Analysis of Desmoplakin in Arrhythmogenic Right Ventricular Cardiomyopathy

Maryam Fish

Supervisors:
Professor Bongani M. Mayosi
Dr Gasnat Shaboodien

Presented for:
MSc (Med) in Cardiovascular Genetics

Department of Medicine
Faculty of Health Sciences
University of Cape Town

August 2010

TABLE OF CONTENTS

PAGE NUMBER

DECLARATION

- This study was performed between 2009-2010 under the supervision of Professor Bongani M Mayosi from the Department of Medicine at the University of Cape Town.
- I hereby certify that this thesis is my work is my own and has not been presented for a degree at any other university.
- I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is one's own.
- Each significant contribution to, and quotation in, this essay/report/project from the work, or works, of other people has been attributed, and has been cited and referenced.
- This essay/report/project is my own work using my own words.

Name (Printed): ……..MARYAM FISH….

Signature: …….

Date: …13 AUGUST 2010..

ABSTRACT

It has been shown that all forms of cardiomyopathy, including the dilated, hypertrophic, restrictive, and right ventricular arrhythmogenic forms, are found in African populations. Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a rare muscle disease characterised by fibrofatty replacement of the right ventricular myocardium, leading to electrical instability and eventual heart failure. Dilated cardiomyopathy (DCM) is a disease characterised by a reduction in ventricular wall thickness which leads to reduced contractility and impaired ventricular function. Mutations that cause ARVC have been reported in five desmosomal and three non-desmosomal genes. Among the desmosomal genes is desmoplakin (*DSP*) which encodes the desmoplakin protein. *DSP* has previously been shown to cause ARVC and DCM. We hypothesise that mutations in *DSP* may be causative of ARVC and familial and idiopathic DCM in our cohort of South African patients

We recruited consecutive probands with ARVC and fDCM from Grootte Schuur Hospital in Cape Town. Informed consent was obtained from 62 well characterised ARVC and 150 well characterised DCM (19 familial, 131 idiopathic) probands. Probands were screened for mutations in *DSP* by means of high resolution melt analysis and DNA sequencing. Population screens were conducted on novel variants to determine the frequency of variants in the South African control population.

The frequency of disease-causing mutations in *DSP* was 3% in cases of ARVC (2/62 probands), 5% in cases of familial DCM (1/19 probands) and 4% in apparently sporadic cases of DCM (5/131 probands). We also detected a number of variants of unknown significance which will require further study to determine their role in disease pathogenesis. We found mutations that were common to both the ARVC and DCM cohorts, and were determined to be causative of both these diseases.

We have shown that mutations in *DSP* are causative of ARVC and DCM in South African cohorts. We have demonstrated that a large number of familial dilated cardiomyopathy probands harbour mutations in the *DSP* gene. We have also shown that the same mutations can cause ARVC and DCM, which suggests that ARVC and DCM may form part of a common spectrum of heart disease.

ACKNOWLEDGEMENTS

- Firstly, to my supervisor, Professor Bongani M. Mayosi, for providing me with support and guidance throughout my MSc

- To my co-supervisor, Dr Gasnat Shaboodien, for always being there for support and for letting me bug her whenever I needed help

- To the rest of the Cardiovascular Genetics and Cardioprotection laboratory team, thank you for all the help provided to me throughout the year

- Finally, to my parents, siblings, fiancé and friends for putting up with me, and for giving me support whenever I needed it

LIST OF FIGURES

PAGE NUMBER

LIST OF TABLES

PAGE NUMBER

LIST OF ABBREVIATIONS AND SYMBOLS

µ	Micro
aa	Amino acid/s
ABI	Applied Biosystems
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
CVD	Cardiovascular Disease
DCM	Dilated Cardiomyopathy
ddNTPs	Dideoxynucleotide Triphosphates
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphates
DSC2	Desmocollin 2
DSG2	Desmoglein 2
DSP	Desmoplakin
Dsp	Murine desmoplakin
DSPI	Desmoplakin protein isoform I
DSPII	Desmoplakin protein isoform II
ECG	Electrocardiography
ESE	Exonic Splice Enhancer
HCM	Hypertrophic Cardiomyopathy
HIV/AIDS	Human Immunodeficiency Virus/AutoImmune Deficiency Syndrome
HRM	High Resolution Melt Analysis
IDP	Inner Dense Plaque
IDT	Integrated DNA Technologies
JUP	Junction plakoglobin

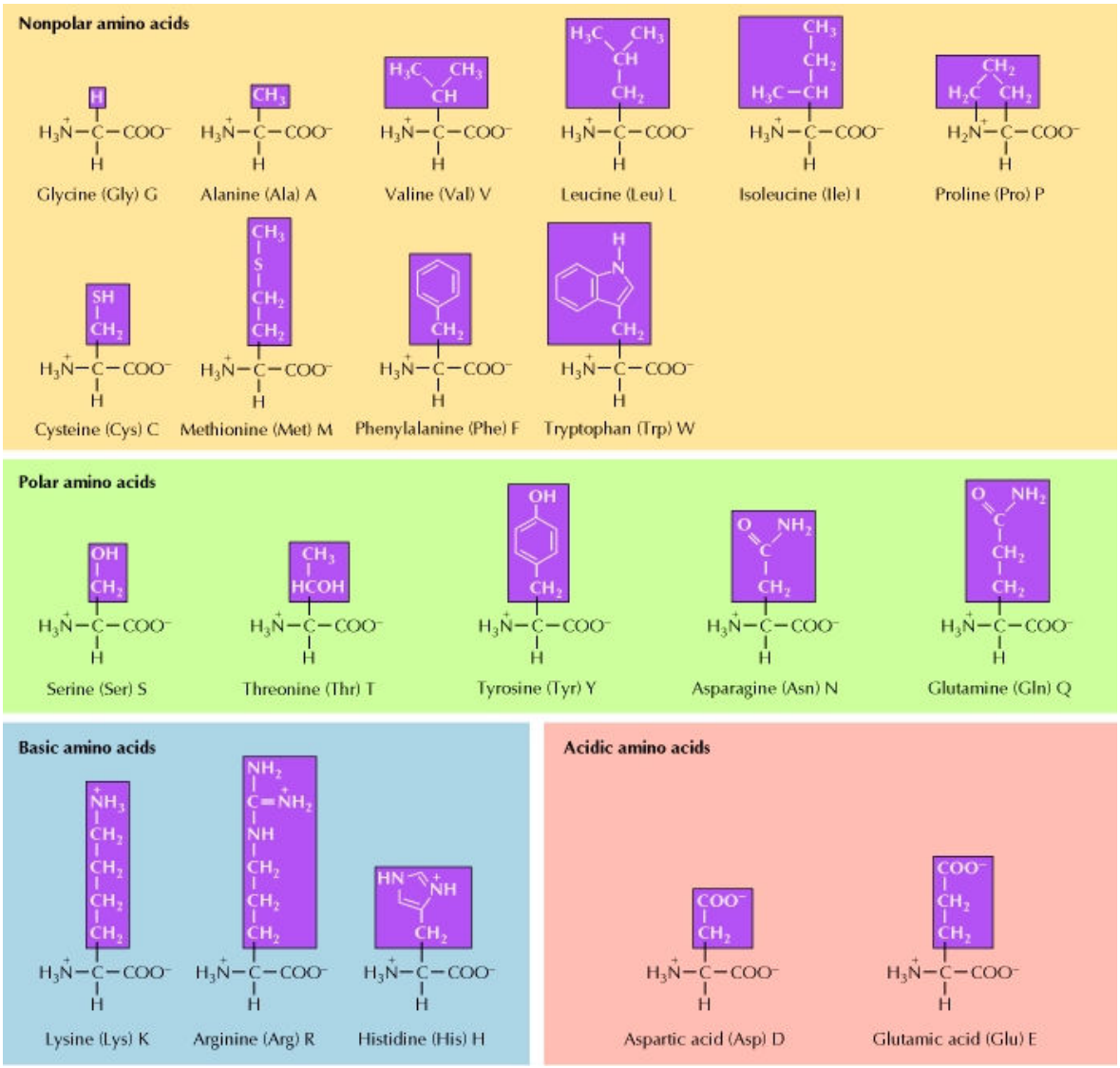
l	Litres
M	Molar (moles per litre)
ml	Millilitres
MLPA	Multiplex Ligation-Dependent Probe Amplification
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
ng	Nanogram
ODP	Outer Dense Plaque
PCR	Polymerase Chain Reaction
PKP2	Plakophilin 2
POLYPHEN	POLYmorphism PHENotyping
RCM	Restrictive Cardiomyopathy
RNA	Ribonucleic Acid
RyR2	Cardiac Ryanodine Receptor gene
SIFT	Sorting Intolerant From Tolerant
TBE	Tris/Borate/EDTA
TGFβ3	Transforming Growth Factor Beta 3
TMEM43	Transmembrane Protein 43
UCT	University of Cape Town
UTR	Untranslated Region
V	Volts

NUCLEIC ACID CODES

A	Adenine
G	Guanine
C	Cytosine
T	Thymine

AMINO ACID CODES

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



1. INTRODUCTION

1.1. STRUCTURE AND FUNCTION OF THE HEART

The heart is responsible for the circulation of blood throughout the body. The heart has four chambers, namely the right and left atria which receive blood from the venous system, and the right and left ventricles which pump blood into the arterial system (Figure 1.1). The right atrium is separated from the left atrium by the muscular wall known as the inter-atrial septum; similarly the right ventricle and left ventricle are divided by the interventricular septum. Valves of the heart include the tricuspid valve (which separates the right atrium and ventricle), the mitral valve (which separates the left atrium and ventricle), the pulmonary valve (which separates the right ventricle from the pulmonary artery) and the aortic valve (which separates the left ventricle from the aorta) (Fox 2006).

Deoxygenated blood enters the right atrium of the heart through the superior and inferior vena cava and then goes to the right ventricle, which subsequently pumps the blood through the pulmonary artery to the lungs where it is oxygenated. The oxygenated blood then travels to the left atrium of the heart via the pulmonary veins, and then to the left ventricle, from where it is pumped through the aorta to the entire body (Fox 2006).

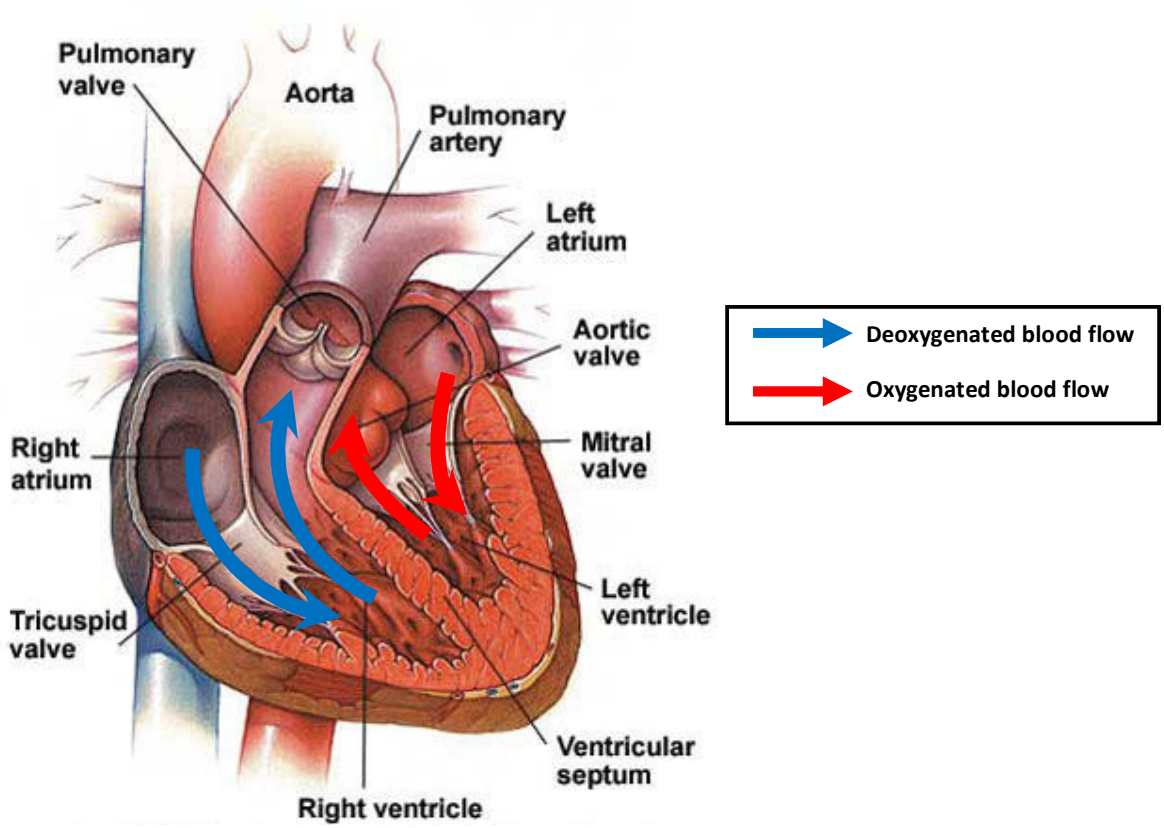


Figure 1.1: Structure of the heart (©Mayo Foundation for Medical Education and Research) (Electronic Resource 1)

The walls of the cardiac chambers are composed of three layers: the epicardium, the myocardium and the endocardium. These layers are surrounded by the pericardium (Figure 1.2). The cells of the myocardium are called cardiomyocytes and are specialised muscle cells that are striated, as they contain actin and myosin filaments arranged in the form of sarcomeres, which allow the contraction of the cardiomyocytes. These cells are short, branched and interconnected by different types of intercellular junctions (Fox 2006). The main functions of the myocardium are the generation of contractile forces (by the sarcomeres) and the transmission of this force to the extracellular matrix.

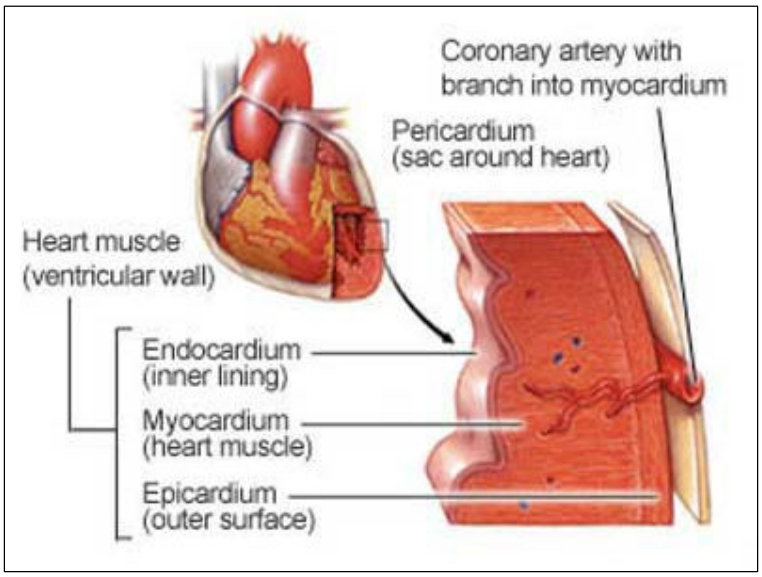


Figure 1.2: Layers of the heart (©Mayo Foundation for Medical Education and Research) (Electronic Resource 2)

1.2. CARDIOVASCULAR DISEASE

Cardiovascular disease (CVD) is the leading cause of death in all developing countries with the exception of sub-Saharan Africa (Opie and Mayosi 2005, Gaziano 2007) and is the second largest cause of mortality (after HIV/AIDS) in South Africa (Bradshaw *et al*. 2003). CVD has an enormous social and economic impact on society, causing great suffering and loss of life as well as posing major economic challenges to developing countries, including costs to the healthcare systems and to the national economy (Gaziano 2005, Gaziano 2007).

CVD can be attributed to many causes, including hypertension, high cholesterol, tobacco and alcohol use and low fruit and vegetable consumption (Gaziano 2005). The major causative factors in Africa, however, are hypertension, rheumatic heart disease and cardiomyopathy (Sliwa *et al*. 2005). Cardiomyopathy poses the greatest challenge of the cardiovascular disease causative factors in Africa because of its prevalence in societies that are still plagued by famine and pestilence and do not have access to proper resources or effective interventions, besides the difficulty of diagnosis and the high mortality associated with the disease (Sliwa *et al*. 2005).

1.3. CARDIOMYOPATHY

Cardiomyopathy is defined as a myocardial disease in which the heart muscle is structurally and functionally abnormal, without hypertension, coronary artery disease, valvular disease or congenital heart disease which is sufficient to cause the observed myocardial abnormality (Elliot *et al.* 2008). Cardiomyopathies often lead to progressive heart failure with significant morbidity and mortality, and can be primary (genetic, mixed or acquired) or secondary (infiltrative, toxic or inflammatory) (Wexler *et al.* 2009). Cardiomyopathies are classified into morphological and functional phenotypes, which are then subclassified into familial and non-familial forms (Elliot *et al.* 2008). Types of cardiomyopathy include dilated cardiomyopathy (DCM), restrictive cardiomyopathy (RCM), hypertrophic cardiomyopathy (HCM), arrhythmogenic right ventricular cardiomyopathy (ARVC) and unclassified cardiomyopathies (Elliot *et al.* 2008). This study will focus primarily on ARVC and DCM.

1.3.1. Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)

ARVC is a rare inherited muscle disease that predominantly affects the right ventricle and is an important cause of sudden death in people aged ≤35 years (Sen-Chowdhry *et al.* 2010). It is characterised by the replacement of cardiomyocytes by fibrofatty tissue, primarily in the right ventricle (Figure 1.3), which leads to electrical instability including ventricular arrhythmias in the early stages and later reduced contractility and heart failure (Awad *et al.* 2008, Kiès *et al*. 2006). Recently, other subtypes of this disease with left ventricular involvement have been recognised, namely left-dominant and biventricular types of arrhythmogenic cardiomyopathy (Sen-Chowdhry *et al.* 2010).

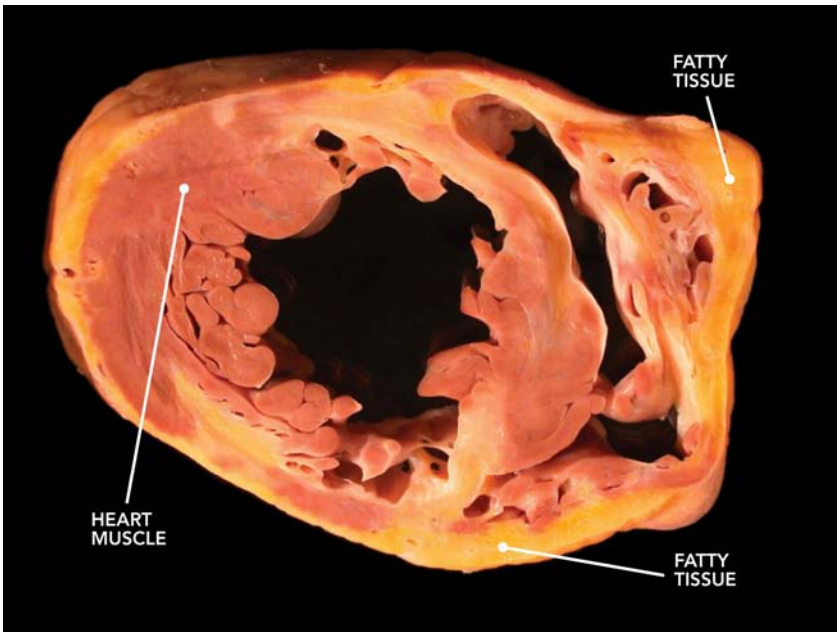


Figure 1.3: Cross section of the ARVC heart (Electronic Resource 3)

The initial phase of the disease is known as the concealed phase and involves a predisposition to ventricular arrhythmia and sudden cardiac death whilst having well-preserved morphology, histology and ventricular function (Sen-Chowdhry *et al.* 2010). As the disease progresses, the myocardial disorder becomes more evident. Myocyte loss occurs, which may be accompanied by fat infiltration, fibrosis and inflammation (Sen-Chowdhry *et al.* 2010). Fibrofatty replacement of the right ventricular myocardium is progressive, and starts from the epicardium or the mid myocardium and eventually leads to wall thinning and aneurysms (Basso *et al.* 2009). Fibrofatty replacement interferes with the conduction of electrical impulses which causes the epsilon waves, right bundle branch block, late potentials and ventricular arrhythmias seen in individuals with ARVC (Basso *et al.* 2009).

Other structural abnormalities that occur in this disease include increased trabeculation, global ventricular dilation and dysfunction and clinical heart failure, though this only occurs in some individuals at late stages of the disease (Sen-Chowdhry *et al.* 2010). Ventricular arrhythmias are worse during or immediately after exercise, and participation in competitive exercise has been associated with an increased risk of sudden death.

ARVC is estimated to have a prevalence ranging from 1 in 1000 to 1 in 5000 and accounts for up to 10% of deaths from undiagnosed cardiac disease in people aged <65 years (Sen-Chowdhry *et al.* 2010, Basso *et al.* 2009). ARVC affects more men than women (in a ratio of 3:1) and usually presents within the second to fourth decade of life (Corrado *et al.* 2009, Basso *et al.* 2009).

Patients who are suspected to have ARVC must undergo clinical assessments which include evaluation of symptoms and family history, physical examination, chest radiograph, 12-lead ECG, 24-hour ambulatory ECG, signal-averaged ECG, stress test and two-dimensional echocardiography. Further assessments include cardiac magnetic resonance imaging, contrast angiography and endomyocardial biopsy (Basso *et al.* 2009). The clinical diagnosis of ARVC is often difficult due to the non-specific nature of the disease features and the broad spectrum of physical manifestation (Corrado *et al.* 2009).

The diagnosis of ARVC is based on the presence of criteria that include structural, histological, electrocardiographic, arrhythmic and genetic factors that are characteristic of the disease (McKenna *et al.* 1994) (Table 1.1). These criteria are divided into major and minor criteria, based on their specificity for ARVC. An initial set of task force criteria was presented in 1994 by McKenna and colleagues, and a revision of these criteria has recently been proposed by Marcus and colleagues (Marcus *et al.* 2010). The 1994 task force criteria were used for the diagnosis of ARVC in this project.

Table 1.1: Task Force Criteria for the Diagnosis of ARVC (McKenna *et al*. 1994)

CRITERIA	DESCRIPTION
Global and/or regional dysfunction and structural alterations	MAJOR Severe dilatation and reduction of right ventricular ejection fraction with no (or only mild) left ventricular impairment Localised right ventricular aneurysms (akinetic or dyskinetic areas with diastolic bulging) Severe segmental dilatation of the right ventricle MINOR Mild global right ventricular dilatation and/or ejection fraction reduction with normal left ventricle Mild segmental dilatation of the right ventricle Regional right ventricular hypokinesia
Tissue characterization of walls	MAJOR Fibrofatty replacement of myocardium on endomyocardial biopsy
Repolarization abnormalities	MINOR Inverted T waves in right precordial leads (V2 and V3) (people aged more than 12 yrs; in absence of right bundle branch block)
Depolarisation/conduction abnormalities	MAJOR Epsilon waves or localised prolongation (> 110 ms) of the QRS complex in right precordial leads (V1-V3) MINOR Late potentials (signal averaged ECG)
Arrhythmias	MINOR Left bundle branch block type ventricular tachycardia (sustained and non-sustained) (ECG, Holter, exercise testing) Frequent ventricular extrasystoles (more than 1000/24 h) (Holter)
Family history	MAJOR Familial disease confirmed at necropsy or surgery MINOR Familial history of premature sudden death (<35 yrs) due to suspected right ventricular dysplasia Familial history (clinical diagnosis based on present criteria)

A familial background has been demonstrated in at least 50% of cases and the disease is typically transmitted as an autosomal dominant trait (Corrado *et al.* 2009, Basso *et al.* 2009). The penetrance of ARVC is reduced and frequently age-dependent, while affected relatives often have variable and incomplete disease expression (Sen-Chowdhry *et al.* 2010).

1.3.2. Dilated Cardiomyopathy (DCM)

DCM is a common form of the cardiomyopathy that is characterised by dilated cardiac chambers and reduced left ventricular systolic function (Moolman-Smook 2003, Hughes *et al.* 2005; Luk *et al*. 2009) (Figure 1.4). Symptoms and signs include dyspnoea, fatigue, angina, pulmonary congestion, arrhythmias and thromboembolic events (Luk *et al*. 2009).

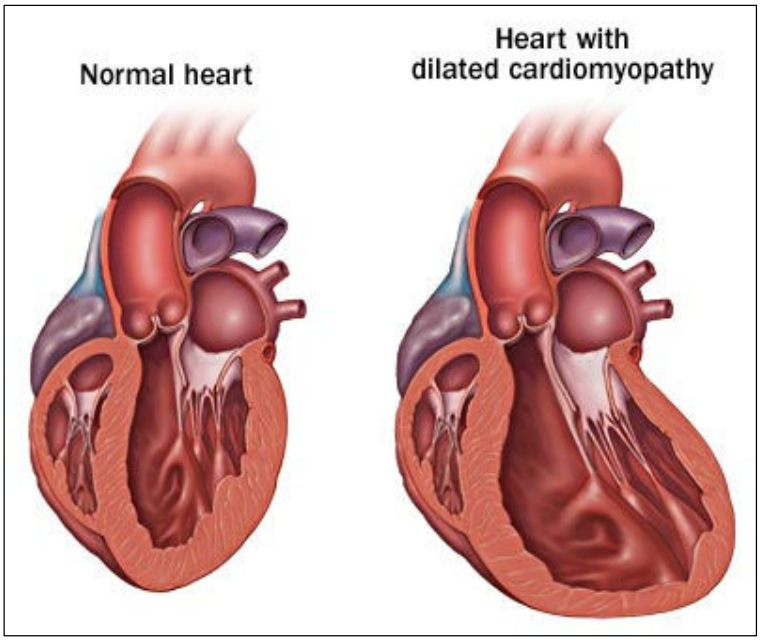


Figure 1.4: Comparison of the dilated cardiomyopathy heart and the normal heart (©Mayo Foundation for Medical Education and Research) (Electronic Resource 4)

In South Africa and Uganda, DCM accounts for 10-17% of all cardiac conditions which are found at autopsy, while in many other parts of Africa, DCM accounts for 17-48% of patients who are hospitalised for heart failure (Sliwa *et al*. 2005). The incidence of DCM in the United States and elsewhere has been reported to be 4-8/100 000 person-years, while the prevalence of this disease is reported to be 36.5/100 000 individuals (Sliwa *et al*. 2005). However, there is no population-based data on the burden of this disease in Africa as yet.

The diagnosis of DCM is confirmed by non-invasive cardiac imaging, particularly with two-dimensional echocardiography (Schönberger and Seidman, 2001). Diagnosis is made by the exclusion of any underlying cardiac diseases such as coronary artery disease, valvular heart disease and arterial hypertension (Fourlas *et al*. 2004).

