation in exon 11.

2.5.3 Genotyping of the markers

STR markers were genotyped in all four ARVC families (N=13) and 50 ethnically matched controls using *PKP2* labeled (FAM/ HEX florescence) primers (Appendix 2: Table 5).

2.5.4 Haplotype and 'assumed' haplotype construction

Construction of the most likely haplotype was designed with reference to the gene and STR marker order on chromosome 12, in the context of the family. The most probable haplotypes were constructed manually based on pedigree data and marker positions. In the two cases for which only the proband was available for genotyping, an "assumed haplotype", based on the haplotype identified by family mapping, was generated (Moolman-Smook, De Lange *et al.* 1999).

2.6 CIS-EFFECT ON GENE EXPRESSION

Tests for the cis-effect of SNPs on gene expression were performed in two different experiments: total gene expression (section 2.6.6) and allelic expression imbalance (section 2.6.7).

2.6.1 Cohorts

For this study peripheral blood RNA and DNA were taken from 192 North East and 307 South African healthy population controls.

2.6.2 RNA sampling

The blood (2.5ml) was collected in PAXgene® Blood RNA (PreAnalytiX) tubes for RNA extraction.

2.6.2.1 RNA handling procedures

RNA tasks were performed in the Microflow Laminar workstation using dedicated equipment designed for said tasks. Equipment and working surfaces were treated with RNase away (Sigma- Aldrich, UK) or 100% alcohol prior to RNA extraction to eliminate the possibility of contamination. Only designated pipettes and sterile RNase-free labware were used in RNA processing.

2.6.2.2 RNA sample collection, extraction and quality control

For the South African control samples, peripheral blood was collected in PAXgene RNA tubes and was stored at -80°C. The blood samples were placed at room temperature for 2-3 hours before extraction. All the extractions were performed according to the manufacturer's instructions. All samples were eluted into a total of volume of 80µl distilled water (dH_2O) and were stored at -80°C after extraction. RNA concentrations were measured using the NanoDrop ND-3000 spectrophotometer (NanoDrop Technologies, USA).

All RNA samples were *DNase I* treated using RQ1 RNase- free *DNase* (Promega, USA) prior to reverse transcription following the manufacturer's standard protocol to eliminate genomic DNA contamination. Fragment sizes and quality of RNA and cDNA were tested for selected samples using the RNA 6000 LapChip Kit (Agilent, USA) and the Agilent 2100 Bioanalyzer following manufacturer's instructions.

2.6.2.3 Reverse transcription

A total amount of 50ng to 2µg RNA was reverse-transcribed using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, USA) and Omniscript III reverse transcription kit (Qiagen, Germany). To reverse-transcribe RNA, the manufacturer's protocol was followed and all samples were eluted in a total volume of 20µl dH_2O. All cDNA was stored at -80^{0}C. The dilution of cDNA to working concentrations was performed assuming a 1:1 ratio of RNA to reverse-transcribed cDNA.

2.6.3 Genomic DNA sampling

Blood (8.5ml) was collected in PAXgene® DNA tubes for gDNA (PreAnalytiX, BD company).

2.6.3.1 gDNA handling

Genomic DNA tasks were performed in the Microflow Laminar workstation using dedicated equipment designed for said tasks. Equipment and working surfaces were treated with 100% alcohol prior to gDNA extraction to eliminate the possibility of contamination. Only designated pipettes and sterile DNase-free labware were used in DNA processing.

2.6.3.2. DNA sample collection, extraction and quality control

Please refer to section 2.2 above

2.6.4 Selection of single nucleotide polymorphisms for gene expression investigation

Three SNPs, in three different desmosomal genes, were selected for this study: rs1046116 in *PKP2*, rs79241126 in *DSG2* and rs868333 in *DSC2*. These SNPs all had a minor allele frequency greater than 5% in the healthy population controls.

2.6.4.1 Plakophilin-2 (PKP2); rs1046116 SNP

The rs1046116 SNP in *PKP2* was found in both ARVC and DCM probands. The rs1046116 SNP was identified in four ARVC probands and was segregating with a founder mutation in the coding region of *PKP2* (chapter 4). The rs1046116 SNP had a minor allelic frequency (MAF) that was greater than 0.05, a minimum criterion for inclusion of a SNP in the experiments on allelic expression imbalance (AEI) and total gene expression.

2.6.4.2 Desmoglein-2 (DSG2); rs79241126 SNP

The rs79241126 SNP was detected in the coding region of exon 14 of *DSG2* and was found only in ARVC index cases and population controls. It was also reported in NCBI from Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) where the minor allelic frequency was 0.05 while there was no available information on the HapMap-Yoruba (YRI) cohort.

2.6.4.3 Desmocollin-2 (DSC2); rs868333 SNP

The rs868333 SNP was detected in only one ARVC index case in exon 1 of *DSC2* but was very common in our population controls and was also published in the ARVC database (www.arvcdatabase.info) and NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). The minor allele frequency for this SNP was 0.11 in the CEPH families and 0.46 in a YRI low coverage panel. The MAF for both ethnicities met the requirement for the SNP inclusion in the experiments on allelic expression imbalance and total gene expression.

2.6.5 Primer design for gene expression in peripheral blood

A set of primers was designed, using Primer3 (v.0.4.0) (www.primerdesign.co.uk), in each gene in such a way that a forward primer will be situated inside one exon and a reverse primer inside the next exon using a cDNA sequence from Ensembl.org (Table 2.5). Specificity of primers was investigated by quantitative PCR (qPCR). The primers were tested in both cDNA and gDNA products. gDNA contamination in cDNA would be easily detected as we know that cDNA has no intronic regions and should not produce an amplicon with the intronic primers. The successful primers were used to amplify both cDNA and gDNA of all genes of interest using the amplification method and visualized in agarose gel.

Table 2.6: Desmosomal gene expression primer information

Primer name	Primer seq.	Annealing temp.	Exon	gDNA (bps)	cDNA (bps)
DSC2 F-primer	TCCAACACTGAGAACCAAGAAA	60.0	3	1093	140
DSC2 R-primer	AGCATCGAACAAGGAATTGG	60.0	4		
PKP2 F-primer	ATTTTGTTGGAGGCCGTTC	59.0	5	643	267
PKP2 R-primer	GCCTGCTTTCTTGGTGGT	59.0	6		
DSG2 F-primer	CTGGAAAAGGGATTACAGAGC	58.0	3	305	171
DSG2 R-primer	TGCGTAGCTCTAAGGGTTTC	58.0	4		

2.6.6 Total gene expression studies using peripheral blood

gDNA was used to genotype the controls for the associated SNPs of interest as described below. Total gene expression in genes of interest was measured by quantitative real time PCR (qRT-PCR) using Applied Biosystem TaqMan gene expression assays. The reactions were performed in replicates of four on an ABI PRISM 7000HT sequence detection system. The expression was measured relative to the housekeeping genes (beta actin (*ACTB*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and RNA, 18S ribosomal 1 (*RN18S1*)). For each gene, the Delta Ct values for each control were plotted against each SNP genotype and linear regression was used to determine if gene expression (relative to genotypes) varied significantly from null.

2.6.6.1 SNP genotype discrimination

Genotyping of the SNPs of interest was carried out by Applied Biosystem TaqMan genotyping assays.

TaqMan SNP genotyping was carried out in a Microflow Laminar flow workstation with pre-PCR pipettes to prevent any possibility of contamination. The genotyping was performed using predesigned TaqMan SNP Genotyping Assays (Applied Biosystems, USA). In PCR,

20-25ng of DNA template was used in a total volume of 5µl reaction in 384-well optical reaction plates. The PCR master mix for each sample was made of 2.5µl PCR Genotyping Master Mix (Applied Biosystems, USA), 0.125µl of 40xAssay Mix (Applied Biosystems, USA), 1.375µl of nuclease-free water, and 1µl of DNA template. PCR conditions used during Taqman SNP genotyp ng are summarized in Table 2.6. Non-template controls were included for quality control to exclude any possible contamination.

Data were analysed using the allelic discrimination algorithms in SDS Software v2.3 (Applied Biosystems, USA) according to the manufacturer's instructions. The data were also used to calculate both genotype and allelic frequencies in the South African cohort and to compare their frequencies with dbSNP frequencies in NCBI in order to determine if the frequencies identified were similar to those of other Africans.

Table 2.7: PCR cycling conditions for TaqMan SNP genotyping reactions

Temperature (°C)	Time	Number of Cycles
95	10 minutes	1
92	15 seconds	40
60	1 minute	

2.6.6.2 Quantitative real-time PCR

Quantitative Real-time PCR reactions were done using TaqMan gene expression probe assays and reagents from Applied Biosystems (USA) and were performed using the 7900HT Real-time PCR System (Applied Biosystems, USA). Gene expression PCR was performed using 5.0 ng/µl of cDNA in a 15µl reaction volume and reactions were performed in a 384 well reaction plate. Multiplex reaction mix was used to accommodate a FAM-labelled target gene assay (*PKP2*) and VIC-labelled control gene assay (*ACTB, GAPDH* and *RN18S1*) in the same reaction. TaqMan assays are validated by the manufacturer to have close to 100% amplification efficiency. PCR mix for each sample was made of 7.5µl 2x TaqMan Gene

expression master mix (Applied Biosystems, USA), 0.75µl of 20x target gene primer/probe mix (Applied Biosystems, USA), 0.75µl 20x control gene primer/probe mix (Applied Biosystems, USA); 1.0µl of nuclease-free water, 5.0µl of cDNA sample. PCR conditions used for gene expression are summarized in Table 2.7. Four replicates of each sample were performed. Non-template controls were included for quality control. Relative total gene expression was analysed using the comparative cycle threshold (Ct) method using SDS Software v2.3 and RQ Manager 1.2 software (Applied Biosystems, USA) according to the manufacturer's instructions. Ct values were normalised to the mean Ct value of the reference gene. Normalised Ct values were used in all analyses (Ct=mean= target gene Ct value – mean reference gene value).

Table 2.8: Total gene expression PCR conditions

Temperature (°C)	Time	*Number of Cycles*
50	2 minutes	1
95	10 minutes	1
95	15 seconds	45
60	1 minute	

The association between total gene expression and the *PKP2* SNP, rs1046116, was tested using a linear regression of the log transformed expression normalized values on the genotypes of the same *PKP2* SNP. The analysis was performed using R and SigmaPlot v11.0 where the effect size of the expression is given as the slope of the regression analysis. The association between total gene expression and SNP genotypes was tested using a linear regression of the log transformed normalized expression values on the genotypes of the same SNP.

2.6.7 Allelic gene expression imbalance

2.6.7.1 SNP genotyping discrimination

Genotyping of 307 South African (SA) population controls were performed to identify heterozygote samples using real-time PCR as described above.

2.6.7.2 Artificial Heterozygous

Artificial heterozygotes were also made with genomic DNA from homozygotes of both the major allele and the minor allele to test the reliability of the Sequenom assays. An artificial herterozygote is a heterozygote sample produced by mixing two samples at certain ratios, one homozygous sample DNA from wild type and one sample from homozygous with minor allele which contains the SNP of interest. The gDNA homozygotes were mixed in the following ratios (major allele: minor allele): 8:1; 4:1; 2:1; 1:1; 1:2; 1:4 and 1:8. Finally the ratio mixtures were diluted to make a total concentration of 25ng/ul. The assay was performed on the artificial heterozygote to demonstrate that there is a reliable correlation

between the known ratio of the two alleles and the ratio of the alleles detected by the Sequenom assay.

2.6.7.3 Quantification of allelic expression using Sequenom platform

In both mixed gDNA (artificial heterozygotes) and cDNA samples, a concentration of 25ng/ul for each sample of our SA cohort was sent to the High Throughput Array Facility (High throughput genomics) at the Welcome Trust Centre for Human Genetics (Oxford (UK)) to quantify the samples using the Sequenom platform. Quantification of the allelic expression ratio was done by primer extension and MALDI-TOF mass spectrometry. Amplification of gDNA with purification of products was performed in replicates of four.

2.6.7.4 Allelic expression analysis

Spectra from the Sequenom assay were manually reviewed. All poor quality spectra with high cluster background were excluded. Allelic ratios were calculated as the ratios of the area under the peak representing allele 1 to the area under the peak representing allele 2. Measurements were performed in four replicates. Since artificial heterozygote (gDNA) samples were taken in a small number of genomic samples to provide normalization, average allelic ratio for all genomic DNA samples were selected to determine the normalization factor and then used to normalize each of the cDNA ratios using the following formula:

Normalized allelic ratio = Average ratio of cDNA/Average allelic ratio of gDNA

2.7 EXOME SEQUENCING

The principle behind exome sequencing is to sequence the coding genomic regions of the whole genome and align them with the reference sequence. The process involves: 1) Preparing the genomic DNA – library preparation (section 2.7.1), 2) Cluster Generation (CG) – to clonally amplify the library to several million copies (section 2.7.2), and 3) Sequencing

by Synthesis (SBS) – this is forward and reverse sequencing of the clonally amplified library clusters (section 2.7.3).

2.7.1 Sample Library preparation

Exome sequencing samples were prepared using Illumina's Paired End Sample Preparation Kit (PE-102-1001, Illumina) and 50Mb SureSelect Human All Exon Capture kit (Agilent Technologies) and following instructions of the Illumina Paired End Sample Preparation Guide (Part 1005063, RevD, Feb2010, Illumina).

2.7.1.1 gDNA quantification and fragmentation

The concentrations for all gDNA samples were obtained using a Qubit Fluorometer and dsDNA Broad Range Assay kit following kit instructions (Invitrogen, Life Technologies). Then 6µg of high quality gDNA was diluted with 1X Low TE Buffer in 1.5ml LoBind tube to make a total volume of 100µl. The 100µl gDNA dilution was transferred to a Covaris tube and all air bubbles were removed at the bottom of each tube. This tube was then loaded onto a Covaris S2 machine, (Covaris) and the gDNA was fragmented with shear settings as follows: duty cycle was 20%; intensity was 4; cycles per bust were 200; time was 118 seconds; set mode was put in frequency sweeping and temperature was at 4°C. The step was repeated to ensure complete fragmentation. Then the fragmented gDNA was transferred back to 1.5ml LoBind tubes. To purify the fragmented gDNA, 500µl PB buffer was added to fragmented gDNA in the 1.5ml Lobind tube and mixed by pipette. A total of 600µl was transferred to a QIAquick PCR column and centrifuged for 1 minute at 13000 rpm. The flow-through was discarded and 750µl of PE buffer was added to wash the column, centrifuged for 1 minute and the flow-through was discarded and centrifuged for another 1 minute to

completely remove residual PE buffer. Then the QIAquick columns were placed in 1.5ml LoBind tubes and left open for 5 minutes at room temperature to allow evaporation of any residual ethanol. The gDNA was eluted into 30µl of Nuclease-free water and the elution step was repeated twice to make a total of 60µl. Using a Savant DNA120 SpeedVac Concentrator (Thermo Scientific), DNA samples were concentrated to 30µl and fragment sizes were checked using a 2100 Bioanalyzer and DNA 1000 lab chip (Agilent Technologies), to see if sizes were in the correct range of 150 to 200 base pairs.

2.7.1.2 DNA Library preparations

2.7.1.2.1 End Repair

All reagents were supplied in the Illumina Paired End Sample Preparation kit (PE-102-1001, Illumina). The gDNA fragments achieved by the Covaris shearing method results in 3' and 5' overhangs and blunt ends. The purpose of End Repair is to convert all the overhanging fragments to blunt ends. For this, the following reagents were added to each sample: 45µl of water; 30µl of gDNA sample; 10µl of T4 DNA Ligase Buffer with 10mM ATP; 4µl 10mM dNTP Mix; 5µl of T4 DNA polymerase; 1 µl Klenow Enzyme and 5µl of PNK. They were pipette mixed and incubated in a thermal cycler (BioRad tetrad) for 45 minutes at 20°C. Then the 100µl End repair mix was purified. The purification step was carried out using the QIAquick PCR kit and was performed as mentioned above and the eluate was concentrated to 32 µl using the SpeedVac (Thermo Scientific).

2.7.1.2.2 Adenylation of 3' ends

The adenylation step was performed. A single 'A' nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a

complementary overhang for ligating the adapter to the fragment. The following reagents were added in each sample (sample volume 32ul): 5µl of Klenow buffer; 10µl of 1mM dATP and 3µl of Klenow Exo with a total of 50µl. The reactions were mixed by pipette and incubated for 30 minutes at 37°C. Reactions were then purified in Qiagen MinElute PCR Purification columns (Qiagen) per instructions. The fragments were eluted in nuclease-free water and concentrated (with SpeedVac) to 10µl before proceeding to Adapter ligation.

2.7.1.2.3 Paired End adapter ligation

Adapters were ligated to all adenylated fragments by adding the following reagents to the 10µl gDNA: 25µl 2X DNA ligase Buffer; 10µl PE adapter Oligo Mix and 5µl DNA Ligase and this made a total volume of 50µl. The mixture was mixed by pipette and incubated on a thermal cycler for 15 minutes at 20°C.

Following the incubation, the ligation reactions were purified with Agencourt AMPure XP beads (Beckman Coulter). The AMPure XP beads were allowed to remain at room temperature for at least 30 minutes and the reagent was mixed. 90µl of the homogenous beads were added to the ligation reaction and mixed well by pipette. The mixture was then incubated for 10 minutes at room temperature. The reactions were performed in a PCR plate. The plate was then placed in a magnetic stand for approximately 3 to 5 minutes in order to separate the beads. The clear solution (supernatant) was discarded from the wells of the plates. Leaving the PCR plate on the magnetic stand, the beads with the ligated products bound to them were washed. This was performed by adding 200µl of 70% ethanol over the surface of the beads and then removing the ethanol. The wash step was repeated and the plate was left for a few minutes at room temperature until residual ethanol had evaporated. 30µl Nuclease-free water was then added to the beads after removing the plate from the magnetic stand to bring the beads back into solution. This resulted in the detachment of the ligated

fragments from the beads back into the water. To aid this, the mixture was left at room temperature for 5 minutes and the plate was then placed back on the magnetic stand. This time the supernatant was transferred to a clean 1.5ml Lobind tube. This elution step was repeated and the volume brought down to 30µl with the SpeedVac.

2.7.1.3 PCR enrichment of Paired End adapter ligated DNA fragments

2.7.1.3.1 Polymerase chain reaction (PCR)

The next step was to perform a PCR with two primers supplied in the Illumina prep kit (PE-102-1001, Illumina). This was done to selectively enrich fragments with adapters on both ends and to amplify the amount of DNA in the library. This PCR step performed four key functions: 1) by using the two primers, additional sequences are added to the ends of the adapters so that it aids the hybridization of the fragments to the oligonucleotides of the flow cell surface for cluster generation, 2) enriches for fragments with adapters on both ends as other species such as fragments with an adapter on only one end would not work for cluster generation, 3) enriches fragments to eliminate adapter dimers, and 4) amplifies the amount of DNA. The PCR mixture was as follows: 12µl library from previous step; 11µl Nuclease free water; 1µl PCR primer PE 2.0; 1µl PCR primer PE 1.0 and 25µl 2X Phusion Master Mix. The reactions were mixed by pipette and placed in a thermal cycler. The reaction conditions are listed in Table 2.8. After amplification, the PCR products were purified using the AMPure XP beads as detailed previously and the volume brought down to 30µl with the SpeedVac.

2.7.1.3.2 Purification of the PCR.

Following PCR clean up, fragment sizes were checked using a 2100 Bioanalyzer and DNA 1000 lab chip (Agilent Technologies), in order to see if sizes had shifted from those at the

start as additional bases (adapters) have been added and for accurate quantification as 500ng is required for the next step, exome enrichment.

Table 2.9: PCR cycling conditions

Temperature (°C)	Time	Number of Cycles
98	30 seconds	
98	40 seconds	4
65	30 seconds	
72	30 seconds	
72	**5 minutes**	

2.7.1.4 Exome Enrichment: The Hybridization step

A 50Mb SureSelect Human All Exon Capture kit (Agilent Technologies) was used. The kit contained all the components required to set up the hybridization and to clean up the reaction after hybridization. The procedure has 3 distinct parts (labelled 1 to 3) and all parts were combined to give a mixture of prepped library (collection of DNA fragments prepared for exome sequencing), baits, and PE blockers which was then incubated on a thermal cycler for 24 hours at 65 °C. All detailed instructions are contained in the SureSelect guide that accompanies the kit.

2.7.1.4.1 Part 1: preparation of DNA capture plate

The first RNAase block was diluted to make a ratio of 1:1 with Nuclease-free water and placed on ice. 1µl of Nuclease-free water was added to a well of a 96 well PCR plate; 5 µl of the bait was added to it. Then 1µl RNAase block was added to that whilst keeping the plate on an ice block. The resulting 7µl of water/bait/RNAse block mixture was mixed using a pipette and the wells covered with a cap and the plate, labelled as the "Capture Plate"; this was left aside in the ice block.

2.7.1.4.2 Part 2: Hybridization buffer preparation

For every reaction that was performed, the following buffers from the SureSelect kit were added in 1.5 ml Eppendorf tubes: 25µl of SureSelect Hyb #1; 1µl of SureSelect Hyb #2; 10 µl of SureSelect Hyb #3 and 10µl of SureSelect Hyb #4. They were mixed by vortexing and spun down by centrifugation. To avoid precipitation, the buffer mixture was kept at 65°C until ready to use.

2.7.1.4.3 Part 3: Prepped library

A 'prepped' library is a collection of DNA fragments prepared for the experiment.The library was quantified in the last step. From that, the volume of each library was adjusted, using the SpeedVac, to contain 500ng in 3.4µl of Nuclease-free water. This was done in a 1.5ml Lobind tube. To this volume, 2.5µl of block # 1, 2.5µl of block # 2 and 0.6µl of block # 3 (PE) were added to give a total volume of 9µl. They were mixed by pipetting and left on ice for 10 minutes.

After the 10 minutes, the library was transferred to a PCR plate, not on ice, and in a separate well, 40µl of hybridization buffer was added. The plate was then put in a thermal cycler and

following program was used: 95°C for 5 minutes and a hold at 65°C. This was to denature the libraries. After the plate was at 65°C for 2 minutes, the "Capture plate" was removed from the ice block and put in another block of the thermal cycler to run at 65°C. When both plates were at 65°C for at least 5 minutes, the 9µl of prepped library was added to the capture plate well, and 13µl of buffer added to a final volume of hybridization mixture of 29µl. This was then left at 65°C for 24 hours.

After the 24 hours, there were steps to remove fragments that hadn't hybridized to the baits, i.e. non-exonic regions; RNA baits were then digested from the library fragments which were from the exon regions. Full details are in the guide from Agilent Technologies accompanying the exome kit.

2.7.1.5 Post-hybridization amplification

Next, post-hybridization amplification was performed to amplify the library samples that have been through the exome enrichment step. The reaction setup was the same as before: 12µl enriched library; 11µl Nuclease-free water; 1µl PCR primer PE 2.0; 1µl PCR primer PE 1.0 and 25µl 2X Phusion Mix were added. They were mixed by pipetting and the reaction mixture was placed in a thermal cycler as before (Table 2.8) but the number of cycles was increased (Table 2.9). Following amplification, the libraries were cleaned up with the AMPure XP beads as described previously and the eluent was concentrated to 30µl.

Table 2.10: Post-hybridization cycling conditions

Temperature (°C)	Time	*Number of Cycles*
98	30econds	

98 65	40 seconds 30 seconds	11
72	30 seconds	
72	**5 minutes**	

2.7.1.6 Paired End library quantification

Finally, the libraries were accurately quantified and profiles checked before being sequenced and this was done using the 2100 Bioanalyzer and High Sensitivity lab chips (Agilent Technologies). Samples were ran in duplicate and an average of the concentrations and library sizes taken to calculate the amount in nM which is critical for sequencing and optimal cluster generation

2.7.2 Cluster Generation

The cluster generation and sequencing was performed by a technician at the Institute of Genetic Medicine, Newcastle University, UK, and involved the amplification of DNA libraries in an Illumina Flow Cell. The flow cell is made up of eight lanes and each lane has a dense lawn of oligonucleotides attached to the inside surface. These are complimentary to the Paired end adapters flanking the gDNA libraries. The first step was to denature the

libraries to single-stranded fragments at a set concentration (2nM) using 1N sodium hydroxide and then loading 12pM of the denatured library in each lane so that the template hybridized to the flow cell via the lawn of oligonucleotides. Then unlabeled nucleotides and enzyme were added to initiate solid-phase amplification which is termed "Bridge PCR". The enzyme incorporates nucleotides to build double-stranded DNA on the flow cell. Once completed and at the next denaturation step, the single stranded fragment loops over and as it has a complimentary sequence, it binds to another of the oligonucleotides bound on the flow cell surface, i.e. a "bridge" is formed, and amplification is repeated. In this way fragments were clonally amplified and several million clusters per lane were generated. Finally, cluster generation was completed. The reverse strands were cleaved and washed away. Ends were blocked and the sequencing primer was hybridized to the DNA template and the flow cell was then ready to load on the GAIIx for sequencing.

2.7.3 Sequencing by Synthesis (SBS)

Following cluster generation, the flow cell was sequenced on the Illumina GAIIx sequencer with a 75 cycle Paired End recipe. This included 75 cycles of sequencing in the forward direction (termed Read 1) followed by 75 cycles in the reverse direction (termed Read 2).

On the Genome Analyser, hundreds of millions of clusters are sequenced simultaneously. The DNA templates are sequenced base by base, in parallel, using four fluorescently-labelled reversibly terminated nucleotides. All four bases compete with each other to bind to the template. This natural competition ensures the highest accuracy. After each round of synthesis, the clusters are excited by a laser and show a colour that identifies the newly added base. The fluorescent label and blocking group are then removed, allowing for the addition of

the next base. This proprietary chemistry reads the single base added in each cycle, thus enabling accurate sequencing through difficult regions such as homo polymers and repetitive sequences.

After sequencing 75 cycles in the forward direction the clusters were processed for reverse strand sequencing in situ using the Paired End Module. This was exactly the same as Cluster Generation except that afterwards the forward strands were cleaved and washed away, ends blocked, reverse sequencing primer added and 75 cycles of sequencing performed. In this way 55 to 82 million reads per lane were generated.

2.8 Bioinformatic analysis

2.8.1. Alignment and variant calling

Raw data was analyzed by Professor Bernard Keavney's Bioinformatics Unit at the Institute of Human Genetics (Centre for Life) at the University of Newcastle upon Tyne, (UK) where I spent nearly a year working on this project under his supervision. Novoalign (http://www.novocraft.com) was used to align the sequence reads to the UCSC human reference genome (http://genome.ucsc.edu/; hg19) and SAMtools (Li, Handsaker *et al.* 2009) was used to call the variants.

2.8.2 Variant filtering

Initially all off-target variants were removed and only those shared by two affected individuals were selected for further downstream analysis. Following this, all SNPs within the dbSNP 135 database were removed. Finally, only those variants with a total base coverage $\geq$10x and a quality score $\geq$ 10 were selected.

Variant pathogenicity was assessed using the MutationTaster, Polyphen-2 and Sift prediction programs as mentioned in section 2.4.5.

The base coverage, quality scores and dbSNP135 database were decreased in order to identify variants present in genes known to cause/influence cardiomyopathy.

2.8.3. Public Genome browsers

UCSC (http://genome.ucsc.edu/) and Ensembl (http://www.ensembl.org) genome browser programs were used to map amino acids affected by the variants, positions of the affected amino acids and to retrieve the cDNA sequences of the specific genes of interest.

2.8.4 Public databases

1000 Genomes (http://www.1000genomes.org/), Exome variant sever (EVS) (http://evs.gs.washington.edu), dbSNP (www.ncbi.nlm.nih.gov/SNP/) and ARVC databases (www.arvcdatabase.info) were used to identify whether the variant had been previously identified and to obtain allelic frequency information.

2.8.5 Validation of potentially disease causing mutations

All genes harbouring identified mutations were annotated and a set of primers was designed to amplify the product containing the variant (Table 2.10). Genomic DNA was amplified and PCR products were visualized on an agarose gel. The frequency of variants was investigated in population controls and was performed using HRM followed by sequencing.

Table 2.11: Primer information of three genes with variants share by affected individuals.

Gene	Primer sequence	Amplicon size	Melting temp.

		(bp)	(°C)
PARVA	F-CAGGAGGAGGGAATGAAC	248	55
	R-CGCTCCTGATTTCTTCTTAC		
HMGXB3	F-CAGGCATGGCAGAGCAGT	221	61
	R-AGTGGGGAGGTCTGGTGAG		
UBR4	F-GCAAGAGCAGGAAGAATTTG	300	57
	F-CTGGGCCTCTGGAGTTACTA		

2.8.6 Investigation of genes harbouring possible mutations in heart tissue

Primers specific to cDNA were designed (Table 2.11) as mentioned in the investigation of gene expression in peripheral blood (see section 2.7.4). PCR products were visualized in an agarose gel.