Patients may present with the disease during childhood, but most patients present in the fourth or fifth decade of life (Schönberger and Seidman, 2001). Modes of inheritance of the familial form of this disease include autosomal dominant, autosomal recessive and X-linked inheritance (Luk *et al*. 2009). Autosomal dominant forms of DCM are caused by mutations in cytoskeletal, sarcomeric protein/Z-band, nuclear membrane and intercalated disc protein genes. The X-linked form of DCM is associated with diseases such as muscular dystrophies (Elliot *et al*. 2008). The disease may also occur in patients with mitochondrial cytopathies

and inherited metabolic disorders. Disease progression and severity varies, as family members with identical mutations may show variable penetrance (Luk *et al*. 2009). Although 25-35% of patients have a genetically inherited (familial) form of DCM, the majority of individuals with this disease have an acquired form which may be caused by nutritional deficiencies, endocrine dysfunction and the administration of cardiotoxic drugs (Elliot *et al*. 2008).

1.3.3. Molecular Genetics of ARVC and DCM

ARVC is caused by mutations in desmosomal protein genes in up to 50% of symptomatic individuals (Awad *et al*. 2008), whereas the causes of DCM are largely unknown, although several genes have been implicated in this disorder (Burkett *et al*. 2005). To date, 12 different chromosomal loci have been associated with ARVC (Table 1.2) (Awad *et al*. 2008). Eight genes have been identified at these loci, five of which encode desmosomal proteins. The fact that the genes encoding the desmosomal components have been implicated in ARVC suggests that ARVC is essentially a disease of the desmosome. The first desmosomal gene mutation found to be associated with ARVC was a mutation in the *JUP* gene which encodes the plakoglobin protein (McKoy *et al*. 2000). This mutation was found in patients with Naxos disease, which is characterised by ARVC, palmoplantar keratoderma and woolly hair (McKoy *et al*. 2000, Corrado *et al.* 2006, Awad *et al*. 2008). The genes encoding the desmoplakin (*DSP*), plakophilin (*PKP2*), desmoglein (*DSG2*) and desmocollin (*DSC2*) desmosomal proteins have also been implicated in ARVC, with *PKP2* being the main contributor to the genetic causes of ARVC, representing 11-43% of ARVC causative mutations (Fowler *et al*. 2010).

Table 1.2: Chromosomal Loci and Disease-causing Genes in ARVC (From Awad *et al*. 2008)

DESIGNATION	LOCUS	SYMBOL	DISEASE GENE
ARVD1	14q23-q24	*TGFβ3*	Transforming Growth Factor- β3
ARVD2	1q42-q43	*RyR2*	Cardiac Ryanodine Receptor
ARVD3	14q12-q22	?	Unknown
ARVD4	2q32.1-q32.3	?	Unknown
ARVD5	3p23	*TMEM43*	Transmembrane protein 43
ARVD6	10p12-p14	?	Unknown
ARVD7	10q22	?	Unknown
ARVD8	6p24	*DSP*	Desmoplakin
ARVD9	12p11	*PKP2*	Plakophilin 2
ARVD10	18q12.1- q12.2	*DSG2*	Desmoglein 2
ARVD11	18q12.1	*DSC2*	Desmocollin 2
ARVD12 / Naxos	17q21	*JUP*	Plakoglobin

However, three non-desmosomal genes have also been implicated in this disease. These include *RyR2* (which encodes the ryanodine receptor) (Tiso *et al*. 2001), *TGFβ3* (which encodes transforming growth factor β3 (Beffagna *et al*. 2005) and *TMEM43* (which encodes the transmembrane protein 43) (Merner *et al*. 2008). A summary of the genes involved in ARVC and the prevalence of mutations identified in these genes are described below (Table 1.3).

Table 1.3: Summary of Desmosomal and Non-Desmosomal Genes Implicated in ARVC and the prevalence of mutations in these genes (from Sen-Chowdhry *et al*. 2007, Fowler *et al*. 2010)

GENE (SYMBOL)	LOCUS	ESTIMATED PREVALENCE (%)
Plakoglobin (*JUP*)	17q21	<1
Desmoplakin (*DSP*)	6p24	1-20
Plakophilin (*PKP2*)	12p11	11-43
Desmoglein (*DSG2*)	18q12-q12.2	12-40
Desmocollin (*DSC2*)	18q12	~1.5%
Transforming Growth factor beta 3 (*TGFβ3*)	14q24	<1
Cardiac Ryanodine Receptor (*RyR2*)	1q42.1–q43	<1
Transmembrane protein 43 (*TMEM43*)	3p25.1	<1

At least 25 different chromosomal loci have been associated with familial DCM (Table 1.4). Genes encoding proteins that have cytoskeletal, contractile and calcium regulatory functions have been discovered at these loci, reflecting the genetic heterogeneity of the disease (Ahmad *et al*. 2005). Additional phenotypes such as skeletal myopathy have also been discovered in conjunction with DCM in patients with various gene mutations.

Table 1.4: Chromosomal Loci and Disease-causing Genes in DCM (From Ahmad *et al*. 2005)

LOCUS	SYMBOL	DISEASE GENE	ADDITIONAL PHENOTYPE
1p1–q21	*LMNA*	Lamin A/C	Conduction disease, skeletal myopathy
1q32	*TNNT2*	Cardiac troponin T2	None
2q14–q22	?	Unknown	Conduction disease
2q31	*TTN*	Titin	None
2q35	*DES*	Desmin	Skeletal myopathy
3p22–p25	?	Unknown	Conduction disease
5q33	*SGCD*	Delta sarcoglycan	Skeletal myopathy
6p23–p24	*DSP*	Desmoplakin	Woolly hair, keratoderma
6q12–q16	?	Unknown	None
6q22.1	*PLN*	Phospholamban	None
6q23–q24	?	*EYA4*	Skeletal myopathy, sensorineural hearing loss
9q13–q22	?	Unknown	None
9q22–q31	?	Unknown	None
10q22–q23	*MVCL*	Metavinculin	Mitral valve prolapse
10q22.3–q23.2	*LDB3*	Cypher/ZASP	Left ventricular noncompaction
11p11.2	*MYBPC3*	Cardiac myosin binding protein C	None
11p15.1	*CLP*	Cardiac muscle LIM protein	None
12p12.1	*ABCC9*	ATP-sensitive K channel	Arrhythmias
14q12	*MYH7*	Beta myosin heavy chain	None
15q14	*ACTC*	Cardiac actin	None
15q22	*TPM1*	Alpha tropomyosin	None

LOCUS	SYMBOL	DISEASE GENE	ADDITIONAL PHENOTYPE
16p11	*CTF1*	Cardiotrophin 1	None
19q13.2	?	Unknown	Conduction disease
Xp21	*DMD*	Dystrophin	Skeletal myopathy
Xq28	*TAZ*	Tafazzin	Skeletal myopathy

1.4. DESMOSOMES

Desmosomes are specialised intercellular junctions that anchor intermediate filaments to the cytoplasmic membranes in adjacent cells, thus forming a supportive network. They play a key role in imparting mechanical strength via cell-cell adhesion and transmission of force between the junctional complex and the intermediate filaments in the cytoskeleton (Sen-Chowdhry *et al*. 2007). Desmosomes are found most abundantly in regions that undergo a high amount of mechanical stress, such as the heart and the skin (Sen-Chowdhry *et al*. 2007, Stokes 2007).

Desmosomes are similar to adherens junctions and hemi-desmosomes in structure, the difference being that these junctional complexes have different protein components (Stokes 2007). The principal sites of adhesion are in the extracellular domains of transmembrane proteins which belong to the cadherin family, that are characterised by five tandem extracellular domains (Stokes 2007). The cadherins which form part of the desmosomal complex are desmocollin-2 (DSC2) and desmoglein-2 (DSG2) (Sen-Chowdhry *et al*. 2007).

The intracellular portion of the desmosome consists of a plaque which can be divided into the outer dense plaque (which is closer to the cell membrane) and the inner dense plaque (which is located further away from the cell membrane) (Figure 1.5). The outer dense plaque consists of the intracellular portions of the cadherin proteins as well as members of the armadillo protein family, namely junctional plakoglobin (JUP) and plakophilin-2 (PKP2). The inner dense plaque consists of desmoplakin (DSP) which connects the desmosomal complex to the intermediate filaments, which form part of the cytoskeleton (Stokes 2007).

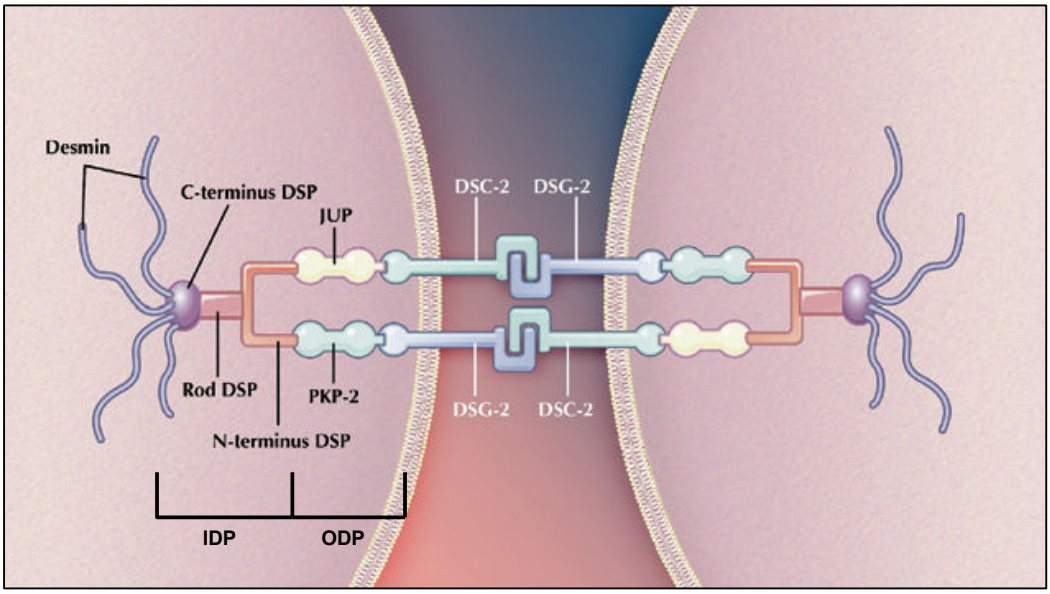


Figure 1.5: Structure of the desmosome (from Sen-Chowdhry *et al*. 2007)

ODP – Outer Dense Plaque; IDP – Inner Dense Plaque; DSP – Desmoplakin; JUP – Junctional Plakoglobin; PKP-2 – Plakophilin-2; DSC-2 – Desmocollin-2; DSG-2 – Desmoglein-2

Mutations which alter the structure and function of the desmosome could lead to myocyte detachment and death and subsequent fibrofatty tissue infiltration (Sen-Chowdhry *et al*. 2007, Awad *et al*. 2008). However, mutations in responsible genes demonstrate incomplete penetrance and variable expressivity, suggesting that genetic modifiers and environmental factors play a role in the etiology of these disorders (Awad *et al*. 2008).

In this study, the focus is on the *DSP* gene which encodes the desmoplakin protein. This gene accounts for 1-20% of ARVC mutations (Marcus *et al.* 2007, den Haan *et al* 2009, Fowler *et al*. 2010, Fressart *et al*. 2010) and has also been implicated in a syndromic form of DCM, known as Carvajal syndrome, which is characterised by DCM, woolly hair and palmoplantar keratoderma (Norgett *et al*. 2000). Therefore, the *DSP* gene is a good candidate for investigations of the genetic aetiology of familial and apparently idiopathic forms of DCM.

1.5. DESMOPLAKIN

1.5.1. Organisation of the *DSP* gene

The *DSP* gene is located on the short arm of chromosome 6, in the region 6p24.3 (Lai Cheong *et al*. 2005) and has orthologues in a number of species including chimpanzee (*Pan troglodytes*), dog (*Canis familiaris*), gorilla (*Gorilla gorilla*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*). It is 45077 base pairs (bp) in length and has 24 exons, the last two of which are particularly large in comparison to the others (Figure 1.6). These exons (exon 23 and exon 24) are 2295 and 4072 bp in length respectively.

1.5.2. The desmoplakin protein

Desmoplakin is one of the largest desmosomal plaque proteins. The desmoplakin protein is expressed in two isoforms, namely DSPI and DSPII, which are 2871 and 2272 aa in length respectively (Bolling *et al*. 2010). The DSP protein has three domains: the N-terminus, the rod domain and the C-terminus (Yang *et al*. 2006) (Table 1.5). The N-terminus and the C-terminus are globular domains and they have the central rod domain between them (Lai Cheong *et al*. 2005). The desmoplakin N-terminus contains six α-helical bundles named NN, Z, X, Y, W and V while the C-terminus contains A, B and C repeat domains that are important for the interaction of desmoplakin with intermediate filaments (Figure 1.7) (Lai Cheong *et al*. 2005).

Table 1.5: Domains of the DSP protein and their functions

DOMAIN	SIZE	FUNCTION	REFERENCE
N-terminus	1056 aa	- Localisation of desmoplakin to the desmosome - Interaction of desmoplakin with plakoglobin and plakophilin	Yang *et al*. 2006
Rod domain	889 aa (DSPI) or 290 aa (DSPII)	Homodimerisation of the desmoplakin protein	Stokes 2007
C-terminus	926 aa	Interaction of desmoplakin with the intermediate filaments in the cell	Yang *et al*. 2006

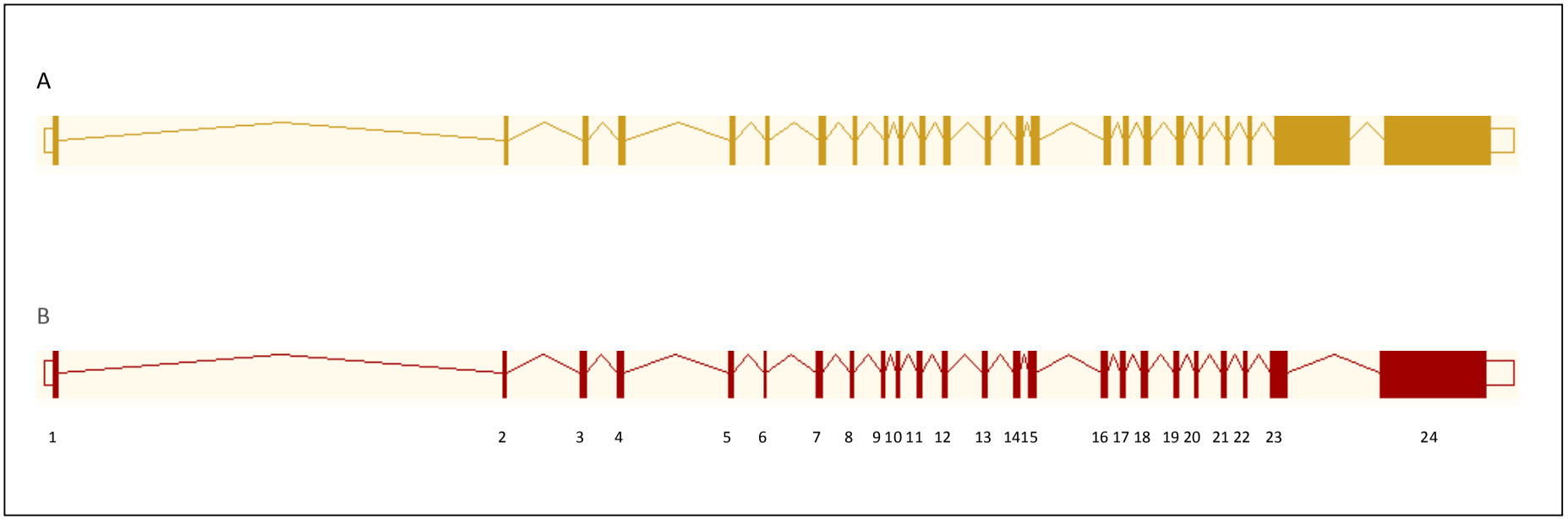


Figure 1.6: Organisation of DSP exons: A: Organisation of DSPI exons (from http://www.ensembl.org/)

B: Organisation of DSPII exons (from http://www.ensembl.org/)

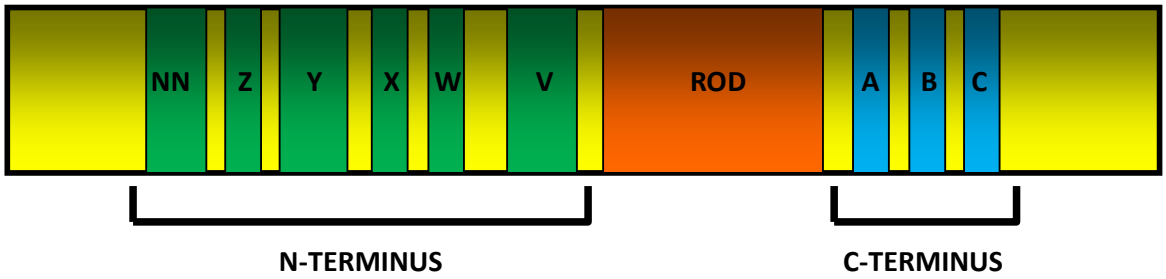


Figure 1.7: Schematic representation of the domains of desmoplakin and their subdivisions (from Lai Cheong *et al*. 2005)

The desmoplakin protein is known to form a homodimer (Lai Cheong *et al*. 2005). The difference in length of the DSPI and DSPII proteins is due to DSPII having a shortened rod domain compared to DSPI (290 vs 889 aa). DSPI is 322 kD in size while DSPII is 259 kD in size (Lai Cheong *et al*. 2005).

Both the DSPI and DSPII isoforms are expressed abundantly in the skin, but while DSPI is expressed abundantly throughout the heart, DSPII is only expressed in certain areas of the heart (left atrium and ventricle, interventricular septum, left auricular appendage and apex of the heart and at much lower levels than DSPI) (Uzumcu *et al*. 2006). Desmoplakin is also found outside desmosomes, in cellular junctions between endothelial cells which line capillaries (Gallicano *et al*. 1998) and is important for the formation of blood vessels *in vitro* (Zhou *et al*. 2004). Desmoplakin is important in the stabilisation of desmosomes and the establishment of tissue development, as demonstrated by the fact that *Dsp* null mice die in embryogenesis (Gallicano *et al*. 1998, Lai Cheong *et al* 2005).

1.5.3. *DSP* mutations

To date, a number of mutations have been described in the *DSP* gene in patients with ARVC (Basso *et al*. 2006, Bauce *et al* 2005, Rampazzo *et al*. 2002, Yang *et al*. 2006) and syndromic DCM associated with woolly hair and palmar plantar keratosis (Norgett *et al*. 2000). These mutations include splice site mutations, nonsense mutations, missense mutations, insertions, deletions and duplications (van der Zwaag *et al*. 2009). An ARVC database has been established which lists all mutations that have been found in patients with ARVC to date (http://www.arvcdatabase.info/) (van der Zwaag *et al*. 2009). This list of mutations includes published and unpublished mutations. Most disease-causing mutations have been identified in exons 23 and 24; exon 23 encodes part of the rod domain, while exon 24 encodes part of the rod domain and the entire C-terminus. A detailed list of the pathogenic *DSP* mutations which have been previously discovered follows in Table 1.6 (from: http://www.arvcdatabase.info/).

Desmoplakin has also been implicated in a number of other diseases including lethal acantholytic epidermolysis bullosa (Jonkman *et al*. 2005), keratosis palmoplantaris striata (Armstrong *et al*. 1999) and skin fragility-woolly hair syndrome (Whittock *et al*. 2002).

Table 1.6: Summary of pathogenic *DSP* mutations associated with ARVC and Carvajal syndrome

EXON	MUTATION	TYPE	DOMAIN	CONDITION	REFERENCE
1	V30M (c.88G>A)	Missense	N-terminus	ARVC	Yang *et al*. 2006
2	Q90R (c.269A>G)	Missense	N-terminus	ARVC	Yang *et al*. 2006
Intronic	c.423-1 G>A	Splice site	N-terminus	ARVC	Bauce *et al*. 2005
Intronic	c.542+5 G>A	Splice site	N-terminus	ARVC	Basso *et al*. 2006
5	W207X (c.620G>A)	Nonsense	N-terminus	ARVC	Xu *et al*. 2010
5	D230N (c.688G>A)	Missense	N-terminus	ARVC	Xu *et al*. 2010
5	W233X (c.699G>A)	Nonsense	N-terminus	ARVC	Yang *et al*. 2006
7	R270X (c.808C>T)	Nonsense	N-terminus	ARVC	Unpublished
7	Q273X (c.817C>T)	Nonsense	N-terminus	ARVC	Unpublished
7	E274fsX288 (c.818_819insA)	Insertion; frameshift	N-terminus	ARVC	Unpublished
7	S299R (c.897C>G)	Missense	N-terminus	ARVC	Rampazzo *et al*. 2002
7	c.939+1G>A	Splice site	N-terminus	LDAC	Sen-Chowdhry *et al*. 2008
9	A354fsX368 (c.1060_1061delCT)	Deletion; frameshift	N-terminus	ARVC	Unpublished
9	N375I (c.1124A>T)	Missense	N-terminus	ARVC	De Bortoli *et al*. 2010
11	S442F (c.1325C>T)	Missense	N-terminus	LDAC	Sen-Chowdhry *et al*. 2008
11	I445V (c.1333A>G)	Missense	N-terminus	ARVC	den Haan *et al*. 2009
11	N458Y (c.1372A>T)	Missense	N-terminus	ARVC	Unpublished
11	K470E (c.1408A>G)	Missense	N-terminus	ARVC	Basso *et al*. 2006
12	S507F (c.1520C>T)	Missense	N-terminus	LDAC	Sen-Chowdhry *et al*. 2008
13	I553T (c1598T>C)	Missense	N-terminus	ARVC	Xu *et al*. 2010
13	A566T (c.1696G>A)	Missense	N-terminus	ARVC	Basso *et al*. 2006
14	T586fsX594 (c.1755_1756insA)	Insertion; frameshift	N-terminus	ARVC	Norman *et al*. 2005
14	I608ins10 (c.1823_1824ins30)	Insertion	N-terminus	Carvajal Syndrome	Norgett *et al*. 2006
15	Q673X (c.2017C>T)	Nonsense	N-terminus	ARVC	Asimaki *et al*; 2009
18	L839fsX862 (c.2516_2519delACTC)	Deletion; frameshift	N-terminus	ARVC	Tanaka *et al*. 2008

18	N871fsX887 (c.2611_2614delGATA)	Deletion; frameshift	N-terminus	ARVC	Unpublished
21	S987P (c.2959T>C)	Missense	N-terminus	ARVC	Dalal *et al*. 2009
22	Arg1015SerfsX3 (c.3045delG)	Deletion; frameshift	N-terminus	ARVC	Sen-Chowdhry *et al*. 2008
23	Glu1068ValfsX19 (c.3203_3204delAG)	Deletion; frameshift	Rod domain	ARVC	Unpublished
23	R1113X (c.3337C>T)	Nonsense	Rod domain	LDAC	Sen-Chowdhry *et al*. 2008
23	R1255K (c.3764G>A)	Missense	Rod domain	ARVC	Bauce *et al*. 2005
23	R1267X (c.3799C>T)	Nonsense	Rod domain	Carvajal Syndrome	Uzumcu *et al* 2006
23	Asn1324ArgfsX24 (c.3971_3974delATAA)	Deletion; frameshift	Rod domain	ARVC	Tanaka *et al*. 2008
23	E1337del (c.4010_4012delAGG)	Deletion	Rod domain	ARVC	Unpublished
23	Q1446X (c.4336C>T)	Nonsense	Rod domain	ARVC	Asimaki *et al*. 2009
23	L1486X (c.4457T>A)	Nonsense	Rod domain	ARVC	Unpublished
23	K1583R (c.4748A>G)	Missense	Rod domain	ARVC	Basso *et al*. 2006
23	L1654P (c.4961T>C)	Missense	Rod domain	ARVC	Xu *et al*. 2010
23	R1775I (c.5324G>T)	Missense	Rod domain	ARVC	Bauce *et al*. 2005
24	Q1810X (c.5428C>T)	Nonsense	Rod domain	ARVC	Unpublished
24	R1934X (c.5800C>T)	Nonsense	Rod domain	ARVC	Fressart *et al*. 2010
24	Ser2000TrpfsX33 (c.5999_6000delinsG)	Ins/del; frameshift	C-terminus	ARVC	Fressart *et al*. 2010
24	S2108X (c.6323C>A)	Nonsense	C-terminus	ARVC	Unpublished
24	R2284X (c.6850C>T)	Nonsense	C-terminus	ARVC	Fressart *et al*. 2010
24	G2375R (c.7123G>C)	Missense	C-terminus	ARVC	Alcalai *et al* 2003
24	c.7503_7508delTGAATG	Deletion	C-terminus	ARVC	Barahona-Dussault 2009
24	R2541K (c.7622G>A)	Missense	C-terminus	ARVC	Bauce *et al*. 2010
24	Lys2542Serfs19 (c.7622delG)	Deletion; frameshift	C-terminus	ARVC	Unpublished
24	T2595I (c.7784C>T)	Missense	C-terminus	ARVC	Unpublished
24	c.7901delG	Deletion	C-terminus	Carvajal Syndrome	Norgett *et al*. 2000
24	R2639Q (c.7916G>A)	Missense	C-terminus	ARVC	Yu *et al*. 2008
24	Q2667X (c.7999C>T)	Nonsense	C-terminus	ARVC	Fressart *et al*. 2010
24	R2834H (c.8501G>A)	Missense	C-terminus	ARVC	Yang *et al*. 2006

1.6. MECHANISMS OF DISEASE

Several mechanisms have been proposed to explain the link between desmosomal gene mutations and ARVC. One mechanism is the degeneration-inflammation model of desmosomal dysfunction. In this model, desmosomal mutations that result in desmosomal dysfunction could impair cell-cell adhesion or intermediate filament function or both, depending on the effect of the mutation on protein structure and function (Sen-Chowdhry *et al*. 2010). The tissue would then have reduced ability to withstand mechanical stress, which would lead to cellular detachment and necrosis. This may be accompanied by inflammation. When the regenerative capacity of the myocardium is exceeded, the myocytes are instead replaced with fibrous and/or fatty tissue (Sen-Chowdhry *et al*. 2010, Awad *et al*. 2008).

Another disease model is the transdifferentiation model. This model involves the Wnt/β-catenin signalling pathway which has been implicated in the regulation of cell fate, proliferation and apoptosis. Usually, the Wnt/β-catenin signalling pathway promotes gene expression which causes the cell to differentiate into a cardiomyocyte and suppresses the expression of genes that promote adipogenesis and fibrogenesis. Plakoglobin, also known as γ-catenin, is structurally and functionally similar to β-catenin. In the event of a mutation that affects the localisation of plakoglobin (e.g., a mutation that suppresses the expression of desmoplakin), plakoglobin translocates to the nucleus where it can function in a similar manner to β-catenin and cause upregulation of adipogenic and fibrogenic genes (Sen-Chowdhry *et al*. 2010, Awad *et al*. 2008).

Recent studies have shown evidence of gap junction remodelling and aberrant calcium homeostasis in the presence of desmosomal gene mutations (Awad *et al*. 2008). It is not known which of these mechanisms are responsible for the phenotype of ARVC, or whether there is interplay between these mechanisms (Sen-Chowdhry *et al*. 2010).

In Cape Town, a unique opportunity exists for the application of powerful molecular techniques for studies that could improve the understanding of the pathogenesis of ARVC and DCM. Groote Schuur Hospital in Cape Town has the only cardiac electrophysiology facility in the public sector in Africa and has been selected as the coordinating centre for the national registry of patients with ARVC in South Africa. The combination of unique patient resources and sophisticated molecular genetics laboratory at UCT makes it possible to propose a detailed study of the molecular genetics and relationships between genetic factors and the outcome in ARVC and DCM. We thus propose to extend the knowledge about the cardiomyopathy in the South African patients by screening ARVC and familial and idiopathic DCM patients for mutations in the *DSP* gene.

1.7. HYPOTHESIS

Mutations in desmosomal genes may be associated with a phenotype that is characteristic of both ARVC and DCM (Kolar *et al*. 2008), as in the case of *DSP* gene mutations which have been associated with an ARVC or DCM phenotype with palmoplantar keratoderma (Lai Cheong *et al*. 2005, Norgett *et al*. 2000). The *DSP* gene is one of the important genetic causes of ARVC, accounting for 1-20% of mutations which are responsible for ARVC (Marcus *et al.* 2007, den Haan *et al* 2009, Fowler *et al*. 2010, Fressart *et al*. 2010). We hypothesise that mutations in *DSP* may be causative of ARVC and familial and idiopathic DCM in our cohort of South African patients.

1.8. AIMS

The aims of this study were to determine the prevalence of *DSP* mutations in a cohort of South African patients with ARVC, and to determine whether these mutations could be causative in familial and idiopathic DCM. By screening large numbers of affected individuals, we aimed to provide a definitive perspective on the contribution of *DSP* mutations to the aetiology of ARVC and DCM in South Africa.

1.9. OBJECTIVES

- To screen for DNA mutations within the *DSP* gene in probands with ARVC, familial and idiopathic DCM
- To determine the prevalence of novel *DSP* mutations in controls drawn from the South African population
- To screen the family members of the probands with novel *DSP* mutation that are not observed in the general population
- To use *in silico* methods of predicting whether mutations that are not polymorphisms could be possibly disease-causing

2. MATERIALS AND METHODS

2.1. STUDY DESIGN

This was a study of cases with ARVC, familial DCM and idiopathic DCM. The ARVC cases were ascertained from the South African national registry for patients with ARVC which was established in 2003 (the ARVC Registry of South Africa) (Watkins *et al*. 2009). The ARVC patients were referred to the ARVC Registry of South Africa by physicians from the major hospitals in Cape Town, Durban, Johannesburg, and Port Elizabeth. The cases of familial and idiopathic DCM were also referred to the Cardiomyopathy Clinic at Groote Schuur Hospital by physicians from all over the country.

This project is part of ongoing studies of the genetics of cardiomyopathy that have been approved by the Research Ethics Committee of the University of Cape Town (REC REF 197/1996 and 016/2002 (Khogali *et al*. 2001; Matolweni *et al*. 2006).

2.2. PATIENT SELECTION

ARVC cases

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 enrolment in the registry.

DCM cases

The patients with DCM were referred to the Cardiomyopathy Clinic in the Cardiac Clinic of Groote Schuur Hospital for evaluation of cause and advice on management by Professor Bongani Mayosi.

2.2.1. Diagnostic evaluation

A standardised case report form is completed for all participants and available family members with ARVC, DCM and other cardiomyopathies. Information on presenting symptoms, family history, electrocardiographic findings, cardiac imaging studies, histology, and outcome is collected. Participants are followed up annually to identify recurrent symptoms, hospitalisation, use of medications and implanted devices and vital status. Informed consent was obtained from all participants (Appendix 1).

2.2.1.1. Inclusion criteria

ARVC cases

A diagnostic panel consisting of a group of cardiologists determined, by consensus, whether or not the referred cases met the diagnostic criteria for ARVC 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 *et al*. 1994). The participants were classified as having definite ARVC (if Task Force criteria were met), probable or possible ARVC (if some criteria were met and no

alternative diagnosis was found), or not having ARVC (if there was no evidence of ARVC and/or an alternative diagnosis was present). The diagnosis of ARVC in first-degree relatives of affected individuals was made based on the modified criteria of Hamid and colleagues (Hamid *et al.* 2002).

DCM cases

The diagnosis of DCM was made on the basis of the definition of the European Society of Cardiology (Elliott *et al*. 2008).

2.2.1.2. Exclusion criteria

Patients with secondary causes of cardiomyopathy or left ventricular dysfunction (such as hypertension, valvular heart disease, diabetes, coronary artery disease, or sepsis) were excluded from the study.

2.3. CONTROL SELECTION

Controls were chosen from a South African population comprising anonymous blood donors (self-reported ethnicity for these individuals was: Cape Mixed Ancestry, Black African, Indian and Caucasian). Informed consent was obtained from all participants.

2.4. DNA ISOLATION AND STORAGE

The DNA was extracted from 5-10ml blood using the PureGene^{TM} DNA Purification System (Gentra systems) according to the instructions of the manufacturer. Written informed consent was obtained from all individuals enrolled in the study. Extracted DNA samples were stored in 1.5ml Eppendorf tubes at -80°C for longterm storage. Individuals were given codes to anonymise the samples and ensure the confidentiality of the patients.

2.5. MUTATION SCREENING OF *DSP*

High Resolution Melt (HRM) analysis with the RotorGene 6000 (Corbett Life Sciences) was used for the mutation screening of *DSP*. The predicted promoter region and the 5' and 3' untranslated regions of this gene as well as all the exons were amplified by polymerase chain reaction (PCR). All amplicons showing changes with HRM were sequenced to determine the nature of the change.

2.5.1. High Resolution Melt (HRM) Analysis

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

2.5.1.1. The principle of HRM Analysis

HRM is used to characterise double stranded DNA according to its dissociation behaviour. This is accomplished by monitoring the transition from double stranded DNA to single stranded DNA with increasing temperature. It is used for the detection of mutations, polymorphisms and epigenetic differences between DNA samples. When compared to other methods for detecting these DNA alterations, HRM analysis has been proven to be more cost-effective, rapid, straightforward and powerful.

HRM is the process of warming of the amplicon of DNA from ~50°C to ~95°C, or an alternative temperature range of choice. Before HRM analysis, the target sequence must be amplified to a high copy number. This is achieved by means of a DNA amplification reaction such as PCR in the presence of an intercalating dye. This dye does not interact with single stranded DNA and intercalates only with double stranded DNA, and when bound, fluoresces. This shift in fluorescence can be used to measure DNA concentration in a pre-HRM amplification, and can also be used to monitor the thermally-induced DNA dissociation by HRM.

At the beginning of HRM there is a high level of fluorescence, due to the presence of billions of copies of the amplicon. At some point during HRM, the melting temperature of the two strands is reached and the DNA strands melt apart (Figure 2.1). As these strands melt apart, the amount of double stranded DNA in the solution decreases. Consequently, the fluorescence of the solution decreases. The RotorGene 6000 measures the fluorescence and plots the data as a melting curve that reflects the change in the fluorescence of the solution with the increasing temperature, as shown in Figure 2.2. The midpoint of the melting curve, where the rate of change in fluorescence is the greatest, is known as the melting temperature (T_m) of the particular DNA fragment under analysis. The melting temperature and melting behaviour of a DNA fragment is dependent on the DNA sequence, and thus sequence differences due to mutations would change the melting temperature and be reflected in the melting curve.

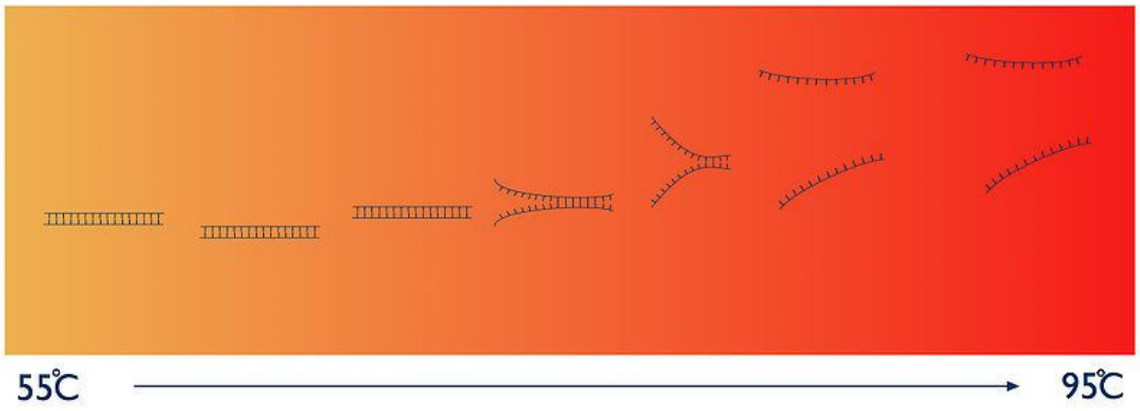


Figure 2.8: Separation of DNA strands with amplicon heating (Electronic Resource 5)

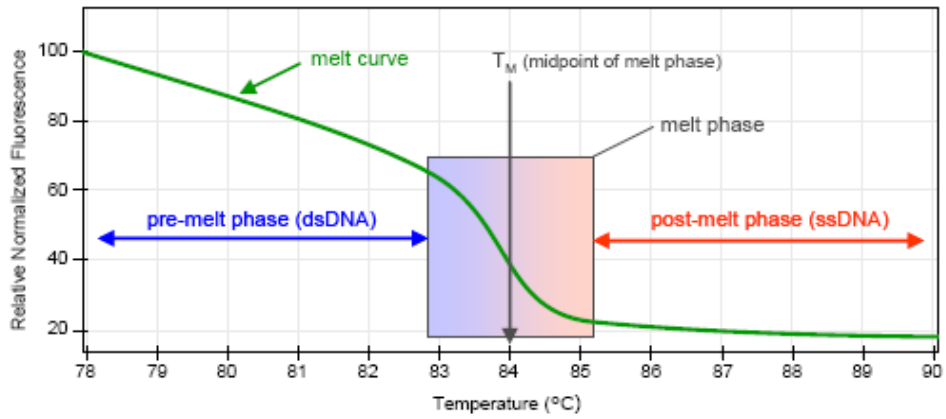


Figure 2.9: Melt Curve for High Resolution Melt Analysis (HRM Assay Design and Analysis CorProtocol™ 6000-1-July06)

However, as organisms possess two copies and thus two alleles of each gene, there are three genotype possibilities: firstly, that neither allele possesses a mutation (also known as wildtype), secondly, that one of the alleles possesses a mutation (known as a mutation heterozygote) and thirdly that both alleles possess a mutation (known as a mutation homozygote). These different cases are clearly distinguishable on the melt curve, as shown in Figure 2.3. In this way, homozygous and heterozygous mutations can be clearly distinguished from control samples by means of this technique.

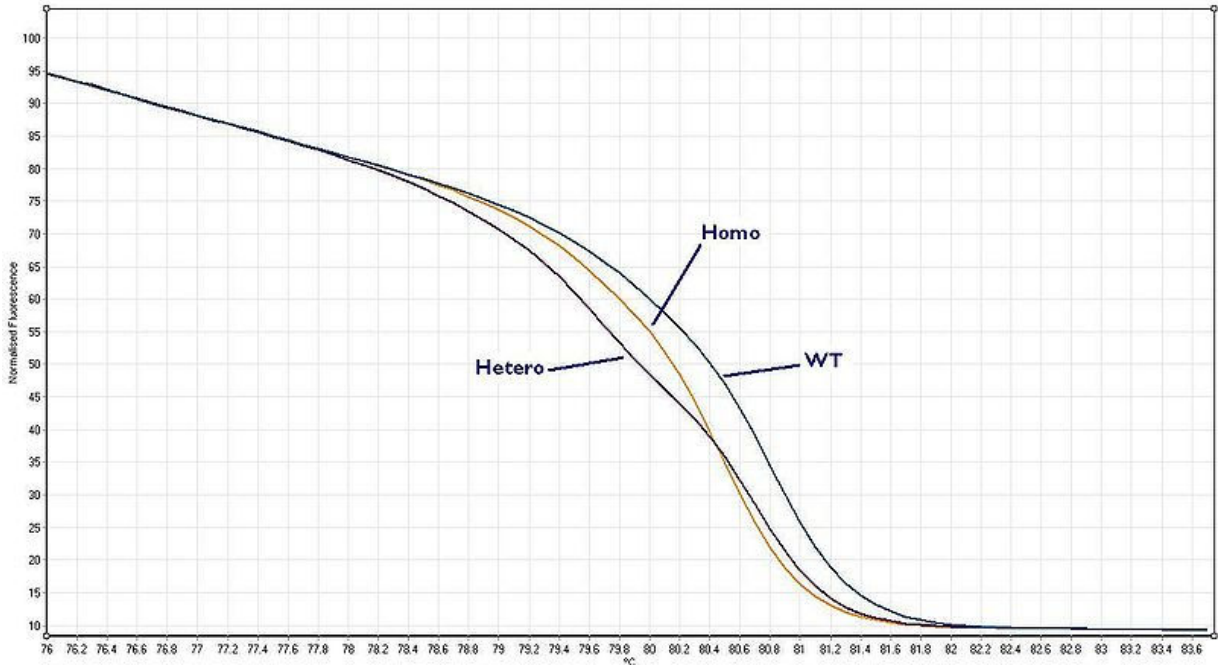


Figure 2.10: Melt curve indicating difference between homozygous and heterozygous samples (Electronic Resource 5)

In this study, probands with ARVC and DCM were screened for mutations by comparing them to control samples known to be negative for the disease.

2.5.2. Polymerase Chain Reaction (PCR)

DNA segments of interest are amplified by means of PCR prior to HRM in the presence of an intercalating dye.

2.5.2.1. The principle of PCR

PCR is a technique that allows the amplification of specific regions of DNA. Two complementary oligonucleotide sequences which flank the target DNA segment, known as primers, direct the specific amplification of the desired DNA fragment, making use of *Taq* polymerase and deoxynucleotide triphosphates (dNTPs), under specific temperature cycling conditions. The concentration of target DNA in the solution is essentially doubled with each cycle and ultimately generates multiple target sequence copies which are used in downstream applications (Garrett *et al.* 2005).

In this project, sample DNA was amplified using real-time PCR, which allows the amplification and simultaneous quantification of a DNA sequence. This PCR is done in the presence of an intercalating dye that does not interact with single stranded DNA and intercalates with only double stranded DNA, and fluoresces when bound. The dye used for the pre-HRM PCR in this project is the EvaGreen dye. The increase in the amount of DNA in the solution as the PCR progresses causes a shift in the fluorescence of the solution, which can be used to measure DNA concentration at different times during the reaction.

2.5.2.2. Primer design

Primers were designed to amplify *DSP,* the 5' and 3' untranslated regions (UTRs) as well as the promoter region as predicted by Gene2Promoter (Electronic Resource 6). Primer sequences were chosen manually by (1) referring to the annotated DSP gene sequence, and (2) analysing the chosen sequences using IDT OligoAnanlyzer (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx) and NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) web-based tools in order to determine the optimal primer sequences for the amplification of the gene.

The primers were required to have: (1) good sequence diversity with minimal sequence repeats and secondary structure, (2) a length of between 18 and 25 base pairs, (3) a melting temperature of between 50°C and 60°C, (4) a GC content of between 45% and 65%, and (5) a G or C clamp at the 3' end.

The amplicon sequences were also restricted to a 250-300bp length due to the fact that with HRM, increased amplicon size leads to decreased resolution in amplicon screening and mutation detection. Using these criteria, fifty sets of primers were designed for the combined PCR and HRM amplification and analysis of the 24 *DSP* exons, while 8 sets of primers were designed for the amplification of the predicted promoter region, and 5' and 3' UTRs (Tables 2.1 and 2.2).

Table 2.7: Details of primers for the PCR and HRM of the 24 *DSP* exons

EXON	PRIMER	SEQUENCE	SIZE (bp)	Tm (°C)	GC CONTENT (%)	PRODUCT SIZE (bp)
1	F	CTGAGCCGCTCTCCCGATTG	20	60.4	65.0	280
	R	AGTCTCCCAGTCGTCCAG	18	56.0	61.1	
2	F	GGGTAAAGGGTCTCACAGG	19	54.8	57.9	288
	R	CTAACCAGAGGACAGTCGG	19	54.4	57.9	
3	F	GTCATGTTTACAGTGGAAGC	20	51.4	45.0	336
	R	AGGAAGCCCTGGGTCATG	18	57.2	61.1	
4	F	CGGGTTTTCATAGGCTGTTTTC	22	54.5	45.5	286
	R	CCATCGTGCTGGAACGTCTG	20	58.5	60.0	
5	F	GTGATGGTGTGCGTAAGTG	19	53.8	52.6	285
	R	AGAAACTCCCTGGGATTGAG	20	54.0	50.0	
6	F	CTATGGAGGGATCTGAGGC	19	54.1	57.2	250
	R	CCCGCCAGGATGTGCTTG	18	59.4	66.7	
7	F	GCAGAGAACACCAGTCAC	18	53.1	55.6	303
	R	CCAATCATTCCCCGAGAC	18	52.8	55.6	
8	F	CTGTGGGGGTGATGGAAG	18	55.4	61.1	273
	R	GTTGGCACGGCTTTAAGTC	19	54.4	52.6	
9	F	GCAGGCTCACCTATCTCTTG	20	55.1	55.0	322
	R	GCAAATCTATGCTACTCTGCACTC	24	55.7	45.8	
10	F	GCCACCTGTCAAAGAGATG	19	52.6	53.4	316
	R	CAGAGCCAACTGCATCCAAG	20	56.4	55.0	
11	F	ATCTCCTCTAAAACTCACAGGG	22	53.7	45.5	329
	R	CAAAGACATTGACTGTGGTG	20	51.6	45.0	
12	F	CTTATGAATAAAGCCAAACCCTG	23	52.1	39.1	273
	R	GGTATAAACACCTGCCCTC	19	52.6	52.6	
13	F	AGTGGTGTGAGCGTGTCCAG	20	59.7	60.0	327
	R	TTGACGAGACACAAAAGCAC	20	53.5	45.0	
14	F	GATTGCATTGTGGGGCAGTC	20	57.0	55.0	382
	R	TGCACGAAGAACACTGGC	18	55.9	55.6	
15	F1	AGCCAGTGTTCTTCGTGC	18	55.6	55.6	227
	R1	CTATCTGCCTGCGAATCTTC	20	53.1	50.0	
	F2	GAAACATGGATGCTGATGG	19	51.4	47.4	318
	R2	CCAGCCTGATAACCGTCTAC	20	54.8	55.0	
16	F1	CCATGGAAGTTGACTGATGTG	21	52.0	48.0	305
	R1	GCCATCTGTAACACCATTG	19	51.3	47.4	
	F2	GATATGCTTGCCAACTTCAG	20	51.6	45.0	256
	R2	CATGCTCTCCAAACCACAG	19	53.7	52.6	

17	F	AGCTGTGAGAGGCAAATCTTC	21	55.1	47.6	300
	R	ATGAAAACCCAGTGAGGC	18	52.6	50.0	
18	F	TGGTGACTTGTAGTGTAGC	19	51.5	47.4	317
	R	GAACAACCAAAAGGGGAGAAG	21	53.9	47.6	
19	F	CTGGGTGATTCTATGTTAC	19	47.2	42.1	318
	R	GCACTGATACAAAAACCAG	19	49.1	42.1	
20	F	TGCTCATCTCCTAAGCTG	18	51.2	50.0	346
	R	TTGCTATCCTCAGGAAGTG	19	51.4	47.4	
21	F	CTGTGGAACTGTAGCTGTTAG	21	52.8	47.6	311
	R	TGGGATGAAATCAGCAATGG	20	53.1	45.0	
22	F	GGCCCTTTGTTTTGCTGTAC	20	52.0	48.0	376
	R	CAGTTTCACCAGGATGTGTC	20	52.0	50.0	
23	F1	TGCACATTGGTCTGGGAG	18	55.0	55.6	356
	R1	CATAAGTCAGTCGGGTGATC	20	52.5	50.0	
	F2	CTGAAGAGACAGGCTGAG	18	52.0	55.6	348
	R2	GTTTCTTACCTTTGCCAGCTC	21	54.1	47.6	
	F3	TGAAAGGTTGAGGGTTCTAC	20	51.8	45.0	343
	R3	CATTATCTCAGAGCCACAGGC	21	55.4	52.4	
	F4	AGCATCTTGCAGGCCACTG	19	58.4	57.9	261
	R4	TTCATATTCCCAGCGGCG	18	55.3	55.6	
	F5	CACAAGCAGTCCCTGGAG	18	55.6	61.1	326
	R5	GTTAGAGTGTTCTTCAGCC	19	50.3	47.4	
	F6	AGTGGCTACCGGGCTCAG	18	59.7	66.7	343
	R6	CATCTTCCAGGCATTTCCG	19	53.8	52.6	
	F7	ATCTCTTGATGATGCTGCC	19	52.4	47.4	346
	R7	CCTCTTCAGCCGATTCAGC	19	55.9	57.9	
	F8	TCCCAGGAGAGGACTGTG	18	55.8	61.1	289
	R8	CATCCAGCACGTCCCTCTG	19	58.0	63.2	
	F9	GCTGAAGGAGCAAGCCATC	19	56.5	57.9	262
	F9	ATCTTCTATCGCCTTCTGG	19	51.2	47.4	
	F10	CCAGGAACAGGAAAGTGTC	19	53.0	52.6	260
	R10	GCTGGCTCCTTAGTTCCAAG	20	55.6	55.0	
	F11	GGAGTACGATGACCTGAG	18	51.5	55.6	278
	R11	GCAGAAAGAGCACAGTTC	18	51.2	50.0	
24	F1	GCTCACAGTGTATCCAGGGAC	21	57.3	57.1	347
	R1	CTGTTTCTCACACTCCAGGC	20	55.7	55.0	
	F2	GGAGGATGAGCTGAATCG	18	52.5	55.6	301
	R2	GCGTTCTGTTTCTAACTGTGAC	22	54.0	45.5	
	F3	CGAAAGACTCCAAGCAGAG	19	53.0	52.6	346
	R3	ACCCCGAAGGAATGGCTG	18	57.7	61.1	
	F4	CGACAAAACAACCTTGGAC	19	51.9	47.4	300

	R4	ATCTGCTGACGGTCATCG	18	54.6	55.6	
	F5	CATCGGAATGAGAAGCTGAC	20	53.5	50.0	264
	R5	CAAGGCGACATCTTTTGGC	19	54.8	52.6	
	F6	TCAGGGGGTGTAGTAGAC	18	52.5	55.6	318
	R6	CAGTTCATTGACAGTGGACGG	21	55.9	52.4	
	F7	TCCAAGGAATCAGACAACC	19	51.9	47.4	344
	R7	CTCAATGCCCACCAGACC	18	56.0	61.1	
	F8	AAGCCCAAGCAGCTACTGG	19	57.8	57.9	293
	R8	GCTATGTCAACTGGTAAACG	20	51.2	45.0	
	F9	CTTATTAGAAGCACAGATCGC	21	51.0	42.9	289
	R9	CCTGAGGGTATTCTTTTGTGATG	23	53.5	43.5	
	F10	CAGGGCTCTGTCTTCTGC	18	55.6	61.1	314
	R10	GAACTTCCTGTCAACAAGGC	20	53.9	50.0	
	F11	GGTGGTCCTGGTAGATAG	18	50.9	55.6	318
	R11	ATGGCTGCAATGGGGCTC	18	58.8	61.1	
	F12	CACCATATCCAGCGTCAG	18	52.8	55.6	321
	R12	ACACCCTCGAAGCCTATGAAGG	22	58.7	54.5	
	F13	CTGTCACTTCAGGACGCAGT	20	57.1	55.0	326
	R13	GGGCAGGTCAGGATTTTGG	19	56.3	57.9	
	F14	GGAAGGGGTTCATAGATGGC	20	55.3	55.0	217
	R14	CCCGGAGCCGAAGACATG	18	58.5	66.7	
	F15	TATCACTGGGCTGCGCCTTC	20	60.0	60.0	336
	R15	TGTCCTATCACTCTACTGC	19	50.5	47.4	

Table 2.8 : Details of primers for the PCR and HRM of the *DSP* predicted promoter region and the 5’ and 3’ UTR

REGION	PRIMER	SEQUENCE	SIZE (bp)	Tm (°C)	GC CONTENT (%)	PRODUCT SIZE (bp)
PROMOTER	F1	AGGAAGTCGTGGTTTTCTG	19	52.4	47.4	301
	R1	TCGGGCTGGGAATGCGTG	18	61.1	66.7	
	F2	TTCCAGCTATTTCCGCCC	18	55.1	55.6	319
	R2	CACCTGGCCGGTTTCTTC	18	56.4	61.1	
	F3	CACCCTGGGAAGAAACCG	18	55.9	61.1	273
	R3	CAAGGATAGGCGGACGGC	18	58.5	66.7	
5'UTR	F	TCCGAGCCACAGCTTTCC	18	57.8	61.1	277
	R	TCATGCGGCCCAGAGTGTTG	20	60.4	60	
3' UTR	F1	GAGTGGTTGCTATACCTTG	19	50.1	47.4	216
	R1	CACTTCAGACGCACTGCATC	20	56.4	55.0	
	F2	AGGCTGTTCTGGCTTTTTATC	21	53.2	42.9	373

3'UTR	R2	GAGTGTGGATTGTATAACCC	20	50.2	45.0	
	F3	AGTCATTCTGCTTCTCATC	19	49.2	42.1	280
	R3	CATAAAGACACCTTAAAGTACATC	24	49.4	33.3	
	F4	TGCATGACAGCGGCAATC	18	56.4	55.6	337
	R4	CATCGGGTACATGACAGG	18	52.4	55.6	

2.5.3. Combined PCR and HRM Protocol

The reagents used for the combined real-time PCR and HRM of the *DSP* amplicons are shown in Table 2.3 and the reaction conditions are shown in Table 2.4. SensiMix HRM Kits (Quantace) were used for HRM analysis in our study.

Table 2.9: Reagents and concentrations in PCR and HRM of the 24 *DSP* exons, the promoter region and the 5' and 3' UTRs

REAGENT (STOCK CONCENTRATION)	FINAL CONCENTRATION / VOLUME IN SOLUTION
2X Reaction Mix (SensiMix Kit - Quantace)	1X
Forward Primer (20 µM)	0.2 µM
Reverse Primer (20 µM)	0.2 µM
DNA	100 ng
EvaGreen dye (SensiMix Kit - Quantace)	1X
FINAL REACTION VOLUME	25 µl

Table 2.10: Optimised temperature cycling conditions for the PCR and HRM of the 24 *DSP* exons, the promoter region and the 5' and 3' UTRs

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

2.5.4. Sequence Analysis

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

2.5.4.1. The principle of DNA sequence analysis

DNA sequencing allows the determination of the order of base pairs in a specific DNA molecule. The procedure requires the DNA of interest to be mixed in solution with DNA polymerase, deoxynucleotide triphosphates (dNTPs), four labelled dideoxynucleotide triphosphates (ddNTPs) and a sequence-specific oligonucleotide primer. The oligonucleotide primer directs the amplification of the DNA strand of interest, using dNTPs, ddNTPs and *Taq* polymerase. ddNTPs are incorporated into the forming DNA strand in a similar manner to dNTPs, but this incorporation causes the termination of the strand. Temperature cycling similar to that used in PCR is used for the sequencing reaction (Hartwell *et al*. 2004)

A complete series of single-stranded sub-fragments which are complementary to parts of the DNA strand of interest is generated, which can be separated by capillary electrophoresis in order to determine the DNA sequence of this strand (Luckey *et al.* 1990).