Table 2.12: primer information for gene expression in heart tissue

Gene	Primer sequence	Amplicon size (bp)	Melting temp. (°C)
PARVA	F- CACTCATCACTTTCGTGAAC	170	54
	R- CTTCTGTTCAAAGCTGTCC		
HMGXB3	F- AGGTAACTGTCGTGATGGAG	300	50

Chapter 3

Prevalence of desmosomal gene mutations in Arrhythmogenic Right Ventricular Cardiomyopathy and Dilated Cardiomyopathy

3.1 INTRODUCTION

The first report of desmosomal gene mutations as a cause of dilated cardiomyopathy (DCM) in addition to arrhythmogenic right right ventricular cardiomyopathy (ARVC) was the

discovery of recessive desmoplakin mutations as a cause of the Carvajal syndrome (Norgett, Hatsell *et al.* 2000). Carvajal syndrome is characterised by palmoplantar keratoderma with left ventricular cardiomyopathy and woolly hair. Subsequently, Elliot *et al.* identified autosomal dominant desmosomal gene mutations as a cause of DCM in 5% of English patients (Elliott, O'Mahony *et al.* 2010). These observations suggested that desmosomal gene mutations cause both ARVC and DCM outside Africa. However, little is known about the frequency of desmosomal gene mutations in Africans with ARVC and DCM (Sliwa and Mayosi 2013, Sliwa and Mayosi 2013).

Prior to my work, there were no studies of desmosomal gene mutations in African patients with cardiomyopathy. My initial work included genetic screening of the first 36 participants in the ARVC Registry of South Africa for mutations in the plakophylin-2 (*PKP2*) gene (Watkins, Hendricks *et al.* 2009). In this study we demonstrated that *PKP2* mutations are found in a quarter of African patients with ARVC. In this work, I have extended the genetic screening of desmosomal gene mutations in cardiomyopathy in three significant ways. First, I screened an additional 29 ARVC patients for mutations in *PKP2*. Second, I screened 65 ARVC patients enrolled in the ARVC registry of South Africa for mutations in three additional desmosomal genes: desmocollin 2 (*DSC2*), desmoglein 2 (*DSG2*), and plakoglobin (*JUP*). Finally, I extended the desmosomal genetic screening to 112 cases of DCM based on the findings of Elliot *et al* (Elliott, O'Mahony *et al.* 2010) in order to delineate the prevalence of desmosomal gene mutations in Africans with DCM.

Hypothesis

We hypothesized that in our cohort of South Africans with cardiomyopathies, both ARVC and DCM are caused by desmosomal gene mutations, and that the same mutation may result in either in an ARVC or DCM phenotype (i.e., phenotypic heterogeneity).

3.2 STUDY AIM

The main aims of this chapter were to determine:

1. The prevalence of desmosomal gene mutations in South Africans with ARVC as compared with DCM;
2. Whether participants with ARVC and DCM share similar mutations, and thus lending support to the hypothesis of phenotypic heterogeneity of desmosomal gene mutations;

3.3 METHODS

DNA samples were available for 177 index cases with cardiomyopathy which were classified into ARVC (n=65) and DCM (n=112). I screened for mutations in the following desmosomal genes: plakophilin-2 (*PKP2*), desmoglein-2 (*DSG2*), desmocollin-2 (*DSC2*), and plakoglobin (*JUP*). The fifth desmosomal gene, desmoplakin (*DSP*) has been screened by another student in the laboratory (Ms Maryam Fish) and the findings have been reported in her MSc thesis (Fish 2010). All methods details for phenotyping and genotyping are described in Chapter 2 (Section 2.4)

3.3.1 Routine mutation screening in desmosomal genes

Two genes, *DSC2* and *PKP2*, were screened using a WAVE dHLPC (denatured high-liquid performance chromatography) technology (Transgenomic, Irvington, NE). *DSG2* and *JUP* were screened using HRM (High resolution melt) with Rotorgene (Corbett Life Science). All

amplicons with melting temperature alterations were sequenced on the ABI 3100 automated sequencer using the BigDye Terminator v3.1 Cycle Sequencing kit and analyzed using BioEdit software (Ibis Biosciences, Carls-bad, CA, USA).

3.3.2 Identification of novel variants

There were four online databases used to verify the frequency of variants found in this study : 1000 Genomes (www.1000genomes.org (25 January2014)); Exome Variant Server (EVS) evs.gs.washington.edu; ARVC database (www.arvcdatabase.info 25 January 2014) and dbSNP database (www.ncbi.nlm.nih.gov/SNP/25January 2014). All of these databases were used to identify whether the variant had been previously identified and to obtain allelic frequency information.

3.3.3 Determination of pathogenic significance of variants

The pathogenicity of rare novel variants was assessed using the polyphen-2 (http://genetics.bwh.harvard.edu), Sorting intolerant from intolerant (SIFT) (http://sift.bii.a-star.edu.sg/), and MutationTaster (http://www.mutationtaster.org/ 25 January 2014)) prediction programmes. Only those variants predicted as disease causing in a protein of interest by at least two of the prediction tools were considered for further analysis. I also used the Berkeley splice site prediction by neural network tool (http://www.fruitfly.org/seq_tools/splice.html 29 january 2014)) and alternate splice site predictor (http://wangcomputing.com/assp/index.html 29 January 2014)) to predict splice site mutations.

3.3.4 Statistical analysis

Categorical variables are expressed as frequency (percentage) and continuous variables are expressed as mean ± SD or median (interquartile range). Gender; age and ethnicity were

determined using a Microsoft excel (2010) statistical formulas. Statistical significance was performed using Chi-square test; or Student's t-test where appropriate. Statistical significance was taken as a two-sided $P < 0.05$.

3.3.5 Analysis of SNPs and disease-causing mutations

We defined a polymorphism as a variant with a minor allele frequency of ≥1%, using either a local study of 232 local population controls or using previously published studies. Individual sequence variants were considered disease causing if (1) they were absent in our control population and public databases listed under section 3.3.2 above; (2) they segregated with the disease where family information was available; (3) they coded for an evolutionarily conserved amino acid (in certain cases); and (4) they were predicted to be pathogenic by at least two of the following prediction tools: Mutation taster, Polyphen2 and SIFT prediction software; or (5) they had been reported previously to cause ARVC and/or DCM. Sequencing variants not complying with these criteria were classified as genetic variants of unknown significance. All sequence variants were cross-referenced with an online database of ARVC mutations (www.arvcdatabase.info 16 February 2014)

3.4 RESULTS

3.4.1 Clinical characteristics of participants with arrhythmogenic right ventricular cardiomyopathy

Between January 2004 and December 2011, 65 unrelated individuals with a definite diagnosis of ARVC were recruited for molecular genetic analysis (Table 3.1). The ethnic origin of these 65 index cases was as follows: white South Africans 37 (56.9%), mixed ancestry South Africans 16 (24.6%), black South African 8 (12.3%), and Indian South Africans 4 (6.2%). The mean age at presentation was 40 ± 15 years and the majority (60.0%) were male. Palpitations were the most frequent symptom (67%), followed by dizziness

(40%), chest discomfort (26%), and syncope (20%). Table 3.1 compares the baseline characteristics of participants with familial disease to those with no history of familial disease. There was no significant difference in the clinical characteristics of probands with or without a family history of ARVC.

Table 3.1: Clinical characteristics of participants with arrhythmogenic right ventricular cardiomyopathy

Characteristic	ARVC Patients			*P*-value
	All	**History of Familial Disease**	**No History of Familial Disease**	
No. of Patients (Index cases) (%)	65 (100.0)	16 (24.6)	49 (75.4)	
Gender:				0.72
Male (%)	39 (60.0)	9 (56.3)	30 (61.2)	
Female (%)	26 (40.0)	7 (43.8)	19 (38.7)	
Average age in years (±SD)	40±15	36±14	41 ±16	0.45
Ethnicity:				
Black (%)	8 (12.3)	2 (12.5)	6 (12.2)	0.19
Mixed Ancestry (%)	16 (24.6)	4 (25.0)	12 (24.5)	
White (%)	37 (56.9)	10 (62.5)	27 (55.1)	
Indian (%)	4 (6.20)	0 (0.00)	4 (8.20)	
Presenting symptoms:				
Palpitations (%)	44 (67.7)	9 (56.3)	35 (71.4)	0.23
Dizziness (%)	26 (40.0)	8 (50.0)	18 (36.7)	0.35
Chest discomfort or chest pain (%)	17 (26.6)	6 (37.5)	11 (22.5)	0.39
Syncope (%)	13 (20.0)	4 (25.0)	9 (18.4)	0.83
Dyspnoea (%)	5 (7.70)	0 (0.00)	5 (10.2)	NS
Suddencardiac death (%)	9 (13.8)	2 (12.5)	7 (14.3)	0.81
Asymptomatic (%)	5 (7.70)	1 (6.30)	4 (8.20)	0.77

Values are given as mean ± SD; SD: standard deviation; ARVC: arrhythmogenic right ventricular cardiomyopathy; NS, not significant. History of familial disease: families with more than one members with information of ARVC diseases

3.4.2 Clinical characteristics of participants with dilated cardiomyopathy

One hundred and twelve unrelated DCM cases were made up of 9 probands (8.1%) with familial disease and 103 probands (91.9%) with no family history of DCM. The participants with DCM were predominantly male (69.6%) and of black African ancestry (63.4%). The vast majority (75.9%) were in New York Heart Association class III (Ahmed, Aronow *et al.* 2006) done or IV heart failure at presentation. As outlined in Table 3.2, the participants with familial DCM were significantly younger ($p = 0.001$) with smaller cardiac chambers compared to those without a family history of DCM.

Table 3.2: Clinical characteristics of patients with dilated cardiomyopathy

Characteristic	DCM Patients			*P*-value
	All	**History of Familial Disease**	**No History of Familial Disease**	
No. of Patients (%)	112	9 (8.1)	103 (92.7)	
Gender:				
Male (%)	78 (69.6)	6 (66.7)	72 (69.9)	
Female (%)	43 (31.1)	3(33.3)	40 (30.1)	0.260
Average age in yrs (mean) (SD)	36±14	28 ±11	39 ±12	0.001
Ethnicity:				
Black (%)	71 (63.4)	5 (55.5)	66 (64.1)	
White (%)	6 (5.40)	1 (11.1)	5 (4.90)	0.140
Indian (%)	4 (3.60)	1(11.1)	3 (2.90)	
Mixed ancestry (%)	16 (14.3)	2 (22.2)	14 (13.7)	
Unknown (%)	15 (13.3)	0 (0.00)	15 (14.7)	
NYHA FC (N (%)):				
Class I and II	27 (24.1)	2 (22.2)	25 (24.3)	0.110
Class III and IV	85 (75.9)	7 (77.8)	78 (75.7)	
Heart rate at initial presentation (beats/ min)(mean(SD))	95.14(19.3)	96.02 (17.9)	94.89 (19.5)	0.490
Blood pressure, systolic (mmHg) (mean(SD))	102.43 (18.5)	102.37 (15.1)	101.78 (18.1)	0.390
ECG LBBB (N (%)	34(30.4)	5 (55.6)	29 (28.4)	0.120
Echocardiography:				
LVEDD(cm) (mean(SD))	6.39(1.80)	6.28 (1.20)	6.84 (1.40)	0.030
LVEF (cm) (mean(SD))	24.68(11.5)	28.10 (10.9)	24.68 (11.5)	0.050

DCM: dilated cardiomyopathy; SD: standard deviation; LBBB: left bundle-branch block; RBBB: right bundle branch block; LVEDD: left ventricular end-diastolic dimension; LVEF: left ventricular ejection fraction and NYHA FC: New York heart association functional class

3.4.3 Desmosomal gene abnormalities in ARVC

The summary of desmosomal gene mutations found in ARVC patients is presented in Table 3.3. Of the 65 ARVC patients, 14 (21.5%) carried pathogenic mutations in *PKP2* (n = 9), *DSG2* (n = 2), *DSC2* (n = 1) or *JUP* (n = 2).

Table 3.3: Desmosomal gene variants found in 65 cases of ARVC

Gene	Index case ID	Familial	Ethnicity	Gender	Nucleotide change	Amino acid	Exon	Type	Frequency	Classification	Reported previously
PKP2											
	22.1	U	B	M	2146-1G>C	-	Intron 10	Splice site	0.0	Pathogenic	(Dalal, James *et al.* 2006)
	19.2	F	W	F	c.2197-2202delCACAinsG	A733fsX740	11	Insertion/Deletion	0.0	Pathogenic	(Dalal, James *et al.* 2006)
	1.2	F	W	M	c.1132C>T	p.Q378X	4	Nonsense	0.0	Pathogenic	(Watkins, Hendricks *et al.* 2009)
	5.1 ; 12.1; 19.2 & 38.3	F,U,F& F	W	F; M; F& M	c.1162C>T	p.R388W	4	Missense	0.0	Pathogenic	(Watkins, Hendricks *et al.* 2009)
	51.1	U	W	M	c.2509delA	S837VfsX94	13	Frame shift	0.0	Pathogenic	(Watkins, Hendricks *et al.* 2009)
	34.5	U	W	M	c.2540T>C	p.L847P	13	Missense	0.0	Pathogenic	(Watkins, Hendricks *et al.* 2009)
	39.5	F	B	F	c.1465G>A	p.G489R	6	Missense	0.0	Pathogenic	(Watkins, Hendricks *et al.* 2009)
DSG2											
	6.1 & 34.5	F&U	W	M	c.1303G>A	p.D435N	4	Missense	0.0	Non-pathogenic (no segregation with disease)	(Cox, van der Smagt *et al.* 2010)as GVUS
	7.1	U	W	F	c.1435A>G	p.K479Q	11	Missense	0.0	Pathogenic	Novel
	48.1	U	B	M	c.1477A>G	p.N493S	11	Missense	0.0	Pathogenic	Novel
DSC2											
	53.1	U	W	M	c.2587G>A	p.G863R	16	Missense	0.0	Pathogenic	(Cox, van der Zwaag *et al* 2011)
	15.1	U	B	M	c.685C>A c.2446G>A	p.L227I p.V816M	6 15	Missense Missense	0.0 0.0	GVUS GVUS	Novel Novel
JUP											
	31.1	U	W	M	c.496G>A	p.V166M	4	Missense	0.0	Pathogenic	Novel
	37.1	U	W	M	c.533C>T c.542G>C	p.A178V p.G181A	4	Missense Missense	0.0 0.0	Pathogenic GVUS	Novel
	16.1	U	W	M	c.1910G>C c.741+2 C>T	p.R637P -	11 14	Missense UTR	0.0 0.0	GVUS GVUS	Novel rs112879398
	16.1 & 30.1	U& F	W	M& F	c.741+2 C>T	-	14	UTR	0.0	GVUS	rs112879398

3.4.3.1 Plakophilin 2 (PKP2)

There were 9 (13.9%) probands who had one of 7 pathogenic mutations in the *PKP2*. Four of the probands were of white ancestry and demonstrated a recurrent c.1162T>C change (Table 3.3). One out of 9 (11.1%) probands had compound heterozygous mutations (c.1162T>C in exon 4 and c.2197-2202delCACAinsG in exon 11) in the *PKP2* gene (Watkins, Hendricks *et al.* 2009). The *PKP2* mutations outlined in Table 3.4 were considered to be pathogenic based on the criteria listed above. Figure 3.1 illustrates the sequence analysis of the c.1162C>T mutation (A), the conservation of the wild type amino acid in evolution (B), and the segregation of the mutation with disease in family ACM 38 (C).

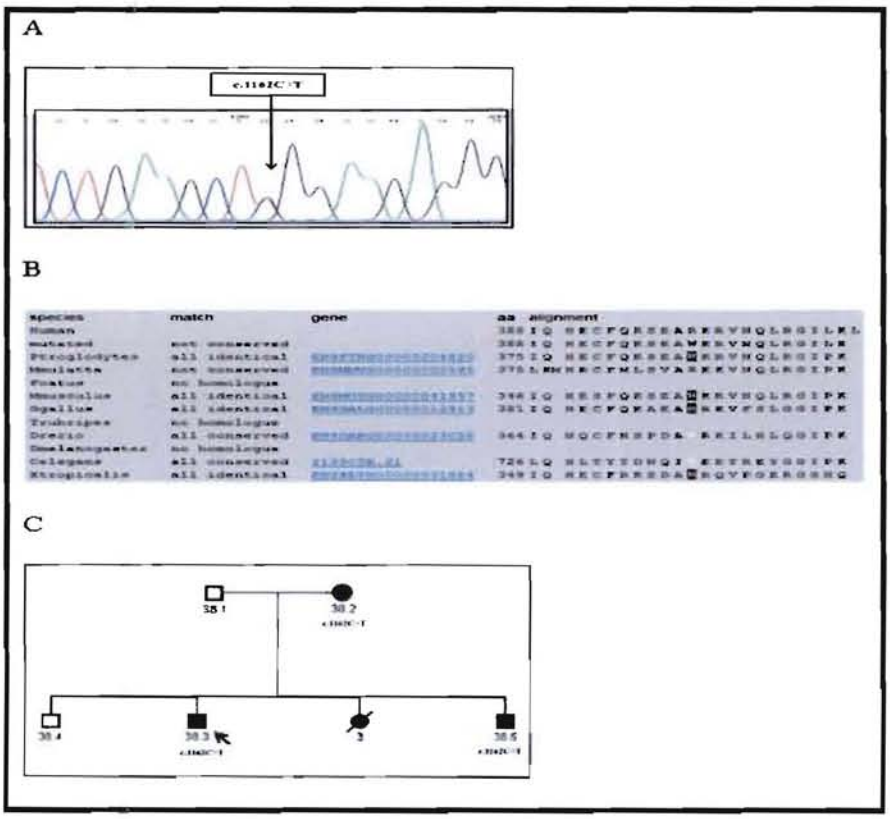


Figure 3.1: Data that demonstrate a disease-causing plakophilin-2 mutation in a proband with ARVC. A, an electropherogram showing an A>T nucleotide substitution; B, amino acids alignment showing mutation conservation information of the amino acid affected; C, the segregation of the mutation with disease in family ACM 38. Ptroglodytes (chimpanzee); Mmulatta (rhesus macaques); Mmusculus (House mouse); Ggallus (chicken); Trubripes (puffer fish); D rerio (Zebra fish); D melanogaster (Drosophila); C elegans (Nematode); X tropicalis (frogs).

Another variant we identified was c.1465G>A (p.R489P), located in exon 6 of *PKP2*. It is not present in *PKP2A* isoform, which is highly expressed in human heart tissue (Gandjbakhch, Charron *et al.* 2011), but occurs in the *PKP2B* isoform.

All 7 *PKP2* mutations detected in this study are located in the armadillo repeat domain, which is known to play an important role in protein folding; disruption in protein folding results in alteration of protein function which may result in cardiomyopathy (Kirchner, Schuetz *et al.* 2012).

3.4.3.2 Desmoglein 2 (DSG2)

In *DSG2*, two novel mutations were identified in two unrelated probands (3.1%); these were missense mutations in exon 11 (c.1435A>G and c.1477A>G). Unfortunately, the probands had no family members available for screening. The mutations were not found in ARVC database, 1000 genomes, NCBI dbSNPs and 232 population controls. They were predicted to be damaging and disease causing by three bioinformatic prediction tools: Polyphen2; SIFT and Mutation taster. The mutations both code for amino acids located in extracellular cadherin (EC) region of the protein

The c.1435A>G mutation codes for p.K479Q, and is associated with the substitution of a conserved polar uncharged amino acid (lysine) with a polar basic (glutamine) type.

The c.1477A>G mutation, results in a change of asparagine (a polar amino acid) to serine (a neutral amino acid) (p.N493S). The extracellular region is responsible for desmoglein/desmocollin protein interactions (Gehmlich, Asimaki *et al.* 2010), and the alterations caused by these amino acids changes were predicted to have a deleterious effect on protein function.

Additionally, a missense mutation (c.1303G>A; p.D435N) was identified in exon 10 of *DSG2* in two unrelated individuals. Although this variant affects a conserved amino acid and was not found in 232 population controls, it did not segregate with disease in a family that was studied. It was reported previously (Cox, van der Smagt *et al.* 2010) to be a variant of unknown significance and was predicted to be non-pathogenic by the Mutation Taster

prediction tool. Therefore, we provide additional evidence that c.1303G>A is not a disease causing variant in ARVC.

3.4.3.3 Desmocollin 2 (DSC2)

One of the 65 probands (1.5%) with ARVC had a missense mutation c.2587G>A (p.G863R) in exon 16 of *DSC2*. This mutation was reported as disease causing (Cox, van der Zwaag *et al.* 2011). It was also predicted by Mutation taster, Polyphen2, and SIFT to be causing deleterious mutation and was not found in 232 population controls. Family members for this proband were not available to assess segregation with the disease. This mutation is located in the intracellular cadherin segment (ICS), which is well conserved in cadherins and is responsible for the interactions of the other desmosomal proteins with cadherins (Gehmlich, Asimaki *et al.* 2010). The extracellular domains contribute to the interactions of cadherins with other cadherin proteins, e.g. interaction of desmoglein with desmocollin proteins (Kirchner, Schuetz *et al.* 2012).

Two novel genetic variants of unknown significance (GVUS) were identified: a missense variant c.685C>A (p.L227I) in exon 6 and another missense variant c.2446G>A (p.V816M) in exon 15 in the same index case. The c.685C>A variant was predicted by Mutation taster to be a polymorphism, by Polyphen2 to be damaging, and by SIFT to be tolerated. The c.2446C>A variant was predicted by Mutation taster to be a polymorphism, by Polyphen2 to be benign, and by SIFT to be tolerated. The variants were not present in 232 population controls and databases. The second variant, c.2446C>A, was predicted not to have a damaging effect by all prediction tools, and there were no samples from family members to investigate regarding segregation with disease. Thus we classified both variants as GVUS.

3.4.3.4 Plakoglobin (JUP)

In *JUP*, there were two pathogenic mutations identified in two separate probands (3.1%): a c.496G>A missense mutation (p.V166M) in exon 4, and a c.533C>T missense mutation (p.A178V) in exon 4 in 2 unrelated individuals. Both of these mutations were novel, coded for a conserved amino acid, and were absent in 232 population controls as well as SNP databases. The p.V166M mutation changes valine to methoinine and both amino acids belong to non-polar amino acids. The variant was predicted by Mutation taster to be disease causing; Polyphen2 and SIFT also predicted the change to be damaging. The p.A178V mutation changes a polar amino acid (arginine) to non polar amino acid (valine). The arginine is conserved among modern animals and the variant was not detected in 464 control chromosomes, dbSNP, 1000 Genomes and ARVC databases. The variant was predicted to be disease causing by Mutation taster; predicted to be damaging by Polyphen2 and SIFT. There were no family members available to test for segregation of the variant with the disease. The two mutations are located in the armadillo repeat domain of *JUP*.

Additionally, three non-pathogenic variants in *JUP* were detected. The first was a p.G181A missense mutation identified in exon 4 of *JUP* in ACM 37.1 was predicted by Mutation taster to be a polymorphism; Polyphen2 predicted it to be benign and SIFT predicted it tolerated change. The second variant was a p.R637P missense mutation in exon 11 of *JUP* and was predicted by Mutation taster to be polymorphism, by SIFT to be tolerated, and by Polyphen2 to be benign. The last variant was a reported c.*2C>T (rs112879398) variant identified in exon 14 of untranslated region (UTR) of *JUP* in two unrelated probands, ACM 16.1 and ACM 30.1. The variant is not predicted by splice site prediction tool by neural network to

cause any effect on the splice site region. In the 1000 Genomes database and dbSNP database, it is described as having a minor allelic frequency of 0.005. These variants were not found in our 232 population controls. Thus we concluded that p.G181A and p.R637P were GVUS and rs112879398 was likely to be a polymorphism.

3.4.4 Desmosomal gene abnormalities in DCM

The summary of desmosomal gene mutations found in DCM patients is presented in Table 3.4. In 112 probands, there was only one mutation in *PKP2*, a c.2540T>C (p.L847P), which was considered by us to be pathogenic in ARVC (Watkins, Hendricks 2009). However, this mutation did not show segregation with disease within this DCM family (Figure 3.2). We provide strong evidence that this mutation is not pathogenic in DCM, and its pathogenicity in ARVC needs further confirmation through family studies. A missense GVUS was identified in *DSC2* (p.V816M), the proband had no history of familial DCM. Finally, a known rare polymorphism, c.741+2 C>T (c.*2C>T) (rs112879398) was identified in the 3' UTR of *JUP* in three probands with DCM. It was not predicted to alter the splice site by splice site prediction tool.

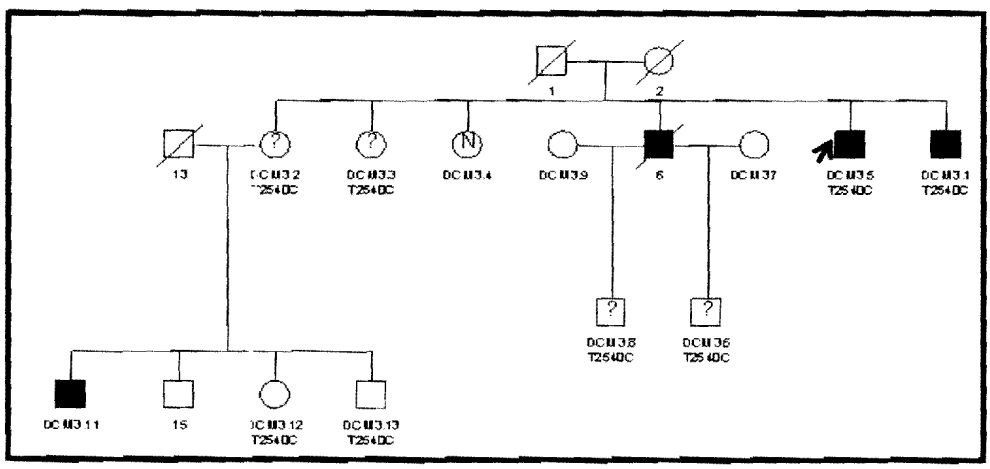


Figure 3.2: Family DCM 3 illustrate no segregation of variant T2540C

3.4.5 Desmosomal gene polymorphisms common to ARVC and DCM

There were 14 polymorphisms identified in DCM patients as well as ARVC patients, including: *PKP2* (n=3), *DSG2* (n=4), *DSC2* (n=5) and *JUP* (n=2) (Table 3.5). The two novel SNPs in *PKP2* and *JUP* are in the process of being registered in the SNP database (dbSNP).

Table 3.4: Desmosomal gene mutation and genetic variants of unknown significance (GVUS) found in 112 cases with DCM

Gene	Index case ID	Familial	Ethnicity	Gender	Nucleotide change	Amino acid	Exon	Type	Frequency	Classification	Reported previously
PKP2											
	3.5	F	B	M	c.2540T>C	p.L847P	13	Missense	0.0	Non-pathogenic (no segregation with disease)	**(Watkins, Hendricks *et al.* 2009)**
DSC2											
	55.1	U	B	M	c.2446G>A	p.V816M	15	Missense	0.0	GVUS	**Novel**
JUP											
	82.1;303.1&333.1	U	B	M	c.741+2 C>T	-	UTR	Splice site	0.0	GVUS	**rs112879398**

F: familial; and U: undertermined

Table 3.5: Desmosomal gene polymorphisms common to both ARVC and DCM cohort members (SNPs)

Gene	Intronic/ exonic	Polymorphisms	Type	Frequency	Reported previously
Plakophilin (*PKP2*)	Exonic	c.1097C>T; p.L366P	Missense SNP	42%	**rs1046116**
	Intronic	IVS11+7C- T	Intronic SNP	35%	**rs74072938**
	Intronic	IVS13+83G-A	Intronic SNP	25%	**Novel**
Desmoglein (*DSG2*)	Exonic	c.3314C>T; p.T1105I	Missense SNP	13%	**rs3211319**
	Exonic	c.2137G>A; p.E713K	Missense SNP	12%	**rs79241126**
	Exonic	c.1051A>G; p.S351G	Missense SNP	0.0%	**rs13936669**
	Intronic	c.217-5G-C	Intronic SNP	1%	**1000 Genomes**
Desmocolin (*DSC2*)	Exonic	c.2327A>G; p.I776V	Missense SNP	7%	**rs1893963**
	Exonic	c.2393G>A;p.R798Q	Missense SNP	7%	**rs61731921**
	Exonic	c.32A>G;p.N11S	Missense SNP	23%	**rs868333**
	Intronic	22709-22713del TTAA	Deletion polymorphism	15%	**rs35172389**
	Exonic	c.2686_2687insGA	Insertion polymorphism	8%	**(Syrris, Ward *et al.* 2006)**
Plakoglobin (*JUP*)	Exonic	c.213T>C; p.D71D	Synonymous SNP	2.6%	**rs7405731**
	Exonic	c.552G>A; --p.Q184Q	Synonymous SNP	2.6%	**Novel**

3.5 DISCUSSION

This is the first report of the frequency of desmosomal gene mutations in South Africans with ARVC and DCM. This chapter shows that desmosomal gene mutations are common in patients with ARVC (at least 21% of cases) than those with DCM (no pathogenic mutation observed in *PKP2*, *DSC2*, *DSG2* and *JUP*). The c.2540T>C (p.L847P) mutation in *PKP2* that has been associated with ARVC proved not to be pathogenic in DCM, suggesting that its pathogenic role in ARVC needs confirmation by further testing. Our study detected 3 out of 8 black South Africans (37.5%) with potentially disease mutations in ARVC.