2.5.4.2. Purification of HRM products

After analysis by HRM, the samples were purified using the *Shrimp Alkaline Phosphatase* and *Exonuclease I* enzymes. The reagents used are shown in Table 2.5 and the reaction conditions are shown in Table 2.6.

Table 2.11: Cleanup protocol for *DSP* HRM products

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

Table 2.12: Cycling conditions for the cleanup reactions of *DSP* HRM products

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

2.5.4.3. Sequencing of DSP exons

The reagents used for the direct sequencing reaction of the *DSP* exons are described in Table 2.7. The temperature cycling conditions used for the sequencing reactions are as described in Table 2.8.

Table 2.13: PCR protocols for the sequencing reactions of *DSP* amplicons

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

Table 2.14: Optimised cycling conditions for the sequencing reactions of *DSP* amplicons

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

2.5.4.4. Capillary electrophoresis

The sequencing products were analysed using capillary electrophoresis at the DNA Sequencing Unit (Department of Genetics, Stellenbosch University), by means of the ABI PRISM® 3130xl Genetic Analyser (Applied Biosystems) or the ABI PRISM® 3730xl Genetic Analyser (Applied Biosystems).

2.5.4.5. Sequence alignment and analysis

The DNA sequences generated in these experiments were aligned with the wildtype sequence obtained from the Ensembl database (http://www.ensembl.org/) using the BioEdit Sequence Alignment Editor (© Tom Hall).

2.6. POPULATION SCREENING

Novel variants found using HRM analysis were confirmed by DNA sequencing. Population screening was conducted to determine the prevalence of these variants in the normal control population. The cohort used for population screening included individuals from four South African ethnic groups: Caucasian, Mixed Ancestry, Black African and Indian.

For the ARVC novel variants, population screening was done by (1) restriction digests, using various restriction enzymes, or (2) by means of HRM.

For the DCM novel variants, population screening was conducted using HRM analysis as described in section 2.5.

2.6.1. Restriction Enzyme Digests

Restriction enzyme digests were performed for population screening of novel variants found in the ARVC cohort.

2.6.1.1. The principle of restriction enzyme digests

Restriction enzymes are proteins made by bacteria that recognise specific, short nucleotide sequences and cut DNA at those sites. Restriction enzymes are divided into three types: I, II and III. Types I and III are ATP-dependent while type II is not. Type II restriction enzymes cut within or near specific nucleotide sequences that they recognise.

In this project, type II restriction enzymes were used to determine the presence of a particular DNA sequence variant. They were used to determine the presence of single nucleotide changes e.g. if the change is present, and there is an A nucleotide base at the position of interest, the restriction enzyme would recognise that sequence and cleave the DNA at that site or a site near to it. In the control sample, where there is a G base at that position, the restriction enzyme will not recognise any sequence and will thus not cleave the DNA at that position. Hence, DNA sequence changes can be distinguished using this method.

The DNA sequence of interest is first amplified using a method such as PCR, and the PCR product is then restriction digested by the enzyme selected for the experiment. If a restriction enzyme is not available at this site, a primer is designed that specifically incorporates a restriction site. After the restriction digest reaction, the samples are run on an agarose gel to separate the fragments by size, and hence determine whether the DNA sequence change is present or not, based on the presence or absence of DNA cleavage by the restriction enzymes.

2.6.1.2. Restriction enzyme digest protocol

The PCR protocol is described in Tables 2.9 – 2.11. The amplicons generated during PCR were digested with *VspI* (Promega), *FokI* (Promega), *PvuII* (Fermentas) or *HphI* (New England Biolabs) as per the manufacturer's protocol.

Table 2.15: PCR Amplification protocol for *DSP*

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

Table 2.16: Optimised cycling conditions for the PCR of the *DSP* amplicons to be restriction digested by *VspI*, *FokI* and *PvuII*

CONDITION	TEMPERATURE (TIME)	
Initial Denaturation	94°C - 5 min	
Denaturation Primer Annealing Template Elongation	94°C - 30 seconds * - 45 seconds 72°C - 50 seconds	33 cycles
Final Template Elongation	72°C - 7 min	

* Refer to Table 2.11

Table 2.17: Primer annealing conditions for the PCR of amplicons for restriction digests

ENZYME	ANNEALING TEMPERATURE FOR PCR OF AMPLICON TO BE RESTRICTION DIGESTED
VspI	55^0C
FokI	55^0C
PvuII	55^0C
HphI	60^0C

2.6.1.3. Agarose gel electrophoresis

After the PCR was completed, the products were run on a 1.5% agarose gel at 160V to verify the success of the PCR amplifications. The samples were loaded into the wells of agarose gels with loading buffer (Appendix 2). Gel electrophoresis is used to separate DNA fragments of different sizes in an agarose gel (Appendix 2). The DNA is negatively charged and migrates towards a positive electrode, allowing the separation of the DNA fragments according to size

(Hartwell *et al*. 2004). The gel is submerged in 1X TBE buffer (Appendix 2) and run with a marker alongside the DNA fragments of interest. The marker is a solution containing DNA fragments of known sizes, and is used as a guide in determining the size of the relevant DNA fragments. The marker used in this study was the 100bp DNA Ladder (Promega) (Appendix 3). Restriction digested products were run on a 3% agarose gel (Appendix 2) using the same protocol used to run PCR products on an agarose gel, in order to determine the sizes of fragments resulting from enzyme cleavage.

2.7. MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION

Multiplex ligation-dependent probe amplification (MLPA) is a variation of the polymerase chain reaction that allows the detection of abnormal copy numbers of up to 50 different DNA or RNA sequences. In this study, MLPA was performed on ARVC samples which were screened for *DSP* changes, using the SALSA MLPA Kit P168 ARVC-PKP2.

2.7.1. The principle of MLPA

In MLPA, it is not the target sequences that are amplified but instead the MLPA probes that hybridise to the target sequence. A single pair of primers is used to amplify multiple targets.

The MLPA procedure can be divided into five major steps: (1) DNA denaturation and hybridisation of MLPA probes, (2) Ligation reaction, (3) PCR reaction, (4) Separation of amplification products by capillary electrophoresis, and (5) Data analysis.

In the first step, the sample DNA is denatured and incubated overnight with a mixture of MLPA probes (Figure 2.4). The MLPA probes consist of two separate oligonucleotides each of which contain one of the PCR primer sequences. The probes recognise adjacent target sites on the DNA. One of the probes contains the sequence recognised by the forward primer and the other contains the sequence recognised by the reverse primer. Only when both of the probes are hybridised can they be ligated into a single probe in the ligation reaction (step 2). Only the ligated probes will be amplified exponentially during the PCR reaction (step 3), and hence the number of the probe ligation products is a measure of the number of target sequences present in the sample. The amplification products are then separated by capillary electrophoresis (step 4). A universal primer pair is used to amplify all the MLPA probes, and each MLPA probe has a specific size (due to the unique stuffer sequence for each probe) which allows the different MLPA products to be separated by capillary electrophoresis.

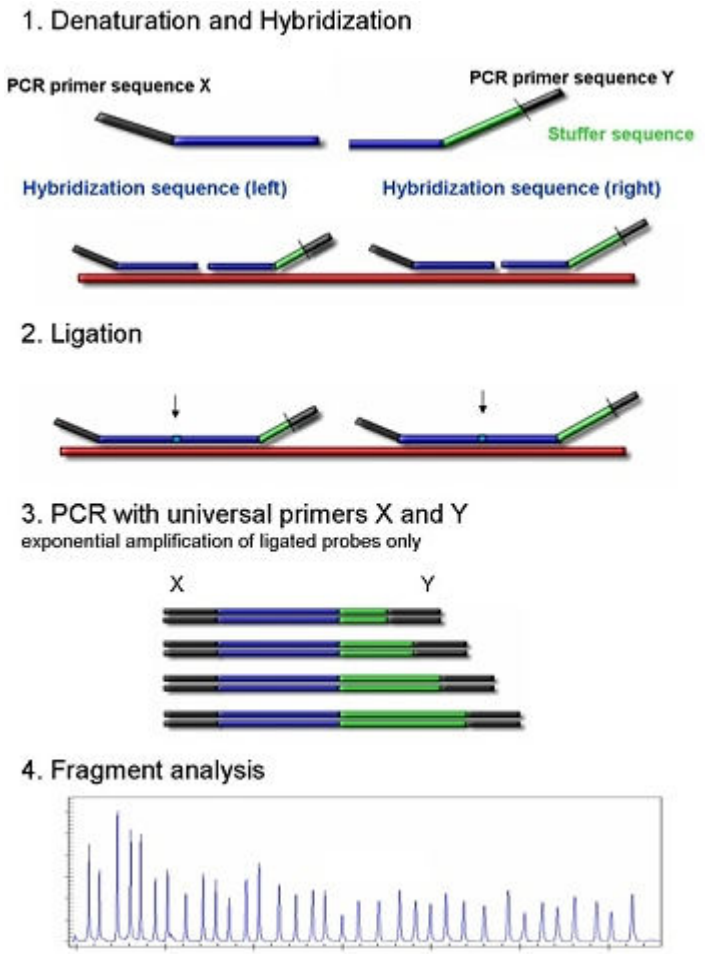


Figure 2.11: MLPA procedure (Info from http://www.mrc-holland.com)

The MLPA kit used in this project was the SALSA MLPA Kit P168 ARVC-PKP2. This kit is designed to detect deletions/duplications of one or more exons of the *PKP2*, *DSP*, *JUP*, *TGFβ3* and *RyR2* genes. MLPA can however also detect point mutations very close to the probe ligation site. The probemix in this kit contains 18 probes for *PKP2,* 7 probes for *DSP* and 3 probes each for *JUP*, *TGFβ3* and *RyR2* (Appendix 4).

2.7.2. MLPA Protocol

The MLPA reactions were carried out according to the manufacturer's instructions (MRC Holland). The thermocycler used was a LabNet MultiGene™ thermocycler.

1. DNA denaturation and hybridisation of the MLPA probes:

The DNA sample is diluted in dH_2O. 3 µl of DNA (100 ng/ µl) and 2 µl of dH_2O are added to a PCR tube, and this mixture is then heated to 98°C for 5 minutes, then cooled to 25°C before opening the thermocycler. Thereafter, 1.5 µl SALSA probemix (MRC Holland) and 1.5 µl MLPA buffer

(MRC Holland) are added to the tube, and mixed thoroughly. This is then incubated at 95°C for 1 minute and then left to hybridise overnight at 60°C for 16-20 hours.

2. Ligation reaction

Ligase-65 mix is prepared. This includes 3 µl Ligase-65 buffer A (MRC Holland), 3 µl Ligase-65 buffer B (MRC Holland), 25 µl dH_2O and 1 µl Ligase-65 (MRC Holland). The thermocycler is cooled to 54°C and the Ligase-65 mix is added. This is then incubated at 98°C for 5 minutes.

3. PCR reaction

In new tubes, 4 µl SALSA PCR buffer (MRC Holland) and 4 µl of dH_2O are added. 10 µl of the MLPA ligation reaction product is added to these new PCR tubes and incubated at 60°C. Polymerase mix is prepared and contains 2 µl SALSA PCR-primers (MRC Holland), 2 µl SALSA Enzyme Dilution Buffer (MRC Holland), 5.5 µl dH_2O and 0.5 µl SALSA Polymerase. This polymerase mix is then added to the PCR tube and the PCR reaction is immediately started. The PCR conditions are listed in Table 2.12.

Table 2.18: Cycling conditions for the PCR reaction forming part of the MLPA protocol

CONDITION	TEMPERATURE (TIME)	
Denaturation	95°C - 30 seconds	35 cycles
Primer Annealing	60°C - 30 seconds	
Template Elongation	72°C - 60 seconds	
Template Elongation	72°C - 20 minutes	

4. Running MLPA products on the ABI 3100 Genetic Analyzer

8.6 µl Hi-Di™ Formamide (Applied Biosystems) and 0.4 µl GeneScan™ 500 ROX™ Size Standard (Applied Biosystems) is added to 1 µl MLPA product. This mixture is denatured at 95°C for 5 minutes then snap-frozen on ice and run on the ABI 3100 Genetic Analyzer (Applied Biosystems). The data is then analysed using GeneMapper® version 3.0 (Applied Biosystems) and Coffalyzer version 8 (MRC-Holland) software.

2.8. BIOINFORMATIC ANALYSIS

Samples that were found to have sequence changes not reported in NCBI's dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/) after sequencing analysis were analysed using various bioinformatic tools in order to determine whether they were likely to affect mRNA splicing, secondary structure of the DSP RNA or secondary structure of the DSP protein.

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

Once mRNA is transcribed from DNA, it undergoes specific post-transcriptional modifications that ensure correct gene expression. One of these modifications is mRNA splicing, in which the introns are removed from the mRNA and the exons are joined to form the processed transcript that will be translated into the protein. Correct splicing of the exon and intron requires a 5' splice site at the beginning of the intron (with the sequence GT, known as the splice donor site), a 3' splice site at the end of the intron (with the sequence AG, known as the splice acceptor site) and a branch site. Sequence changes which alter these sites could alter mRNA splicing.

In order to determine whether novel sequence changes caused a change in mRNA splicing, the sequences were analysed using the Berkeley Drosophila Genome Project NNSPLICE 0.9 (http://www.fruitfly.org/seq_tools/splice.html), SplicePort (http://spliceport.cs.umd.edu/SplicingAnalyser2.html) (Dogan *et al*. 2007) and NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) (Hebsgaard *et al*. 1991) web-based tools. These tools allow the prediction of donor and acceptor splice sites in a particular sequence.

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

In addition to the canonical splice sites mentioned above, other *cis*-elements exist in the form of splicing enhancers and silencers (Cartegni *et al*. 2003). These elements are especially important in the event of alternative splicing. Mutations that disrupt these elements could cause skipping of the mutant exon, with dramatic effects on the gene product.

In order to determine whether novel sequence changes could disrupt existing ESEs, sequences were analysed using the ESEFinder web-based tool (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) (Cartegni *et al*. 2003). This tool allows the prediction of ESEs that act as binding sites for Ser/Arg-rich proteins (SR proteins), a family of conserved splicing factors that participate in the splicing pathway.

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

Changes in the DNA sequence, and consequent changes in the transcribed RNA, could affect RNA folding and RNA secondary structure. Sequences were analysed using the Mfold (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) (Zuker 2003) and RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (Hofacker 2003) web-based tools to predict RNA folding. Mfold can be used to predict optimal and suboptimal secondary structures of RNA molecules using energy minimisation criteria developed by the creators. RNAfold performs a similar function, in that it can be used to predict the minimum free energy structure of a single sequence. It can however also be used to calculate the equilibrium base-pairing probabilities of this sequence.

2.8.4. Bioinformatic Analysis of the Effect of Amino Acid Substitutions on Protein Structure

Mutations that lead to a change in the protein sequence could affect the secondary structure of protein. In the case of mutations that affect the protein sequence, the PSIPRED protein structure prediction server (http://bioinf.cs.ucl.ac.uk/psipred/) was used to predict the secondary structure of the protein based on the protein sequence (McGuffin *et al*. 2000). PSIPRED incorporates two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST (Position-Specific Iterated Basic Local Alignment Search Tool) (McGuffin *et al*. 2000).

The Align-GVGD (http://agvgd.iarc.fr/agvgd_input.php) (Tavtigian *et al.* 2006), POLYPHEN (POLYmorphism PHENotyping) (http://genetics.bwh.harvard.edu/pph/) (Ramensky *et al*. 2002) and SIFT (Sorting Intolerant From Tolerant) (http://sift.jcvi.org/) (Ng and Henikoff 2003) bioinformatic tools were used to predict whether non-synonymous mutations had deleterious effects on the protein. The Align-GVGD program uses calculations that are an extension of the Grantham Difference and take into account the polarity, position and volume of amino acids (Mathe *et al*. 2006). It calculates the biochemical difference between a mutant amino acid and the observed variation at that position in order to determine whether or not the variant is deleterious. The POLYPHEN tool predicts whether a mutation is disease-causing based on a set of rules that are applied to the sequence, structural and phylogenetic information characterising the amino acid substitution (Ramensky *et al*. 2002). The SIFT tool calculates the probability of an amino acid being found at a specific position relative to the most frequent amino acid at that position, as well as the effect it would have on the protein (Mathe *et al*. 2006).

3. RESULTS

3.1. MUTATION SCREENING OF *DSP*

The ARVC and DCM probands selected for this study were screened for mutations in the 24 exons of the *DSP* gene, the 5' and 3' untranslated regions (UTRs) as well as the promoter region (as predicted by the Gene2Promoter tool) by means of high resolution melt (HRM) and multiplex ligation-dependent probe amplification (MLPA) analyses. We performed mutation screening in 62 probands with ARVC, 19 index cases with familial DCM and 131 unrelated cases of apparently sporadic DCM (Table 3.1) (a total of 212 individuals).

In the ARVC cohort, 32% were female while 68% were male. In the DCM cohort 37% were female while 63% were male. The average age (±SD) of the probands in the ARVC cohort was 41 ± 16 years, while the average age of the DCM probands was 47 ± 16 years. In the ARVC cohort, the majority of patients were of Caucasian ethnicity, while in the DCM cohort, the majority were of Black African ethnicity. These samples were screened in conjunction with ethnicity matched population control samples from anonymous blood donors from the blood transfusion bank of the Western Cape Province of South Africa.

Table 3.19: Age, Gender and Diagnosis of Study Participants

CHARACTERISTIC	ARVC PATIENTS	DCM PATIENTS
No. of Patients	62 (%)	150 (19 familial, 131 idiopathic) (%)
Gender:		
Male	42 (68)	95 (63)
Female	20 (32)	55 (37)
Average age in yrs (±SD)	41 ±16	47 ± 16
Ethnicity:		
Black African	5 (8)	97 (65)
Caucasian	26 (42)	2 (1)
Mixed Ancestry	12 (19)	17 (11)
Indian	2 (3)	4 (3)
Not recorded	17 (28)	30 (20)

3.1.1. Mutation screening of the ARVC Cohort

ARVC samples were screened using high resolution melt analysis (see Appendix 5 for detailed explanation of the various experiments). Samples showing differences in melting behaviour relative to the control samples were sequenced in order to determine the nature of these changes. A number of DNA sequence variants were found in the ARVC cohort (Table 3.2). These included six novel variants, one known rare variant and 22 known common variants/polymorphisms.

RNAfold and MFOLD bioinformatic analysis tools were used to predict whether the novel *DSP* variants and the known rare variant caused structural changes in the mRNA, while the ESEFinder tool was used to predict whether variants affected mRNA splicing, by determining whether variants affected exonic splice enhancers (ESEs). The PSIPRED bioinformatic tool was used to predict if these variants caused changes in the secondary structure of the DSP protein, while the POLYPHEN, SIFT and Align GVGD bioinformatic tools were used to determine whether these variants had a deleterious effect on the DSP protein (Fressart *et al*. 2010).

Population screening was conducted on all novel variants and the known rare ARVC variants in order to determine their prevalence in the South African population. High resolution melt analysis and DNA sequencing, or restriction digests were used. The control cohort consisted of 200 individuals (400 chromosomes) who belonged to the four main South African ethnic groups (Cape Mixed Ancestry (98), Black African (63), Indian (29) and Caucasian (10)).

The variants identified in this cohort were classified as (1) Causal mutations, (2) Genetic Variants of Unknown Significance, or (3) Polymorphisms using a system based on the criteria described by Fressart and colleagues (Fressart *et al*. 2010); this system was extended to include bioinformatic tools for the prediction of mRNA secondary structure, ESE consensus sequences and protein secondary structure.

A genetic variant was considered to be a causal mutation on the basis of (1) the absence of the variant in the control population (and in SNP databases), (2) alteration of conserved amino acids (in certain cases), (3) prediction by bioinformatic tools that the variant would cause changes in mRNA folding, mRNA splicing or protein secondary structure, (4) prediction by the POLYPHEN, SIFT and Align GVGD tools that the variant would be deleterious to the DSP protein, and (5) segregation with disease where family information was available. A genetic variant was considered as a Genetic Variant of Unknown Significance (GVUS) in the presence of a missense variant which was absent or rare (i.e., <1%) in the control population but without a causal effect predicted by at least one software or with a mutated residue that is not highly conserved among species/isoforms. A genetic polymorphism was defined as a variant observed in the control population or in polymorphism databases at a frequency of ≥1%.

Table 3.20: Sequence variants found in the ARVC cohort

REGION	*NO.*	SNP ID	NO. OF CARRIERS	NO. HOMOZYGOUS/ NO. HETEROZYGOUS	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	REGION	MINOR ALLELE FREQUENCY IN POPULATION CONTROLS (%)	MINOR ALLELE FREQUENCY IN ARVC COHORT (%)	MINOR ALLELE REPORTED FREQUENCY (dbSNP) (%)
Novel Variants										
Exon	7	-	1 (ACM 6.1)	0/1	c.860A>G	p.N287S	N-terminus	0	0.81	-
Exon	11	-	1 (ACM 8.3)	0/1	c.1344G>A	p.L448	N-terminus	0	0.81	-
Exon	12	-	1 (ACM 56.1)	0/1	c.1534A>G	p.I512V	N-terminus	0	0.81	-
Exon	23	-	1 (ACM 34.5)	0/1	c.4022G>A	p.R1341H	Rod domain	0.75	0.81	-
Exon	24	-	1 (ACM 33.2)	0/1	c.8199G>C	p.T2733	C-terminus	0	0.81	-
		-	1 (ACM 34.5)	0/1	c.8539A>C	p.R2847	C-terminus	0	0.81	-
Reported Variants										
Rare variant										
Exon	1	-	1 (ACM 33.2)	0/1	c.105G>A	p.35G	N-terminus	0	0.81	-
Polymorphisms (reported in dbSNP database)										
Promoter	-	rs2076298	1	1/0	c.1-538G>C	-	-	-	1.61	50
Intron	1	rs9392904	16	0/16	c.171-97G>C	-	-	-	12.90	30.8
Intron	2	rs10484326	18	2/16	c.274-31T>C	-	-	-	16.13	25.6
Exon	6	rs2806234	1	1/0	c.741T>G	p.A247	N-terminus	-	1.61	-
Exon	7	rs17604693	6	0/6	c.913A>T	p.I306F	N-terminus	-	4.84	2.1
Intron	9	rs2076296	22	1/21	c.1141-44C>T	-	-	-	18.55	45.7
Intron	14	rs926411	16	0/16	c.1903+57G>A	-	-	-	12.90	23.7
Intron	15	rs2076303	26	0/36	c.1904-49T>A	-	-	-	29.03	50.0
Exon	15	rs2076304	24	1/23	c.2091A>G	p.G697	N-terminus	-	20.16	35.9
Intron	16	rs7741957	2	0/2	c.2297+51T>C	-	-	-	1.61	25.9
Exon	20	rs2064217	17	2/15	c.2862C>T	p.C954	N-terminus	-	15.32	38.4
Intron	20	rs2064218	9	0/9	c.2877+79C>A	-	-	-	7.26	46.6
Intron	21	rs6942260	18	3/15	c.2984+70G>A	-	-	-	16.94	48.1
Intron	22	rs6905839	12	0/12	c.3084-72G>A	-	-	-	9.68	33.2

Exon	23	rs28763964	1	0/1	c.3510G>A	p.E1170	Rod domain	-	0.81	3.9
		rs61731476	2	0/2	c.3963G>A	p.Q1321	Rod domain	-	1.62	-
		rs28763968	1	0/1	c.4773G>A	p.R1591	Rod domain	-	0.81	-
		rs6929069	13	0/13	c.5213G>A	p.R1738Q	Rod domain	-	10.48	35.5
Exon	24	rs11558731	2	0/2	c.8175C>A	p.R2725	C-terminus	-	1.62	14.8
		rs2744380	27	1/26	c.8472G>C	p.G2824	C-terminus	-	22.58	46.2
3'UTR	-	rs11558732	1	0/1	c.8616+9T>A	-	-	-	0.81	-
		rs12250	20	2/18	c.8616+197C>T	-	-	-	17.74	37.6

3.1.1.1. Novel variants in the ARVC cohort

The six novel changes detected in the ARVC cohort were found to be exonic: three non-synonymous and three synonymous changes (Table 3.2). For probands where novel variants were found, their first degree relatives were screened for these specific variants (if the DNA samples were available) to determine the segregation pattern of the variants with disease.

3.1.1.1.1. c.860A>G variant

In exon 7 of *DSP,* we detected a novel c.860A>G transition that caused the amino acid to change from asparagine (N) to serine (S) (p.N287S) (Figure 3.1). This mutation occurred in one ARVC proband (ACM 6.1), who was a male of Mixed Ancestry. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. This missense mutation was predicted to have a deleterious effect on the DSP protein by the SIFT and Align-GVGD bioinformatic tools. Population screening for this variant was conducted by means of restriction digest by the *VspI* enzyme. The results showed that the variant was absent in the control population. Based on the fact that this variant (i) was not found in the control population, (ii) was shown to alter a conserved amino acid and (iii) was predicted to have a deleterious effect on the protein by SIFT and Align-GVGD, we concluded that this variant is causative of ARVC.

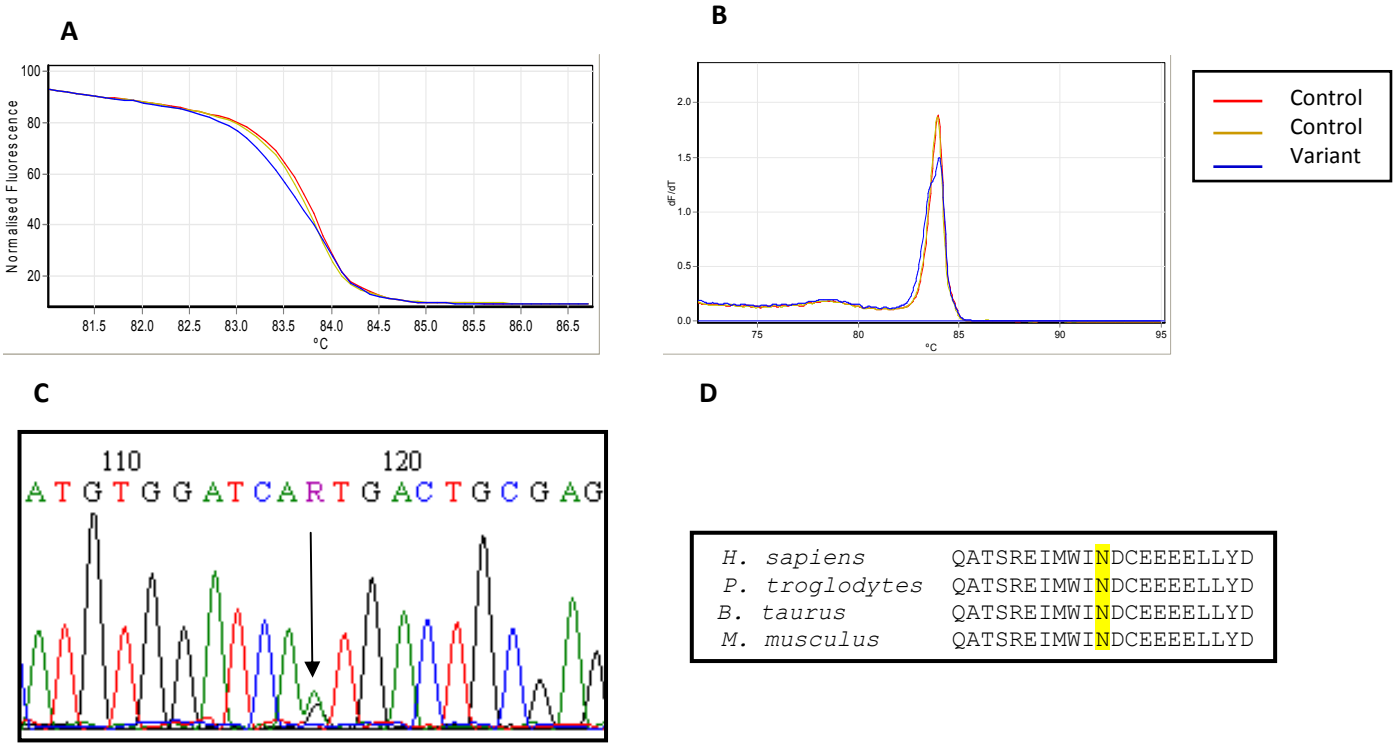


Figure 3.12: Results of c.860A>G: (A) HRM graph showing the exon 7 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A)(C) Sequencing electropherogram showing the c.860A>G sequence change (D) Multiple species protein alignment of this sequence

3.1.1.1.2. c.1344G>A variant

In exon 11 of *DSP,* we detected a novel c.1344G>A transition; this did not result in an amino acid change (p.L448) (Figure 3.2). This mutation occurred in one ARVC proband (ACM 8.3), who was a female of Caucasian origin. RNAfold predicted that the c.1344G>A variant destabilized the mRNA secondary structure, which could result in mRNA degradation (Figure 3.3). Population screening for this variant was conducted by means of restriction digest by the *PvuII* enzyme, and revealed the variant to be absent in the control population.

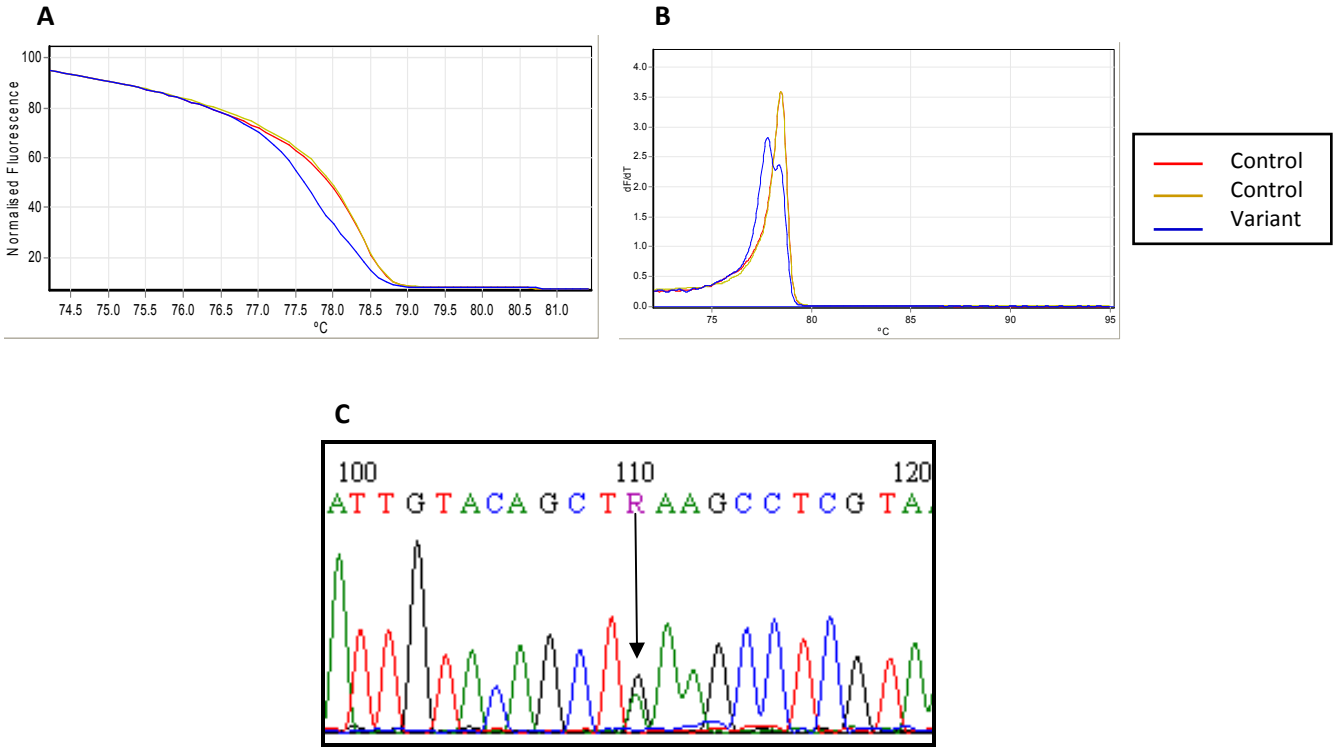


Figure 3.13: Results of c.1344G>A: (A) HRM graph showing the exon 11 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A); (C) Sequencing electropherogram showing the c.1344G>A sequence change

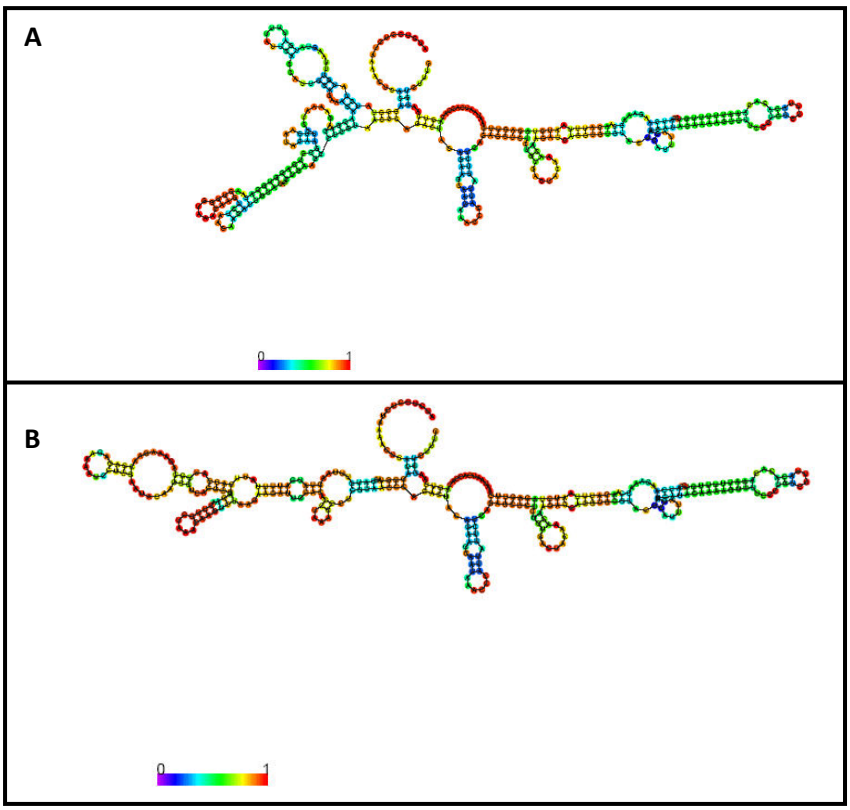


Figure 3.14: RNAfold results of c.1344G>A: (A) Exon 11 control and (B) Exon 11 with c.1344G>A variant mRNA folding structures

The discovery of this variant resulted in the screening of the available family members of this proband (Family ACM 8); this included an affected brother, both parents, the mother's sister and her husband and two daughters (ACM 8.1, 8.2 and 8.4 - 8.8) (Figure 3.4). These family members, with the exception of the proband's brother, were considered to be clinically unaffected with ARVC at the time of genetic testing. Of these individuals, only the proband's mother was found to carry this variant. Although this variant (i) was absent in the control population and (ii) was predicted to alter mRNA secondary structure by the RNAfold tool, it (i) did not alter an amino acid, and (ii) does not segregate with disease in this family. Therefore, we concluded that this variant was unlikely to be disease-causing, but it could serve as a modifier of gene function. We classified this mutation as a Genetic Variant of Unknown Significance at this stage.

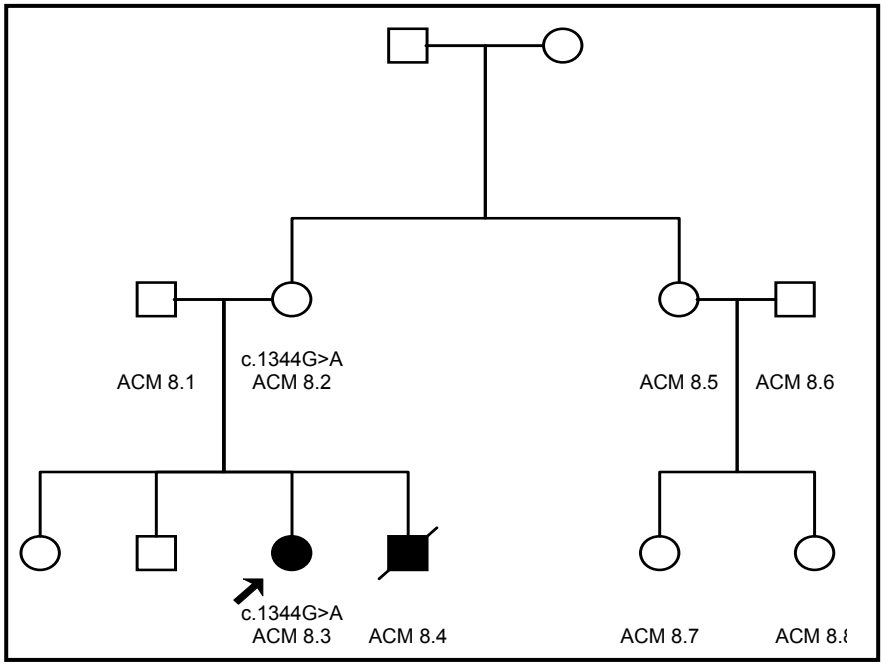


Figure 3.15: Pedigree of family ACM 8 showing variant c.1344G>A and individuals with ARVC

3.1.1.1.3. c.1534A>G variant

In exon 12 of *DSP,* we detected a novel c.1534A>G transition that caused the amino acid to change from isoleucine (I) to valine (V) (p.I512V) (Figure 3.5). This mutation occurred in one ARVC proband (ACM 56.1), who was a female of Black African ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found this amino acid to be highly conserved. MFOLD and RNAfold revealed that the c.1534A>G sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation (Figure 3.6).

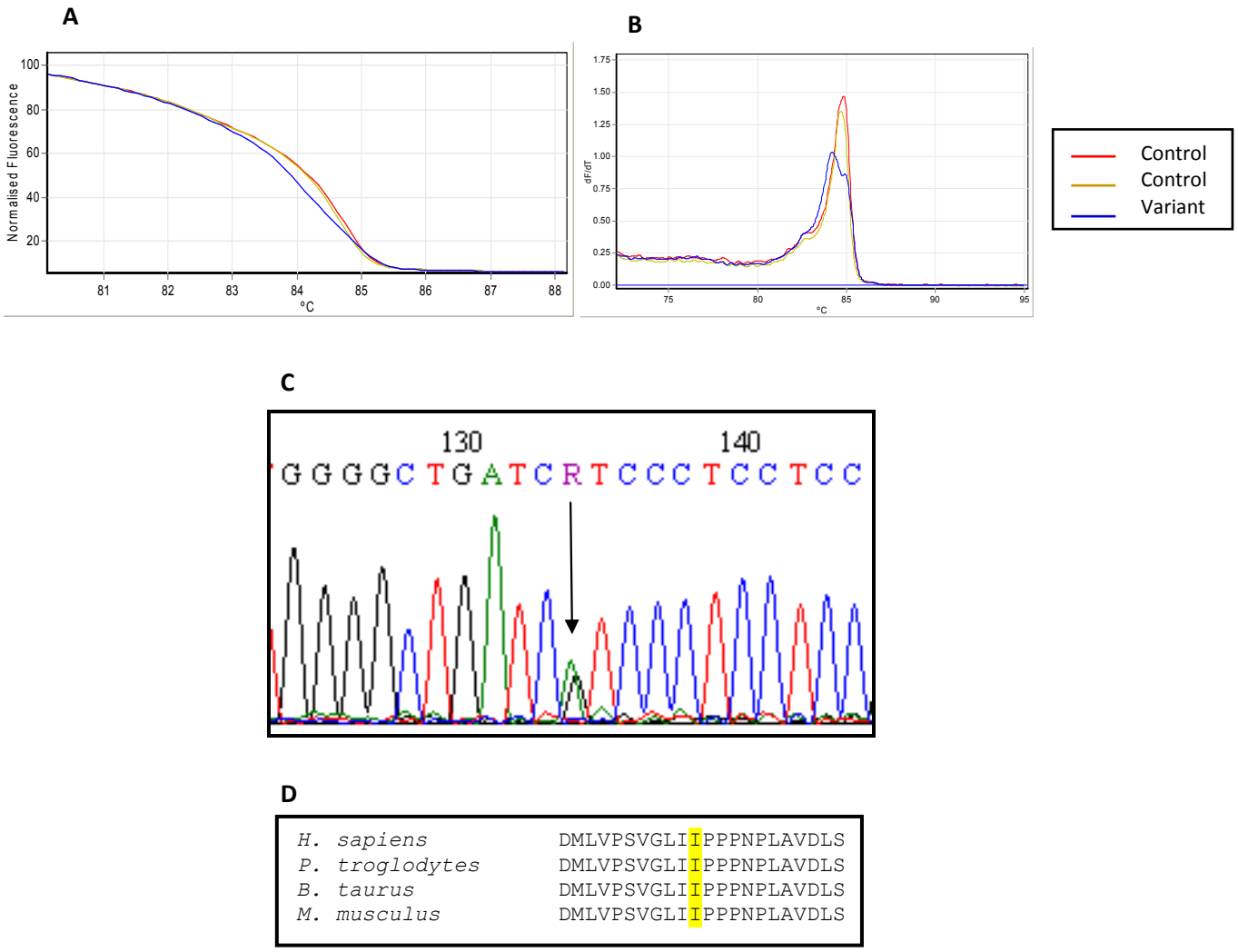


Figure 3.16: Results of c.1534A>G: (A) HRM graph showing the exon 12 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A); (C) Sequencing electropherogram showing the c.1534A>G sequence change; (D) Multiple species protein alignment of this sequence.

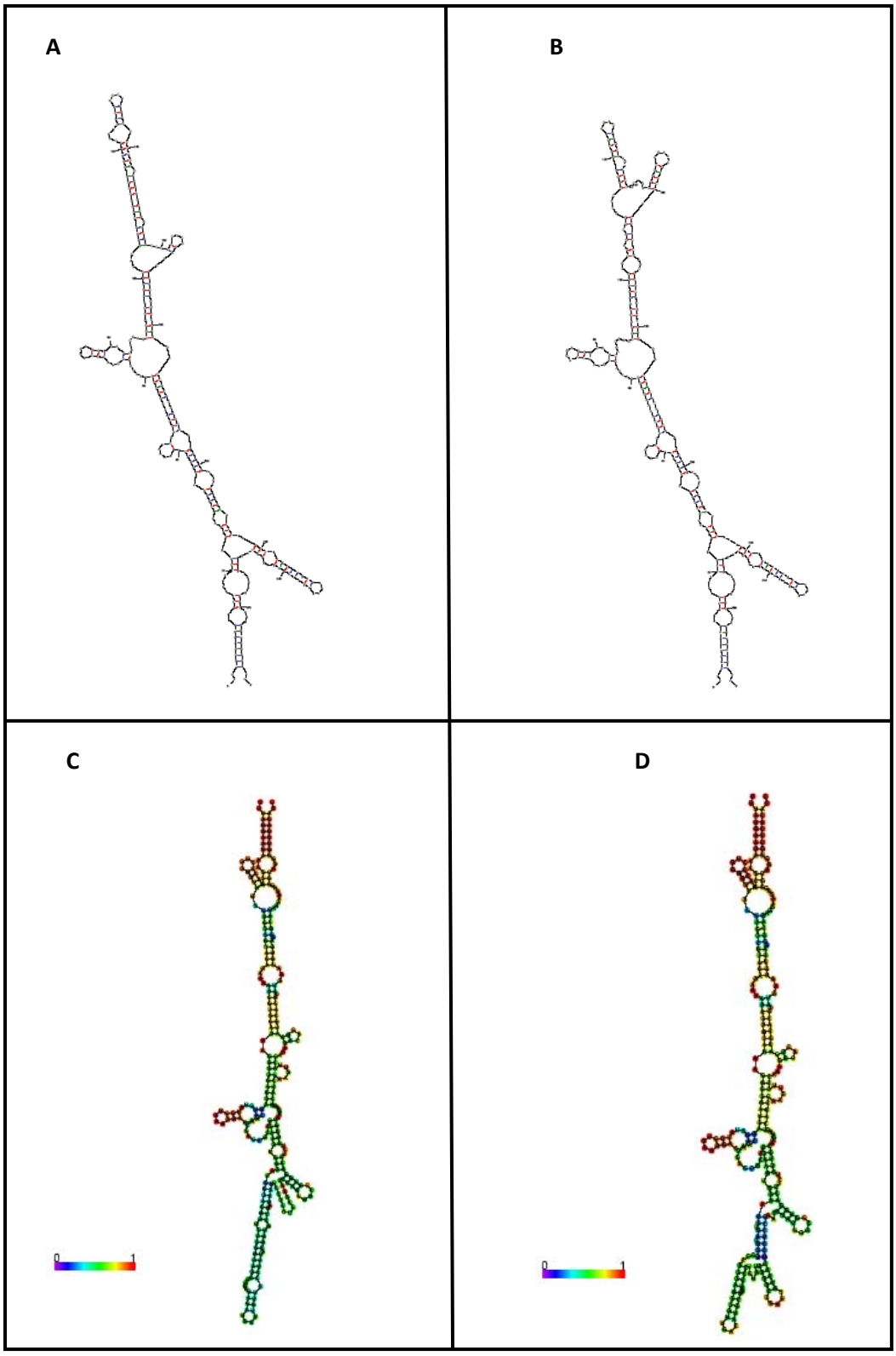


Figure 3.17: MFOLD and RNAfold results of c.1534A>G: (A) Exon 12 control and (B) Exon 12 with c.1534A>G variant mRNA folding structures predicted by mFold. (C) Exon 12 control and (D) Exon 12 with c.1534A>G variant mRNA folding structures predicted by RNAFold.

Population screening for this variant was conducted by means of restriction digest by the *FokI* enzyme, and revealed the variant to be absent in the control population. Based on the fact that (i) the variant is absent in the control population, (ii) it alters a conserved amino acid and (iii) was predicted to alter mRNA secondary structure by MFOLD and RNAfold, we conclude that this variant is causative of ARVC.

3.1.1.1.4. c.4022G>A variant

In exon 23 of *DSP,* we detected a novel c.4022G>A transition that caused the amino acid to change from an arginine (R) to histidine (H) (p.R1341H) (Figure 3.7). This missense mutation occurred in one ARVC proband (ACM 34.5), who was a male of Mixed Ancestry. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. ESEFinder found that the c.4022G>A variant in exon 12 is predicted to disrupt an SRp40 consensus sequence (Figure 3.8).

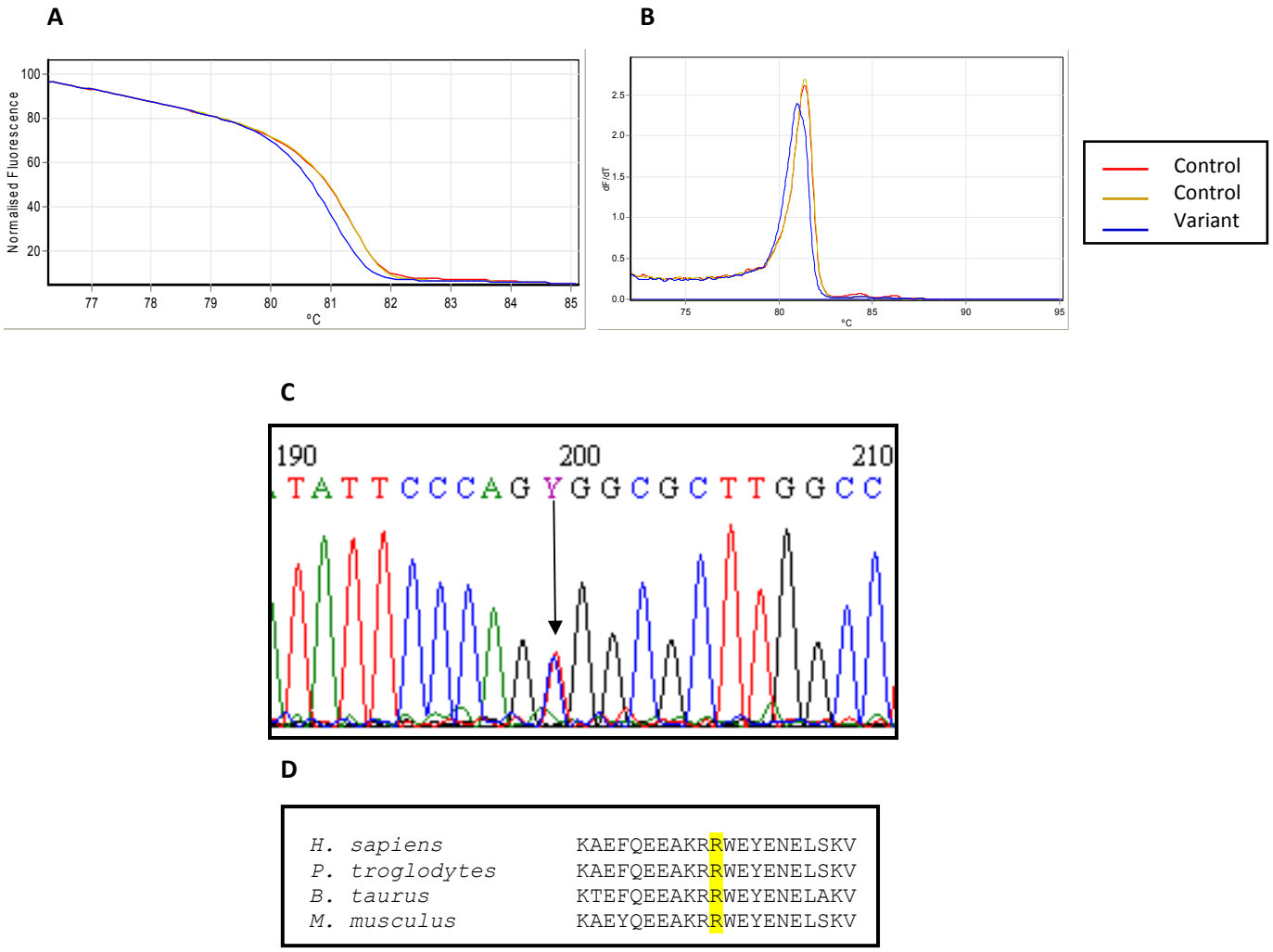


Figure 3.18: Results of c.4022G>A: (A) HRM graph showing the exon 12 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A); (C) Sequencing electropherogram showing the c.4022G>A sequence change (sequenced with reverse primer); (D) Multiple species protein alignment of this sequence

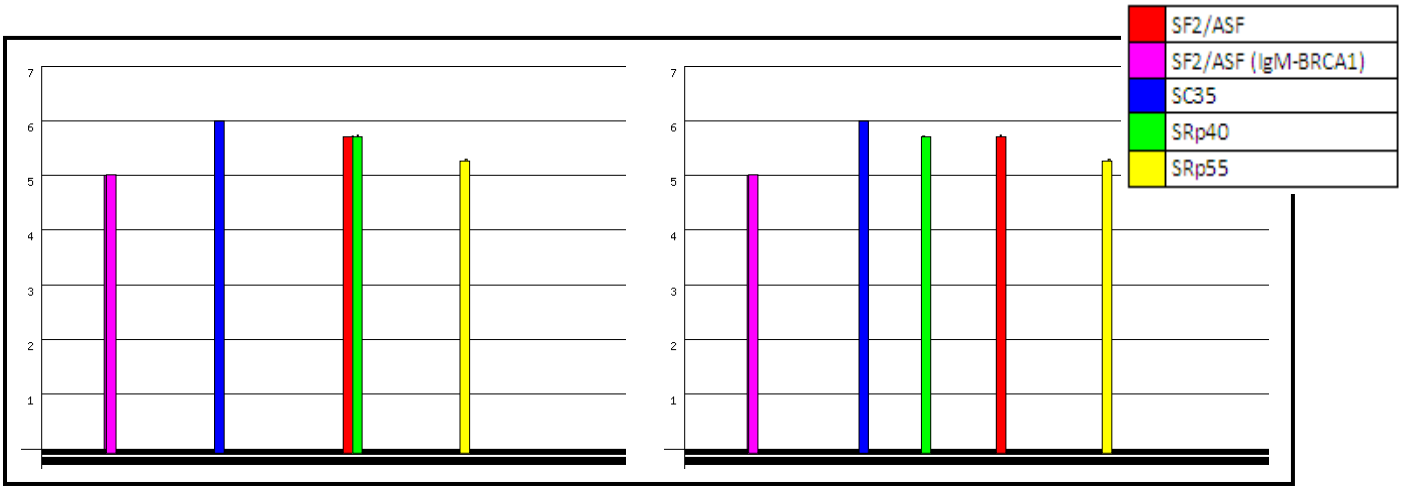


Figure 3.19: ESE results of c.4022G>A: (A) Exon 23 control ESE analysis; (B) Exon 23 c.4022 G>A variant ESE analysis

This variant is predicted to have a deleterious effect on the DSP protein by the SIFT and Align-GVGD bioinformatic tools but and was found at a frequency of 0.75% in the control population. Although the variant was predicted to (i) alter a conserved amino acid, (ii) alter an ESE consensus sequence by the ESEFinder tool and (iii) have a deleterious effect on the DSP protein by the SIFT and Align GVGD tools, it was found at a low frequency of 0.75% in the control population. Based on the detection of the variant in apparently normal controls, albeit below the 1% frequency of a polymorphism, we classified this variant as a Genetic Variant of Unknown Significance at this stage.