3.5.1 The frequency of desmosomal gene variants in ARVC compared with other studies

PKP2 mutations have been reported to be more common as compared to the other desmosomal gene mutations in patients with ARVC. Recent studies have shown that *PKP2* (which in this study had a prevalence rate of 13.9%) is the most common gene implicated in ARVC, followed by *DSP* or *DSG2*, then *DSC2* and *JUP* (Bauce, Nava *et al.* 2010, Rigato, Bauce *et al.* 2013), the prevalence of DSC2 and JUP is no statistically significant but it is mentioned to shown a picture of the direction of these genes results. The frequency of *DSP* mutations was 3% (3/62) in our ARVC cohort (Fish 2010). In the present study, both *DSG2* and *JUP* genes accounted for 3.1% of cases, and *DSC2* 1.5% of cases. We did, however, find *JUP* variants to be more common than *DSC2* variants, in contrast to other studies (Bauce, Nava *et al.* 2010, Rigato, Bauce *et al.* 2013), although our results are not statistical significant

.

3.5.2 The frequency of desmosomal gene variants in DCM compared with other studies

A previous study from our laboratory revealed a deleterious mutations in the desmoplakin (*DSP*) gene in 6/150 (4%) patients with DCM (Fish 2010). The screening of four additional

desmosomal genes –*PKP2*, *DSG2*, *DSC2*, and *JUP* – did not identify an unequivocally pathogenic mutation that causes DCM. The frequency of desmosomal gene mutations in DCM of approximately 3% is similar to the findings of Elliott *et al* in the UK (Elliott, O'Mahony *et al.* 2010). However, *PKP2* mutations were the dominant cause of DCM in the UK cohort, whereas *DSP* mutations were the dominant cause in our South African cohort (Fish 2010).

3.5.3 Phenotypic variation arising from identical desmosomal gene mutations

It is well established that mutations in one gene can result in different types of cardiomyopathies (i.e., phenotypic heterogeneity). It has been shown that sarcomeric gene mutations may be associated with both hypertrophic cardiomyopathy (HCM), DCM and restrictive cardiomyopathy (RCM) (Watkins, Ashrafian *et al.* 2011). Identical tropin I mutations can lead to a phenotype of RCM and HCM in the same family (Mogensen, Kubo *et al.* 2003). There was no evidence, however, of phenotypic heterogeneity in our study.

Recently, other investigators have shown that mutations detected in DCM patients have been reported with the ARVC phenotype, but their clinical features did not fulfil the recently-modified criteria for the diagnosis of ARVC (Garcia-Pavia *et al* 2013). In their patients with familial DCM who had desmosomal mutations, (Garcia-Pavia, Cobo-Marcos et al. 2013)Garcia-Pavia *et al*. found no correlation between presence of a mutation and fibrofatty changes in the myocardium, which led them to challenge the histological criteria for the diagnosis of ARVC (Garcia-Pavia *et al* 2013). Furthermore, other investigators reported lethal desmosomal mutations in sudden death patients with none of the pathological features

of ARVC, and they suggested that the link between desmosomal mutations and ARVC be re-evaluated (Zhang, Tavora et al. 2012).

3.5.4 Mutation hotspots in ARVC gene mutations

The location of disease-causing mutations in *DSG2, DSC2*, and *JUP* follow the expected sites of mutation hotspots in ARVC. Pilichou and his colleagues have shown mutation hotspots are located in extracellular domains of *DSG2* (Pilichou, Nava *et al.* 2006); likewise the mutations we identified were located in the same region of the gene (Figure 3.3A).

In *DSC2*, reported mutation hotspots are located in the region of the gene coding for the extracellular domain and the C-terminus (Awad, Calkins *et al.* 2008). The mutations that were identified in our study were detected in extracellular domain, and there was also one reported mutation that was found in the intracellular cadherin specific (ICS) region as described above (Figure 3.3B).

In *JUP*, mutation hotspots are located in the region of the gene coding for the N-terminus and armadillo repeat units; (Awad, Calkins *et al.* 2008) our mutation was identified in the armadillo repeat units (Figure 3.3C). No specific mutation hotspots have been described in *PKP2*; this study identified mutations in the N-terminus and armadillo repeat units. Other authors have suggested that these are the functional domains of the gene (Awad, Calkins *et al.* 2008) (Figure 3.3D). It is remarkable that the GVUS and rare 3' UTR polymorphism were also observed with both cases of cardiomyopathies (Figure 3.2).The desmosomal gene mutations implicated in DCM are found in the same regions of the genes where ARVC mutations are also located.

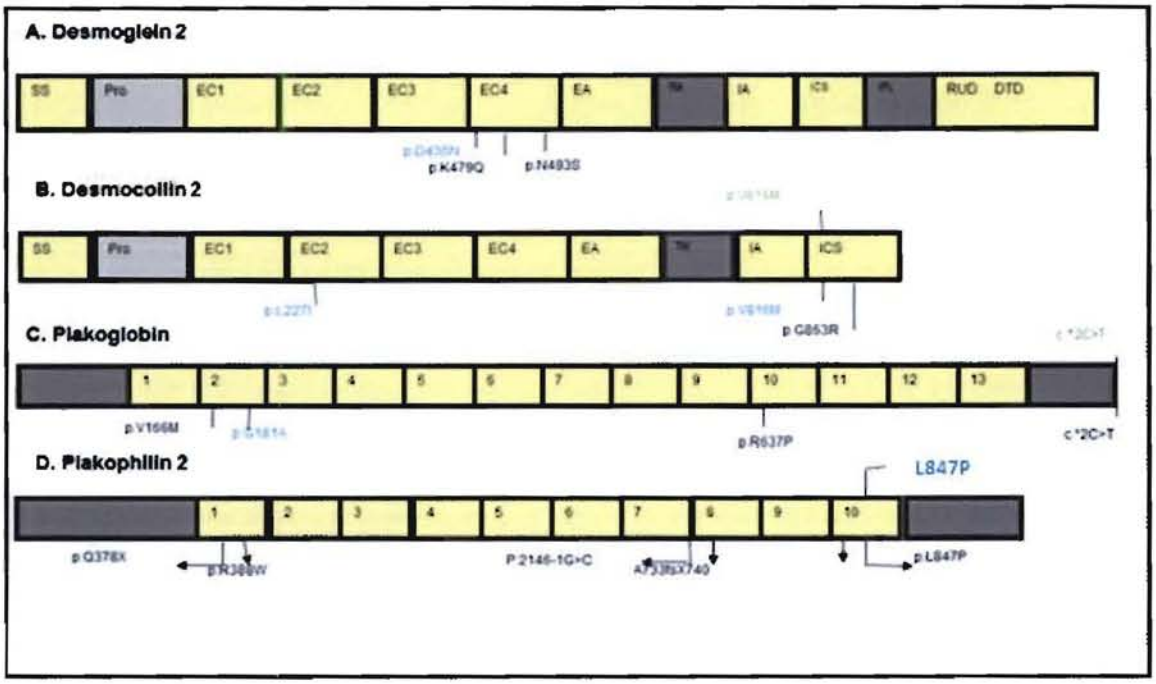


Figure 3.3: Desmosomal gene domains affected by mutations in cardiomyopathy

SS: signal peptide sequence; Pro: propeptide; EC1-4: extracellular domains 1-4; EA: extracellular anchor; TM: transmembrane domain; IA: intracellular anchor; ICS: intracellular cadherin segment; IPL: intracellular proline-rich linker; RUD: 6 repeated-unit domains; DTD: desmoglein-specific terminal domain and in plakophilin-2 and plakoglobin, 1-13 stands for armadillo repeats units

Black arrows: ARVC mutations; blue variants: ARVC genetic variants of unknown; green variants: DCM genetic variants of unknown significant; red variant: DCM mutation

3.5.5 Potential roles for polymorphisms in disease pathogenesis

I have clarified the status of two GVUS in this study (results section 3.4.3.1 and 3.4.3.2). The first, a c.1303G>A (p.D435N) variant in *DSG2*, was previously reported as a GVUS associated with ARVC (Cox, van der Smagt *et al.* 2010). In our study, there was no segregation of this GVUS with ARVC, which suggests that it is a non-pathogenic mutation in ARVC. The second, a c.1051A>G (p.S351G) variant in *DSG2*, was previously reported (Quarta, Muir *et al.* 2011) as a GVUS. We determined this to be a SNP according to data from the 1000 Genomes and dbSNP databases (rs139326669).

There were 12 SNPs out of 14 polymorphisms identified in desmosomal genes in this study (Table 3.5). Many SNPs in certain genes have been associated with the disease phenotypes by modifying gene expression and imparting allelic expression imbalance (Stranger, Nica *et al.* 2007, Cunnington, Santibanez Koref *et al.* 2010, Schaub, Boyle *et al.* 2012). However, to date, no studies of the impact of SNPs on gene expression have been performed in desmosomal genes that cause ARVC. In chapter 5, we describe the results of the first allelic expression imbalance study based on 3 of these SNPs. Since c.1465G>A (p.R489P) in *PKP2* was detected exclusively in ARVC patients, we propose this variant to be pathogenic via other mechanisms, such as aberrant splicing or allelic- specific down regulation at the mRNA level (Gandjbakhch, Charron *et al.* 2011).

3.5.6 Limitations and future directions

We have not yet performed functional studies on the putative pathogenic gene mutations reported in this work. Incorporation of individual mutations in cellular and animal models could help to demostrate their role in disease pathogenesis. An additional limitation is that to date we have only screened cardiomyopathy patients who fulfilled the Task Force Criteria for ARVC; it is possible that other patients who do not meet the criteria could have desmosomal gene mutations. Furthermore, this study was too small to establish whether there are any differences in the distribution of mutations in the desmosomal genes between the two types of cardiomyopathies. Finally, there is a need for extension of the pedigrees and genetic screening in cardiomyopathy families and to compare clinical features and outcomes over time. This will be especially helpful in cases where mutations did not clearly segregate with disease or where family member samples were unavailable for the present study.

3.6 CONCLUSIONS

There are three major observations from this chapter. First, taken together with the findings of *DSP* screening which were reported elsewhere, desmosomal gene mutations account for about 25% of ARVC cases and 3% DCM cases. Elsewhere, desmosomal gene mutations are found in up to 50% of probands with ARVC (Basso, Corrado *et al.* 2012). The relatively low yield in this South African cohort for both ARVC and DCM may not be sufficient to justify clinical testing for these mutations in routine practice. Second, pathogenic mutations in desmosomal genes are an uncommon cause of DCM in South Africa. Finally, we have identified four apparently unrelated individuals with an identical mutation in *PKP2*. The occurrence of recurrent mutations in the probabnds with ARVC and DCM, provides an opportunity for the study of a potential founder effect, as described in the following chapter 4. In the future, the identification of large numbers of individuals with an identical mutation on a common haplotype background could provide a resource for genotype-phenotype studies.

Chapter 4

Investigation of the recurrent PKP2 c.1162C>T variant as a founder mutation

4.1 INTRODUCTION

In the previous chapter, I identified a recurrent *PKP2* c.1162C>T (p.R388W) mutation in four unrelated probands. The mutation was first reported previously by our group (Watkins, Hendricks *et al.* 2009) and subsequently by (Xu, Yang *et al.* 2010); in all instances, the mutation occurred in individuals of white ancestry. The recurrent c.1162C>T (p.R388W) mutation in exon 4 of *PKP2* codes for an amino acid located in the armadillo repeat unit 2 of *PKP2* (Figure 4.1). The four unrelated individuals with the recurrent *PKP2* mutation were white South Africans of European origin. This clustering of the recurrent mutations in this ethnic group raised the possibility that the mutation may have arisen on a common haplotype background due to a common ancestor (i.e., a founder effect). Afrikaans-speaking white South Africans and people of mixed ancestry are known to have a high prevalence of inherited conditions due to founder effects, such as Fanconi anemia caused by mutations in the *FANCC* gene (Tipping, Pearson *et al.* 2001); hypertrophic cardiomyopathy due to mutations in the *MYH7* gene (Moolman-Smook, De Lange *et al.* 1999); long QT syndrome due to mutations in the *KCNQC1* gene (Brink and Schwartz 2009); and familial hypercholesterolemia due to mutations in the LDLR gene (Jelassi, Slimani et al. 2011).

Recurrent mutations (i.e. the exact same mutation observed in two or more unrelated patients) are either the result of repeated *de novo* events (usually in the context of 'mutational hotspots' in the genome) or amplification in a population through vertical transmission from a common ancestor, known as a founder effect (Merner, Hodgkinson *et al.* 2008). One of the most recently discovered ARVC gene (*TMEM43*)(Merner, Hodgkinson *et al.* 2008) was

identified on the basis of linkage to chromosome 3p25 and a recurrent missense mutation (c.1073C>T: p.S358L) that was exclusively found in patients across 15 unrelated ARVC families from the island of Newfoundland, a Canadian founder population of ~510 000 people of English and Irish extraction (Mannion 1977). This strong founder effect was instrumental in pinpointing a missense mutation as causative, given that *TMEM43* was previously unassociated with human disease and there was no available functional assay (Haywood, Merner *et al.* 2013). The disease haplotype was reconstructed by genotyping polymorphic markers spanning the critical region on 3p25 in patients, which identified a 2.36 Mb shared ancestral haplotype across these 15 families. This powerful approach allowed clinical studies on 257 patients with the same mutation, which show that ARVC due to c.1073C>T:p.S358L is particularly lethal (Merner, Hodgkinson *et al.* 2008).

If another founder effect with a large number of affected families is identified through this work, then it will provide an extended platform for genotype-phenotype correlations in ARVC.

Hypothesis

I hypothesized that the recurrent c.1162C>T (p.R388W) *PKP2* mutation identified in these four ARVC probands of white South African ancestry was a founder mutation.

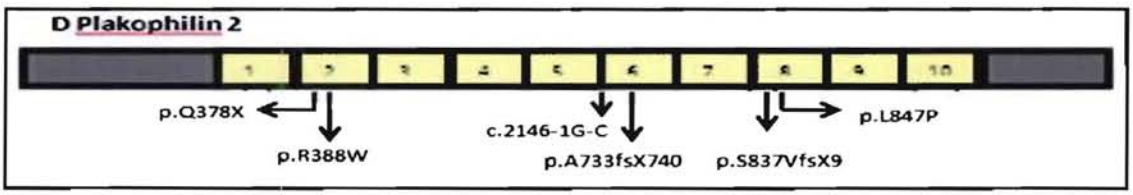


Figure 4.1: Schematic representation of *PKP2* domains and mutations described in this thesis. Grey boxes represents N and C domains, respectively, and yellow boxes represent armadillo repeat units.

4.1.1 Study aim

The aim of the study was to use microsatellite markers and single nucleotide polymorphisms (SNPs) to derive haplotypes for 4 probands and their families with the recurrent *PKP2* mutation c.1162C>T (p.R388W) in order to determine whether this mutation occurred on a common haplotype background or not (i.e., represents a founder effect).

4.2 METHODS

4.2.1 Phenotyping

The phenotyping and diagnostic evaluation of participants in the ARVC Registry of South Africa are described in full in chapter 2 (Section 2.4). Of the 4 probands with the recurrent *PKP2* c.1162C>T (p.R388W) mutation (i.e., ACM 5.1, ACM 12.1, ACM 19.2 and ACM 38.3), two had first-degree relatives who were available for further study (i.e., ACM19.2 and ACM 38.3). In family ACM 19, the recurrent mutation was transmitted by the affected mother (ACM 19.6) to affected children (ACM 19.1; 19.2 and 19.4) (Figure 4.2).

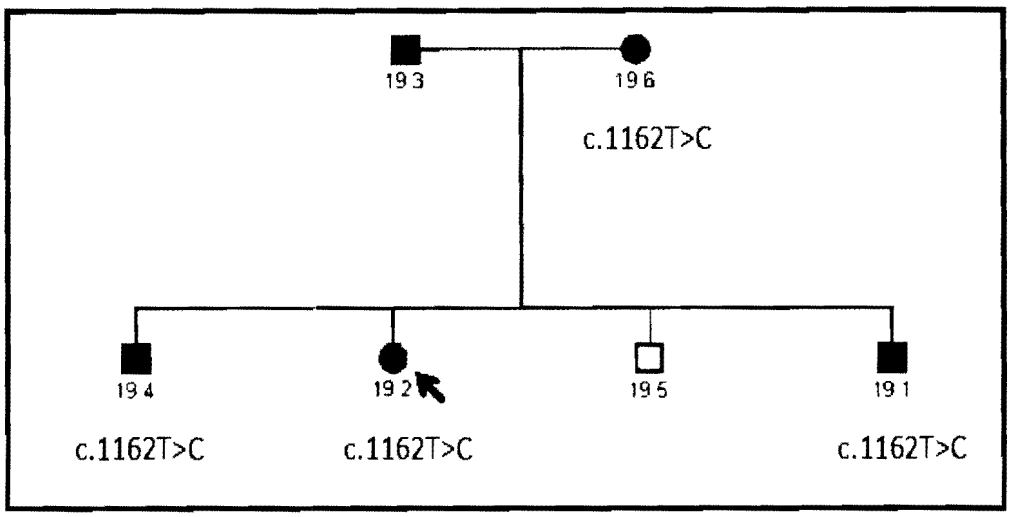


Figure 4.2: ACM 19 family pedigree

In family ACM 38, the recurrent mutation was transmitted by the affected mother (ACM 38.2) to the affected children (ACM 38.3 and 38.5) (Figure 4.3).

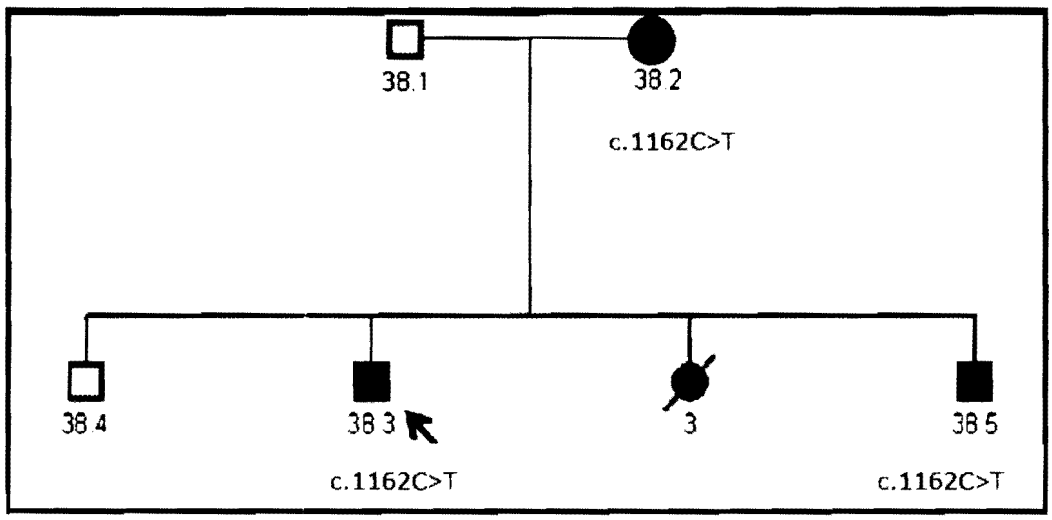


Figure 4.3: ACM 38 family pedigree

4.2.2 Haplotyping

I sought to determine whether individuals carrying the recurrent mutation in *PKP2* share a common ancestral haplotype. Microsatellite markers across *PKP2*, spanning a 300-kb region on chromosome 12 (NC_000012.10), were genotyped in the four affected ARVC index cases

as well as in the first-degree relatives of two index cases with family members available for genotyping. A single, published, intragenic marker, present in exon 7 (D12S1692) of *PKP2*, was genotyped along with two unpublished flanking markers (PKP2_5'_CA1 and PKP2_3'_TG1) on the ABI sequencer. Two other variants (T1097C and 2197-2202del/CACACCinsG mutation) were genotyped, for a total of five markers across the *PKP2* locus. Results were analyzed using GeneMapper software v3.0 (Applied Biosystems). The most probable haplotypes were constructed manually based on pedigree data and marker positions. In the two cases for which only the proband was available for genotyping, an "assumed" haplotype (Moolman-Smook, De Lange *et al.* 1999), based on the haplotype identified by family mapping, was generated. Fifty ethnically matched controls selected from 232 population controls were screened for this disease haplotype.

4.3 RESULTS

The results of the haplotyping are presented in Figure 4.4. A specific haplotype, based on five markers (three microsatellites and two variants), was present in the unrelated index cases (ACM 5.1, ACM12.1, ACM19.2 and ACM 38.3). The common PKP2_5'_CA1_12-1097_T-1162_T-D12S1612_2-DEL/INS_WT PKP2_3'_TG1_3 haplotype is represented in yellow in Figure 4.4 A-D. This disease-associated haplotype was not detected in the 50 ethnically matched population controls. These data suggest that the recurrent mutation is identical by descent in two families. According to the assumed haplotype, the 2 unrelated probands with no relatives (ACM 5.1 and ACM 12.1) appeared to be identical by descent to the other two families (ACM 19 and ACM 38). This common haplotype with the recurrent mutation accounts for 28.6% of ARVC chromosomes in the nine apparently unrelated individuals with a *PKP2* mutation. The father's phenotypes information has been added in my thesis. He presented shown no symptomps of ARVC at first and as the times goes he he developed a

mild phenotypes and he was found having a reported *PKP2* mutation (A733fsX740) in exon 11.

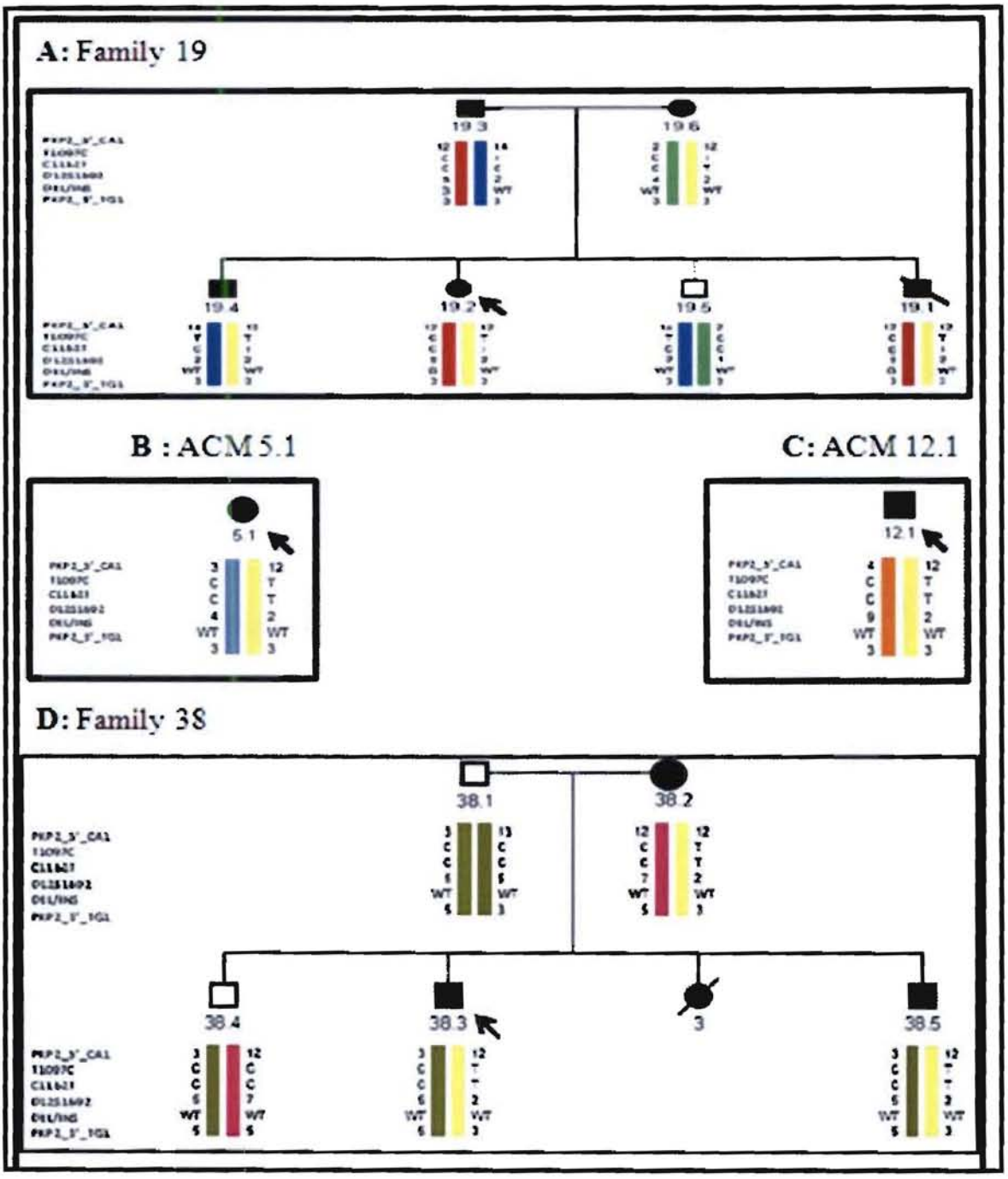


Figure 4.4: Haplotype results for ACM families with recurrent mutations. Black boxes and circles represent severely affected males/females (including stroke marks for deceased patients), and the yellow bar represents the haplotype that segregates with the recurrent mutation.

4.4 DISCUSSION

I have identified an identical mutation with a common haplotype that was present in four apparently unrelated individuals with a recurrent *PKP2* mutation. These four individuals were white South Africans of European ancestry and probably descended from a common unidentified ancestor (i.e., a founder effect). To date, this is the first report of a founder mutation in South Africans with ARVC. The observation of founder mutations in *PKP2* was first made in the Netherlands, where four different recurrent mutations were present on identical haplotype backgrounds (van Tintelen, Entius *et al.* 2006), and subsequently in Newfoundland (Merner, Hodgkinson *et al.* 2008). It is of interest that the majority of white South Africans are of Dutch origin, but the founder mutation that is present in the South African cohort is unique to the South African population.

The founder effect in South Africa likely results from the fact that a small group of Dutch people harbouring a variety of rare, single gene diseases originally settled in the Western Cape Province in 17th century (Moolman-Smook, De Lange *et al.* 1999) and later migrated to the hinterland of South Africa. Those single gene diseases were kept in their population which behaved as a population isolate due to their perceptions of origin, culture and religion (Brink and Schwartz 2009).

The c.1162C>T (p.R388W) mutation, initially identified in ACM 5.1; 12.1 and 19.2 (Chapter 3 Section 3.3.5.1), is a rare mutation and was not observed in 464 chromosomes. The mutation is located in the armadillo repeat domain and is known to play an important role in protein folding, due to a missense c.1162C>T, p.R388W recurrent mutation *PKP2* may

become unstable due to *PKP2* unfolding (Wimley and White 1996). Arginine is hydrophilic whereas tryptophan is hydrophobic, a change that may adversely affect *PKP2* protein function. This study revealed that haplotype analysis across *PKP2* (12p21) in two families (ACM 19 and ACM 38) and assumed haplotype (Moolman-Smook, De Lange *et al.* 1999) in individuals (ACM 5.1 and ACM 12.1) carried the same mutation also shared either a disease-associated haplotype or a haplotype that was assumed to be disease-associated. This disease-associated haplotype was not detected in 50 ethnically matched population controls. Our data are consistent with the possibility that recurrent mutation is identical by descent in two families and we assumed that two unrelated individuals with no relatives had the same disease-associated haplotype by descent.

The discovery of the founder effect has important implications for the molecular investigation and study of ARVC in South Africa. In the present study, the *PKP2* c.1162C>T (p.R388W) mutation accounts for 28.6% (4/14) ARVC cases with a desmosomal protein gene mutation. A simple restriction enzyme test for restriction fragment length polymorphisms is available for the identification of the founder mutation. Therefore, all new index cases entering the ARVC Registry of South Africa will be screened for this particular founder mutation prior to embarking upon full examination of *PKP2* and other desmosomal protein genes.