3.1.1.1.5. c.8199G>C variant

We detected a novel c.8199G>C transversion in exon 24 of ACM33.2; this substitution did not cause an amino acid change (p.T2733) (Figure 3.9). This mutation occurred in one ARVC proband (ACM 33.2), who was a male of Caucasian ethnicity. Population screening for this variant was conducted by means of restriction digest by the *HphI* enzyme, and revealed the variant to be absent in the control population. Furthermore, it did not alter an amino acid and was not predicted to alter mRNA secondary structure, ESE consensus sequences or protein secondary structure or have a deleterious effect on the DSP protein. We have classified this rare mutation as a Genetic Variant of Unknown Significance pending functional studies.

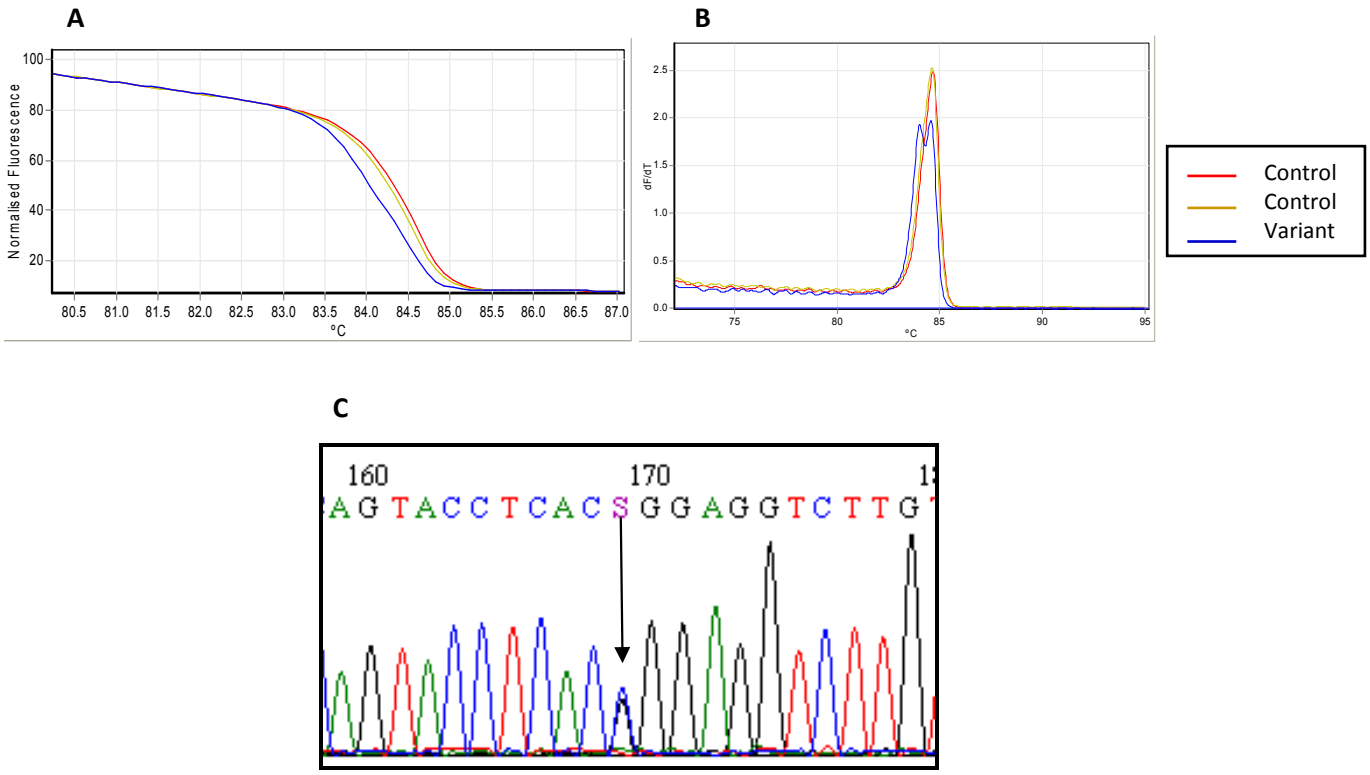


Figure 3.20: Results of c.8199G>C: (A) HRM graph showing the exon 24 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A); (C) Electropherogram showing the c.8199G>C sequence change

3.1.1.1.6. c.8539A>C variant

We also detected a second novel sequence variant in ACM 34.5. In exon 24 of *DSP,* we detected a novel c.8539A>C transversion that did not cause the amino acid to change (p.R2847) (Figure 3.10). This mutation occurred in one ARVC proband (ACM 34.5). MFOLD predicted that the mutated mRNA sequence would cause a change in the mRNA secondary structure of this gene (Figure 3.11).

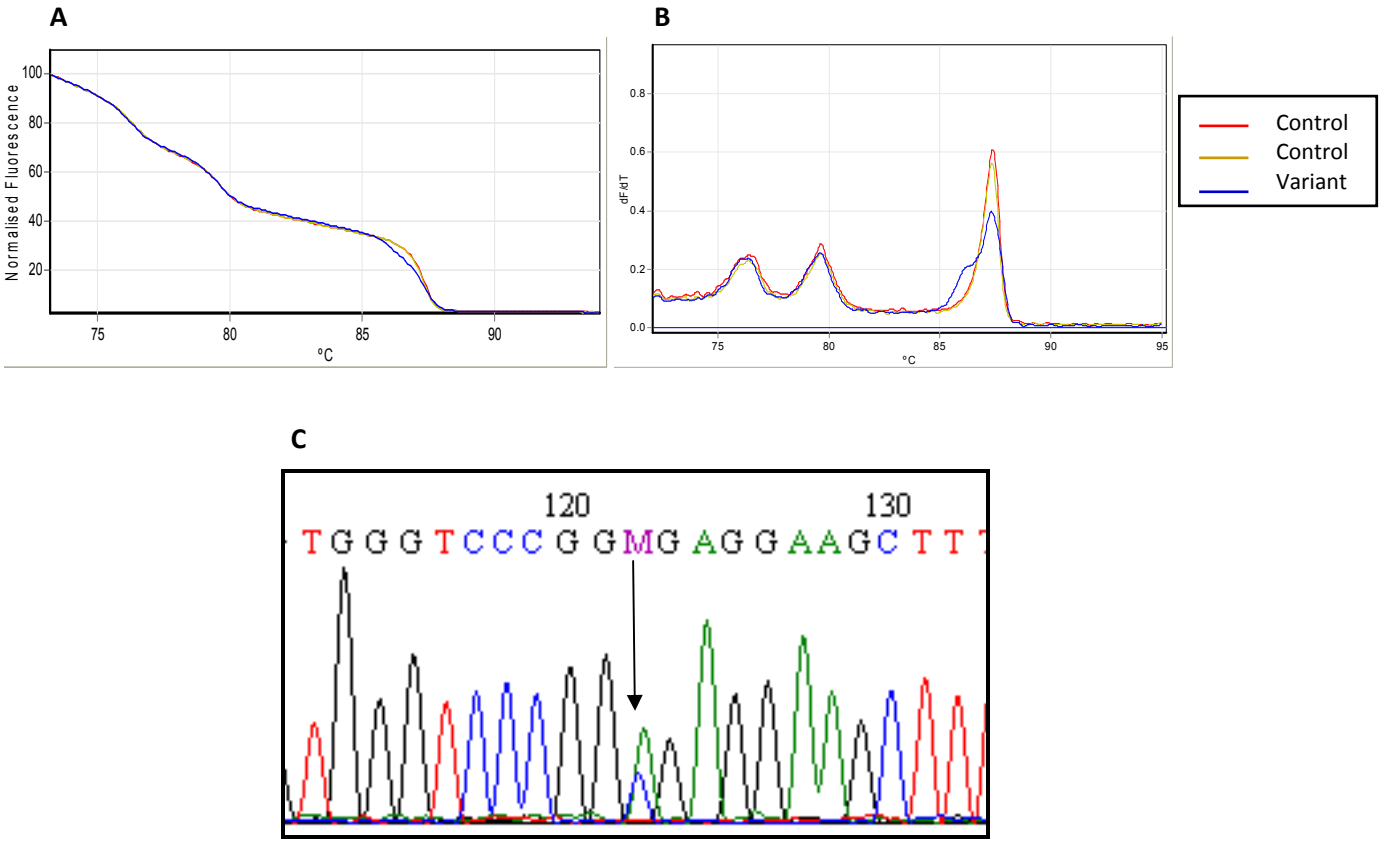


Figure 3.21: Results of c.8539A>C: (A) HRM graph showing the exon 24 melting curve of *DSP* in the control samples and patient sample (B) Derivative graph of (A); (C) Sequencing electropherogram showing the c.8539A>C sequence change

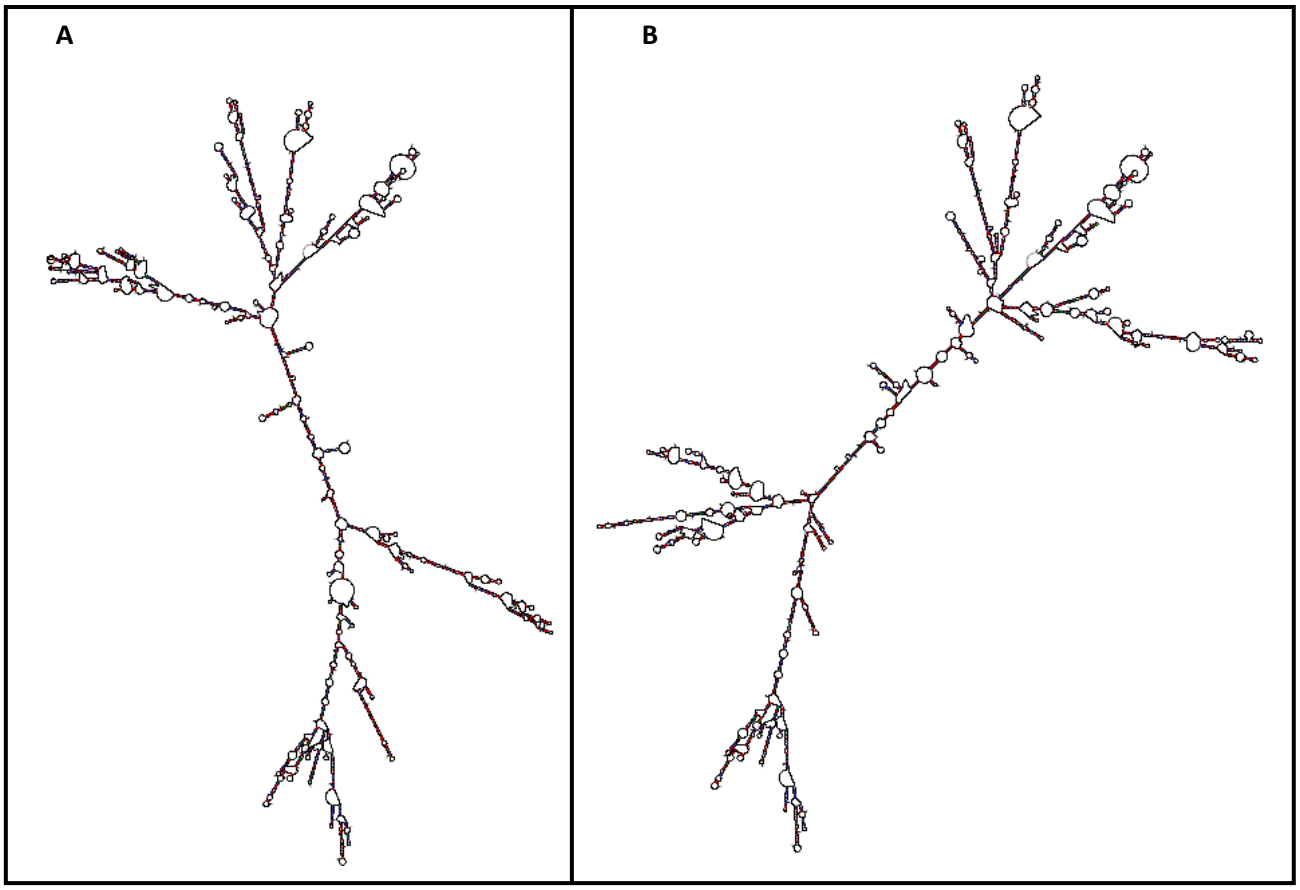


Figure 3.22: MFOLD results of c.8539A>C: (A) Exon 24 control and (B) Exon 24 with c.8539A>C variant mRNA folding structures

Population screening revealed that this variant was absent in the control population. Although this variant (i) was absent in the control population and (ii) was predicted to alter mRNA secondary structure by the MFOLD tool, it was a silent mutation that did not alter an amino acid. We have not yet been able to conduct direct analysis of the mRNA of the proband to rule out a cryptic splice site that may be introduced by this variant due to unavailability of an appropriate sample (Awad *et al.* 2006). Therefore, we concluded that this was a Genetic Variant of Unknown Significance at this stage.

3.1.1.2. Known variants in the ARVC cohort

Twenty three reported variants were found in the ARVC cohort. These included one rare variant, 12 common intronic variants and 10 common exonic variants (eight synonymous and two non-synonymous changes).

3.1.1.2.1. Rare known variants

The one known rare DNA sequence change (c.105G>A transition) was detected in exon 1 of *DSP* (Figure 3.12). Even though the variant had been reported in the ARVC database (http://www.arvcdatabase.info/), no further information was available about its causal role or otherwise in ARVC. The sequence change did not result in an amino acid change (p.35G). This mutation occurred in one ARVC proband (ACM 33.2). MFOLD revealed that the c.105G>A variant caused a predicted change in mRNA secondary structure (Figure 3.13).

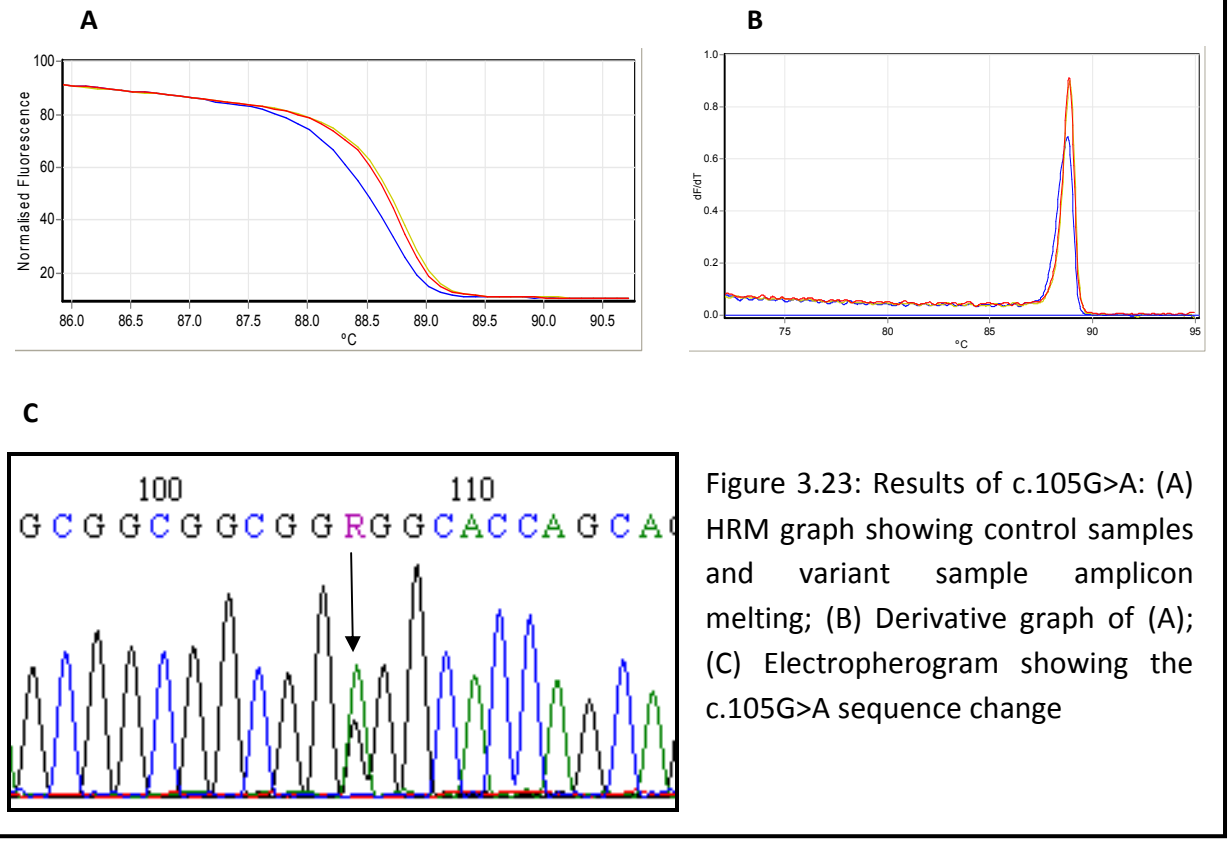


Figure 3.23: Results of c.105G>A: (A) HRM graph showing control samples and variant sample amplicon melting; (B) Derivative graph of (A); (C) Electropherogram showing the c.105G>A sequence change

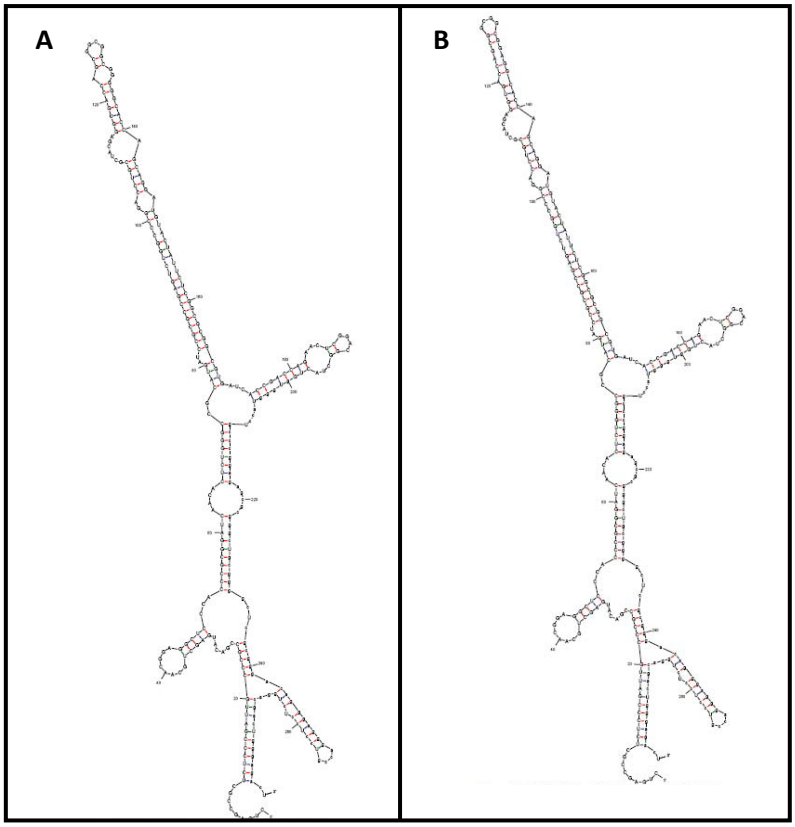


Figure 3.24: MFOLD results of c.105G>A: (A) Exon 1 control and (B) Exon 1 with the c.105G>A variant mRNA folding structures predicted by mFold

This variant was absent in the control population. Although (i) this variant was absent in the control population and (ii) was predicted to alter mRNA folding by the MFOLD tool, it has not been possible to conduct direct analysis of mRNA to rule out a cryptic splice site due to unavailability of appropriate samples (Awad 2006). We have therefore classified it as a Genetic variant of Unknown Significance at this stage.

3.1.1.2.2. Common known variants / polymorphisms

The 22 common known variants found in the ARVC cohort included 12 intronic variants and 10 exonic variants (eight synonymous and two non-synonymous changes) (Table 3.3). As these variants are reported in NCBI's single nucleotide polymorphism (SNP) database (http://www.ncbi.nlm.nih.gov/projects/SNP/), they were not screened for in our control population. Of these, eight variants had a minor allele frequency of <5% in our ARVC cohort, while the other 14 had a minor allele frequency of >5%.

All variants discovered in the ARVC cohort were summarised in Table 3.3 using the extended Fressart criteria (Fressart *et al*. 2010) as described in section 3.1.1.

Table 3.21: Classification of ARVC Variants according to extended Fressart criteria

REGION	*NO.*	SNP ID	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	FREQUENCY IN CONTROL POPULATION (%)	TYPE
Novel Variants						
Exon	7	-	c.860A>G	p.N287S	0	Causal variant
Exon	11	-	c.1344G>A	p.L448	0	Genetic Variant of Unknown Significance
Exon	12	-	c.1534A>G	p.I512V	0	Causal variant
Exon	23	-	c.4022G>A	p.R1341H	0.75	Genetic variant of Unknown Significance
Exon	24	-	c.8199G>C	p.T2733	0	Genetic Variant of Unknown Significance
		-	c.8539A>C	p.R2847	0	Genetic Variant of Unknown Significance
Reported Variants						
Rare Variant						
Exon	1	-	c.105G>A	p.35G	0	Genetic Variant of Unknown Significance
Polymorphisms (reported in dbSNP database)						
Promoter	-	rs2076298	c.1-538G>C	-	-	Polymorphism
Intron	1	rs9392904	c.171-97G>C	-	-	Polymorphism
Intron	2	rs10484326	c.274-31T>C	-	-	Polymorphism
Exon	6	rs2806234	c.741T>G	p.A247	-	Polymorphism
Exon	7	rs17604693	c.913A>T	p.I306F	-	Polymorphism
Intron	9	rs2076296	c.1141-44C>T	-	-	Polymorphism
Intron	14	rs926411	c.1903+57G>A	-	-	Polymorphism
Intron	15	rs2076303	c.1904-49T>A	-	-	Polymorphism
Exon	15	rs2076304	c.2091A>G	p.G697	-	Polymorphism
Intron	16	rs7741957	c.2297+51T>C	-	-	Polymorphism
Exon	20	rs2064217	c.2862C>T	p.C954	-	Polymorphism
Intron	20	rs2064218	c.2877+79C>A	-	-	Polymorphism
Intron	21	rs6942260	c.2984+70G>A	-	-	Polymorphism
Intron	22	rs6905839	c.3084-72G>A	-	-	Polymorphism
Exon	23	rs28763964	c.3510G>A	p.E1170	-	Polymorphism
		rs61731476	c.3963G>A	p.Q1321	-	Polymorphism
		rs28763968	c.4773G>A	p.R1591	-	Polymorphism
		rs6929069	c.5213G>A	p.R1738Q	-	Polymorphism
Exon	24	rs11558731	c.8175C>A	p.R2725	-	Polymorphism
		rs2744380	c.8472G>C	p.G2824	-	Polymorphism
3'UTR	-	rs11558732	c.8616+9T>A	-	-	Polymorphism
		rs12250	c.8616+197C>T	-	-	Polymorphism

3.1.2. Mutation screening of the DCM Cohort

DCM samples were screened using high resolution melt analysis and DNA sequencing in the same manner as the ARVC samples. A number of DNA sequence variants were found in the DCM cohort. These included 24 novel variants, three known rare variants and 16 known common variants/polymorphisms (Table 3.4).

RNAfold, MFOLD, ESEFinder, PSIPRED, POLYPHEN, SIFT and Align GVGD bioinformatic analysis software tools were used to predict if any of these variants caused structural changes in the RNA, splicesomes, transcriptomes or protein structure/function.

Population screens, for the novel and reported rare variants found in the DCM cohort, were conducted using high resolution melt analysis and DNA sequencing; for this, the same control cohort used for the ARVC population screens was used.

The variants identified in this cohort were classified as (1) Causal mutations, (2) Genetic Variants of Unknown Significance or (3) Polymorphisms using the extended Fressart criteria described in section 3.1.1.

Table 3.22: Sequence Variants found in the DCM Cohort

REGION	NO.	SNP ID	NO. OF SAMPLES	NO. HOMOZYGOUS/ NO. HETEROZYGOUS	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	REGION	MINOR ALLELE FREQUENCY IN POPULATION CONTROLS (%)	MINOR ALLELE FREQUENCY IN DCM COHORT (%)	MINOR ALLELE REPORTED FREQUENCY (dbSNP) (%)
Novel Variants										
Intron	3	-	2	0/2	c.422+23C>T	-	-	0	0.67	-
Exon	5	-	2	0/2	c.687C>T	p.G229	N-terminus	1	0.67	-
Exon	7	-	2	0/2	c.816G>A	p.L272	N-terminus	1	0.67	-
Intron	8	-	2	0/2	c.1043+32A>T	-	-	0	0.67	-
Exon	10	-	1	0/1	c.1222A>C	p.N408H	N-terminus	0	0.33	-
		-	1	0/1	c.1228C>T	p.P410S	N-terminus	0	0.33	-
Exon	12	-	3	0/3	c.1512T>C	p.L504	N-terminus	1.50	1	-
		-	1	0/1	c.1534A>G	p.I512V	N-terminus	0	0.33	-
Intron	13	-	3	0/3	c.1701+54G>A	-	-	1.75	1	-
Intron	14	-	1	0/1	c.1904-54C>T	-	-	0	0.33	-
		-	4	0/4	c.1904-12T>G	-	-	1	1.33	-
Exon	15	-	1	0/1	c.1997C>T	p.T666I	N-terminus	0.25	0.33	-
		-	1	0/1	c.2121C>T	p.N707	N-terminus	0.25	0.33	-
Intron	15	-	3	0/3	c.2130+63A>T	-	-	0	1	-
Exon	16	-	1	0/1	c.2288A>G	p.Y763C	N-terminus	0	0.33	-
Exon	19	-	3	0/3	c.2773C>T	p.R925W	N-terminus	1.25	1	-
Intron	22	-	1	0/1	c.3085-2A>G	-	-	0	0.33	-
Exon	23	-	1	0/1	c.3260A>C	p.E1087A	Rod domain	0	0.33	-
		-	1	0/1	c.3806G>A	p.R1269N	Rod domain	0	0.33	-
		-	2	0/2	c.4022G>A	p.R1341H	Rod domain	0.75	0.67	-
		-	4	0/4	c.4455G>T	p.R1485S	Rod domain	0.25	1.33	-
		-	1	0/1	c.5218G>A	p.E1740K	Rod domain	0.50	0.33	-
Exon	24	-	1	0/1	c.7116C>T	p.I2172	C-terminus	0.25	0.33	-

		-	1	0/1	c.8481C>T	p.S2827	C-terminus	0.25	0.33	-
Reported Variants										
Rare Variants										
Intron	3	-	1	0/1	c.422+32C>G	-	-	0	0.33	-
Exon	20	-	1	0/1	c.2815G>A	p.G939S	N-terminus	1	0.33	-
Exon	23	-	1	0/1	c.5178C>A	p.N1726K	Rod domain	0	0.33	-
Polymorphisms (reported in the dbSNP database)										
Intron	2	rs56148603	1	0/1	c.273+10C>T	-	-	-	0.33	2.7
		rs10484326	38	1/37	c.274-31T>C	-	-	-	13.0	25.6
Exon	7	rs17604693	2	0/2	c.913A>T	p.I306F	N-terminus	-	0.67	2.1
Intron	9	rs2076296	54	6/48	c.1142-34C>T	-	-	-	20	45.7
Intron	14	rs2076303	63	0/63	c.1904-49T>A	-	-	-	21	50.0
Exon	15	rs2076304	57	8/49	c.2091A>G	p.G697	N-terminus	-	21.67	35.9
Intron	16	rs7741957	18	0/18	c.2296+51T>C	-	-	-	6	25.9
Exon	20	rs2064217	24	0/24	c.2862C>T	p.C954	N-terminus	-	8	38.4
Intron	20	rs2064218	10	8/2	c.2877+79C>A	-	-	-	6	46.6
Intron	21	rs6942260	36	1/35	c.2985+70G>A	-	-	-	12.33	48.1
Intron	22	rs6905839	52	0/52	c.3085-66G>A	-	-	-	17.33	33.2
Exon	23	rs28763966	5	0/5	c.4578C>A	p.N1526K	Rod domain	-	1.67	2.6
		rs6929069	57	10/47	c.5213G>A	p.R1738Q	Rod domain	-	22.33	35.5
Exon	24	rs2076300	17	0/17	c.7122C>T	p.T2374	C-terminus	-	5.67	33.2
		rs2744380	23	0/23	c.8472G>C	p.G2824	C-terminus	-	7.67	46.2
3'UTR		rs12250	68	13/55	c.8616+197C>T	-	-	-	27	37.6

3.1.2.1. Novel variants in the DCM cohort

In the DCM cohort, 24 sequence changes were found that had not been previously reported (Table 3.4). These variants included seven intronic variants and 17 exonic variants (six synonymous and 11 non-synonymous variants). The majority of these changes were found in patients with idiopathic DCM. Only one of the individuals with a novel sequence change (DCM 296.1) was diagnosed with familial DCM. A few of the novel variants are discussed in more detail but information on the rest of the novel variants can be found in Table 3.4. If a novel variant was detected in the proband and we had the DNA available of the nuclear family, these members were genetically tested for segregation of the variant with disease.