4.5 Limitation and future work

In two of the four probands, there were no family members available for the study, so haplotypes had to be created based on assumptions (imputing). Extension of screening to those probands' family members will be necessary to confirm our results. It will also be necessary to perform functional studies to investigate the effect of this recurrent mutation on

gene expression and protein function. In the future, screening of a large ARVC cohort to find more desmosomal founder effects could also be performed and could simplify the molecular diagnosis for clinical use. Finally, a large number of families with founder effects will be needed to create a platform for genotype-phenotype studies in ARVC (Merner, Hodgkinson *et al.* 2008).

4.6 Conclusion

I have identified a new founder mutation in *PKP2* in four families with ARVC. All patients with ARVC of white South African ancestry should be screened for this mutation prior to embarking upon the study of the genetic mutations that cause ARVC. These four families should be extended and new families sought in order to provide a large panel of cases of ARVC with the same mutation on a common haplotype background for genotype-phenotype studies.

Chapter 5

Investigation of desmosomal gene allelic expression in peripheral blood

5.1 INTRODUCTION

In chapter 3, I identified 14 single-nucleotide polymorphisms (SNPs) in both ARVC and DCM patients. While not directly pathogenic, some SNPs are known to modify the phenotype of a disease (Cunnington, Kay *et al.* 2009). Previous studies have demonstrated that heterozygous SNPs have a cis-effect that is indicated by unequal amounts of transcript from each allele, (Cunnington, Kay *et al.* 2009) e.g. in diseases such as cancer and cardiovascular disease (Chen, Weaver *et al.* 2008, Cunnington, Kay *et al.* 2009). In some cardiomyopathies, an allelic expression imbalance (AEI), which is an allelic imbalance in gene expression caused by a variant in one of the alleles, has been reported; e.g., a case-control study of DCM patients identified a rs1739843 polymorphism in intron 2 of *HspB-7*, which encodes the small heat shock protein cyHSP associated with DCM susceptibility (Brenner, Johnson *et al.* 2000, Marian and Belmont 2011). To study the effect a SNP has on gene expression, it must be common in population controls, with a minor allelic frequency of at least 5% (Stranger, Nica *et al.* 2007, Cunnington, Kay *et al.* 2009, Cunnington, Santibanez Koref *et al.* 2010). In the 14 cardiomyopathy SNPs I identified, 3 single nucleotide polymorphisms (SNPs) in 3 desmosomal genes (*PKP2*; *DSG2;* and *DSC2*) met this criterion.

Hypothesis

I hypothesized that the eligible *PKP2* SNPs I detected in 3 cardiomyopathy cases (rs1046116; *DSG2*, rs79241126 and *DSC2*, rs686333) have a cis-regulatory effect that may modify the ARVC phenotype.

5.1.1 Study aims

This chapter seeks to investigate the *cis*-effects and total gene expression of the above mentioned exonic SNPs in peripheral blood to understand the association between these SNP genotypes and gene expression.

5.2 METHODS

Detailed methods and selection of cardiomyopathy patients SNPs who met the criteria for gene expression is presented in chapter 2 (section 2.6).

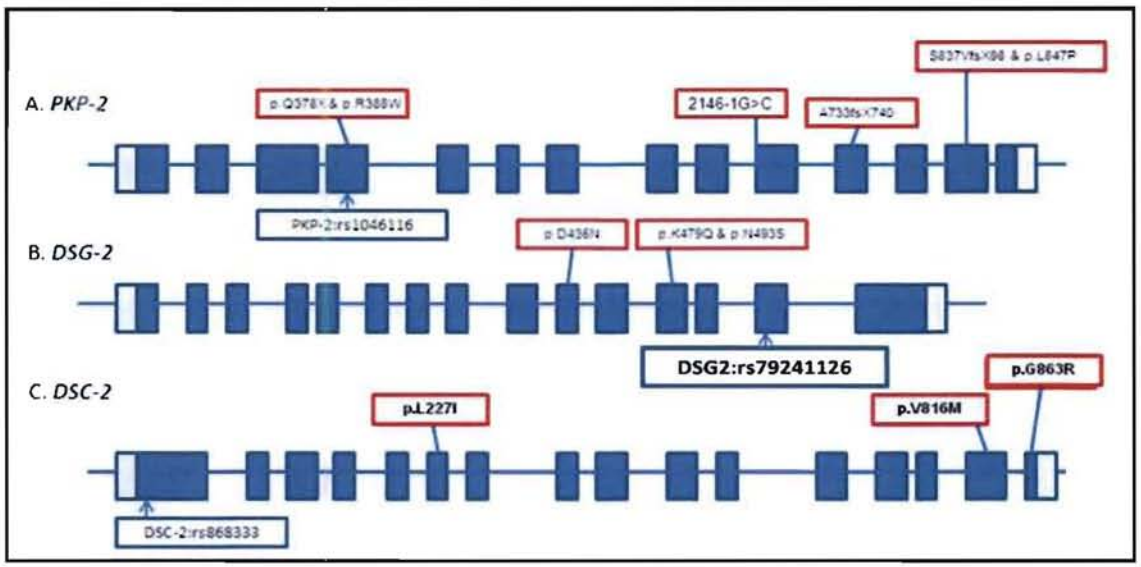
5.2.1 Selection of participants

For this study, two different sets of population controls were used: the South African (SA) cohort recruited in Cape Town, South Africa, and North East (NE) cohort recruited from Newcastle in United Kingdom. The SA participants were anticipated to exhibit greater genetic diversity as compared to European subjects. (Cunnington, Kay *et al.* 2009, Tishkoff SA, A *et al.* 2009). The cohort recruitment process is described in greater detail in Chapter 2 (section 2.6).

5.2.2 Selection of heterozygous SNPs

Three of the desmosomal genes that were screened had 3 SNPs with a minor allelic frequency that was greater than 5%. These SNPs were selected for investigation of allelic expression imbalance. In *PKP2*, located on chromosome 12, rs1046116 identified in exon 4 was selected (minor allele frequency = 42%); in *DSG2*, located in chromosome 18, rs79241126 in exon 14 was chosen (minor allele frequency = 12%) and in *DSC2*, also located in chromosome 18,

rs868333 was detected in exon 1 and chosen for the study (minor allele frequency = 23%). The location of desmosomal SNPs of interest is illustrated in Figure 5.1.



PKP2: Plakophilin-2; DSG2: Desmoglein-2; DSC2: Desmocollin-2; white box: untranslated region; blue box: exons and rs no.: reference no. for a SNP of interest, Red boxes denote identified mutations and blue boxes denote SNPs of interest in this study

Figure 5.1: Desmosomal structural genes showing mutations and regions of SNPs of interest

5.2.3 Investigation of gene expression in peripheral blood

In order to investigate allele expression imbalance (AEI), I only used samples from study subjects who were heterozygotes for the SNP in question. Because I used peripheral blood samples rather than cardiac tissues to obtain RNA for my genetic studies, it was important to confirm that the genes harbouring the SNPs of interest are expressed in RNA found in the peripheral blood.

In respect to total gene expression using North East cohort, genotyping of the SNPs of interest was carried out by Applied Biosystems TaqMan genotyping assays. Quantitative Real-time PCR reactions were carried out using TaqMan gene expression probes assays and reagents from Applied Biosystems (USA) and were performed through the 7900HT Real-time PCR System (Applied Biosystems, USA). Relative total gene expression was analysed

using the comparative cycle threshold (Ct) method using SDS Software v2.3 and RQ Manager 1.2 software (Applied Biosystems, USA) according to the manufacturer's instructions.

AEI genotyping of the SA population controls were performed to discriminate heterozygotes samples using real-time PCR as described above. Artificial heterozygotes were also made with genomic DNA from homozygotes of both major allele and minor allele to test the reliability of the Sequenom assays. The gDNA homozygotes were mixed in the following ratios (major allele: minor allele): 8:1; 4:1; 2:1; 1:1; 1:2; 1:4 and 1:8. In both mixed gDNA (artificial heterozygotes) and cDNA samples, a concentration of 25ng/ul for each sample in our SA cohort was sent to Dr Christine Blancher who is Project Manager in High Throughput Array Facility, (High throughput genomics) in Welcome Trust Centre for Human Genetics (Oxford,UK) to quantify the samples using the Sequenom platform. Quantification of the allelic expression ratio was done by primer extension and MALDI-TOF mass spectrometry. Amplification of gDNA with purification of products was performed in replicates of 4.

5.3 RESULTS

5.3.1 Investigation of gene expression in peripheral blood

In *DSC2*, both cDNA and gDNA were amplified and products were detected at position 1093 base pairs (bps) for gDNA and at position 140bp for cDNA PCR products. In *DSG2*, cDNA and gDNA were amplified and only gDNA was detected (indicated by a band at 305bp) but cDNA amplification revealed no band at position 171bp. These findings suggest a lack of expression of *DSG2* in blood and that led to its exclusion from the experiment.

In the case of *PKP2*, four gDNA bands were seen, including an expected band at position 643bp, showing non-specificity of *PKP2* primers. These primers were optimised in cDNA

and only one band was detected at the expected size. cDNA amplification products revealed a faint or weak band in the agarose gel at position 267 bps. Gene expression results are presented in Figure 5.2.

Figure 5.2: Gene expression in peripheral blood**:** In this figure, three genes are investigated and represented by numbered boxes: 1) *DSC2*; 2) *DSG2* and 3) *PKP2*, Lane 1 of each gene contains a marker to measure the size of PCR products; lane 2 is a blank to eliminate contaminators; lane 3 contains the cDNA PCR product; lane 4 contains the gDNA PCR product.

5.3.2 Whole gene expression investigation in PKP2 rs1046116

5.3.2.1 Genotyping results

Information on the selected desmosomal gene SNPs is summarised in Table 5.1. There were 182 NE cohort participants who were genotyped successfully for *PKP2* at the rs1046116 SNP: 115 were homozygous for the major allele, 59 were heterozygous, and 8 were homozygous for the minor allele. The minor allele frequency (MAF) found in white controls was similar to that of the Caucasian (CEPH) HapMap cohort (CEU) published in NCBI; SNP genotypes had a MAF of 0.20 while CEU, NCBI dbSNP had a MAF of 0.24.

All 192 white individuals from the NE cohort who were genotyped for *DSC2* rs868333 demonstrated only the wi d type alleles (i.e., the SNP was not present). *DSC2* custom probes from Applied Biosystem failed to work in the SA Black ethnicity cohort, so no results could be obtained. *DSG2* was not detected in the blood during the gene expression investigation, so no further studies were conducted on the gene. Thus, total gene expression studies were only conducted on the *PKP2* SNP.

Table 5.1: Identified minor allelic frequencies (MAF) from allelic genotyping

				South African		Caucasian	
Gene	**SNP ID**	**Chr. no.**	**Genotype**	**SA MAF**	**NCBI YRI MAF**	**NE MAF**	**NCBI CEU MAF**
PKP2	**rs1046116**	12p11	T:C	0.190	0.195	0.200	0.240
DSC2	**rs868333**	18q12.1	T:C	0.12	0.466	0.0	0.110
DSG2	***Rs79241126***	18q12.1	A:G	-	-	-	0.05

SNP: single nucleotide polymorphism; Chr: chromosome; no.: number; SA: South Africa; MAF: minor allelic frequency; NE: North East United kingdom; PKP2: plakophilin-2; DSC2: desmocollin-2; DSG2: desmoglein-2; NCBI: national centre for biotechnology information; YRI: Yoruba in Abadan, Nigeria and CEU: Utah residents with ancestry from Northern and Western Europe

5.3.2.2 Variability of PKP2 total gene expression

Total gene expression for *PKP2* indicated variability within individuals by 64-fold for the SNP rs1046116. A plot of the normalised total gene expression crossing threshold (Ct) values is shown in Figure 5.3. Ct value is a number of cycles taken for each reaction to reach an arbitrary amount of fluorescence that in turn is used to generate a relative expression level (VanGuilder, Vrana *et al.* 2008). Furthermore, the standard error on total gene expression was high, suggesting that *PKP2* gene expression was weak in blood cDNA.

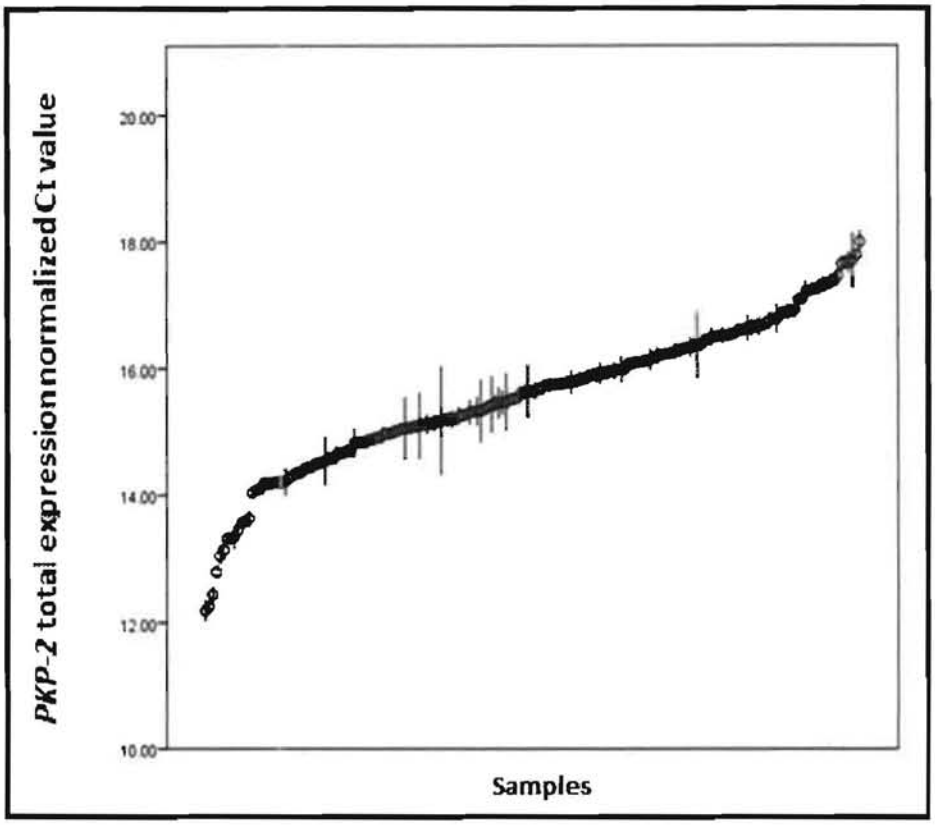


Figure 5.3: Normalized total gene expression in the North East UK cohort. Graph shows normalised total gene expression Ct values relative to reference gene. Each black circle represents an individual, and vertical lines represent standard error.

5.3.2.3 Correlation of PKP2 total gene expression with their genotypes

Genotypes of rs1046116 SNP had a normalised mean that was not statistically different in all genotypes. The normalised mean (N-mean) of the major allele was 15.53 ± 0.22; the N-mean of heterozygotes was 15.59 ± 0.29, and homozygotes for minor allele had a N-mean of 15.3 ± 0.819. The coefficient of correlation was r = -0.04, indicating that total gene expression had no correlation with the genotypes, and the p-value of this coefficient was 0.595, indicating that the findings were not statistically significant. Furthermore, Figure 5.4 shows a graph of *PKP2* genotypes against *PKP2* N-mean. The graph indicates that there is no influence of rs1046116 SNP genotypes on the *PKP2* gene expression, as the middle line appears horizontal rather than skewed.

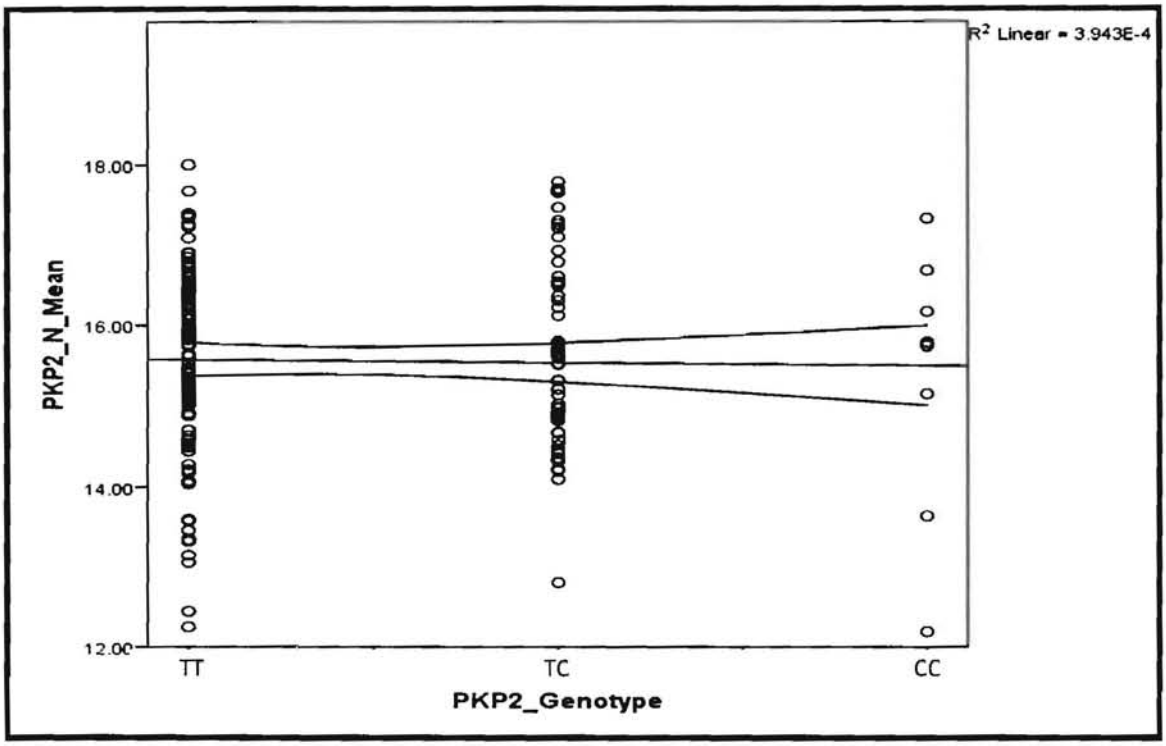


Figure 5.4: Effect of genotype on total gene expression of PKP-2 for rs1046116. The x– axis shows genotype for rs1046116 SNP. The y- axis shows the normalised total gene expression value for PKP-2. Linear regression lines are shown as middle lines; side lines indicates the 95% confidence interval

5.3.3 Allelic expression imbalance

5.3.4.1 Genotypes of PKP-2 rs1046116 SNP

Allelic expression imbalance studies were performed to confirm the total gene expression results. As the *DSC2* SNP was located at the beginning of the exon, primers could not be designed within this region, and the *DSC2* SNP had to be excluded from allelic expression imbalance study. A MAF of 0.19 of the *PKP2* SNP, rs1046116, was obtained in this cohort and was not different from reported HapMap MAF in NCBI Yoruba in the Ibadan Nigeria HapMap cohort (YRI) (MAF = 0.195).

5.3.4.2 Comparison of two gDNA and two cDNA product peaks

There were no spectral differences between artificial gDNA heterozygous and cDNA heterozygous performed to normalize the expression experiment. Figure 5.5 illustrates spectral peaks with their mass position indicating genotypes of rs1046116 SNP in both gDNA and cDNA. This shows an accuracy of experiment and that allelic expression ratio analysis can be performed.

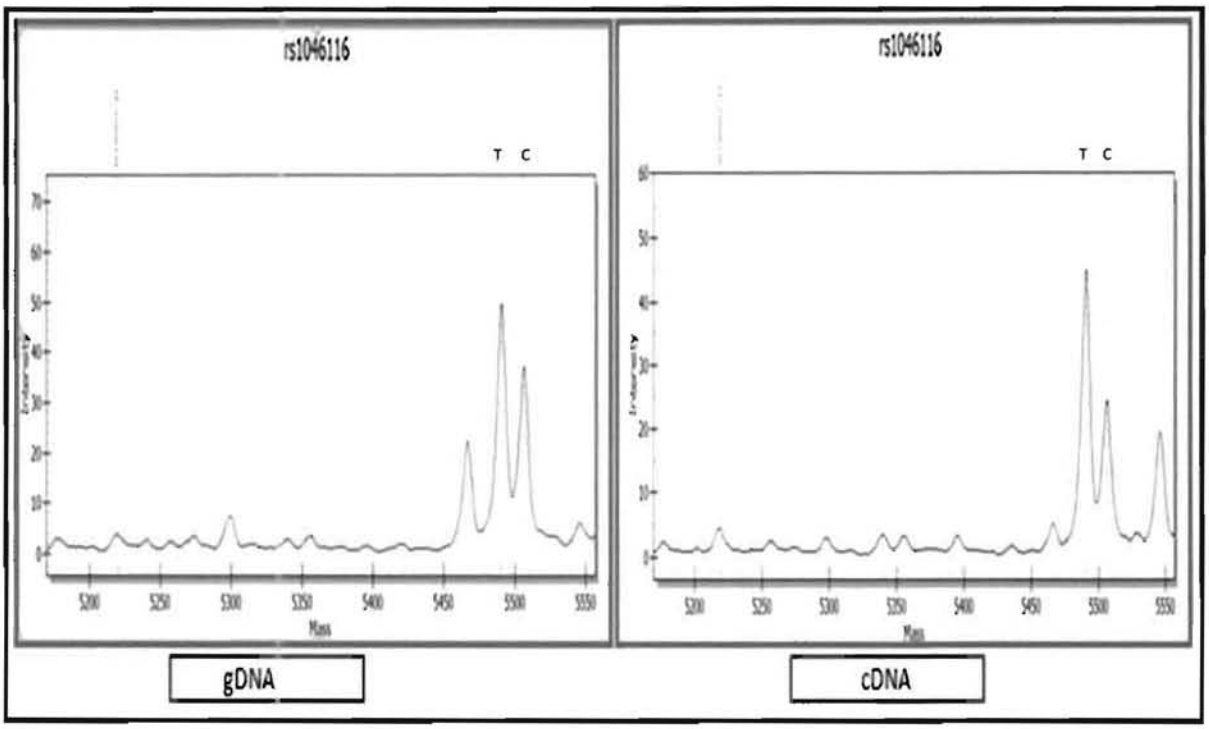


Figure 5.5: Spectra comparison of gDNA and cDNA quantification. The two products are detected and quantified by MALDI-TOF mass spectrometry. The relative mounts of each allele are used to quantify the amount of allelic expression imbalance using the allelic expression ratio (AER). The unlabeled peaks are the results of the background and spectra peaks multiplexed with our SNP of interest.

5.3.4.3 Linearity and normality of allelic expression ratios

Genomic DNA (gDNA) from two individuals homozygous for the minor and major alleles of rs1046116 in *PKP2* were mixed in varying ratios (8:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:8). The gDNA of varying ratios were used as the template for the allelic expression assays to investigate the

appropriateness of genomic normalization ratios and the linearity of the allelic expression imbalance (AER) response. This experiment confirmed that allelic expression response is linear and that allelic expression imbalance analysis can be performed. Figure 5.6 shows a relationship between log AER and log mixture ratio.

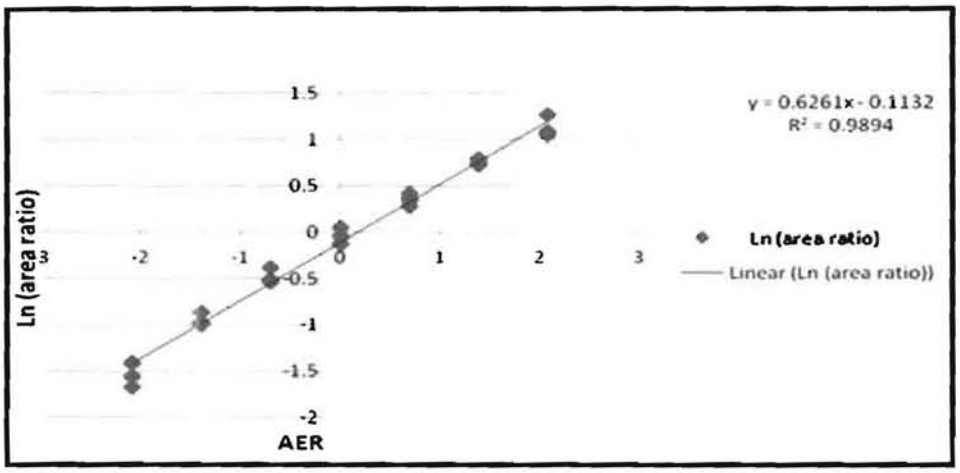


Figure 5.6: Linear relationship between measured and expected allelic expression ratios for alleles mixed in known ratios

Plots of the normalised allelic ratios relative to homozygous gDNA ratio mix are shown in Figure 5.7. The plot indicates inter-individual variations and the figure also shows accuracy of the experiment at gDNA allelic ratio of 1:1. The graph revealed most individual allelic expression to be at zero (as expected), although there were few variations in each individual sample.

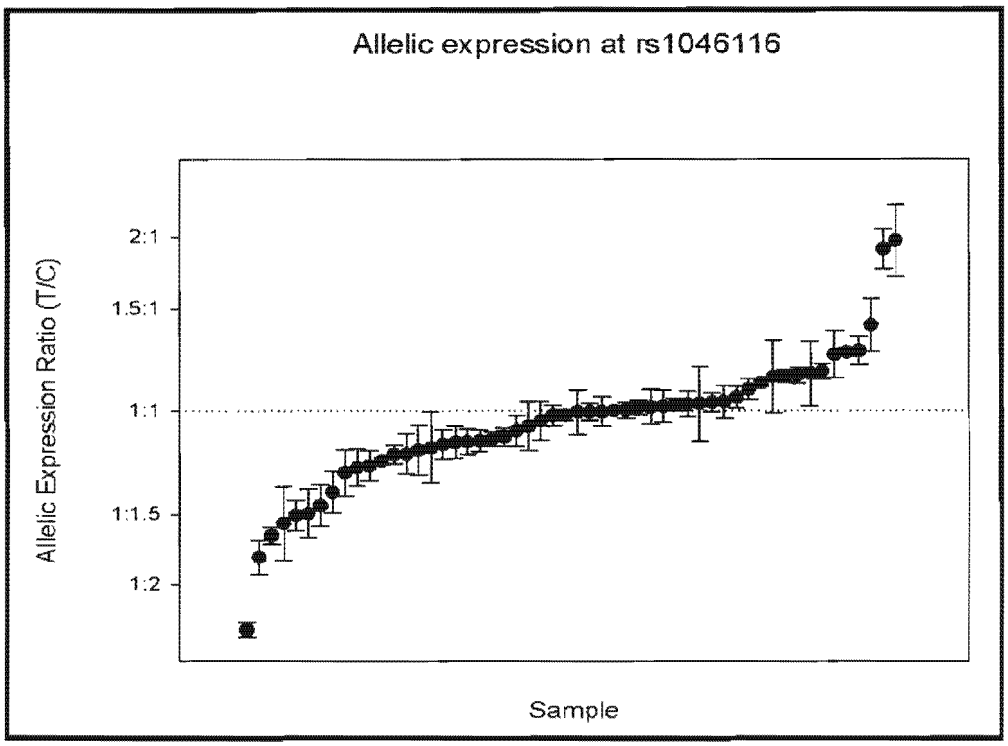


Figure 5.7: Normalised Allelic expression values in the SA cohort. The graph show normalised allelic expression ratio (x-axis) relative to reference homozygous gDNA ratio mix (y-axis). Each black circle represents an individual, vertical lines represent standard error.

5.3.4.4 Influence of cis-acting alleles on allelic expression

There were 54 heterozygous spectra peaks left after outlier exclusions, with a mean of -0.03907. The 95% confidence interval of the mean intersects zero, suggesting that the rs1046116 SNP does not exert a cis-effect on allelic expression.

5.4 DISCUSSION

This study demonstrated that a SNP in the *PKP2* gene is expressed in the peripheral blood, however there does not appear to be an association between the SNP genotype and peripheral blood gene expression, nor does there appear to be a cis-acting effect of the SNP on gene expression. AEI techniques are regarded to be a powerful tool to investigate gene expression patterns (Cunnington, Santibanez Koref *et al.* 2010), and to date, this is the first investigation ARVC/DCM SNPs using samples that combine South African and British population individuals.

In total gene expression, our data demonstrated that there is no association between genotypes and gene expression in NE (UK) Caucasian cohort. Total gene expression showed a great variability within individuals, and this may be due to an influence from both trans and cis factors. Trans factors include environmental factors such as age, sex and ethnicity that may influence the way a gene can be expressed (Cunnington, Kay *et al.* 2009). However, in this case, ethnicity did appear to have an influence, since the members of this cohort were all of Caucasian ethnicitywhile in other genes,e.g. *DSC2*, there were no genotyped found. Cis factors can be due to some nearby cis-acting SNP that may have an influence in gene expression; e.g., SNPs in promoter regions and coding regions may be close enough to have an influence on the SNP under investigation (Pham, Bonello *et al.* 2012). Total gene expression can also be influenced by both cis-trans effects and the natural errors in inter-individual comparison.

Allelic expression investigation confirmed the total gene expression results that rs1046116 does not have any cis-effect on gene expression. In allelic expression, inter-individual variability was not investigated, since the objective of this study was to investigate if the rs1046116 SNP demonstrates a cis-effect on gene expression – not to investigate effects on expression of the SNP. Because *PKP2* demonstrated that expression of both alleles can be

measured in peripheral blood but not *DSG2*, as a results the *DSG2* SNP of interest had to be excluded from the gene expression study. To study gene expression in *DSG2*, cDNA would need to be obtained from tissues where *DSG2* gene is highly expressed (in this case, the heart). On the other hand, *DSC2* was expressed in the peripheral blood, but we lacked enough working material to conduct an AEI experiment. The failure of the Applied Biosytem custom probe assays to genotype cDNA extracted from black South Africans also contributed to the exclusion of the gene from AEI experiments.

5.4.1 Limitations and future work

To investigate both allelic and total gene expression in both *PKP2* and *DSG2* genes, cDNA from a gene-specific tissue such as heart tissue and skin tissue will be required in order to conduct an informative study. *DSC2* appeared to be a good candidate gene for study in peripheral blood, but enough material will be required.

5.4.2 Conclusions

The *PKP2* rs1046116 SNP does not appear to influence the expression of the *PKP2* gene; total gene expression revealed that there might be both trans-acting regulator acting in both pair of chromosomes and cis-acting regulator acting on a single chromosome harbouring the SNP of interest (Grundberg, Small et al. 2012). Weak expression and no expression at all of *PKP2* and *DSG2* genes, respectively, suggest that they do not have a vital function in circulating blood cells. These results also indicate that blood is a suboptimal source of DNA to study *PKP2* and *DSG2* gene expression. Thus the *PKP2* rs1046116 SNP does not have any cis-effect in blood, and blood cDNA may not be the optimal substrate to investigate influence of *PKP2* and *DSG2* gene variations on ARVC disease phenotypes.

Chapter 6

A search for known or new genes that cause ARVC by exome sequencing of the most promising families

6.1 INTRODUCTION

In the ARVC Registry of South Africa there are 12 out of 16 (75.0%) affected families who do not have a disease-causing mutation in one of the five desmosomal genes (i.e. *PKP2*, *DSG2*, *DSP*, *DSC2* or *JUP*). In this chapter, I set out to identify known or novel genetic causes of ARVC by exome sequencing of the families without evidence of a causal genetic mutation on routine genetic screening. Family ACM 8 has an autosomal recessive pattern of inheritance and no known causal mutation after screening five desmosomal protein genes (Figure 6.1). This family was prioritised for mutation screening by exome sequencing.

Exome sequencing is a next-generation high-throughput method that uses massive parallel sequencing of coding regions of the whole genome to identify known or novel genetic abnormalities in a single preparation, and is a substantial improvement over Sanger sequencing which screens one candidate gene at a time (Feng, Tavtigian *et al.* 2011). Exome sequencing is especially useful in recessive inheritance patterns, in which parents are expected to be carriers of the variant while affected siblings are expected to be homozygous for the variant (Ng, Buckingham *et al.* 2010).

Hypothesis

I hypothesis that ACM family 8, which is consistent with an autosomal recessive inheritance pattern (Figure 6.1), may have a gene mutation or mutations in which both parents are carriers and the affected offspring are homozygous for the mutation(s) in question.

6.1.1 The aim of the study

The aim of the study was to conduct exome sequencing in the most promising family without evidence of a causal genetic mutation on routine genetic screening of five desmosomal protein genes, in order to determine whether other known or novel genes could be responsible for ARVC in affected family members.

6.2 METHODS

6.2.1 Study participants

A South African ARVC family of white ethnicity with no genetic mutation found on routine screening was selected for investigation of causative gene(s). The family underwent mutation screening in five desmosomal genes (i.e. *PKP2, DSG2, DSP, DSC2* or *JUP*), but no mutation was identified (see chapter 2 (section 2.4) for details of the screening methods). The study was performed in a family of five members (ACM8.1, ACM8.2, ACM8.3, ACM8.4 and ACM8.9). Both parents (ACM8.1 and 8.2) were clinically examined and did not meet the criteria for the diagnosis of ARVC. Two affected family members were diagnosed with ARVC at age 10 (ACM8.3) and 12 (ACM8.4) years. Two other sibs were clinically examined and were determined to be unaffected by ARVC (ACM 8.9 and the first daughter, who was not available for blood sampling).

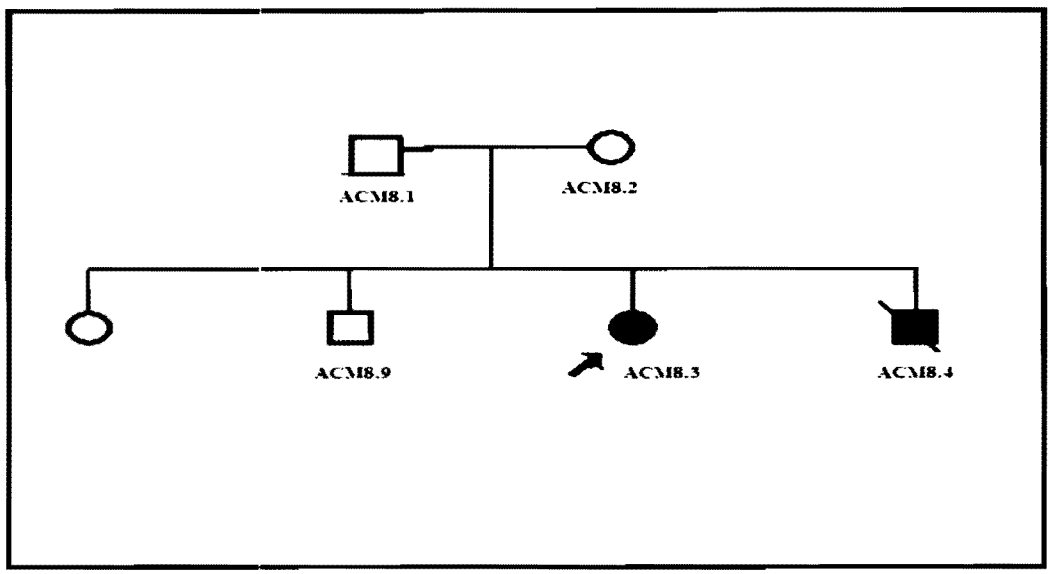


Figure 6.1: ACM 8 family pedigree. ACM is the disease code for ARVC in our laboratory; circles represent females; squares represent males; clear circles and squares represent unaffected family members and black circles and squares represent affected family members; crossed symbols represent deceased members of the family.

6.2.2 Exome sequencing

Genomic DNA from ACM 8.3 and 8.4 was subjected to exome sequencing; detailed methods are provided in chapter 2 (Section 2.7). The sequence data was further processed using bioinformatic tools to identify all variants shared by the two affected individuals.

6.2.3 Sanger sequencing of candidate variants

Sanger sequencing was used to confirm the presence of variants that segregated with the disease. The validation of the variants was carried out on the ABI 3100 Genetic Analyzer (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and analyzed using BioEdit software (Ibis Biosciences, Carls-bad, CA, USA).

6.2.4 Databases used to determine the allele frequency of variants

There were four online databases used to determine the allele frequency of variants as follows: 1000 Genomes (www.1000genomes.org), Exome Variant Server (EVS) (evs.gs.washington.edu), ARVC database (www.arvcdatabase.info) and dbSNP database (www.ncbi.nlm.nih.gov/SNP/). These databases were used to identify whether the variant had been previously identified and to obtain information on allele frequency.

6.2.5 Prediction software used to evaluate pathogenic significance of variants

The pathogenicity of all non-synonymous rare novel variants was assessed using the Polyphen-2 (http://genetics.bwh.harvard.edu; 25 January 2014), SIFT (http://sift.bii.a-star.edu.sg/; 25 January 2014), and MutationTaster (http://www.mutationtaster.org/; 25 January 2014) prediction programmes. Only those variants predicted to be disease causing by at least two of the prediction tools were considered for further analysis. The pathogenicity of rare synonymous variants was assessed using the RNAfolder webserver (http://rna.tbi.univie.ac.at/; 14 February 2014) and MutationTaster tool (http://www.mutationtaster.org/; 13 February 2014).

6.2.6 Amplification of cDNA in heart tissue

Heart tissue was used to verify the expression of candidate genes that were discovered in this study. A measurement of 20 ng of heart tissue cDNA was amplified using primers designed to be specific for only cDNA. The amplification methods are described in chapter 2.

6.3 RESULTS

6.3.1 Exome variant calling quality and coverage

A total of 5798 variants were identified before filtering, and there were 13 variants predicted to be disease-causing and shared by the two individuals (ACM 8.3 and ACM 8.4). The thirteen variants were made of eight heterozygous variants (summarised in Table 6.1) and five homozygous variants (summarized in Table 6.2)..

Table 6.1: Heterozygous variants detected in the two affected family members

Gene	Gene function	Amino acid change
Parvin, Alpha (*PARVA*)	Plays a role in the organization of the actin cytoskeleton; and cell adhesion	p.D131V
Ankyrin 3, Node Of Ranvier (*ANK3*)	Maintenance of ion channels and cell adhesion	p.R993Q
Anaplastic lymphoma receptor tyrosine kinase (*ALK*)	Plays an important role in the genesis and differentiation of the nervous system	p.1346Q
Deleted in lung and esophageal cancer 1 (*DLEC1*)	Act as a tumor suppressor by inhibiting cell proliferation	p.R81H
WW domain containing Adapter with coiled-coil (*WAC*)	Regulates the cell-cycle checkpoint Activation in response to DNA damage	p.K401E
HMG box domain containing 3 (*HMGXB3*)	Plays a role in DNA replication and repair	p.E203K
Neuralized homolog (Drosophila) (*NEURL*)	May function as an E3 ubiquitin-protein ligase to activate monoubiquitination of JAG1	p.M188T
Trichorhinophalangeal syndrome I *(TRPS 1)*	Transcriptional repressor	c.-9_-8insT

Table 6.2: Homozygous variants detected in two affected family members

Gene	Gene function	Amino acid change
Ubiquitin protein ligase E3 component N-recognin 4 *(UBR4)*	It regulates integrin-mediated signalling.	c.15009_229-15009_228inG
SMAD family 5 *(SMAD5)*	It functions as a transctional modulator activated by bone morphogenetic protein.	g.331694-13551308insC
Glutamate receptor ionotropic AMP3*(GRIA3)*	It functions as a glutamate receptor.	g.6586096-656097insG g.122336601-122336602insG
Zinc finger protein 141 (ZNF141)	It is involved in transcriptional regulation as a repressor	g.331694-331695insC

6.3.2 Sanger sequencing validation of variants in family members

6.3.2.1 Heterozygous variants

Sanger sequencing was performed in all family members to validate the variants and to confirm segregation with disease (Figure 6.2).

6.3.2.1.1 Novel variants

There were three novel variants identified in ACM 8.1 (father) that were transmitted to the three offspring: p.R81H in *DLEC1*, p.E203K in *HMGXB3* and p.M188T in Neuralized homolog (*Drosophila* (*NEURL*)). Only p.E203K (*HMGXB3*) segregated with the disease in the offspring. These variants were also not present in the 1000 Genomes and dbSNP databases. They were also predicted to be possibly damaging by the Polyphen-2 prediction tool because they change amino acid groups. They were all detected in an unaffected individual ACM 8.9 (Table 6.3).

6.3.2.1.2 Reported variants

There were five reported variants that were transmitted from ACM 8.2 (mother) to the three offspring: p.D131V in *PARVA*, p.K401E in *WAC*, p.R993Q in *ANK3* and p.R1346Q in *ALK*.

The heterozygous variant *TRPS1*; c.-9_-8insT was not processed further, since this variant was a known polymorphism found in the dbSNP database with a minor allelic frequency of 0.356 in 775 population individuals.

ANK3 p.R993Q (rs141939315) was previously reported as a SNP, with a minor allelic frequency (MAF) of 0.004. This variant changes arginine (polar basic) to glutamine (polar uncharged). The variant was also detected in an unaffected family member.

HMGXB3 p.E203K (rs142114383) in exon 3 was also previously reported as a SNP with a MAF of 0.0038. This variant was transmitted from the father to the two affected children. The variant changes a glutamic acid (polar basic) amino acid to lysine (polar acidic), and it was predicted to be disease causing by MutationTaster and damaging by the Polyphen-2 and SIFT prediction tools.

NEURL p.M188T was transmitted from the father to all offspring (including the unaffected individulals). The variant changes methionine, a hydrophobic amino acid, to threonine which is a neutral amino acid. It was predicted by MutationTaster and Polyphen-2 to be possibly damaging.

PARVA p.D131V was transmitted from the mother to the affected offspring only. The variant is reported in the exome variant server as a variant of unknown significance, and it was found in only one out of 12945 population controls. It changes aspartate (polar acidic) to valine (non-polar).

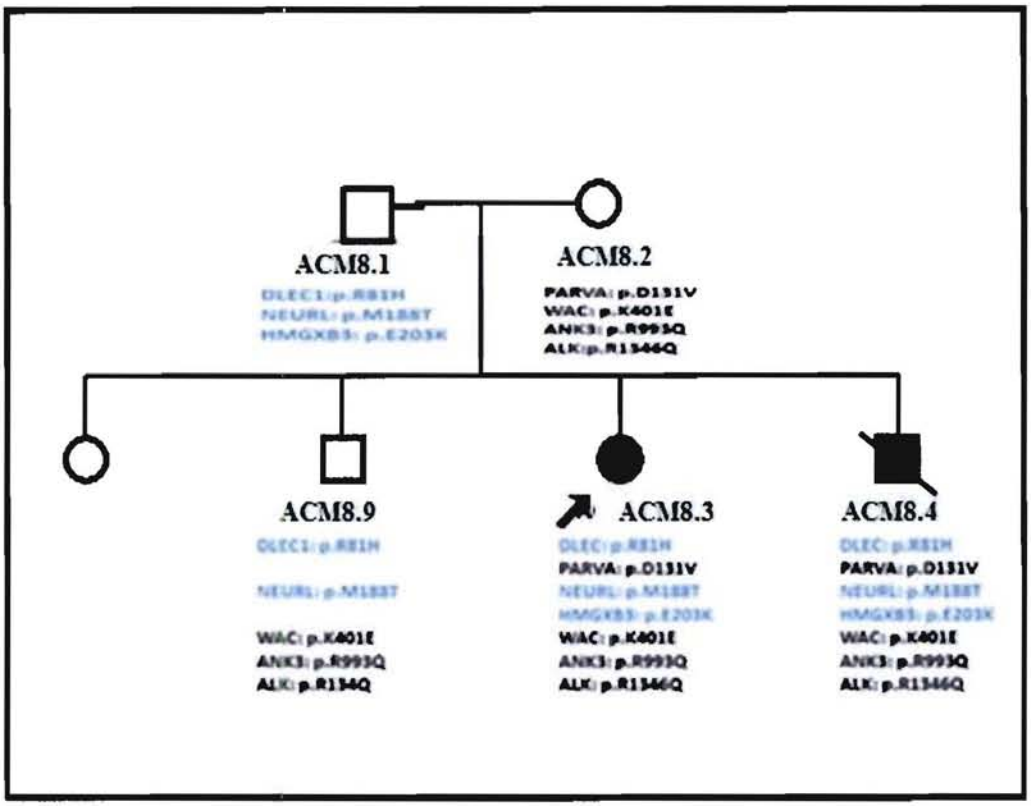


Figure 6.2: Validation of heterozygous variants in ACM 8 family members. Gene abbreviations are listed in capital letters. D, aspartic amino acid; V, valine; R, arginine; H, histidine; G, glycine; K, lysine; M, methionine; T, threonine; E, glutamine acid; Q, glutamine.

6.3.2.2 Homozygous variants

The 5 homozygous variants shared by two affected family members were sequenced. Four of the variants– *ZNF141* (g.331694-331695insC), *SMAD5* (g.135513085-135513086insC) and both variants from *GRIA3*, g.6586096-65866097insG and 122336601-122336602insG – were present in the unaffected sibling and therefore did not segregate with disease. *UBR4* (c.15009_229-15009_228 insC) was the only variant that exhibited an autosomal recessive pattern in the two affected family members; however this was detected in 52% of white population controls. All homozygous variants identified were reported as SNPs according to the dbSNP database.

Table 6.3: Reported heterozygous variants

Gene	Variant	Chromosomal position	Amino acid change	Minor allelic frequency	Reported/published
ANK3	p.R993Q, c.2978C>T	chr10:61868783 C>A	Basic- uncharged	0.004	rs141939315
HMGXB3	p.E203K, c.607G>A	chr5:149389968 G>A	Neutral- basic	0.0038	rs142114383
NEURL	p.M188T, c.563T>C	chr10:105331493 T>C	Hydrophobic-neutral	-	EVS
PARVA	p.D131V, c.392A>T	chr11:12495505A >T	Acidic-non-polar	-	EVS
TRPS1	g.50356_50357insT c.-9_-8ins.T	Ch8:116635872 insT	UTR	-	rs35329862
ALK	p.R1346Q, c.4947G>A	chr2:29416006C >T	Basic-acidic		Novel
DLEC1	p.R81H, c.262C>G	chr3:38080978C >G	Hydrophobic-uncharged		Novel
WAC	p.K401E, c.1465A>G	chr10:28905145A >G	Basic-acidic		Novel

UTR: untranslated region; EVS: exome variant server

6.3.2.3 Variants that are potentially disease-causing

6.3.2.3.1 PARVA: c.392A>T, p.D131V

The *PARVA* gene encodes a focal adhesion protein known as actopaxin. The variant identified, c.392A>T (p.D131V), was absent in the 1000 Genomes database but was identified in the exome variant server in one out of 12227 population controls. The variant was not observed in 232 population controls from our laboratory. This variant changes the polarity of amino acids from aspartic acid (D) (negative polar) to valine (V) (non-polar). All three bioinformatic prediction tools predicted this variant to be probably damaging. Figure 6.3 illustrates the restriction enzyme digest (A), electropherogram (B), and conservation among modern species (C).

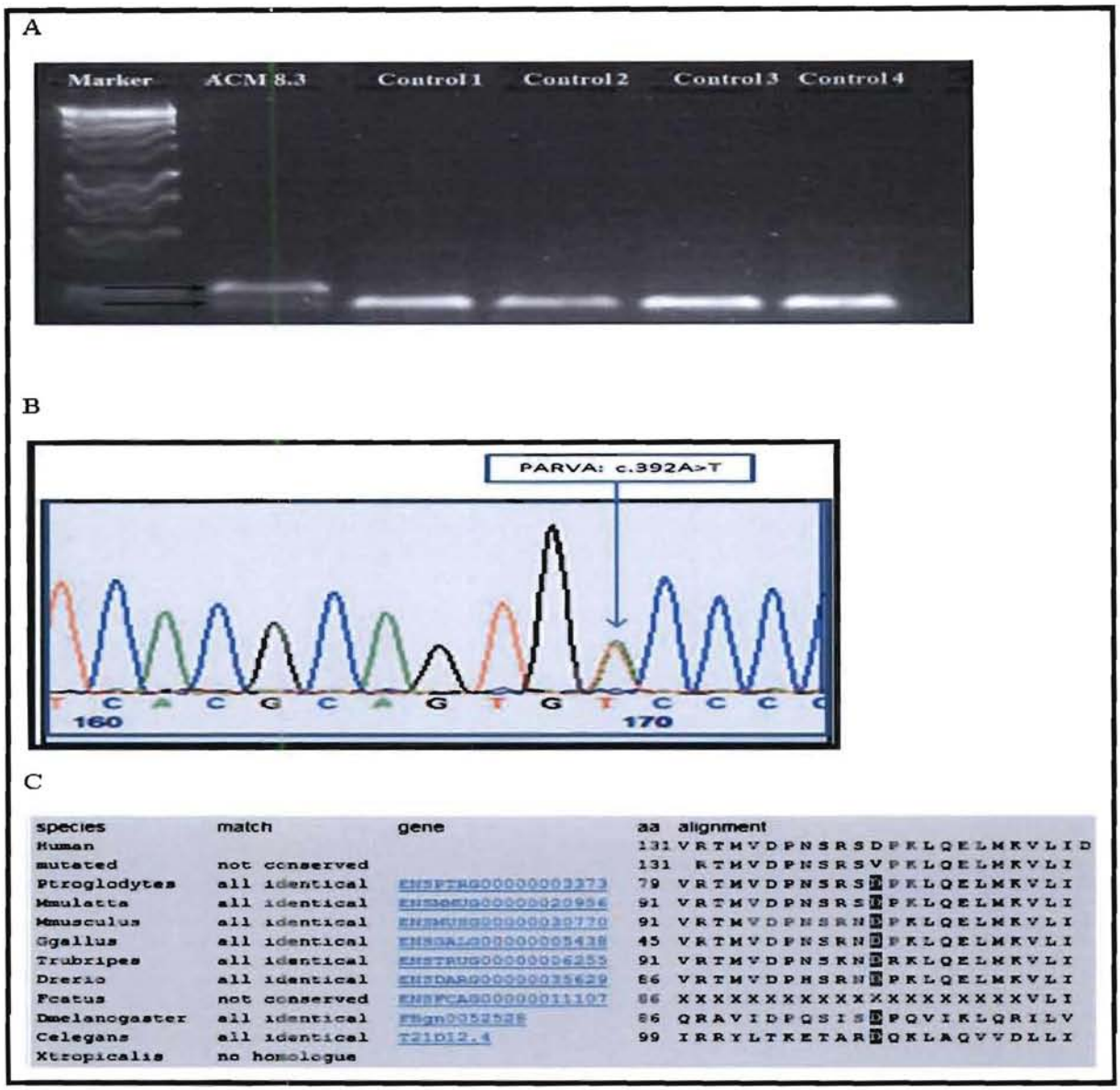


Figure 6.3: c.392A>T, p.D131V variant; A, illustrating Tsp45I restriction enzyme digest fragments B, an electropherogram showing an A>T nucleotide substitution; C, amino acid alignment showing MutationTaster conservation information of the amino acid affected. Ptroglodytes (chimpanzee); Mmulatta (rhesus macaques); Mmusculus (House mouse); Ggallus (chicken); Trubripes (puffer fish); D rerio (Zebra fish); D melanogaster (Drosophila); C elegans (Nematode); X tropicalis (frogs).

6.3.2.3.2 HMGXB3: c.607 G>A, p.E203K

The *HMGXB3* gene encodes for HMG box domain containing 3 and this protein plays a role in DNA replication and DNA repair. The variant identified, c.607G>A (p.E203K), was

reported in the Exome Variant Server database to be present in 36 out of 4530 individuals, thus qualifying as a rare polymorphism. This variant was not observed in 232 population controls from our laboratory. The variant changes glutamic acid (E) (negative polar) to lysine (K) (positive polar). All 3 bioinformatic prediction tools predicted this variant to be probably damaging. The location where the variant is situated harbours a conserved amino acid among animal species; Figure 6.4 illustrates the restriction enzyme digest (A), electropherogram (B), and conservation among modern species (C).

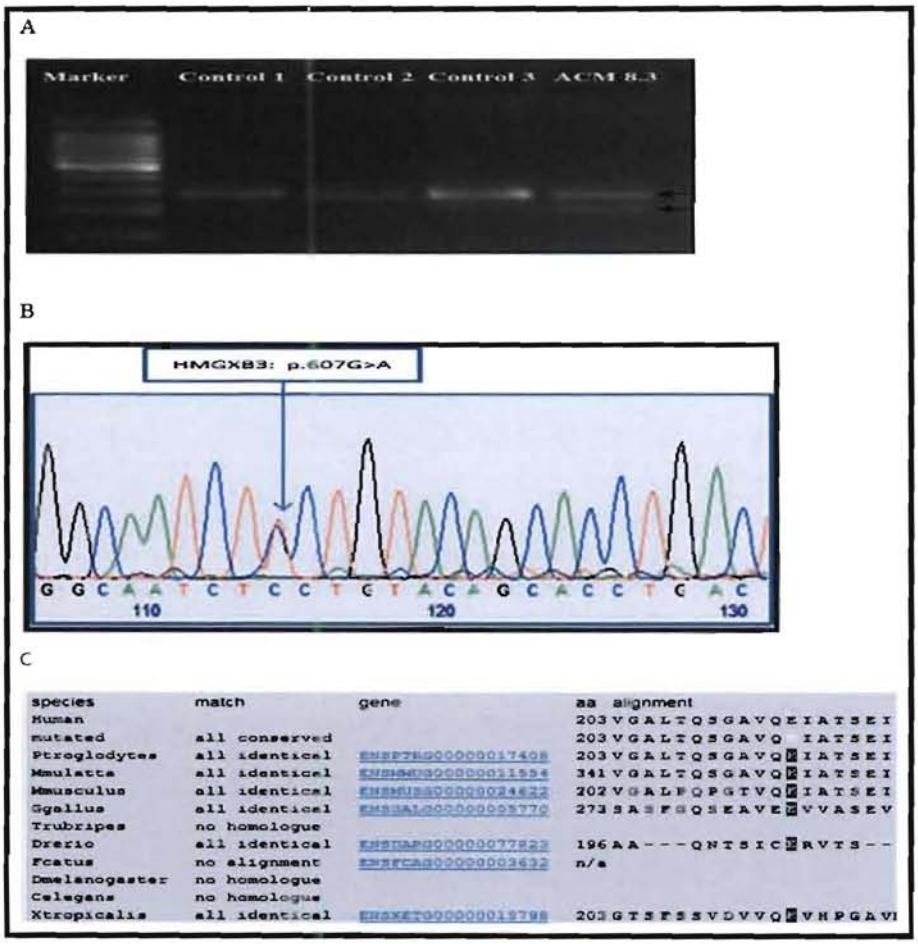


Figure 6.4: c.607G>A (p.E203K) variant; A, illustrating MboII restriction enzyme digest fragments; B, an electropherogram showing an T>C, which is in the reverse form ,nucleotide substitution; and C, amino acid alignment showing conservation information of the amino acid affected. Ptroglodytes (chimpanzee); Mmulatta (rhesus macaques); Mmusculus (House mouse); Ggallus (chicken); Trubripes (puffer fish); D rerio (Zebra fish); D melanogaster (Drosophila); C elegans (Nematode); X tropicalis (frogs).

Thus, both the *PARVA* and *HMGXB3* variants are possibly damaging and segregate with ARVC in the ACM 8 family (Figure 6.5).

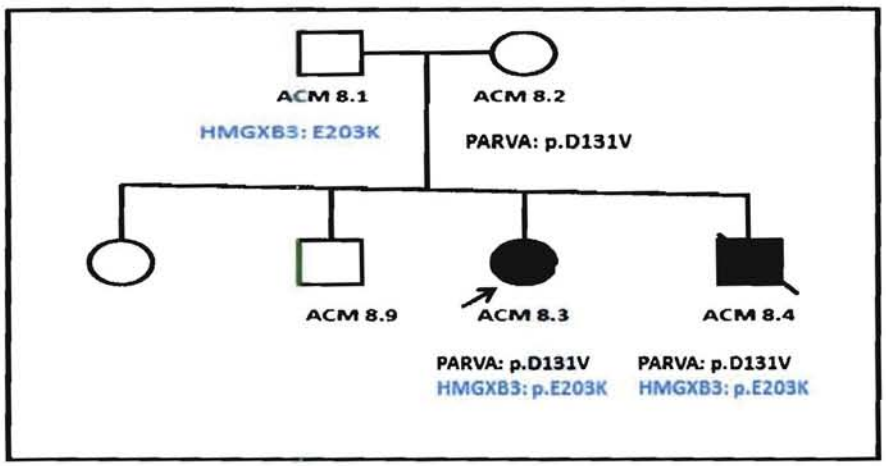


Figure 6.5: Variants segregating in ARVC family members. ACM is an ARVC identification code; HMGXB3 and PARVA are the relevant variants; D-aspartic amino acid; V-valine; R-Arginine; H- Histidine; G- Glycine; K-Lysine; M- Methionine; T- Threonine; E-Glutamine acid; Q- Glutamine; the numbers between the abbreviations refer to the amino acid position in the protein.