3.1.2.1.1. c.1222A>C variant

In exon 10 of the *DSP* gene, we detected a c.1222A>C transversion that caused an amino acid change from asparagine (N) to histidine (H) (N408H). This mutation occurred in only one DCM proband (DCM 173.1), who was a male of Mixed Ancestry ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. MFOLD and RNAfold revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation. ESEFinder found that this variant is predicted to disrupt an SF2/ASF consensus sequence, while the SIFT and Align GVGD tools predicted that this variant had a deleterious effect on the DSP protein. This variant was also found to be absent in the control population. Based on the fact that this variant (I) was absent from the control population, (ii) alters a conserved amino acid, (iii) was predicted to alter mRNA secondary structure by MFOLD and RNAfold, (iv) was predicted to alter an ESE consensus sequence by ESEFinder and (v) was predicted to have a deleterious effect on the DSP protein by SIFT and Align-GVGD, we conclude that this variant is causative of DCM.

3.1.2.1.2. c.1228C>T variant

In exon 10 of *DSP*, we detected a c.1228C>T transition that caused an amino acid change from phenylalanine to serine (p.P410S). This mutation was found in only one DCM proband (DCM 158.1), who was a male of Black African ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. RNAfold revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation. The PSIPRED tool predicted that this variant would cause a change in protein secondary structure, while the Align GVGD tool predicted that this variant would have a deleterious effect on the DSP protein. Based on the fact that this variant (I) was absent in the control population, (ii) alters a conserved amino acid, (iii) was predicted to cause a change in mRNA secondary structure by RNAfold, (iv) was predicted to alter protein secondary structure by PSIPRED and (v)

was predicted to have a deleterious effect on the DSP protein by Align GVGD, we conclude that this variant is causative of DCM.

3.1.2.1.3. c.2288A>G variant

In exon 16 of *DSP*, we detected a c.2288A>G transition that caused an amino acid change from tyrosine (Y) to cysteine (C) (Y763C). This mutation was found in only one DCM proband (DCM 58.1), who was a female of Black African ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. The SIFT, POLYPHEN and Align GVGD tools predicted that this variant would have a deleterious effect on the DSP protein. This variant was absent in the control population. Based on the fact that (i) the variant is absent in the control population, (ii) it alters a conserved amino acid and (iii) it was predicted to have a deleterious effect on the DSP protein by the POLYPHEN, SIFT and ALIGN GVGD tools, we conclude that this variant is causative of DCM.

3.1.2.1.4. c.3085-2A>G variant

In intron 22 of DSP, we detected a c.3085-2A>G transition that did not cause an amino acid change. This mutation occurred in only one DCM proband (DCM 301.1), who was a female of Black African ethnicity. This variant caused an alteration in the sequence at a 3' intronic splice site, which would result in a change in mRNA splicing. RNAfold revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation. This variant was found to be absent in the control population. Based on the fact that (i) the variant is absent in the control population, (ii) it is a splice junction mutation (Shapiro and Senapathy 1987), and (iii) it was predicted to change mRNA secondary structure by RNAfold, we conclude that this variant is causative of DCM.

3.1.2.1.5. c.3260A>C variant

In exon 23 of *DSP*, we detected a c.3260A>C transversion that caused an amino acid change from glutamate (E) to alanine (A) (E1087A). This mutation occurred in only one DCM proband (DCM 301.1). Orthologous alignment of the *Homo sapiens* protein sequence to other modern species found that the amino acid is highly conserved. MFOLD and RNAfold revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation. The POLYPHEN and Align GVGD tools predicted that this variant had a deleterious effect on the DSP protein. This variant was found to be absent in the control population. Based on the fact that (i) the variant is absent in the control population, (ii) it alters a conserved amino acid, (iii) it was predicted to cause a change in mRNA secondary structure by MFOLD and RNAfold and (iv) it was predicted to have a deleterious effect on the DSP protein by the POLYPHEN and Align GVGD tools, we conclude that this variant is causative of DCM.

3.1.2.1.6. c.3806G>A variant

One of the novel changes, a c.3806G>A transition in exon 23 of *DSP,* changed the amino acid from an arginine (R) to asparagine (N) (p.R1269N). This mutation was found in only one DCM proband (DCM 296.1), who was a female of Mixed Ancestry ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to that of modern species found that the amino acid is highly conserved. This variant was predicted to have a deleterious effect on the DSP protein by the POLYPHEN and Align GVGD tools, and was absent in the control population. Based on the fact that (i) the variant is absent in the control population, (ii) it alters a conserved amino acid and (iii) it was predicted to have a deleterious effect on the DSP protein by the POLYPHEN and ALIGN GVGD tools, we conclude that this variant is causative of DCM.

3.1.2.2. Known variants in the DCM cohort

In the DCM cohort, three rare reported DNA sequence variants and 16 common reported variants were identified. The three rare variants included two exonic, non-synonymous variants and one intronic variant. These changes (c.422+32C>T, c.2815G>A and c.5178C>A) had been reported in the ARVC database (http://www.arvcdatabase.info/); however, no further information was available about the causal role or otherwise of these mutations in ARVC.

3.1.2.2.1 Rare known variants

3.1.2.2.1.1. c.422+32C>G variant

In intron 3 of *DSP*, we detected a c.422+32C>G transversion (Figure 3.14) which only occurred in one DCM proband (DCM 35.1), who was a female of Black African ethnicity.

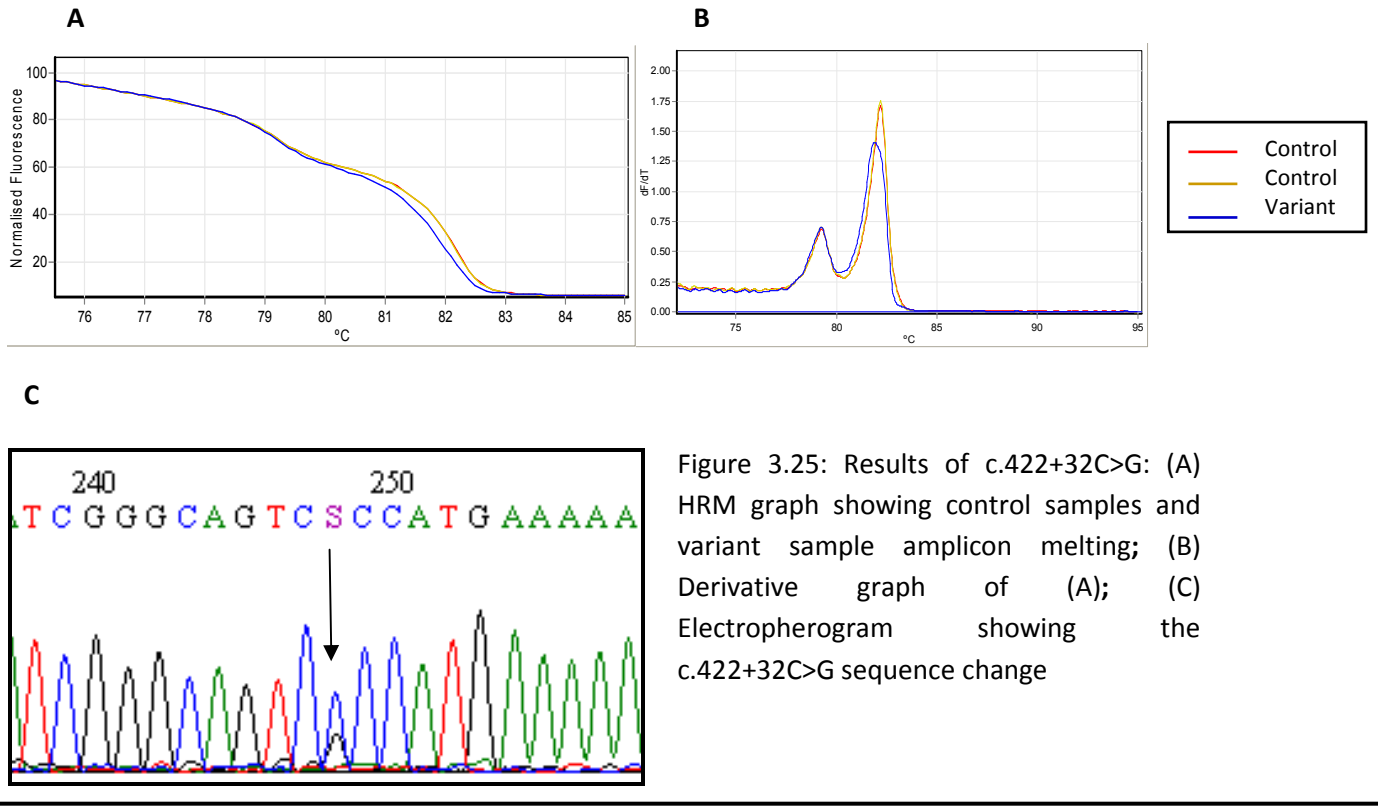


Figure 3.25: Results of c.422+32C>G: (A) HRM graph showing control samples and variant sample amplicon melting; (B) Derivative graph of (A); (C) Electropherogram showing the c.422+32C>G sequence change

RNAfold revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation (Figure 3.15).

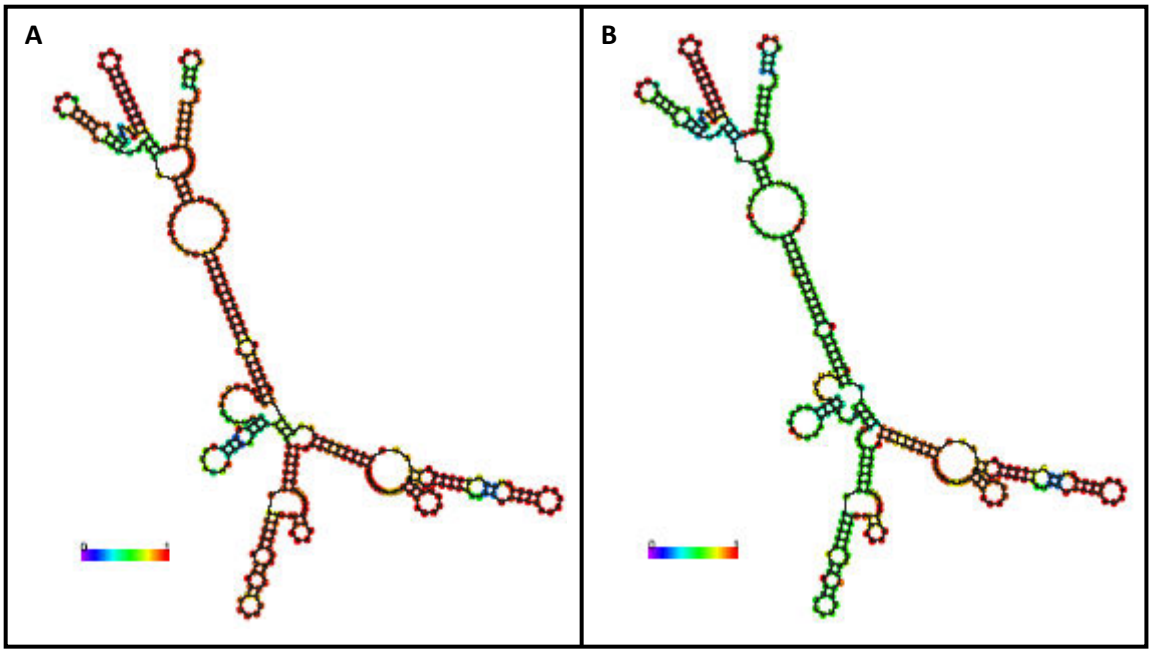


Figure 3.26: RNAfold results of c.422+32C>G: (A) Exon 3 with part of intron 3 control and (B) Exon 3 with part of intron 3 with c.422+32C>G variant mRNA folding structures predicted by RNAFold.

This variant was absent in the control population. Although this variant (i) was not found in the control population and (ii) was predicted to alter mRNA secondary structure by RNAfold, its intronic position raises doubt about its pathogenic significance. We have therefore classified it as a Genetic Variant of Unknown Significance at this stage.

3.1.2.2.1.2. c.2815G>A variant

In exon 20 of *DSP*, we detected a c.2815G>A transition that changed the amino acid from glycine (G) to serine (S) (p.G939S) (Figure 3.16). This change occurred in four DCM probands (DCM 31.1 (Black African male), DCM 35.1 (Black African female), DCM 59.1 (Male of Unknown ethnicity) and DCM 73.1 (Black African male)). Orthologous alignment of the *Homo sapiens* protein sequence to modern animal species found that the amino acid is highly conserved.

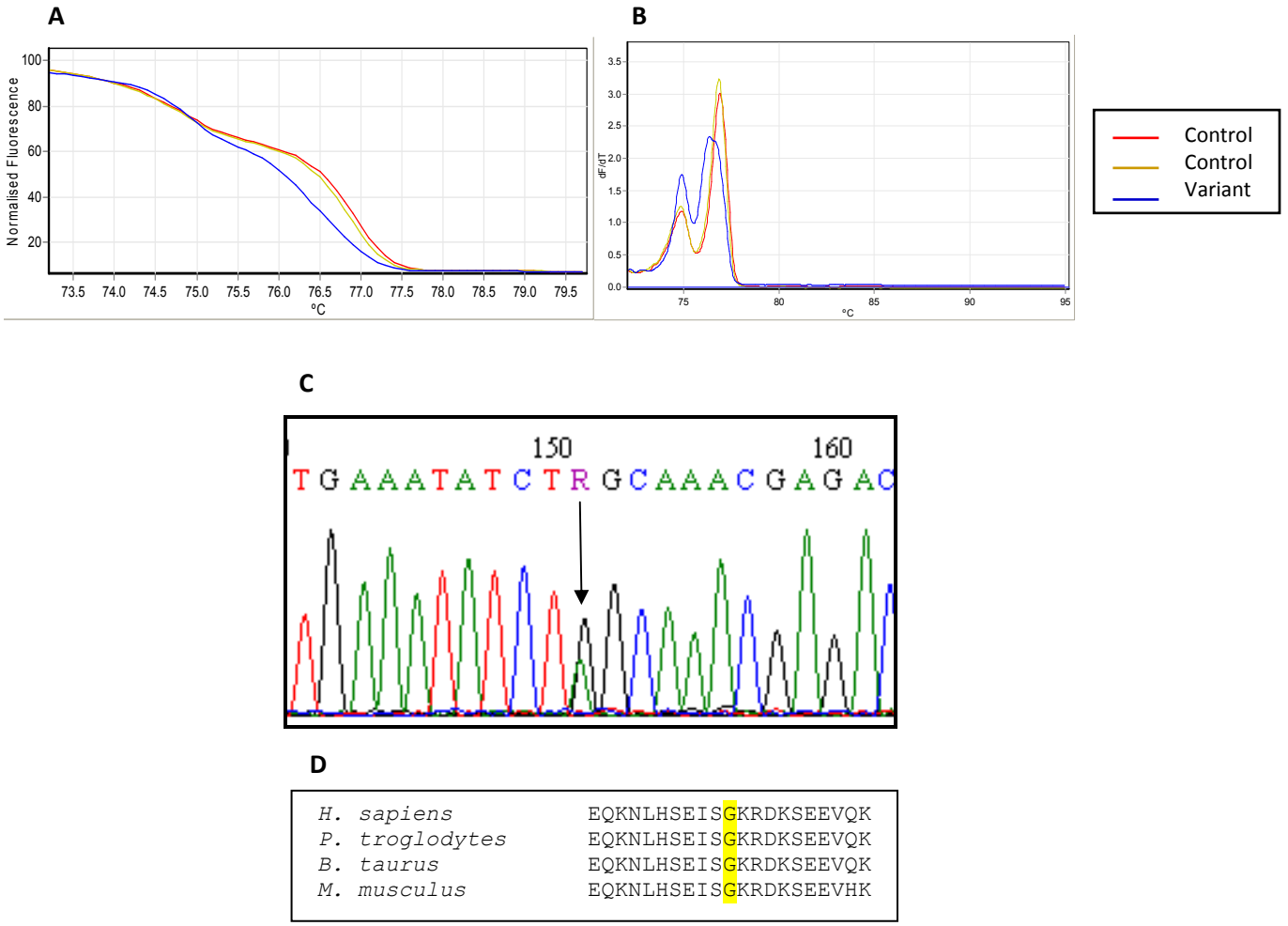


Figure 3.27: Results of c.2815G>A: (A) HRM graph showing control samples and variant sample amplicon melting; (B) Derivative graph of (A); (C) Electropherogram showing the c.2815G>A sequence change; (D) Multiple species protein alignment of this sequence

This variant was predicted to cause a change in protein secondary structure by the PSIPRED tool (Figure 3.17); it was also predicted to have a deleterious effect on the protein by the POLYPHEN and Align GVGD tools. However, this variant was found at a frequency of 1% in the control population, and was thus classified as a polymorphism.

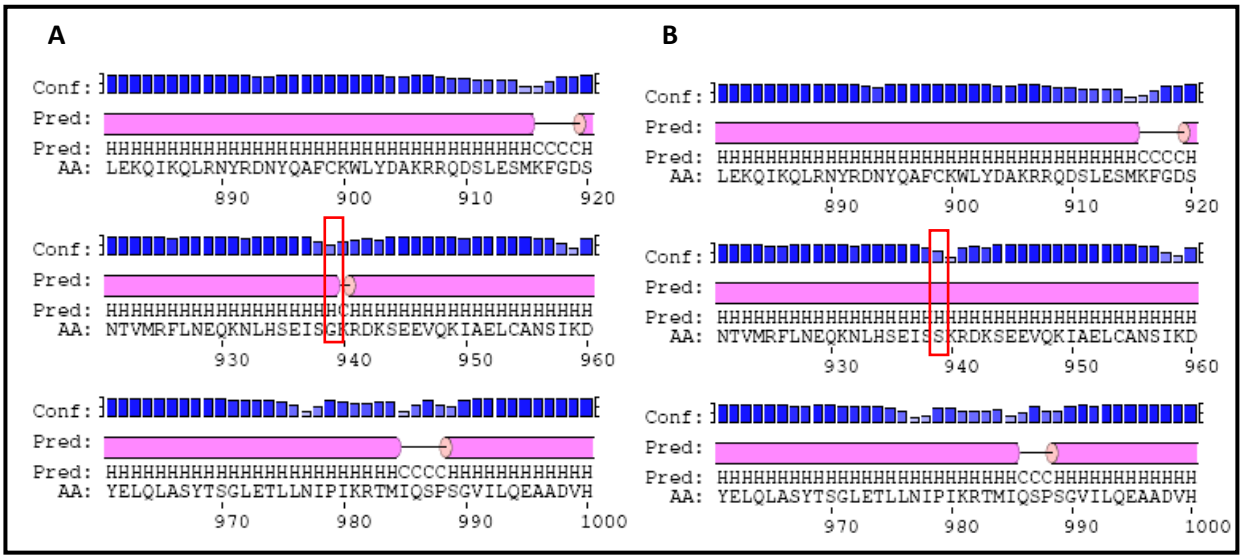


Figure 3.28: PSIPRED results of c.2815G>A: (A) PSIPRED predicted protein secondary structure of control; (B) PSIPRED predicted protein secondary structure of sequence with the c.2815G>A (p.G939S) variant (H – α-helix element; E – β-strand element; C – random coil element)

3.1.2.2.1.3. c.5178C>A variant

In exon 23 of *DSP,* we detected a c.5178C>A transversion that changed the amino acid from an asparagine (N) to lysine (K) (p.N1726K) (Figure 3.18). This mutation was found in only one DCM proband (DCM 241.1), who was a male of Indian ethnicity. Orthologous alignment of the *Homo sapiens* protein sequence to modern animal species found that the amino acid is highly conserved.

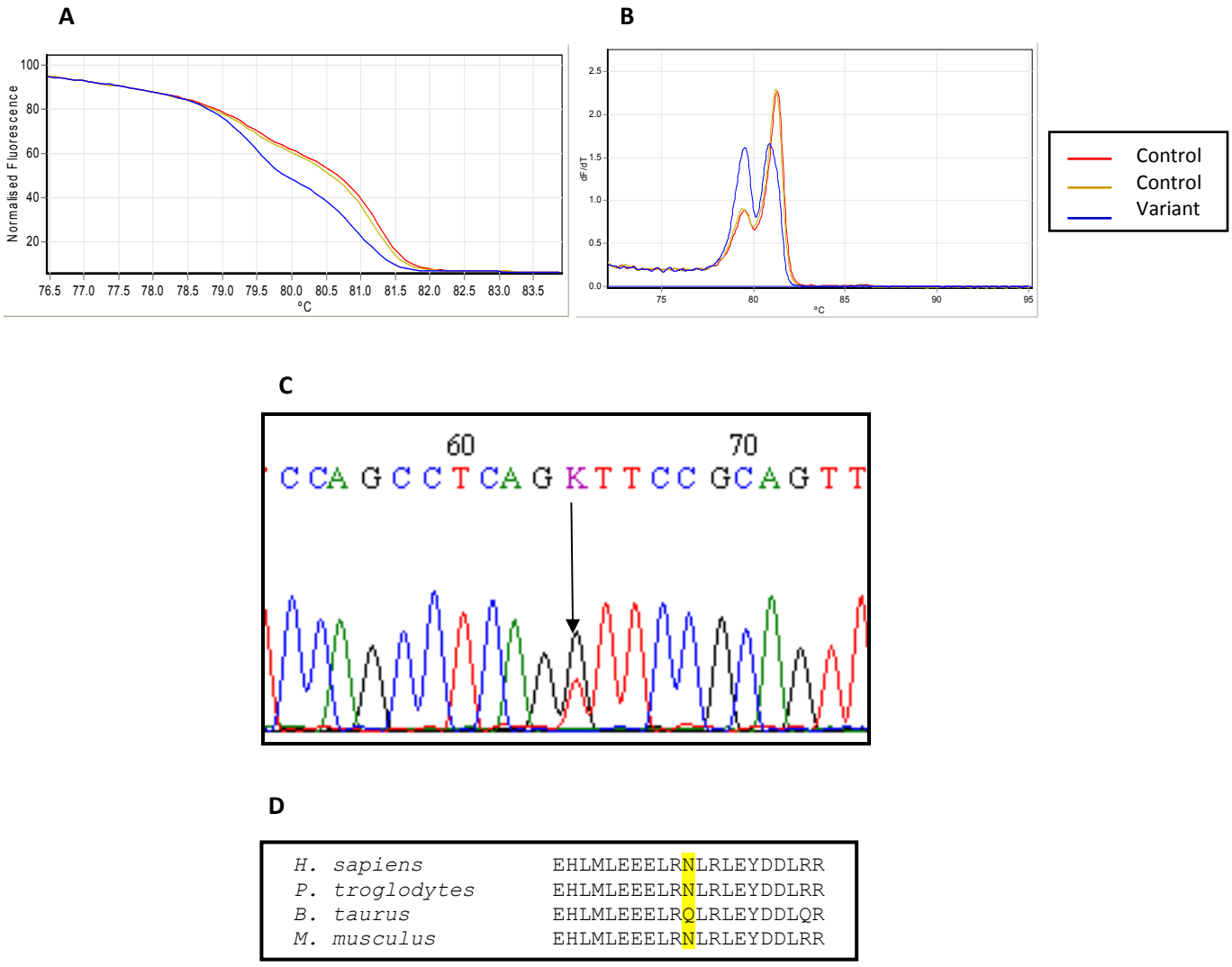


Figure 3.29: Results of c.5178C>A: (A) HRM graph showing control samples and variant sample amplicon melting; (B) Derivative graph of (A); (C) Electropherogram showing the c.5178C>A sequence change; (D) Multiple species protein alignment of this sequence

MFOLD revealed that this sequence variant caused a predicted change in mRNA secondary structure which could result in destabilized mRNA and premature mRNA degradation (Figure 3.19). The POLYPHEN and Align GVGD tools predicted that this variant would have a deleterious effect on the protein. This variant was absent in the control population.