6.3.3 Known cardiomyopathy gene variants

All known cardiomyopathy genes that cause ARVC, DCM and HCM were inspected for mutations in the exome data set. This was done by knocking out all the filters and searching for variants in a list of cardiomyopathy known genes. One SNP was identified in *CALR3* which is associated with familial hypertrophic cardiomyopathy (Chiu, Tebo *et al.* 2007). The SNP was not conserved, not damaging and did not affect amino acids groups. The SNP had a MAF of 0.008.

6.3.4 Expression of *PARVA* and *HMGXB3* in heart tissue

Amplification of cDNA from left and right atrial appendages revealed that both *PARVA* and *HMGXB3* genes are highly expressed in the heart, with *DSC-2* used as a positive control. Figure 6.6 shows amplification products visualisation through ultraviolet agarose gel.

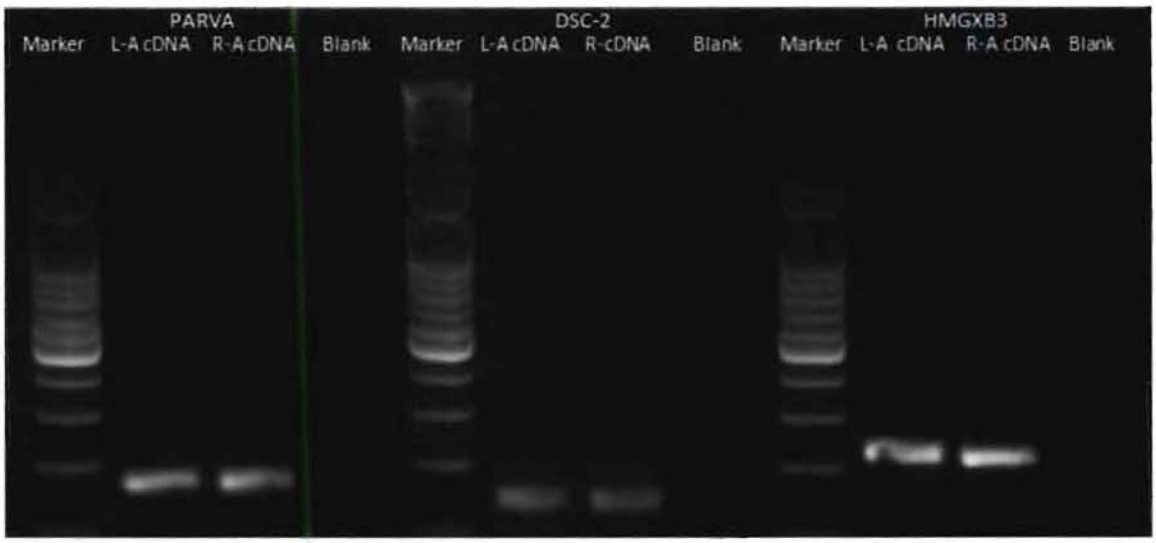


Figure 6.6: Investigation of *HMGXB3* and *PARVA* expression in the heart tissue. PARVA- Parvin alpha; DSC-2- Desmocolin-2; HMGXB3- HMG box domain containing 3; L-A- Left atrium; R-A- Right atrium.

6.4 Screening of *PARVA* gene mutations patients with ARVC

The mutation in *PARVA* was considered to be a more plausible candidate gene for ARVC, with the variant in *HMGXB3* possibly playing a modifying role. This hypothesis is supported first by the biological plausibility of the role of PARVA in cardiomyopathy as a cell adhesion protein, the rarity of PARVA mutations in the general population, evolutionary conservation of the amino acid change and predicted deleterious effect on protein function. Whilst the *HMGXB3* variant is likely to be a rare polymorphism, its role as a modifier gene cannot be ruled out given a potential digenic effect based on the inheritance pattern in this family. Thus *HMGXB3* harbours a GVUS that was not pursued further in this study.

Further evidence for the causal role of *PARVA* mutations in ARVC was pursued by screening 65 patients with ARVC for the presence of mutations in this gene. Two unrelated probands with ARVC were found to have rare mutations in the *PARVA* gene, thus providing additional support for *PARVA* as a new candidate gene for ARVC (Table 6.4).

Table 6.4: Additional *PARVA* mutations based on prediction tools and population controls screening in sixty five ARVC probands

Index case ID	Nucleotide change	Amino acid	Exon /intron	Type	MAF	Classification	Reported
30.1	c.523A>G	p.K156E	4	Missense	0.0	Pathogenic	Novel
32.1	c.597T>G	p.T199	5	Synonymous	0.0	Pathogenic	rs111282823

6.4.1 *PARVA*: c.523A>G, p.K156E

A c.523A>G (p.K156E) mutation, located in exon 4 of *PARVA*, was identified in ACM 30.1. This mutation was absent in all public databases and in our local population controls. The amino acid substitution causes a change in polarity (lysine = positive polar to glutamic acid = negative polar). It was predicted by the 3 prediction tools to be damaging and potentially disease-causing. Figure 6.7 illustrates the restriction enzyme digest (A), electropherogram (B) and conservation among modern species (C).

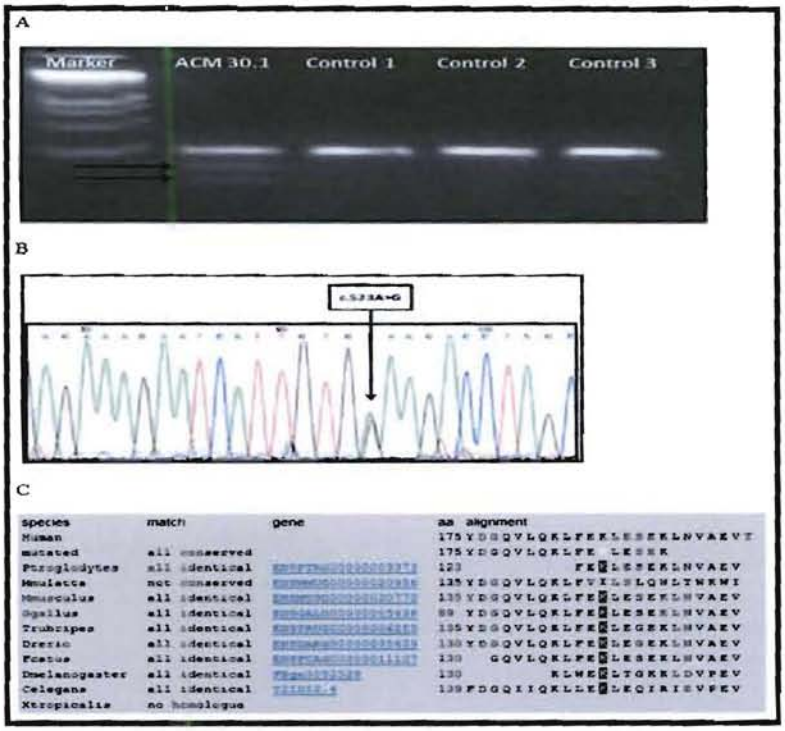


Figure 6.7: c.523A>G, p.K156E variant; A, illustrating *BpiI* restriction enzyme digest fragments; B, an electropherogram showing an A>G nucleotide substitution; and C, amino acid alignment showing conservation information of the amino acid affected. Ptroglodytes (chimpanzee); Mmulatta (rhesus macaques); Mmusculus (House mouse); Ggallus (chicken); Trubripes (puffer fish); D rerio (Zebra fish); D melanogaster (Drosophila); C elegans (Nematode); X tropicalis (frogs).

The mutation was investigated in 2 other family members who had their DNA available for the study. ACM 30.2, a daughter to ACM 30.1, is a carrier of the p.K156E mutation (Figure 6.8).

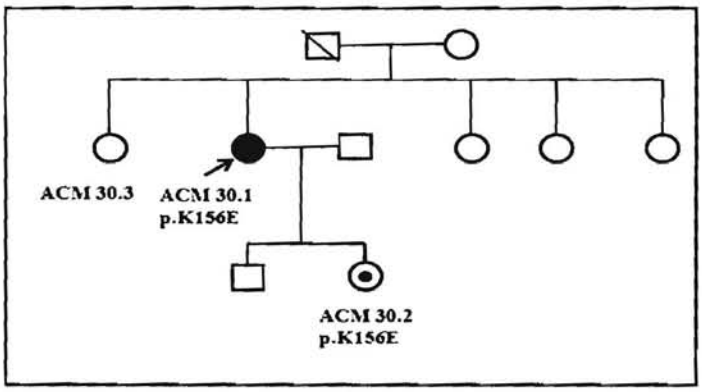


Figure 6.8: ACM 30 family members with *PARVA* c.523A>G, p.K156E mutations. Circle: female, square: male, black circle/square: affected and dot in a circle/square: carrier of the mutation

6.4.2 *PARVA*: c.597T>G, p.T199

A c.597T>G (p.T199) variant, located in exon 5 of *PARVA*, was identified in proband ACM 32.1. It has been reported in the 1000 Genomes database as rs111282823 with no MAF provided; it was not found in our laboratory population controls (of note, these controls were screened by Sanger sequencing, since there was no restriction enzyme available for the variation substitution site). Although it was a silent mutation, it was predicted by MutationTaster to be disease-causing and by the RNAfold webserver to change the mRNA structure. Polyphen-2 and SIFT were not used because they are not suitable for synonymous variants as it does not affect the amino acid. There were no family members available for screening of this variant. Figure 6.9 illustrates the restriction enzyme digest (A), electropherogram (B), and (lack of) conservation among modern species (C). We concluded that this is a variant of unknown significance (GVUS), since it is not highly conserved, rare in 464 chromosomes, and predicted to be disease-causing.

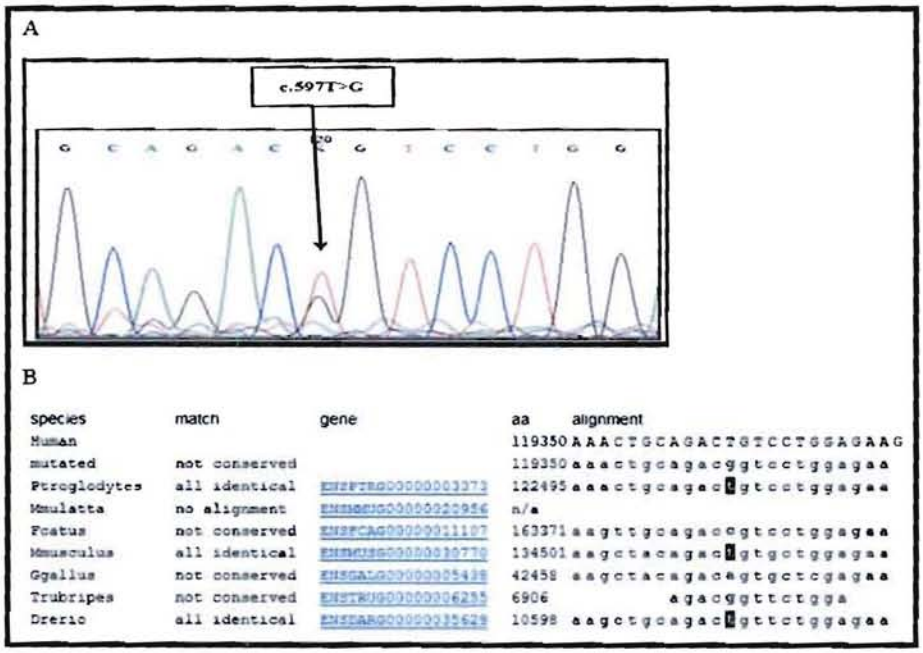


Figure 6.9: c.597T>G, p.T199 variant; A, an electropherogram showing a T>G nucleotide substitution; B, nucleotide alignment showing mutation taster conservation

information of the amino acid affected. Ptroglodytes (chimpanzee); Mmulatta (rhesus macaques); Mmusculus (House mouse); Ggallus (chicken); Trubripes (puffer fish); D rerio (Zebra fish); D melanogaster (Drosophila); C elegans (Nematode); X tropicalis (frogs).

6.5 DISCUSSION

This exome sequencing study has identified *PARVA* as a new candidate gene for ARVC and cardiomyopathy in general. It is also possible that a rare polymorphism in *HMGXB3* may be a modifier of the cardiomyopathy phenotype.

6.5.1 *PARVA*

PARVA encodes parvin alpha, also known as actopaxin, a protein that belongs to a series of focal adhesion proteins that form part of the extracellular matrix and are referred to as focal complexes (Clarke, Brown *et al.* 2004, Stiegler, Draheim *et al.* 2012, Wehrle-Haller 2012). Figure 6.10 summarizes other proteins that bind to parvin alpha indicating their activation and inhibition proteins. These proteins link the extracellular matrix to the actin cytoskeleton (Kanchanawong, Shtengel *et al.* 2010) in the area composite. Focal adhesion proteins bind F-actin through adherens junction proteins, where they provide cell-cell adhesion by linking E-cadherin to the actin cytoskeleton (van Hengel, Calore *et al.* 2013). Northern blot analyses have demonstrated that the highest expression of *PARVA* is in the kidney, heart and brain (http://omim.org/entry/608120).

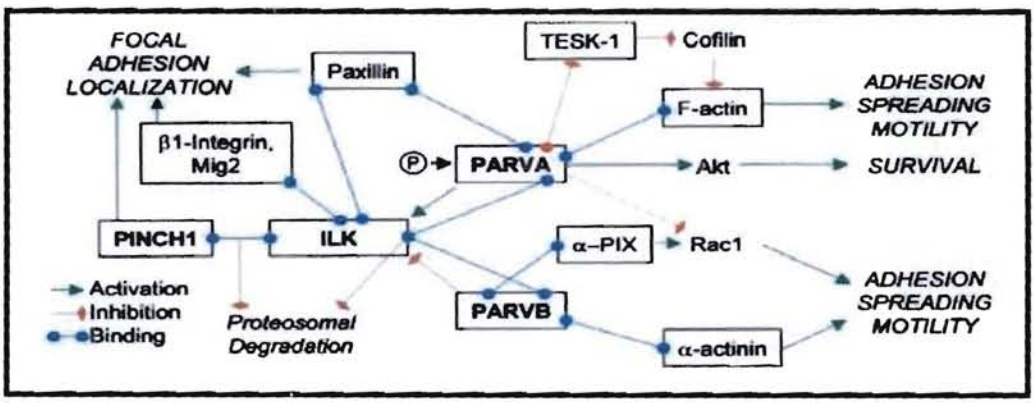


Figure 6.10: Role of *PARVA* among focal adhesion genes. (Image taken from http://www.springerimages.com/Images/).

Parvin alpha is composed of an N-domain, two CH (Calponin homology) domains and a C-domain. The N-domain has 95 amino acids, and while it is not involved in any interactions, it is a major site of phosphorylation and thus conformational change of the whole protein. The N-terminus contains two nuclear localization signals and three SH-binding sites. The two CH domains are the actin-binding sites and are separated by a linker region (http://omim.org/entry/608120). The CH2 domain contains a paxillin-binding subdomain (PBS) that interacts with an integrin-linked kinase.

The mother of the original family under investigation (ACM 8.2) appears to be a carrier for the rare variant p.D131V. This variant (as well as the variants identified in ACM families 30 and 32) is located in the CH1 domain (exons 3-5) and near sites confirmed to interact with F-actin (Nikolopoulos and Turner 2000, Clarke, Brown *et al.* 2004, Lorenz, Vakonakis *et al.* 2008). This region is highly conserved among modern animals. The structure of *PARVA* and parvin alpha and the relevant amino acid substitutions are summarized in Figure 6.11.

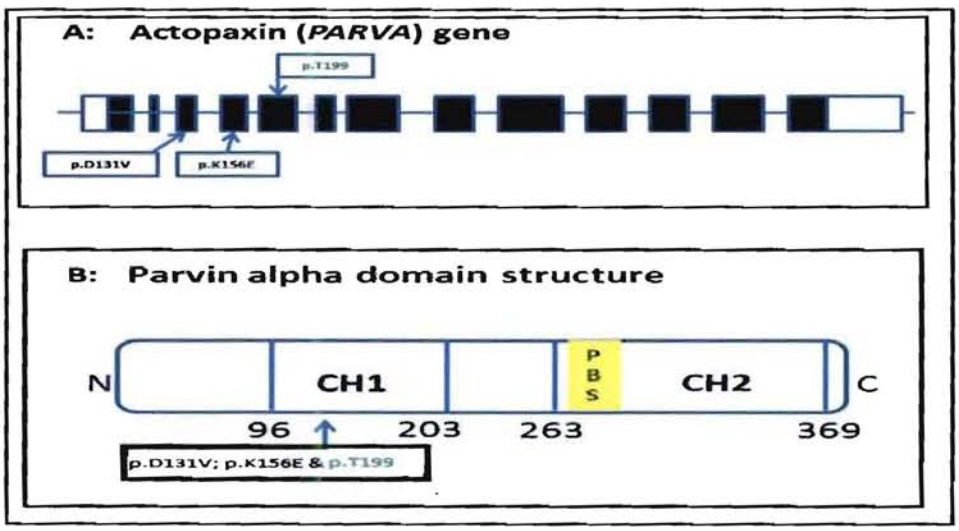


Figure 6.11: Structure of *PARVA* (actopaxin) gene and protein, and amino acid substitutions identified in ARVC patients in this study. (A) exonic structure and mutations; (B) structural domains. *Black boxes represent exons; white boxes represent untranslated regions; N: N-*

terminal domain, CH: calponin homology, PBS: paxillin binding subdomain, C: C-terminal domain, D: aspartic acid, V: valine, K: Lysine, E: Glutamic acid.

6.5.2. *HMGXB3*

HMGXB3 gene is a member of the high mobility group box family that has a DNA-binding high mobility group (HMG). It is thought to have a vital function in DNA replication and repair (Bustin 1999, Reeves and Beckerbauer 2001). The function of this gene has not been previously characterised, and no studies have shown association of this protein with heart disease. Our expression studies have shown that this protein is highly expressed in heart tissue and is thus likely to have an important function in the heart. The variant is conserved and predicted to be damaging and disease-causing by prediction tools, reported to be present at a frequency of less than 1% in the Exome Variant Server cohort and absent in 464 chromosomes of population controls. The father (ACM 8.1) appears to be a carrier for the rare variant (p.E203K) that was transmitted to both affected siblings. Although the variant is reported as a polymorphism in the 1000 Genomes database, it is possible that this SNP could modify the clinical expression of a latent ARVC mutation. In this case, the mechanism may be similar to a previously reported *KCNH2* polymorphism associated with long QT syndrome (Crotti, Lundquist *et al.* 2005). Further studies could test this hypothesis.

6.5.3. Patterns of inheritance

In ACM family 8, neither father nor mother demonstrated the ARVC phenotype, making them carriers for the mutations I have identified and the pattern of inheritance a single dose of each parent for both *PARVA* and *HMGXB3*. Because both affected offspring inherited both gene mutations, it is possible that asingle dose effect from each parent exists (i.e. that one gene mutation is not sufficient to cause ARVC). On the other hand, when *PARVA* was screened in ARVC probands, two variants identified in unrelated probands were found,

suggesting that in these individuals one mutation may be sufficient to cause disease. There are several lines of evidence that make *PARVA* a plausible candidate gene for ARVC. *PARVA* is a focal adhesion gene and there are also focal adhesion gene mutations reported to cause hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM) such as vinculin and metavincullin gene mutations (Vasile, Ommen *et al.* 2006, Vasile, Will *et al.* 2006) . Focal adhesion proteins play a role in cell-cell adhesion, a function that is akin to the desmosomal proteins that are known to play a causal role in ARVC. Furthermore, the variant that is associated with ARVC in this study is rare, having been observed only once previously in over 12,000 individuals. In addition, the mutation, which codes for a conserved amino acid, is predicted to have a deleterious effect on protein function.

The clinical genetics of ARVC in family ACM 8 may be compatible with 3 possible inheritance patterns. The first is that of inherited a single dose of each parent condition. The results of exome sequencing have not identified a causal recessive mutation in this family. The second possibility is that the severe phenotype in the offspring with no discernable disease in the parents may be due to compound heterozygosity on the basis of inheriting mutations of mild effect from each parent. This possibility is supported by the transmission of the *PARVA* mutation from the mother and the *HMGXB3* from the father. However, the *HMGXB3* mutation appears to be a rare polymorphism, which may either have a modifying effect or play no role in disease causation. Finally, the disease could be due to a deleterious mutation in *PARVA*, which has incomplete penetrance in the mother.

Compound and digenic heterozygosity have been previously noted in ARVC (Xu, Yang *et al.* 2010) and compound heterozygosity was also identified in other ARVC families in our cohort. Two unique *PKP2* mutations were identified in ACM family 19, and these mutations demonstrated an allele dose effect resulting in a severe phenotype in the offspring (Mbele 2008, Watkins, Hendricks *et al.* 2009). Nakajima *et al.* suggested that compound and digenic

heterozygosity are more prevalent in ARVC patients in some ethnicities (e.g. Japan compared with some Western countries) (Nakajima, Kaneko *et al.* 2012). Others have suggested that, although variants may be common in desmosomal genes, they might not be sufficient in themselves to cause ARVC but more commonly require a "second hit" – e.g. from compound and digenic heterozygosity, or perhaps environmental factors (Xu, Yang *et al.* 2010, Nakajima, Kaneko *et al.* 2012).

6.6 Limitations and future work

In the present study we only screened 65 ARVC probands with no family information. It is unknown whether *PARVA* mutations are found in other cardiomyopathies (e.g. DCM, HCM). Future work will involve extending screening to more ARVC probands and family members, as well as other cardiomyopathies. Finally, functional studies will need to be performed for *PARVA* and *HMGXB3* to support their causal or modifier role in cardiomyopathy.

6.7 Conclusions

Mutations in the *PARVA* and *HMGXB3* genes are possible associated with the ARVC phenotype, and in the former case, a biologically plausible mechanism exists to explain the association. Future work will include validating these findings in a larger cohort of cardiomyopathy probands and family members, and performing functional studies.

Chapter 7

Summary of Findings and Future work

7.1 Summary of findings

I have conducted a study of the molecular genetics of arrhythmogenic right ventricular cardiomyopathy (ARVC) and dilated cardiomyopathy (DCM) in South Africans. There were major experiments that were performed to delineate the causal genes and the effects of polymorphisms on gene expression in patients with cardiomyopathy. The first experiment surveyed the prevalence of desmosomal protein gene mutations in 65 patients with ARVC and 112 patients with DCM. The second experiment was concerned with the verification of a recurrent mutation in the plakophilin 2 gene (*PKP2*) as a founder mutation. The third experiment sought to determine whether polymorphisms in desmosomal protein genes influenced gene expression. Finally, given the lack of a known genetic cause in the majority of cases studied, exome sequencing was embarked upon to identify other known and novel genetic causes of ARVC in the South African families.

7.1.1 Prevalence of desmosomal protein gene mutations in ARVC and DCM

Fourteen of the 65 participants with ARVC had disease-causing mutations and none of the 112 DCM patients was found to have a disease-causing mutation in the four desmosomal protein genes that were screened, i.e. plakophilin 2 (*PKP2*), desmocollin 2 (*DSC2*), desmoglein 2 (*DSG2*) and plakoglobin (*JUP*) (chapter three). In a separate study in our laboratory in which 150 probands with DCM were screened for mutations in the desmoplakin (*DSP*) gene, 6 (4%) individuals had a disease-causing mutation (Fish 2010)(. Taken together,

these data suggest that desmosomal gene mutations are an uncommon cause of DCM, and *DSP* is associated with DCM in a significant proportion of Africans.

7.1.2 The discovery of a founder mutation in the *PKP2* gene

There were four unrelated probands (ACM 5.1, ACM 12.1; ACM 19.2 and ACM 38.3) identified with a recurrent mutation, c.1162C>T (p.R388W), in *PKP2* who were investigated further (chapter four). The haplotype study has shown an identical mutation with a common haplotype which was present in 28.6% (4/14) of the ARVC probands with disease-causing mutations in the desmosomal protein genes that were screened in this study. The four individuals are white South Africans of European ancestry who inherited the haplotype bearing the recurrent mutation from a common unidentified ancestor.

New index cases entering the ARVC Registry of South Africa are first screened for this particular founder mutation. The advantage of founder mutations is that they offer the possibility of identifying large numbers of individuals with an identical genetic background for the study of genotype-phenotype correlations.

7.1.3 Influence of polymorphisms on gene expression

There were fourteen SNPs identified in the four desmosomal genes from participants with ARVC and DCM. Three SNPs with a minor allelic frequency of ≥5% were selected to investigate their effect on gene expression (chapter five). The three SNPs were rs1046116 in *PKP2*, rs868333 in *DSC2* and rs1893963 in *DSG2*. Due to gene expression technical difficulties in these experiments, only *PKP2* rs 1046116was investigated further. There was no association between genotype and total gene expression nor was there any evidence of

allelic expression imbalance of the rs1046116 *PKP2* SNP in both South African and North East (UK) populations.

7.1.4 Identification of novel candidate genes for ARVC through exome sequencing

There were also a number of families with both ARVC and DCM who did not have a causal mutation in desmosomal genes, and one of these families was selected for exome sequencing to investigate for a known or novel causal genetic mutation for ARVC (chapter six). Family ACM 8 had two heterozygous mutations that segregated with disease in the offspring: *PARVA*, p.D131V and *HMGXB3* p.E203K. The latter was reported as a rare SNP (rs142114383) with a prevalence of less than 1%. *PARVA* gene encodes α-parvin also known as actopaxin, a focal adhesion protein, and *HMGXB3* encodes a high mobility group box containing 3, a transcription protein.

PARVA was prioritized to be screened in sixty five ARVC probands as it belongs to a group of focal adhesion proteins, such as vinculin and metavinculin, which are implicated in the causation of cardiomyopathy (Vasile, Will et al. 2006). The screening of 65 unrelated probands with ARVC identified two cases with mutations in PARVA that may have a deleterious effect on protein function. This study has established PARVA as a strong candidate gene for cardiomyopathy.

7.2 Future directions

This work has laid the foundation for at least four directions of research in South Africa. First, I have shown that desmosomal protein gene mutations account for a minority (about a fifth) of cases of ARVC in South Africa – which is much lower than the proportion of up to 50% found elsewhere in the world. Therefore, there is a need not only to screen for non-

desmosomal gene causes of ARVC, but also for a major effort to identify hitherto unknown genetic causes of ARVC and DCM. Numbers of persons diagnosed with ARVC will also be influenced by the threshold for diagnoses and the other factors. This may also play a role in finding mutations and the proportions with specific mutations.

Second, the discovery of a founder effect provides the basis for the extension of the existing families to establish a large panel of individuals with ARVC due to the same genetic cause. It is likely that the four families identified in this thesis from a likely common founder is a fraction of carriers of the disease-causing mutation in South Africa.The ascertainment and phenotyping of these family members will provide a resource for genotype-phenotype studies of ARVC.

Third, I have shown in a proof of principle study that desmosomal protein genes, such as *PKP2*, are expressed in peripheral blood. Although the exonic variant in *PKP2* had no effect on whole gene expression or allele specific expression, there is a need to study a large number of polymorphisms and variants of unknown significance by this method. This study has shown that it is possible to study gene expression of desmosomal protein genes reliably in peripheral blood, although ideally heart tissue should be used for these experiments.

Finally, I have discovered strong evidence supporting the role of *PARVA* as a candidate gene for ARVC. The screening of large numbers of cases of ARVC and other cardiomyopathies, and functional studies of the impact of *PARVA* mutations are required to establish the causality of this variant. In addition, the potential role of mutations in *HMGX3* as modifier genes for ARVC requires further study.

Publication

Watkins, D. A., N. Hendricks, G. Shaboodien, M. Mbele, M. Parker, B. Z. Vezi, A. Latib, A. Chin, F. Little, M. Badri, J. C. Moolman-Smook, A. Okreglicki, B. M. Mayosi and A. R. o. t. C. A. S. o. S. Africa (2009). "Clinical features, survival experience, and profile of plakophylin-2 gene mutations in participants of the arrhythmogenic right ventricular cardiomyopathy registry of South Africa." Heart Rhythm **6**(11 Suppl): S10-17.

References

Ahmad, F., D. Li, A. Karibe, O. Gonzalez, T. Tapscott, R. Hill, D. Weilbaecher, P. Blackie, M. Furey, M. Gardner, L. L. Bachinski and R. Roberts (1998). "Localization of a gene responsible for arrhythmogenic right ventricular dysplasia to chromosome 3p23." Circulation **98**(25): 2791-2795.

Ahmed, A., W. S. Aronow and J. L. Fleg (2006). "Higher New York Heart Association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function." Am Heart J **151**(2): 444-450.

Arbustini, E., N. Narula, G. W. Dec, K. S. Reddy, B. Greenberg, S. Kushwaha, T. Marwick, S. Pinney, R. Bellazzi, V. Favalli, C. Kramer, R. Roberts, W. A. Zoghbi, R. Bonow, L. Tavazzi, V. Fuster and J. Narula (2013). "The MOGE(S) classification for a phenotype-genotype nomenclature of cardiomyopathy: endorsed by the World Heart Federation." J Am Coll Cardiol **62**(22): 2046-2072.

Awad, M. M., H. Calkins and D. P. Judge (2008). "Mechanisms of disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy." Nat Clin Pract Cardiovasc Med **5**(5): 258-267.

Badenhorst, G., A. van Staden and E. Coetsee (2008). "HIV/AIDS risk factors among residence students at the University of the Free State." Curationis **31**(3): 27-35.

Basso, C., D. Corrado, B. Bauce and G. Thiene (2012). "Arrhythmogenic right ventricular cardiomyopathy." Circ Arrhythm Electrophysiol **5**(6): 1233-1246.

Bauce, B., A. Nava, G. Beffagna, C. Basso, A. Lorenzon, G. Smaniotto, M. De Bortoli, I. Rigato, E. Mazzotti, A. Steriotis, M. P. Marra, J. A. Towbin, G. Thiene, G. A. Danieli and A. Rampazzo (2010). "Multiple mutations in desmosomal proteins encoding genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia." Heart Rhythm **7**(1): 22-29.

Beffagna, G., G. Occhi, A. Nava, L. Vitiello, A. Ditadi, C. Basso, B. Bauce, G. Carraro, G. Thiene, J. A. Towbin, G. A. Danieli and A. Rampazzo (2005). "Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1." Cardiovasc Res **65**(2): 366-373.

Beitelshees, A. L., C. L. Aquilante, H. Allayee, T. Y. Langaee, G. J. Welder, R. S. Schofield and I. Zineh (2012). "CXCL5 polymorphisms are associated with variable blood pressure in cardiovascular disease-free adults." Hum Genomics **6**: 9.

Brenner, S., M. Johnson, J. Bridgham, G. Golda, D. H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S. R. Williams, K. Moon, T. Burcham, M. Pallas, R. B. DuBridge, J. Kirchner, K. Fearon, J. Mao and K. Corcoran (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays." Nat Biotechnol **18**(6): 630-634.

Brink, A. J., M. Torrington and J. J. van der Walt (1976). "Hereditary dysrhythmic congestive cardiomyopathy." S Afr Med J **50**(54): 2119-2123.

Brink, P. A. and P. J. Schwartz (2009). "Of founder populations, long QT syndrome, and destiny." Heart Rhythm **6**(11 Suppl): S25-33.

Brooksbank, R., D. Badenhorst, K. Sliwa, G. Norton and A. Woodiwiss (2008). "The G-308A polymorphism of the TNF-alpha gene does not predict changes in cardiac function in response to medical therapy for idiopathic dilated cardiomyopathy." Cardiovasc J Afr **19**(5): 254-258.

Bustin, M. (1999). "Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins." Mol Cell Biol **19**(8): 5237-5246.

Campino, S., J. Fortor, S. Raj, B. Mohr, S. Auburn, A. Fry, V. D. Mangano, C. Vandiedonck, A. Richardson, K. Rockett, T. G. Clark and D. P. Kwiatkowski (2008). "Validating discovered Cis-acting

regulatory genetic variants: application of an allele specific expression approach to HapMap populations." PLoS One **3**(12): e4105.

Chen, X., J. Weaver, B. A. Bove, L. A. Vanderveer, S. C. Weil, A. Miron, M. B. Daly and A. K. Godwin (2008). "Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk." Hum Mol Genet **17**(9): 1336-1348.

Chiu, C., M. Tebo, J. Ingles, L. Yeates, J. W. Arthur, J. M. Lind and C. Semsarian (2007). "Genetic screening of calcium regulation genes in familial hypertrophic cardiomyopathy." J Mol Cell Cardiol **43**(3): 337-343.

Clarke, D. M., M. C. Brown, D. P. LaLonde and C. E. Turner (2004). "Phosphorylation of actopaxin regulates cell spreading and migration." J Cell Biol **166**(6): 901-912.

Coats, C. J., G. Quarta, A. S. Flett, A. A. Pantazis, W. J. McKenna and J. C. Moon (2009). "Arrhythmogenic left ventricular cardiomyopathy." Circulation **120**(25): 2613-2614.

Coonar, A. S., N. Protonotarios, A. Tsatsopoulou, E. W. Needham, R. S. Houlston, S. Cliff, M. I. Otter, V. A. Murday, R. K. Mattu and W. J. McKenna (1998). "Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21." Circulation **97**(20): 2049-2058.

Cox, M. G., J. J. van der Smagt, M. Noorman, A. C. Wiesfeld, P. G. Volders, I. M. van Langen, D. E. Atsma, D. Dooijes, A. C. Houweling, P. Loh, L. Jordaens, Y. Arens, M. J. Cramer, P. A. Doevendans, J. P. van Tintelen, A. A. Wilde and R. N. Hauer (2010). "Arrhythmogenic right ventricular dysplasia/cardiomyopathy diagnostic task force criteria: impact of new task force criteria." Circ Arrhythm Electrophysiol **3**(2): 126-133.

Cox, M. G., P. A. van der Zwaag, C. van der Werf, J. J. van der Smagt, M. Noorman, Z. A. Bhuiyan, A. C. Wiesfeld, P. G. Volders, I. M. van Langen, D. E. Atsma, D. Dooijes, A. van den Wijngaard, A. C.

Houweling, J. D. Jongbloed, L. Jordaens, M. J. Cramer, P. A. Doevendans, J. M. de Bakker, A. A. Wilde, J. P. van Tintelen and R. N. Hauer (2011). "Arrhythmogenic right ventricular dysplasia/cardiomyopathy: pathogenic desmosome mutations in index-patients predict outcome of family screening: Dutch arrhythmogenic right ventricular dysplasia/cardiomyopathy genotype-phenotype follow-up study." Circulation **123**(23): 2690-2700.

Crotti, L., A. L. Lundquist, R. Insolia, M. Pedrazzini, C. Ferrandi, G. M. De Ferrari, A. Vicentini, P. Yang, D. M. Roden, A. L. George, Jr. and P. J. Schwartz (2005). "KCNH2-K897T is a genetic modifier of latent congenital long-QT syndrome." Circulation **112**(9): 1251-1258.

Cunnington, M. S., C. Kay, P. J. Avery, B. M. Mayosi, M. S. Koref and B. Keavney (2009). "STK39 polymorphisms and blood pressure: an association study in British Caucasians and assessment of cis-acting influences on gene expression." BMC Med Genet **10**: 135.

Cunnington, M. S., M. Santibanez Koref, B. M. Mayosi, J. Burn and B. Keavney (2010). "Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression." PLoS Genet **6**(4): e1000899.

Dalal, D., C. James, R. Devanagondi, C. Tichnell, A. Tucker, K. Prakasa, P. J. Spevak, D. A. Bluemke, T. Abraham, S. D. Russell, H. Calkins and D. P. Judge (2006). "Penetrance of mutations in plakophilin-2 among families with arrhythmogenic right ventricular dysplasia/cardiomyopathy." J Am Coll Cardiol **48**(7): 1416-1424.

Damasceno, A., B. M. Mayosi, M. Sani, O. S. Ogah, C. Mondo, D. Ojji, A. Dzudie, C. K. Kouam, A. Suliman, N. Schrueder, G. Yonga, S. A. Ba, F. Maru, B. Alemayehu, C. Edwards, B. A. Davison, G. Cotter and K. Sliwa (2012). "The causes, treatment, and outcome of acute heart failure in 1006 Africans from 9 countries." Arch Intern Med **172**(18): 1386-1394.

Du Preez, J., L. O. Matolweni, J. Greenberg, P. Mntla, A. A. Adeyemo and B. M. Mayosi (2008). "The alpha 2C Del322-325 adrenergic receptor polymorphism is not associated with heart failure due to idiopathic dilated cardiomyopathy in black Africans." Cardiovasc J Afr **19**(1): 15-16.

Elliott, P., B. Andersson, E. Arbustini, Z. Bilinska, F. Cecchi, P. Charron, O. Dubourg, U. Kuhl, B. Maisch, W. J. McKenna, L. Monserrat, S. Pankuweit, C. Rapezzi, P. Seferovic, L. Tavazzi and A. Keren (2008). "Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases." Eur Heart J **29**(2): 270-276.

Elliott, P., C. O'Mahony, P. Syrris, A. Evans, C. Rivera Sorensen, M. N. Sheppard, G. Carr-White, A. Pantazis and W. J. McKenna (2010). "Prevalence of desmosomal protein gene mutations in patients with dilated cardiomyopathy." Circ Cardiovasc Genet **3**(4): 314-322.

Feng, B. J., S. V. Tavtigian, M. C. Southey and D. E. Goldgar (2011). "Design considerations for massively parallel sequencing studies of complex human disease." PLoS One **6**(8): e23221.

Fernandez, P., J. Moolman-Smook, P. Brink and V. Corfield (2005). "A gene locus for progressive familial heart block type II (PFHBII) maps to chromosome 1q32.2-q32.3." Hum Genet **118**(1): 133-137.

Fish, M. (2010). "Analysis of desmosomal right ventricular cardiomyopathy." MSC theses.

Frances, R., A. M. Rodriguez Benitez and D. R. Cohen (1997). "Arrhythmogenic right ventricular dysplasia and anterior polar cataract." Am J Med Genet **73**(2): 125-126.

Frueh, F. W. and M. Noyer-Weidner (2003). "The use of denaturing high-performance liquid chromatography (DHPLC) for the analysis of genetic variations: impact for diagnostics and pharmacogenetics." Clin Chem Lab Med **41**(4): 452-461.

Gandjbakhch, E., P. Charron, V. Fressart, G. Lorin de la Grandmaison, F. Simon, F. Gary, A. Vite, B. Hainque, F. Hidden-Lucet, M. Komajda and E. Villard (2011). "Plakophilin 2A is the dominant isoform in human heart tissue: consequences for the genetic screening of arrhythmogenic right ventricular cardiomyopathy." Heart **97**(10): 844-849.

Garcia-Pavia, P., M. Cobo-Marcos, G. Guzzo-Merello, M. Gomez-Bueno, B. Bornstein, E. Lara-Pezzi, J. Segovia and L. Alonso-Pulpon (2013). "Genetics in dilated cardiomyopathy." Biomark Med **7**(4): 517-533.

Gehmlich, K., A. Asimaki, T. J. Cahill, E. Ehler, P. Syrris, E. Zachara, F. Re, A. Avella, L. Monserrat, J. E. Saffitz and W. J. McKenna (2010). "Novel missense mutations in exon 15 of desmoglein-2: role of the intracellular cadherin segment in arrhythmogenic right ventricular cardiomyopathy?" Heart Rhythm **7**(10): 1446-1453.

Gerull, B., A. Heuser, T. Wichter, M. Paul, C. T. Basson, D. A. McDermott, B. B. Lerman, S. M. Markowitz, P. T. Ellinor, C. A. MacRae, S. Peters, K. S. Grossmann, J. Drenckhahn, B. Michely, S. Sasse-Klaassen, W. Birchmeier, R. Dietz, G. Breithardt, E. Schulze-Bahr and L. Thierfelder (2004). "Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy." Nat Genet **36**(11): 1162-1164.

Grundberg, E., K. S. Small, A. K. Hedman, A. C. Nica, A. Buil, S. Keildson, J. T. Bell, T. P. Yang, E. Meduri, A. Barrett, J. Nisbett, M. Sekowska, A. Wilk, S. Y. Shin, D. Glass, M. Travers, J. L. Min, S. Ring, K. Ho, G. Thorleifsson, A. Kong, U. Thorsteindottir, C. Ainali, A. S. Dimas, N. Hassanali, C. Ingle, D. Knowles, M. Krestyaninova, C. E. Lowe, P. Di Meglio, S. B. Montgomery, L. Parts, S. Potter, G. Surdulescu, L. Tsaprouni, S. Tsoka, V. Bataille, R. Durbin, F. O. Nestle, S. O'Rahilly, N. Soranzo, C. M. Lindgren, K. T. Zondervan, K. R. Ahmadi, E. E. Schadt, K. Stefansson, G. D. Smith, M. I. McCarthy, P. Deloukas, E. T. Dermitzakis, T. D. Spector and C. Multiple Tissue Human Expression Resource (2012).

"Mapping cis- and trans-regulatory effects across multiple tissues in twins." Nat Genet **44**(10): 1084-1089.

Hamid, M. S., M. Norman, A. Quraishi, S. Firoozi, R. Thaman, J. R. Gimeno, B. Sachdev, E. Rowland, P. M. Elliott and W. J. McKenna (2002). "Prospective evaluation of relatives for familial arrhythmogenic right ventricular cardiomyopathy/dysplasia reveals a need to broaden diagnostic criteria." J Am Coll Cardiol **40**(8): 1445-1450.

Harismendy, O., P. C. Ng, R. L. Strausberg, X. Wang, T. B. Stockwell, K. Y. Beeson, N. J. Schork, S. S. Murray, E. J. Topol, S. Levy and K. A. Frazer (2009). "Evaluation of next generation sequencing platforms for population targeted sequencing studies." Genome Biol **10**(3): R32.

Hassel, D., T. Dahme, J. Erdmann, B. Meder, A. Huge, M. Stoll, S. Just, A. Hess, P. Ehlermann, D. Weichenhan, M. Grimmler, H. Liptau, R. Hetzer, V. Regitz-Zagrosek, C. Fischer, P. Nurnberg, H. Schunkert, H. A. Katus and W. Rottbauer (2009). "Nexilin mutations destabilize cardiac Z-disks and lead to dilated cardiomyopathy." Nat Med **15**(11): 1281-1288.

Haywood, A. F., N. D. Merner, K. A. Hodgkinson, J. Houston, P. Syrris, V. Booth, S. Connors, A. Pantazis, G. Quarta, P. Elliott, W. McKenna and T. L. Young (2013). "Recurrent missense mutations in TMEM43 (ARVD5) due to founder effects cause arrhythmogenic cardiomyopathies in the UK and Canada." Eur Heart J **34**(13): 1002-1011.

Hu, J. and I. Peter (2013). "Evidence of expression variation and allelic imbalance in Crohn's disease susceptibility genes NOD2 and ATG16L1 in human dendritic cells." Gene **527**(2): 496-502.

Itoh-Satoh, M., T. Hayashi, H. Nishi, Y. Koga, T. Arimura, T. Koyanagi, M. Takahashi, S. Hohda, K. Ueda, T. Nouchi, M. Hiroe, F. Marumo, T. Imaizumi, M. Yasunami and A. Kimura (2002). "Titin mutations as the molecular basis for dilated cardiomyopathy." Biochem Biophys Res Commun **291**(2): 385-393.

Jelassi, A., A. Slimani, I. Jguirim, M. Najah, F. Maatouk, M. Varret and M. N. Slimane (2011). "Effect of a splice site mutation in LDLR gene and two variations in PCSK9 gene in Tunisian families with familial hypercholesterolaemia." Ann Clin Biochem **48**(Pt 1): 83-86.

Kanchanawong, P., G. Shtengel, A. M. Pasapera, E. B. Ramko, M. W. Davidson, H. F. Hess and C. M. Waterman (2010). "Nanoscale architecture of integrin-based cell adhesions." Nature **468**(7323): 580-584.

Khogali, S. S., B. M. Mayosi, J. M. Beattie, W. J. McKenna, H. Watkins and J. Poulton (2001). "A common mitochondrial DNA variant associated with susceptibility to dilated cardiomyopathy in two different populations." Lancet **357**(9264): 1265-1267.

Kirchner, F., A. Schuetz, L. H. Boldt, K. Martens, G. Dittmar, W. Haverkamp, L. Thierfelder, U. Heinemann and B. Gerull (2012). "Molecular insights into arrhythmogenic right ventricular cardiomyopathy caused by plakophilin-2 missense mutations." Circ Cardiovasc Genet **5**(4): 400-411.

Kotze, M. J., W. J. De V lliers, K. Steyn, J. A. Kriek, A. D. Marais, E. Langenhoven, J. S. Herbert, J. F. Graadt Van Roggen, D. R. Van der Westhuyzen and G. A. Coetzee (1993). "Phenotypic variation among familial hypercholesterolemics heterozygous for either one of two Afrikaner founder LDL receptor mutations." Arterioscler Thromb **13**(10): 1460-1468.

Kotze, M. J., E. Langenhoven, L. Warnich, L. du Plessis and A. E. Retief (1991). "The molecular basis and diagnosis of familial hypercholesterolaemia in South African Afrikaners." Ann Hum Genet **55**(Pt 2): 115-121.

Kruse, M., E. Schulze-Bahr, V. Corfield, A. Beckmann, B. Stallmeyer, G. Kurtbay, I. Ohmert, E. Schulze-Bahr, P. Brink and O. Pongs (2009). "Impaired endocytosis of the ion channel TRPM4 is associated with human progressive familial heart block type I." J Clin Invest **119**(9): 2737-2744.

La Gerche, A., C. Robberecht, C. Kuiperi, D. Nuyens, R. Willems, T. de Ravel, G. Matthijs and H. Heidbuchel (2010). "Lower than expected desmosomal gene mutation prevalence in endurance athletes with complex ventricular arrhythmias of right ventricular origin." Heart **96**(16): 1268-1274.

Latib, Michaels and Mayosi (2004). Initial Report of the Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa. 15.

Li, D., F. Ahmad, M. J. Gardner, D. Weilbaecher, R. Hill, A. Karibe, O. Gonzalez, T. Tapscott, G. P. Sharratt, L. L. Bachinski and R. Roberts (2000). "The locus of a novel gene responsible for arrhythmogenic right-ventricular dysplasia characterized by early onset and high penetrance maps to chromosome 10p12-p14." Am J Hum Genet **66**(1): 148-156.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lopes, L. R. and P. M. Elliott (2013). "Genetics of heart failure." Biochim Biophys Acta.

Lorenz, S., I. Vakonakis, E. D. Lowe, I. D. Campbell, M. E. Noble and M. K. Hoellerer (2008). "Structural analysis of the interactions between paxillin LD motifs and alpha-parvin." Structure **16**(10): 1521-1531.

Maharaj, B. and M. G. Hammond (1990). "HLA-A, B, DR, and DQ antigens in black patients with idiopathic dilated cardiomyopathy." Am J Cardiol **65**(20): 1402-1403.

Mannion, J. (1977). "The peopling of newfoundland: Essays in historical geography." Institute of Social and Economic Research.

Marcus FI, M. W., Sherrill D, Basso C, Bauce B, Bluemke DA, Calkins H, Corrado D, Cox MG, Daubert JP, Fontaine G, Gear K, Hauer R, Nava A, Picard MH, Protonotarios N, Saffitz JE, Sanborn DM,

Steinberg JS, Tandri H, Thiene G, Towbin JA, Tsatsopoulou A, Wichter T, Zareba W (2010). Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: Proposed modification of the task force criteria. 121.

Marian, A. J. and J. Belmont (2011). "Strategic approaches to unraveling genetic causes of cardiovascular diseases.' Circ Res **108**(10): 1252-1269.

Matkovich, S. J., D. J. Van Booven, A. Hindes, M. Y. Kang, T. E. Druley, F. L. Vallania, R. D. Mitra, M. P. Reilly, T. P. Cappola and G. W. Dorn, 2nd (2010). "Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease." J Clin Invest **120**(1): 280-289.

Matolweni, L. O., S. Bardien, G. Rebello, E. Oppon, M. Munclinger, R. Ramesar, H. Watkins and B. M. Mayosi (2006). "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 Med Genet **7**: 29.

Matolweni, L. O., S. Bardien, G. Rebello, E. Oppon, M. J. Munclinger, R. Ramesar and B. M. Mayosi "Arrhythmogenic right ventricular dysplasia/cardiomyopathy type 6 (ARVC6): corroboration of locus assignment, narrowing of the critical region and mutation screening of three candidate genes." (submitted).

Mayosi, B. M., S. Khogali, B. Zhang and H. Watkins (1999). "Cardiac and skeletal actin gene mutations are not a common cause of dilated cardiomyopathy." J Med Genet **36**(10): 796-797.

Mayosi BM, M. L., Matolweni LO, Lawrenson J, Moolman-Smook JC (2004). "A cardiac troponin T gene mutation causes early-onset familial dilated cardiomyopathy in a South African family." Cardiovasc J S Afr **15**(237).

Mbele, M. (2008). "The prevalence of plakophilin-2 gene mutations in South African patients with cardiomyopathy." Masters thesis.

McKenna, W. J., G. Thiene, A. Nava, F. Fontaliran, C. Blomstrom-Lundqvist, G. Fontaine and F. Camerini (1994). "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." Br Heart J **71**(3): 215-218.

McKoy, G., N. Protonotarios, A. Crosby, A. Tsatsopoulou, A. Anastasakis, A. Coonar, M. Norman, C. Baboonian, S. Jeffery and W. J. McKenna (2000). "Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease)." Lancet **355**(9221): 2119-2124.

Melberg, A., A. Oldfors, C. Blomstrom-Lundqvist, E. Stalberg, B. Carlsson, E. Larrson, C. Lidell, K. E. Eeg-Olofsson, G. Wikstrom, G. Henriksson and N. Dahl (1999). "Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy linked to chromosome 10q." Ann Neurol **46**(5): 684-692.

Merner, N. D., K. A. Hodgkinson, A. F. Haywood, S. Connors, V. M. French, J. D. Drenckhahn, C. Kupprion, K. Ramadanova, L. Thierfelder, W. McKenna, B. Gallagher, L. Morris-Larkin, A. S. Bassett, P. S. Parfrey and T. L. Young (2008). "Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene." Am J Hum Genet **82**(4): 809-821.

Millar, R. N. and B. M. Mayosi (2003). "Utilization of implantable defibrillators in Africa." Card Electrophysiol Rev **7**(1): 14-16.

Mogensen, J., T. Kubo, M. Duque, W. Uribe, A. Shaw, R. Murphy, J. R. Gimeno, P. Elliott and W. J. McKenna (2003). "Idiopathic restrictive cardiomyopathy is part of the clinical expression of cardiac troponin I mutations." J Clin Invest **111**(2): 209-216.

Mohini Joshi, D. J. D. (2010). "POLYMERASE CHAIN REACTION: METHODS, PRINCIPLES AND APPLICATION." International Journal of Biomedical Research **5**(1): 81-97.

Mokhobo, K. P. and P. S. Mntla (1997). "Lone ventricular cardiomyopathy, 1993-1996." S Afr Med J **87**(7): 892-896.

Moolman-Smook, J. C., W. J. De Lange, E. C. Bruwer, P. A. Brink and V. A. Corfield (1999). "The origins of hypertrophic cardiomyopathy-causing mutations in two South African subpopulations: a unique profile of both independent and founder events." Am J Hum Genet **65**(5): 1308-1320.

Munclinger, M. J., J. J. Patel and A. S. Mitha (2000). "Follow-up of patients with arrhythmogenic right ventricular cardiomyopathy dysplasia." S Afr Med J **90**(1): 61-68.

Nakajima, T., Y. Kaneko, T. Irie, R. Takahashi, T. Kato, T. Iijima, T. Iso and M. Kurabayashi (2012). "Compound and digenic heterozygosity in desmosome genes as a cause of arrhythmogenic right ventricular cardiomyopathy in Japanese patients." Circ J **76**(3): 737-743.

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure and M. J. Bamshad (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-35.

Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D. A. Nickerson and J. Shendure (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.

Nikolopoulos, S. N. and C. E. Turner (2000). "Actopaxin, a new focal adhesion protein that binds paxillin LD motifs and actin and regulates cell adhesion." J Cell Biol **151**(7): 1435-1448.

Norgett, E. E., S. J. Hatsell, L. Carvajal-Huerta, J. C. Cabezas, J. Common, P. E. Purkis, N. Whittock, I. M. Leigh, H. P. Stevens and D. P. Kelsell (2000). "Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma." Hum Mol Genet **9**(18): 2761-2766.

Norio, R. (2003). "The Finnish Disease Heritage III: the individual diseases." Hum Genet **112**(5-6): 470-526.

Ntusi, N. B. and B. M. Mayosi (2009). "Aetiology and risk factors of peripartum cardiomyopathy: a systematic review." Int J Cardiol **131**(2): 168-179.

Ntusi, N. B., A. Wonkam, G. Shaboodien, M. Badri and B. M. Mayosi (2011). "Frequency and clinical genetics of familial dilated cardiomyopathy in Cape Town: implications for the evaluation of patients with unexplained cardiomyopathy." S Afr Med J **101**(6): 394-398.

Pham, M. H., G. B. Bonello, J. Castiblanco, T. Le, J. Sigala, W. He and S. Mummidi (2012). "The rs1024611 regulatory region polymorphism is associated with CCL2 allelic expression imbalance." PLoS One **7**(11): e49498.

Pilichou, K., A. Nava, C. Basso, G. Beffagna, B. Bauce, A. Lorenzon, G. Frigo, A. Vettori, M. Valente, J. Towbin, G. Thiene, G. A. Danieli and A. Rampazzo (2006). "Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy." Circulation **113**(9): 1171-1179.

Pilichou, K., A. Nava, C. Basso, G. Beffagna, B. Bauce, A. Lorenzon, G. Frigo, A. Vettori, M. Valente, J. Towbin, G. Thiene, G. A. Danieli and A. Rampazzo (2006). "Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy." Circulation.

Posch, M. G., M. J. Posch, C. Geier, B. Erdmann, W. Mueller, A. Richter, V. Ruppert, S. Pankuweit, B. Maisch, A. Perrot, J. Buttgereit, R. Dietz, W. Haverkamp and C. Ozcelik (2008). "A missense variant in desmoglein-2 predisposes to dilated cardiomyopathy." Mol Genet Metab **95**(1-2): 74-80.

Quarta, G., A. Muir, A. Pantazis, P. Syrris, K. Gehmlich, P. Garcia-Pavia, D. Ward, S. Sen-Chowdhry, P. M. Elliott and W. J. McKenna (2011). "Familial evaluation in arrhythmogenic right ventricular cardiomyopathy: impact of genetics and revised task force criteria." Circulation **123**(23): 2701-2709.

Rampazzo, A., A. Nava, G. A. Danieli, G. Buja, L. Daliento, G. Fasoli, R. Scognamiglio, D. Corrado and G. Thiene (1994). "The gene for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 14q23-q24." Hum Mol Genet **3**(6): 959-962.

Rampazzo, A., A. Nava, P. Erne, M. Eberhard, E. Vian, P. Slomp, N. Tiso, G. Thiene and G. A. Danieli (1995). "A new locus for arrhythmogenic right ventricular cardiomyopathy (ARVD2) maps to chromosome 1q42-q43." Hum Mol Genet **4**(11): 2151-2154.

Rampazzo, A., A. Nava, S. Malacrida, G. Beffagna, B. Bauce, V. Rossi, R. Zimbello, B. Simionati, C. Basso, G. Thiene, J. A. Towbin and G. A. Danieli (2002). "Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy." Am J Hum Genet **71**(5): 1200-1206.

Rampazzo, A., A. Nava, M. Miorin, P. Fonderico, B. Pope, N. Tiso, B. Livolsi, R. Zimbello, G. Thiene and G. A. Danieli (1997). "ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy, maps to chromosome 2 long arm." Genomics **45**(2): 259-263.

Reeves, R. and L. Beckerbauer (2001). "HMGI/Y proteins: flexible regulators of transcription and chromatin structure." Biochim Biophys Acta **1519**(1-2): 13-29.

Reis-Filho, J. S. (2009). "Next-generation sequencing." Breast Cancer Res **11 Suppl 3**: S12.

Rigato, I., B. Bauce, A. Rampazzo, A. Zorzi, K. Pilichou, E. Mazzotti, F. Migliore, M. Perazzolo Marra, A. Lorenzon, M. De Bortoli, M. Calore, A. Nava, L. Daliento, D. Gregori, S. Iliceto, G. Thiene, C. Basso and D. Corrado (2013). "Compound and Digenic Heterozygosity Predicts Life-Time Arrhythmic Outcome and Sudden Cardiac Death in Desmosomal Gene-Related Arrhythmogenic Right Ventricular Cardiomyopathy." Circ Cardiovasc Genet.

Salisbury, B. A., M. Pungliya, J. Y. Choi, R. Jiang, X. J. Sun and J. C. Stephens (2003). "SNP and haplotype variation in the human genome." Mutat Res **526**(1-2): 53-61.

Santos, C. F., V. T. Sakai, M. A. Machado, D. N. Schippers and A. S. Greene (2004). "Reverse transcription and polymerase chain reaction: principles and applications in dentistry." J Appl Oral Sci **12**(1): 1-11.

Schaub, M. A., A. P. Boyle, A. Kundaje, S. Batzoglou and M. Snyder (2012). "Linking disease associations with regulatory information in the human genome." Genome Res **22**(9): 1748-1759.

Sen-Chowdhry, S., P. Syrris and W. J. McKenna (2007). "Role of genetic analysis in the management of patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy." J Am Coll Cardiol **50**(19): 1813-1821.

Severini, G. M., M. Krajinovic, B. Pinamonti, G. Sinagra, P. Fioretti, M. C. Brunazzi, A. Falaschi, F. Camerini, M. Giacca and L. Mestroni (1996). "A new locus for arrhythmogenic right ventricular dysplasia on the long arm of chromosome 14." Genomics **31**(2): 193-200.

Shaboodien, G., M. E. Engel, F. F. Syed, J. Poulton, M. Badri and B. M. Mayosi (2009). "The mitochondrial DNA T16189C polymorphism and HIV-associated cardiomyopathy: a genotype-phenotype association study." BMC Med Genet **10**: 37.

Shaboodien, G., C. Maske, H. Wainwright, H. Smuts, M. Ntsekhe, P. J. Commerford, M. Badri and B. M. Mayosi (2013). "Prevalence of myocarditis and cardiotropic virus infection in Africans with HIV-

associated cardiomyopathy, idiopathic dilated cardiomyopathy and heart transplant recipients: a pilot study: cardiovascular topic." Cardiovasc J Afr **24**(6): 218-223.

Sheu, C. C., R. Zhai, Z. Wang, M. N. Gong, P. Tejera, F. Chen, L. Su, B. T. Thompson and D. C. Christiani (2009). "Heme oxygenase-1 microsatellite polymorphism and haplotypes are associated with the development of acute respiratory distress syndrome." Intensive Care Med **35**(8): 1343-1351.

Sliwa, K., A. Damasceno and B. M. Mayosi (2005). "Epidemiology and etiology of cardiomyopathy in Africa." Circulation **112**(23): 3577-3583.

Sliwa, K., D. Hilfiker-Kleiner, M. C. Petrie, A. Mebazaa, B. Pieske, E. Buchmann, V. Regitz-Zagrosek, M. Schaufelberger, L. Tavazzi, D. J. van Veldhuisen, H. Watkins, A. J. Shah, P. M. Seferovic, U. Elkayam, S. Pankuweit, Z. Papp, F. Mouquet, J. J. McMurray and C. Heart Failure Association of the European Society of Cardiology Working Group on Peripartum (2010). "Current state of knowledge on aetiology, diagnosis, management, and therapy of peripartum cardiomyopathy: a position statement from the Heart Failure Association of the European Society of Cardiology Working Group on peripartum cardiomyopathy." Eur J Heart Fail **12**(8): 767-778.

Sliwa, K. and B. M. Mayosi (2013). "Recent advances in the epidemiology, pathogenesis and prognosis of acute heart failure and cardiomyopathy in Africa." Heart.

Sliwa, K. and B. M. Mayosi (2013). "Recent advances in the epidemiology, pathogenesis and prognosis of acute heart failure and cardiomyopathy in Africa." Heart **99**(18): 1317-1322.

Stiegler, A. L., K. M. Draheim, X. Li, N. E. Chayen, D. A. Calderwood and T. J. Boggon (2012). "Structural basis for paxillin binding and focal adhesion targeting of beta-parvin." J Biol Chem **287**(39): 32566-32577.

Strachan, T. and A. P. Read (1999). Human Molecular Genetics. New York.

Stranger, B. E., A. C. Nica, M. S. Forrest, A. Dimas, C. P. Bird, C. Beazley, C. E. Ingle, M. Dunning, P. Flicek, D. Koller, S. Montgomery, S. Tavare, P. Deloukas and E. T. Dermitzakis (2007). "Population genomics of human gene expression." Nat Genet **39**(10): 1217-1224.

Syrris, P., D. Ward, A. Asimaki, S. Sen-Chowdhry, H. Y. Ebrahim, A. Evans, N. Hitomi, M. Norman, A. Pantazis, A. L. Shaw, P. M. Elliott and W. J. McKenna (2006). "Clinical expression of plakophilin-2 mutations in familial arrhythmogenic right ventricular cardiomyopathy." Circulation **113**(3): 356-364.

Taylor, M., S. Graw, G. Sinagra, C. Barnes, D. Slavov, F. Brun, B. Pinamonti, E. E. Salcedo, W. Sauer, S. Pyxaras, B. Anderson, B. Simon, J. Bogomolovas, S. Labeit, H. Granzier and L. Mestroni (2011). "Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes." Circulation **124**(8): 876-885.

Te Rijdt, W. P., J. D. Jongbloed, R. A. de Boer, G. Thiene, C. Basso, M. P. van den Berg and J. P. van Tintelen (2013). "Clinical utility gene card for: arrhythmogenic right ventricular cardiomyopathy (ARVC)." Eur J Hum Genet.

Tiago, A. D., D. Badenhorst, D. Skudicky, A. J. Woodiwiss, G. P. Candy, R. Brooksbank, K. Sliwa, P. Sareli and G. R. Norton (2002). "An aldosterone synthase gene variant is associated with improvement in left ventricular ejection fraction in dilated cardiomyopathy." Cardiovasc Res **54**(3): 584-589.

Tipping, A. J., T. Pearson, N. V. Morgan, R. A. Gibson, L. P. Kuyt, C. Havenga, E. Gluckman, H. Joenje, T. de Ravel, S. Jansen and C. G. Mathew (2001). "Molecular and genealogical evidence for a founder effect in Fanconi anemia families of the Afrikaner population of South Africa." Proc Natl Acad Sci U S A **98**(10): 5734-5739.

Tishkoff SA, R. F., Friedlaender FR, Ehret C, Ranciaro A, Froment, H. J. A, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma and K. M. AT, Lema G, Moore JH, (2009). "The Genetic Structure and History of Africans and African Americans." cience **324**((5930)): 1035-1044.

Tiso, N., D. A. Stephan, A. Nava, A. Bagattin, J. M. Devaney, F. Stanchi, G. Larderet, B. Brahmbhatt, K. Brown, B. Bauce, M. Muriago, C. Basso, G. Thiene, G. A. Danieli and A. Rampazzo (2001). "Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2)." Hum Mol Genet **10**(3): 189-194.

Tuch, B. B., R. R. Laborde, X. Xu, J. Gu, C. B. Chung, C. K. Monighetti, S. J. Stanley, K. D. Olsen, J. L. Kasperbauer, E. J. Moore, A. J. Broomer, R. Tan, P. M. Brzoska, M. W. Muller, A. S. Siddiqui, Y. W. Asmann, Y. Sun, S. Kuersten, M. A. Barker, F. M. De La Vega and D. I. Smith (2010). "Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations." PLoS One **5**(2): e9317.

van Hengel, J., M. Calore, B. Bauce, E. Dazzo, E. Mazzotti, M. De Bortoli, A. Lorenzon, I. E. Li Mura, G. Beffagna, I. Rigato, M. Vleeschouwers, K. Tyberghein, P. Hulpiau, E. van Hamme, T. Zaglia, D. Corrado, C. Basso, G. Thiene, L. Daliento, A. Nava, F. van Roy and A. Rampazzo (2013). "Mutations in the area composita protein alphaT-catenin are associated with arrhythmogenic right ventricular cardiomyopathy." Eur Heart J **34**(3): 201-210.

van Tintelen, J. P., M. M. Entius, Z. A. Bhuiyan, R. Jongbloed, A. C. Wiesfeld, A. A. Wilde, J. van der Smagt, L. G. Boven, M. M. Mannens, I. M. van Langen, R. M. Hofstra, L. C. Otterspoor, P. A. Doevendans, L. M. Rodriguez, I. C. van Gelder and R. N. Hauer (2006). "Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy." Circulation **113**(13): 1650-1658.

VanGuilder, H. D., K. E. Vrana and W. M. Freeman (2008). "Twenty-five years of quantitative PCR for gene expression analysis." Biotechniques **44**(5): 619-626.

Vasile, V. C., S. R. Ommen, W. D. Edwards and M. J. Ackerman (2006). "A missense mutation in a ubiquitously expressed protein, vinculin, confers susceptibility to hypertrophic cardiomyopathy." Biochem Biophys Res Commun **345**(3): 998-1003.

Vasile, V. C., M. L. Will, S. R. Ommen, W. D. Edwards, T. M. Olson and M. J. Ackerman (2006). "Identification of a metavinculin missense mutation, R975W, associated with both hypertrophic and dilated cardiomyopathy." Mol Genet Metab **87**(2): 169-174.

Watkins, D. A., N. Hendricks, G. Shaboodien, M. Mbele, M. Parker, B. Z. Vezi, A. Latib, A. Chin, F. Little, M. Badri, J. C. Moolman-Smook, A. Okreglicki, B. M. Mayosi and A. R. o. t. C. A. S. o. S. Africa (2009). "Clinical features, survival experience, and profile of plakophylin-2 gene mutations in participants of the arrhythmogenic right ventricular cardiomyopathy registry of South Africa." Heart Rhythm **6**(11 Suppl): S10-17.

Watkins, H., H. Ashrafian and C. Redwood (2011). "Inherited cardiomyopathies." N Engl J Med **364**(17): 1643-1656.

Watkins, H., H. Ashrafian and C. Redwood (2011). "MECHANISMS OF DISEASE Inherited Cardiomyopathies." New England Journal of Medicine **364**(17): 1643-1656.

Wehrle-Haller, B. (2012). "Structure and function of focal adhesions." Curr Opin Cell Biol **24**(1): 116-124.

Wimley, W. C. and S. H. White (1996). "Experimentally determined hydrophobicity scale for proteins at membrane interfaces." Nat Struct Biol **3**(10): 842-848.

Woodiwiss, A. J., D. Badenhorst, K. Sliwa, R. Brooksbank, R. Essop, P. Sareli and G. R. Norton (2008). "Beta1- and alpha2c-adrenoreceptor variants as predictors of clinical aspects of dilated cardiomyopathy in people of African ancestry." Cardiovasc J Afr **19**(4): 188-193.

Xiao, W., D. Stern, M. Jain, C. G. Huber and P. J. Oefner (2001). "Multiplex capillary denaturing high-performance liquid chromatography with laser-induced fluorescence detection." Biotechniques **30**(6): 1332-1338.

Xu, T., Z. Yang, M. Vatta, A. Rampazzo, G. Beffagna, K. Pilichou, S. E. Scherer, J. Saffitz, J. Kravitz, W. Zareba, G. A. Danieli, A. Lorenzon, A. Nava, B. Bauce, G. Thiene, C. Basso, H. Calkins, K. Gear, F. Marcus, J. A. Towbin and I. Multidisciplinary Study of Right Ventricular Dysplasia (2010). "Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy." J Am Coll Cardiol **55**(6): 587-597.

Yang, Y., D. M. Muzny, J. G. Reid, M. N. Bainbridge, A. Willis, P. A. Ward, A. Braxton, J. Beuten, F. Xia, Z. Niu, M. Hardison, R. Person, M. R. Bekheirnia, M. S. Leduc, A. Kirby, P. Pham, J. Scull, M. Wang, Y. Ding, S. E. Plon, J. R. Lupski, A. L. Beaudet, R. A. Gibbs and C. M. Eng (2013). "Clinical whole-exome sequencing for the diagnosis of mendelian disorders." N Engl J Med **369**(16): 1502-1511.

Zhang, M., F. Tavora, J. B. Oliveira, L. Li, M. Franco, D. Fowler, Z. Zhao and A. Burke (2012). "PKP2 mutations in sudden death from arrhythmogenic right ventricular cardiomyopathy (ARVC) and sudden unexpected death with negative autopsy (SUDNA)." Circ J **76**(1): 189-194.

Appendices

Appendix 1: Approval letter from Ethics committee

UNIVERSITY OF CAPE TOWN

Research Ethics Committee
Faculty of Health Sciences
OMB E46 Room 26, GSH
Queries : Xolile Fula
Tel : (021) 406-6492 Fax 406-6411
E-mail : Xfula@curie.uct.ac.za

21 May 2003

REC REF: 047/2003

Dr B Mayosi
Cardiology

Dear Dr Mayosi

ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY REGISTRY OF SOUTH AFRICA

Thank you for submitting your study to the Research Ethics Committee for review.

It is a pleasure to inform you that the Ethics Committee has formally approved the above-mentioned study.

Please quote the Reference number in all correspondence.

Yours sincerely

PROF T ZABOW
CHAIRPERSON

Appendix 2: Solutions and Buffers

10X Tris Borate EDTA (TBE) buffer (2L)

216g (0.89M) Tris base (ICN Biomedicals, Inc.)

110g (0.89M) Boric Acid (ICN Biomedicals, Inc.)

14.8g (0.5M) Ethylenediaminetetraacetic acid, pH 8.0 (Merck)

Made up to 2L with dH_2O

Agarose Loading Dye solution

0.125g (0.25%) Bromophenol blue

20g (40%) Sucrose (Calbiochem)

Made up to 50ml with dH_2O and pH adjusted to 8

1X Buffer O solution (for 100% *MboII* digestion)

50mM Tris-HCL (pH 7.5) (Promega)

10mM $MgCl_2$ (Merck)

100mM NaCl (Sigma-Aldrich)

0.1mg/ml Bovine Serum Albumin (Thermo Scientific

1. **1X Buffer Tango solution (for 100% *NspI* digestion)**

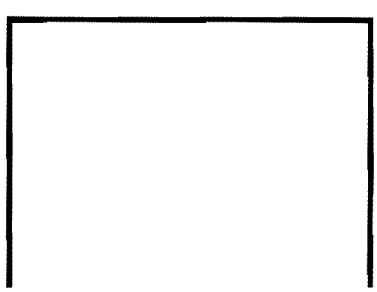
33mM Tris-acetate (pH 7.9)

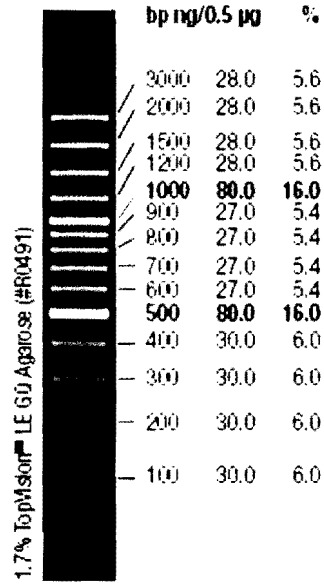
10mM $Mg(C_2H_3O_2)_2$ (Sigma-Aldrich)

66mM CH_3COOK (AnalytiCals)

0.1mg/ml Bovine Serum Albumin (Thermo Scientific)

Appendix 3: GeneRuler™ 100bp DNA Ladder Plus





0.5 µg/lane, 8 cm length gel,
1X TBE, 5 V/cm, 1 h

Appendix 4: Desmosomal gene designed PCR primers

Table 1: *PKP2* primer sequences used and PCR product sizes for each exon

Primer	PCR products Size (bp)	Primer sequence (5'–3')	Annealing Temperature (°C)
1-1F	263	ACT CGA GCG GGG CGG GGC TCG C	68
1-1R		GCG GC CAG ACA GTC AAG AGC CTG	
1-2F	268	GTC CTG GGC CAG CAG ATC CTG	55
1-2R		CTC ACC CCG CGT GCT GGG AGT	
2F1	203	GCC TAC TTG TAT TCT GAG TT	53
2R1		CTA GTG ATC AGA TTA CTA CCA TG	
3-1F1	231	CAG AAA GGG GCA AAC TTC TCG	58
3-1R1		AAG GTC CTT GAG GCA TCC TC	
3-2F	291	CTC TCA TGT GCA GGC TGG C	55
3-2R		GAC ACA TAC CAC AGA CAG TAC	
3-3F	273	GAT ATG CTC GTT CCG AGA TC	55
3-3R		GTC CTC CTG GCA TCA GAG	
3-4F	236	GTC ACT CAG AAC AGG G	55
3-4R		GCA GTT CAG TCA TTG CCA G	
4F	291	AGT ATT CGC TGA GTC GTC TCT	55
4R		CTT CTA TTA TGG TGA CTT TGC	
5-1F	272	GAA AGG TTA TAG TCA GCA TCA G	55
5-1R		GTA CCT CGG CTG CTC CAG GTG	
5-2F	250	CGT TCA GCG AGC TGT GTG	57
5-2R		TCC TGG AGC AAA TGA TTG ATG	
6F	258	TTG CTG TGT TCA TAA AGG AGC C	55
6R		GTC GTG GTC TGC GCC TGT AAT	
7F	291	GGT CCC ACC CTG CAC TGT TTT C	58
7R		GTA GCC CCA AGG AAG TCA G	
8F	250	CAA AGA CCT GTT GGA TAC ACA	58
8R		ATT CTT AAG ATA GCC GCT TGG	
9-1F	305	TAC TCA TTG CAT TTC CCC CAG	57
9-1R		GAT GTT TTG GCA GTC GAA GCA G	
9-2F	173	GCA TTC TTC ATA ACC TCT CCT AC	55
9-2R		GGG AAG TTT TTG AAA GAA TG	
10-1F	222	CAG TAT TTC TGG TCT CCT GG	58
10-1R		CGC AAC TAC ACA CAA GAA GCA TCC	
10-2F	218	GGA GGA AAA GAG CAA CCC CAA G	58
10-2R		GTA GAC ATA CCT ACA TTC AAT TC	
11-1F	191	CAT CAA CCT CTG GTA ATC TAC G	60
11-1R		CCA AGT GTG AAA AAG ACA GCC ATC	
11-2F	219	CTC AGA CAG TTG TCC AGA AG	56
11-2R		GTT GTC TGT ATC ACC TCC CG	
12-1F	210	GGG ATC TTG GGA TTA AGA AAC AAT C	62
12-1R		GAA TGC ACG CGA CCT TCT AAA C	
12-2F	187	CAG CCT CTG CCT GTT ACA CAT TG	59
12-2R		GTG TGT AAA TAA TCC GGT TG	
13F	149	CCC TGA TCT CAG AAT GTC C	53
13R		GAA CAA GGA ATG TGT CAT AG	
14F	240	GTT ACC TTT CAC GTT TCT G	56
14R		GTG GTC CCC TGA ATC CAG AA	

*Number (e.g. 9) denotes the exon; F= forward primer, R= reverse primer

Table 2: *DSG2* primer sequences used and PCR product sizes for each exon

Primer	PCR product size (bp)	Primer sequence (5'–3')	Annealing temperature (°c)
1F	351	CTTTGGGTTGGGCTGGGCTG	59
1R		CGCCCTTGTGCGCGTTAC	
2F	167	CTGAATTGAGCAGTAAATTGGC	55
2R		GTTGACTGGGCTTTACTAG	
3F	337	GAA GCC TCA TAG GAA ATA CG	58
3R		GTG TGA GAG GAC TTT TAT GTC	
4F	298	CAA CTC CAC TGC CAT TCC CC	53
4R		GTA GTT TTT CTG TCA TAA TAA G	
5F	275	CAT AGT ATG GTA ATT TAG TTT TC	51
5R		CTT TAT CCC CAC TGT AAA TAA AC	
6F	399	CAT TCA CGC TTA TGT CCT CAT C	54
6R		TCC TTTCCA GTT TCT CCT CTC	
7F	345	GGA CTA AAA CCA GAA AGC CAG	59
7R		GCA GCT TGA AGG GAT GGT TG	
8F	604	CAA GCT CAA GTT CAG ATT CG	55
8R		GAT TCC TTT CCA GCA CCA AC	
9-1F	287	TCC TGT GCA TTA AAT TAT TGT ATC	50
9-1R		GCA GCG TCA TCT CAA TTT ATG	
9-2F	386	GAA GAA T CTT GAC TTC AGT G	54
9-2R		CAA AAG GTC TAC AAG TTA AAT TTT CC	
10F	307	GAAAGAGCTGTAGATGTTAGAGG	60
10R		CCAACATCAGAAACCATTGATTTGATTG	
11F	438	GGCAAGGGAATTCAAACTATG	58
11R		GGAGTTTATGAAATGAGGTCCTG	
12F	364	CATTTGTGGAATTTAATGTTGCCTC	55
12R		AGTGCACTCCTGGAGATG	
13F	313	GTGAAGACAAGTCCAGGAAGG	54
13R		GTGTTTCTTACATGTTTGGTTG	
14-1F	262	CCAAGGATGAAGGATTCCCTC	59
14-1R		GTAGAGCTACCCAGTTTACAGG	
14-2F	286	CAACATGAGATGTCCGAGATG	52
14-2R		GTATGTGAATGTGATATTATTAGGG	
15-1F	313	GAC CCA GGC AAA CTT CAG	56
15-1R		CAG TCC AGT TGT GTC GTA GG	
15-2F	310	GCT TCT ATT GGT TGT TGC AG	53
15-2R		CAT AGT TTG CTC ACA GAG TG	
15-3F	393	GAG CAG AGA CAA AAA CCT G	57
15-3R		CAG TGG TTG GTG GCA TAG	
15-4F	350	CAC CTC TTC CTG ACC CAA TG	53
15-4R		GTC ACA TTC TGT CCT ACT GC	
15-5F	362	GAA TCC TCT GGA AGG CAC TC	55
15-5R		GCA GCA TCA TCT CTA TGT C	

Table 3: *DSC2* primer sequences used and PCR product sizes for each exon

Primer	PCR product size (bp)	Primer sequence (5'–3')	Annealing Temperature (°C)
1F	461	CTA AGA AAA GCA CCT CTC CGC	54
1R		GTG CGC TGA TTA CAT CTA CC	
2F	679	GGA GGA AGC TAG AAC TCA G	53
2R		GTG GGG AAC AAC ACA AGT G	
3-1F	223	GAA TTT TAA GTT TTT GGC TCT CCC	52
3-1R		CTT CTC TTC TCC GAG GAC	
3-2F	259	GGA GGA TGG TTC AGT CTA TAC	53
3-2R		GCC TCA TGG TTT TCA TTC GTC	
4F	269	CTG AGA GGA GAA AGA TGT GTC	51
4R		GTG TCA TAA TGG TAA GAG ATG G	
5-1F	335	GGA GAA CTA AGA TGC ACT AC	50
5-1R		CGA GTA CAA TAC AAG TTT CCA G	
5-2F	233	CCT GGA GTT GAC CAA GAA C	53
5-2R		CTA TTA GGG AGT AGC CAG AGC	
6F	320	CAG TCT CCA AGT TAA AGC C	51
6R		CTG CTT CTC AAC GGA CAT AG	
7F	692	CTG TAT GGA GAG GTC TGC	51
7R		CCA GAG ATT AGA GAG ACA TAA AC	
8F	334	GCT GTG AAA TGT AAT AAT GCA C	56
8R		CAA CTT TTA ATG TCA CTG GGG CAC	
9F	540	GTG CTA GGG TTT CAA ATC ATG	55
9R		CTG GGC TAG GAA AAG TCT GG	
10F	554	CGT GGT TAT TCC CAA CAT G	51
10R		CTC TCT ATC CAG GCT TCT G	
11F	475	GAA ACA AGA AGT AGC AGT GG	51
11R		GCT ATT AGA AAG CAG ACA TGA G	
12-1F	211	CAG TGC ATA CTT TTG TGG TG	52
12-1R		GGA TAG GCT CAT CAG GAT C	
12-2F	288	CTG CAA ACC CAC CAT GTC	52
12-2R		CAC TAG AGA GAC CTA CTC CC	
13F	621	GCC ATC CCA TCA GGA ATA TC	53
13R		GAT AGA GAC TGG CCT GGC	
14F	397	GAC ACA AGC AAT CCA GGA AAG C	56
14R		CAG GGG ACC CAT GAC ATT C	
15F	901	GTT TCC ATA GTT GGG TAT GC	51
15R		CAC GCA TGT GGC AGC AAG	
16F	615	GAA TCC ATT AGA GGA CAC ACT C	51
16R		GGC TGT TGT CAT TAT ACC C	

Table 4: *JUP* primer sequences used and PCR product sizes for each exon

Primer	PCR product size (bp)	Primer sequence (5'–3')	Annealing Temperature (°C)
1F	186	GTCTCTTCGCCTTTTGTTCGGT	61
1R		CTG TCC CCA ACG ATA CCT G	
2-1F	199	CCT TTG TGC CCC CAG TAG	60
2-1R		GTG GTT TTC TTG AGC GTG TAC	
2-2F	184	CAGCAGCAAGGGCATCATGG	58
2-2R		CAG AGA CCC CCT ACA ATC	
3-1F	223	GTT TGG GCA GGG CAA GAG	59
3-1R		GTG GCC AGC AGA AGC GAG	
3-2F	195	GAGTACCAGATGTCCACAAC	53
3-2R		GCTCAAGTCGGCCATTGT	
3-3F	210	GCTCAAGTCGGCCATTGT	56
3-3R		CAT ACC CTT CCA CAG AGC TGA	
4-1F	299	GAG AGC AGA TGT CCT CAG	58
4-1R		GTG TCC AGG TCG CTG GTA TTC	
4-2F	236	GCT GTC GTG CGT ACC ATG C	62
4-2R		CAC TGG ATA TTT ATG GAA GCT	
5-1F	341	CAG AAG CAG GAG GAG TCA TTG	57
5-1R		GAT AGA AGA TGG CGC AAG GGT G	
5-2F	223	CAT CTC ACA CTC ATT CCC TC	54
5-2R		CAC CCT TGC GCC ATC TTC TAT C	
6F	280	GCG CTT CCT TGT TCC TGT CA	53
6R		GGA GCA TGG CTG ACT GAG	
7F	279	GAGGCTGGTGAGTATGATG	56
7R		CAG ACA GGA GGC TGG ATG	
8F	267	CTT GGA CAT ATT CGA GAA GG	59
8R		CGT GAT GTC GTC CTT GTC	
9-1F	205	CTT CAG ATG TCC AAG AAG TGC	55
9-1R		CGT GTA GGG CTG CTG TGT G	
9-2F	211	CTC GTC CAA CTG CTG GTG AAG	58
9-2R		CCA TGA CAG CCG AAT GAA C	
10F	279	CCA TGT CCA AGG CAC CTG	51
10R		CTC CAA AGA CCT CTT GAT AC	
11-1F	257	GAC GGC CCA TTT TCC ACT ATG	55
11-1R		GAG TGC AGC AAC TCC ATG AGT	
11-2F	239	GTG GAG AAC ATC CAG CGC GTG	56
11-2R		GGT AGT CTG GGT TCT TGT C	
12F	258	GAG GCT GGC CGA CGT TTA AC	58
12R		CTA TGA AGT TGA CAG TAG TAG	
13F	234	GAG CCT GCA TTT TCA AAC CTC	60
13R		CGG AGC CTG CAT TTT CAA AC	
14F	135	GAC CGA GCC CCA CTT TTT GTC	58
14R		CAA AGA GGG GGC CGT ACT	

Table 5: *PARVA* primer sequences used and PCR product sizes for each exon

Primer	PCR product size (bp)	Primer sequence (5'–3')	Annealing Temperature (°C)
1-1F	247	CAG CTC CTT CCA GGG CAG AG	62
1-1R		GTC GCC TTC TGT CCC CAA GTC	
1-2F	417	CGT TTG GAA AGC CGC AGC CTC AG	63
1-2R		CGA CGC GCA CCC CAC ACT GAG	
2F	225	TGT TAC AGA GCT CCA ACC CC	60
2R		GCT GGC ATG GGA AAG CTT AA	
3F	210	CACGATGCTGGGTAACTGTG	59
3R		CTCACTGACCCACCATGTG	
4F	242	GCT GGT CTG TAA GAA AAT ACC CC	60
4R		CAC ACA TCC ATT GCA GCA CT	
5F	246	GCA TGC AGT AGG GTT GTC TG	60
5R		ATG GAT GGC TTC TTG GGG AA	
6F	243	TGC CTA GAA CCT GGA AGA AGA	58
6R		GAC CTA CTC GAC GCA CCA TA	
7F	250	CAG GAA TGC TCA TCA GTG TTT C	60
7R		TCC CTC CTG CAA TCC CAA TA	
8F	201	CAA TAA TTG GAG GCC GTG CA	55
8R		TGA GAG CCT GTT CCT CCC A	
9F	241	GAG GCT TCC CAT CGG TAG TT	58
9R		TCC CAA TGT TCT GTT GCA TCT C	
10F	240	TCC TCT GCT TTC ACC CTC AC	60
10R		AGA TTC CCA TGC CCT CAT CC	
11F	250	GGT AGG CTT CAG GTG GCA TA	52
11R		TCC CTT GAT GTC CCT TCT CC	
12F	207	TGC TGG GCT CCT TCA CTT C	57
12R		TGC TTC CCT CTG CTA CAA AAT C	
13F	238	GAA GGG AGG GGC AGT GTA TG	55
13R		GTA AGG CAG TTC GGA GGA GG	