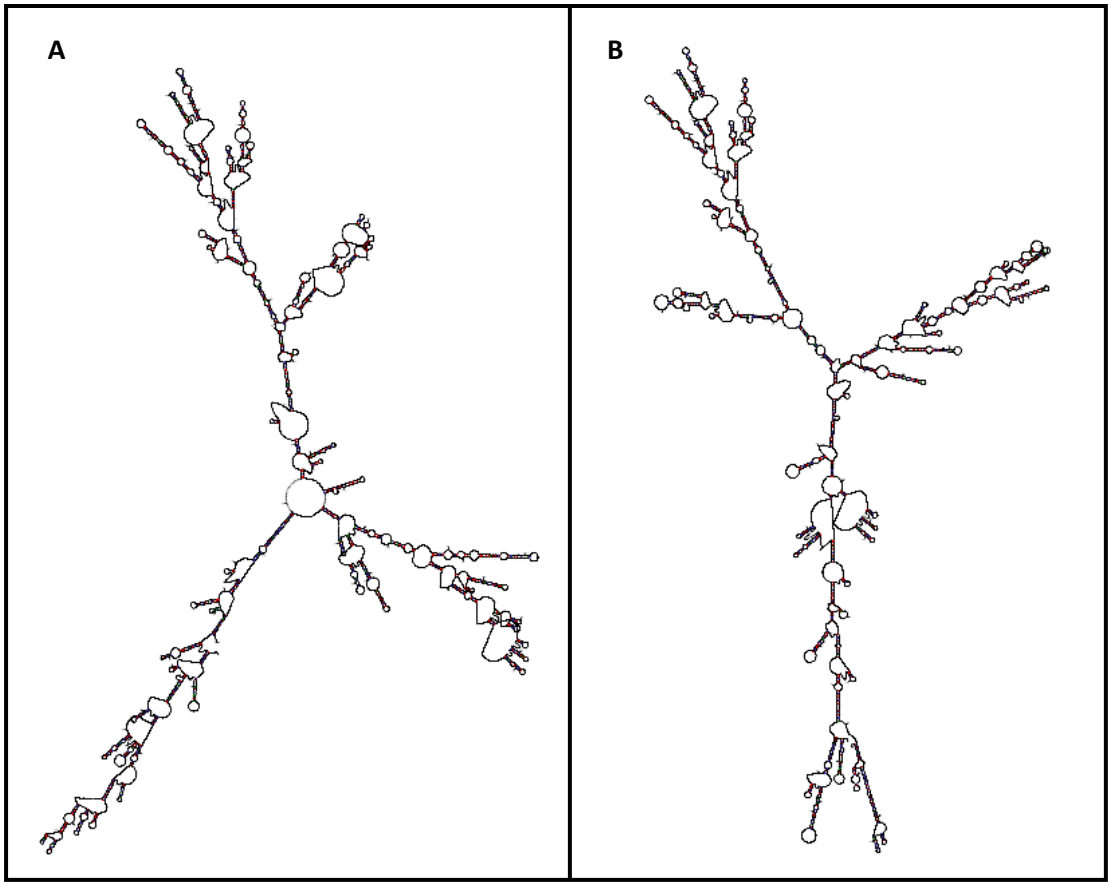


Figure 3.30: MFOLD results of c.5178C>A: (A) Exon 23 control and (B) Exon 23 with c.5178C>A variant mRNA folding structures predicted by mFold.

The discovery of this mutation led to the genetic screening of the DNA of one deceased son (DCM 241.2) of this proband (DCM 241.1) (Figure 3.20).

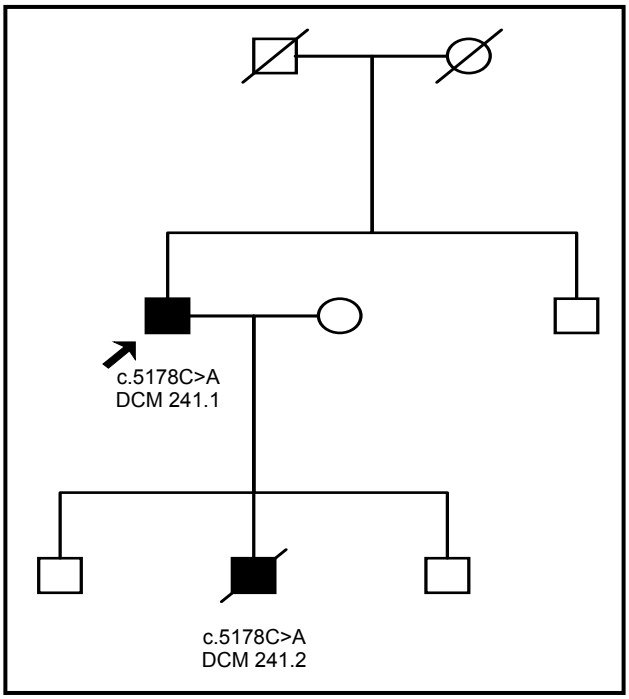


Figure 3.31: Pedigree of family DCM 241 showing variant c.5178C>A and individuals with DCM

This family member was found to carry the c.5178C>A variant, and was found to be clinically affected with DCM at the time of genetic screening. Based on the fact that this variant (i) is absent in the control population, (ii) alters a conserved amino acid, (iii) was predicted to alter mRNA secondary structure by the MFOLD tool, (iv) was predicted to have a deleterious effect on the protein by the POLYPHEN and Align GVGD tools and (v) there is preliminary evidence of segregation with disease in this family, we conclude that this variant is causative of DCM.

3.1.2.2.2 Common known variants

The 16 common polymorphisms detected in the DCM cohort included nine intronic variants and seven exonic variants (four synonymous and three non-synonymous changes). As these variants are reported in the SNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/), they were not screened for in our control population. Of these, three variants had a minor allele frequency of <5% in our DCM cohort, while the other 13 had a minor allele frequency of >5%.

All variants discovered in the DCM cohort are summarised in Table 3.5 using the extended Fressart criteria (Fressart *et al*. 2010) described in section 3.1.1.

Table 3.23: Classification of DCM Variants according to the extended Fressart criteria

REGION	NO.	SNP ID	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	FREQUENCY IN CONTROL POPULATION (%)	TYPE
Novel Variants						
Intron	3	-	c.422+23C>T	-	0	Genetic Variant of Unknown Significance
Exon	5	-	c.687C>T	p.G229	1	Polymorphism
Exon	7	-	c.816G>A	p.L272	1	Polymorphism
Intron	8	-	c.1043+32A>T	-	0	Genetic Variant of Unknown Significance
Exon	10	-	c.1222A>C	p.N408H	0	Causal Variant
		-	c.1228C>T	p.P410S	0	Causal Variant
Exon	12	-	c.1512T>C	p.L504	1.50	Polymorphism
		-	c.1534A>G	p.I512V	0	Causal Variant
Intron	13	-	c.1701+54G>A	-	1.75	Polymorphism
Intron	14	-	c.1904-54C>T	-	0	Genetic Variant of Unknown Significance
		-	c.1904-12T>G	-	1	Polymorphism
Exon	15	-	c.1997C>T	p.T666I	0.25	Genetic Variant of Unknown Significance
		-	c.2121C>T	p.N707	0.25	Genetic Variant of Unknown Significance
Intron	15	-	c.2130+63A>T	-	0	Genetic Variant of Unknown Significance
Exon	16	-	c.2288A>G	p.Y763C	0	Causal Variant
Exon	19	-	c.2773C>T	p.R925W	1.25	Polymorphism
Intron	22	-	c.3085-2A>G	-	0	Causal Variant
Exon	23	-	c.3260A>C	p.E1087A	0	Causal Variant
		-	c.3806G>A	p.R1269N	0	Causal Variant
		-	c.4022G>A	p.R1341H	0.75	Genetic Variant of Unknown Significance
		-	c.4455G>T	p.R1485S	0.25	Genetic Variant of Unknown Significance
		-	c.5218G>A	p.E1740K	0.50	Genetic Variant of Unknown Significance
Exon	24	-	c.7116C>T	p.I2172	0.25	Genetic Variant of Unknown Significance
		-	c.8481C>T	p.S2827	0.25	Genetic Variant of Unknown Significance
Reported Variants						
Rare Variant						
Intron	3	-	c.422+32C>G	-	0	Genetic Variant of Unknown Significance
Exon	20	-	c.2815G>A	p.G939S	1	Polymorphism
Exon	23	-	c.5178C>A	p.N1726K	0	Causal Variant
Polymorphisms (reported in dbSNP database)						
Intron	2	rs56148603	c.273+10C>T	-	-	Polymorphism
		rs10484326	c.274-31T>C	-	-	Polymorphism
Exon	7	rs17604693	c.913A>T	p.I306F	-	Polymorphism
Intron	9	rs2076296	c.1142-34C>T	-	-	Polymorphism
Intron	14	rs2076303	c.1904-49T>A	-	-	Polymorphism

Exon	15	rs2076304	c.2091A>G	p.G697	-	Polymorphism
Intron	16	rs7741957	c.2296+51T>C	-	-	Polymorphism
Exon	20	rs2064217	c.2862C>T	p.C954	-	Polymorphism
Intron	20	rs2064218	c.2877+79C>A	-	-	Polymorphism
Intron	21	rs6942260	c.2985+70G>A	-	-	Polymorphism
Intron	22	rs6905839	c.3085-66G>A	-	-	Polymorphism
Exon	23	rs28763966	c.4578C>A	p.N1526K	-	Polymorphism
		rs6929069	c.5213G>A	p.R1738Q	-	Polymorphism
Exon	24	rs2076300	c.7122C>T	p.T2374	-	Polymorphism
		rs2744380	c.8472G>C	p.G2824	-	Polymorphism
3'UTR		rs12250	c.8616+197C>T	-	-	Polymorphism

3.1.3. Variants found in the ARVC and DCM cohorts

The c.1534A>G and c.4022G>A novel variants discussed above (sections 3.1.1.1.3 and 3.1.1.1.4) were found in patients with ARVC and patients with DCM. The c.1534A>G variant was considered to be causative of ARVC, while the c.4022G>A variant was determined to be a Genetic Variant of Unknown Significance. In the DCM cohort, the c.1534A>G variant was found in one DCM proband, while the c.4022G>A variant was found in two DCM probands. We hypothesise that the c.1534A>G variant is causative of ARVC and DCM, using the criteria discussed in section 3.1.1.1.3. In the case of the c.4022G>A variant, we conclude that this variant is a Genetic Variant of Unknown Significance in the case of ARVC and DCM, using the criteria discussed in section 3.1.1.1.4.

Figure 3.21 below indicates the positions of the variants (novel and reported) detected in the DSP gene in ARVC and DCM probands (that have not been reported in NCBI's dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/)) in an attempt to provide a global overview of the nature and position of the changes detected.

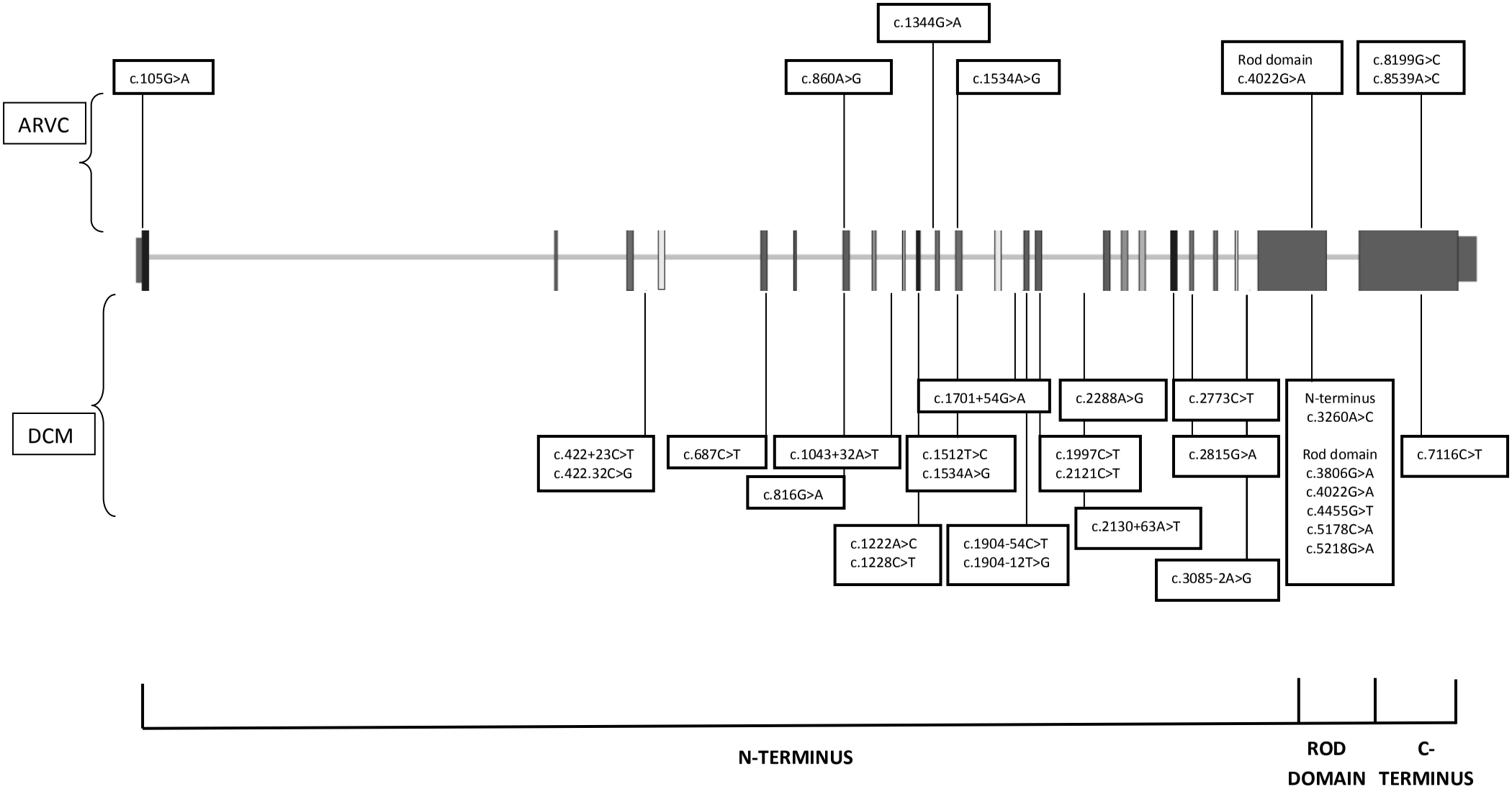


Figure 3.32: A schematic diagram showing the location of all variants identified in the *DSP* gene in ARVC and DCM probands (besides those listed in the dbSNP database)

3.2. MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA)

In this study, the SALSA MLPA Kit P168 ARVC-PKP2 kit was used to screen the DSP gene for all ARVC samples.This kit is designed to detect deletions/duplications of one or more exons of the *PKP2*, *DSP*, *JUP*, *TGFβ3* and *RyR2* genes. MLPA can however also detect point mutations very close to the probe ligation site. Screening of the ARVC samples with MLPA revealed no deletions, insertions or copy number variations of the DSP gene (Figure 3.22). The ratios approximating one in the Coffalyser output in Figure 3.21(B) indicate normal function and distribution of the *DSP* probes. Based on the graphs it appears that *DSP* is largely intact. However, deletions of one or two nucleotides, which would not be detected by MLPA, cannot be excluded.

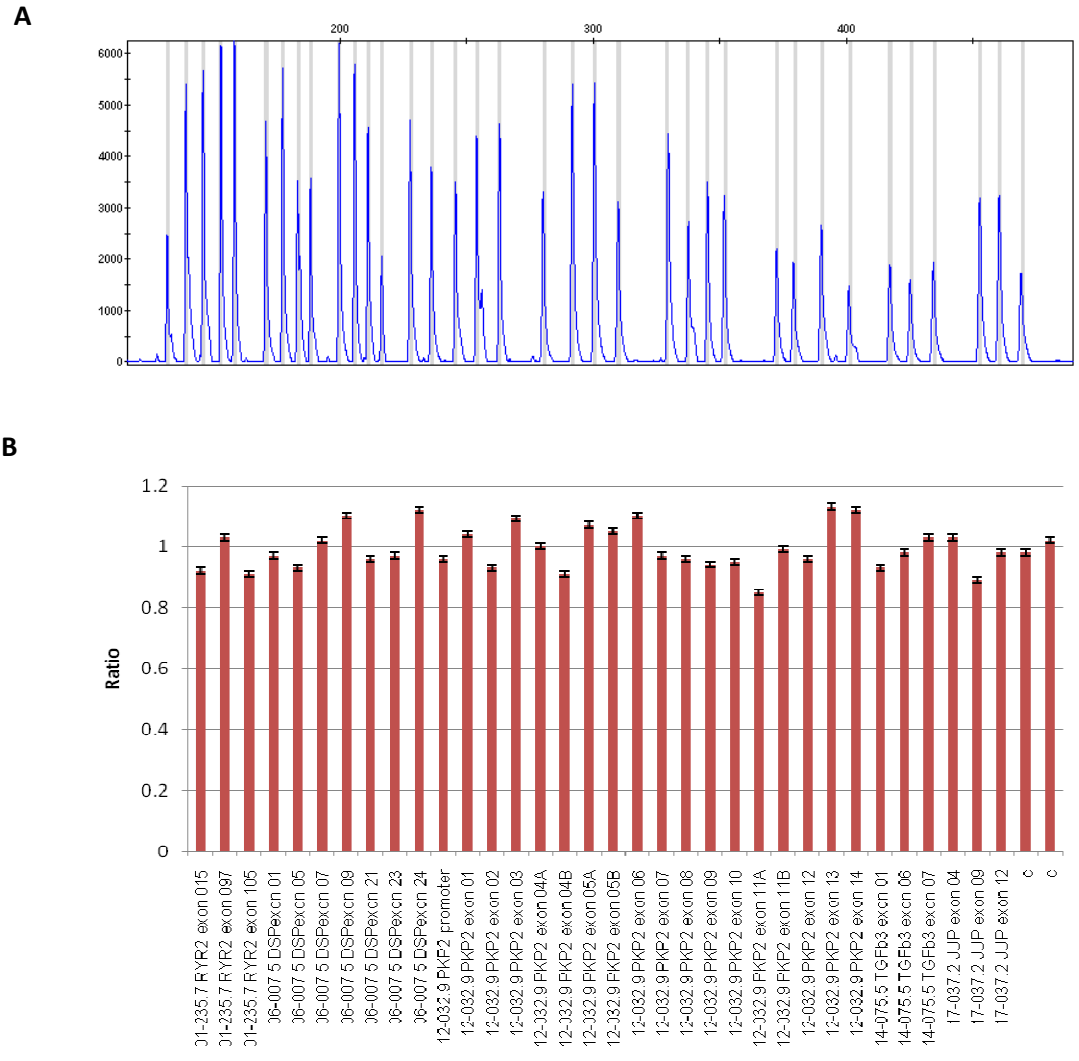


Figure 3.33: MLPA results: (A) MLPA electropherogram and (B) Coffalyser output representative of all ARVC samples screened for deletions/duplications in *DSP*

4. DISCUSSION AND CONCLUSIONS

Genetic mutations in the *DSP* gene are known to cause ARVC. We hypothesised that mutations in the *DSP* gene may also cause familial and idiopathic DCM on the basis of overlap in phenotype between a syndromic form of DSP-associated disease (i.e., Carvajal syndrome) and DCM. We have found evidence to support this hypothesis as several individuals were classified as having causal mutations in the *DSP* gene: 2/62 (3%) of ARVC probands, 1/19 (5%) of index cases with familial DCM and 5/131 (4%) of apparently sporadic cases of idiopathic DCM. In the ARVC cohort, the variants classified as causal included the c.860A>G, c.1534A>G and c.8539A>C variants; in the DCM cohort, the variants classified as causal included the c.1222A>C, c.1228C>T, c.1534A>G, c.2288A>G, c.3085-2A>G, c.3260A>C and c.3806G>A variants. These data suggest that *DSP* mutations may be as important in ARVC as in DCM in South African patients. We also detected 22 known polymorphisms in the ARVC cohort and 23 polymorphisms (six novel, 17 reported) in the DCM cohort.

We found a large number of Genetic Variants of Unknown Significance in this study, a finding that is similar to the experience of other investigators who have screened the desmosomal protein genes for mutations in ARVC (Fressart *et al*. 2010). In the ARVC cohort, these were the c.105G>A, c.1344G>A, c.4022G>A, c.8199G>C and c.8539A>C mutations, while in the DCM cohort, 12 variants were classified as being of unknown significance. Additional studies are required to determine their possible role in disease pathogenesis as previous studies have reported that a change in mRNA folding, and the consequent change in the stability of the mRNA, could lead to altered gene expression levels. mRNA with a more stable secondary structure could be harder for ribosome-associated helicases to unwind, thus limiting mRNA translation and reducing protein levels. On the other hand, more stable structures could cause mRNA persistence, which could potentially increase the rate of gene expression (Parmley and Hurst 2007). Changes in mRNA splicing could also lead to the production of alternative proteins, while changes in protein secondary structure could affect the function of the protein. These changes in the structure and/or function of the DSP protein could ultimately be causative of disease. If the Genetic Variants of Unknown Significance are shown to be pathogenic, the frequency of disease-causing mutations will increase to 10% in ARVC (6/62 probands), and 14% in sporadic DCM (21/150 probands); no Genetic Variants of Unknown Significance were found in patients with familial DCM.

The incidence and prevalence of ARVC is uncertain and varies globally. A literature scan revealed that approximately 39.2%-50% of symptomatic individuals with ARVC possess mutations in one of the genes encoding the desmosomal proteins (*DSP*, *PKP2*, *JUP*, *DSC2* and *DSG2*) (Corrado *et al.* 1998, Awad *et al.* 2008, den Haan *et al.* 2009, Fowler *et al.* 2010, Fressart *et al*. 2010). Based on the reports by den Haan, Marcus, Fressart and Fowler, the frequency of disease-causing mutations in *DSP* ranges from 1% to 20% (Marcus *et al.* 2007, den Haan *et al* 2009, Fowler *et al*. 2010, Fressart *et al*. 2010). The frequency of pathogenic

DSP mutations in our ARVC cohort of 3% is thus comparable to other populations in the world.

We have screened the familial and idiopathic DCM cohorts for pathogenic *DSP* mutations and found causal variants in 5% of index cases with familial DCM and 4% of index cases with idiopathic DCM, and Genetic Variants of Unknown Significance in 9% of patients with idiopathic DCM. To the best of our knowledge, this is the first study of the prevalence of *DSP* mutations in familial and idiopathic DCM. Previous studies have identified *DSP* mutations in a syndromic form of dilated cardiomyopathy that is associated with woolly hair and striate palmoplantar keratoderma (Carvajal syndrome) (Norgett *et al*. 2000, Norgett *et al*. 2006, Uzumcu *et al*. 2006). This is the first study in which a large number of potentially causal genetic mutations were discovered in African patients with familial or idiopathic DCM (Sliwa *et al*. 2005). This is also the first study to demonstrate that *DSP* is causative of non-syndromic familial DCM, as is the case with the c.5178C>A variant, which was found to segregate with disease in a family with DCM.

Another important finding of this study is the allelic heterogeneity demonstrated in *DSP* mutations. Two mutations in *DSP* were found to be causative of both ARVC and DCM. These include the c.1534A>G and c.4022G>A variants which were found in individuals with ARVC and individuals with idiopathic DCM. The fact that the same mutations can lead to different types of cardiomyopathy suggests that ARVC and DCM could form part of a spectrum of heart disease that are caused by defects in desmosomal protein genes.

STUDY LIMITATIONS

One of the limitations of this study is that at the time of completion of this study, pedigree information was not available for the majority of the probands.

Another limitation of this study is that the functional impact of the causal variants and the Genetic Variants of Unknown Significance was not established. Functional studies are required in order to determine the effects these mutations have on the function of the gene and to enable us to postulate the biological pathways by which these mutations could lead to disease.

FUTURE WORK

We intend to extend the pedigrees of the probands, with a view to determining whether the putative disease-causing mutations in the *DSP* gene segregate with disease. Furthermore, we intend to obtain blood samples for RNA extraction, and endomyocardial biopsy samples in affected individuals for studies of protein expression.

The range of functional studies that we intend to perform on the Genetic Variants of Unknown Significance include analysis of mRNA of affected individuals to determine whether silent mutations are associated with cryptic splicing (Awad *et al*. 2006), analysis of

endomyocardial biopsy samples (where indicated on clinical grounds) for expression of desmoplakin, and a yeast-2-hybrid study of protein-protein interactions of mutant DSP protein. Other studies will include immunoblotting and confocal microscopy in human embryonic kidney 293 (HEK293) cell lines transfected to stably express either wild-type or mutant desmoplakin protein.

CONLCUDING REMARKS

We have found a number of disease-causing variants in the *DSP* gene in this study in individuals with ARVC, familial DCM and idiopathic DCM. Future work will focus on confirming the pathogenicity of these variants, as well as determining the functional effects of Genetic Variants of Unknown Significance. Characterisation of the functional effects of these variants would allow us to postulate the way in which these variants may lead to disease. It is also of interest that a number of mutations in *DSP* were deemed to be causative of ARVC and idiopathic DCM, which raises the possibility that ARVC and DCM could form part of a continuous spectrum of cardiomyopathy.

Screening for the causal variants is now available in our laboratory to first degree relatives of affected individuals. Presymptomatic diagnosis by means of genetic screening would allow the early identification of individuals possessing disease-causing variants and who would then require extensive monitoring due to the age-related penetrance of the disease. Genetic screening would also allow the identification of individuals with concealed forms of the disease, and would allow risk stratification for preventative strategies such as the restriction of physical exercise, treatment with antiarrhythmic drugs and therapy with implantable cardioverter-defibrillators.

REFERENCES

Ahmad F., Seidman J. G., Seidman C. E. 2005. *The Genetic Basis for Cardiac Remodelling*. Annual Review of Genomics and Human Genetics 6:185–216

Alcalai R., Metzger S., Rosenheck S., Meiner V., Chajek-Shaul T. 2003. *A Recessive Mutation in Desmoplakin Causes Arrhythmogenic Right Ventricular Dysplasia, Skin Disorder, and Woolly Hair.* Journal of the American College of Cardiology 42:2 319–327

Armstrong D. K. B., McKenna K. E., Purkis P. E., Green K. J., Eady R. A. J., Leigh I. M., Hughes A. E. 1999. *Haploinsufficiency of desmoplakin causes a striate subtype of plamoplantar keratoderma*. Human Molecular Genetics 8 143-148

Asimaki A., Syrris P., Ward D., Guereta L. G., Saffitz J. E., McKenna W. J. 2009. *Unique epidermolytic bullous dermatosiswith associated lethal cardiomyopathy related to novel desmoplakin mutations*. Journal of Cutaneous Pathology 36: 553–559

Awad M. M., Dalal D., Tichnell C., James C., Tucker A., Abraham T., Spevak P. J., Calkins H., Judge D. P. 2006. *Recessive arrhythmogenic right ventricular dysplasia due to novel cryptic splice mutation in* PKP2. Human Mutation 27(11): 1157.

Awad M. M., Calkins H., Judge D. P. 2008. *Mechanisms of Disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy*. Nature Clinical Practice – Cardiovascular Medicine 5:5 258-267

Barahona-Dussault C., Benito B., Campuzano O., Iglesias A., Leung T. L., Robb L., Talajic M., Brugada R. 2009. *Role of genetic testing in arrhythmogenic right ventricular cardiomyopathy/dysplasia*. Clinical Genetics 77: 37–48

Basso C., Czarnowska E., Barbera M. D., Bauce B., Beffagna G., Wlodarska E. K., Pilichou K., Ramondo A., Lorenzon A., Wozniek O., Corrado D., Daliento L., Danieli G. A., Valente M., Nava A., Thiene G., Rampazzo A. 2006. *Ultrastructural evidence of intercalated disc remodelling in arrhythmogenic right ventricular cardiomyopathy: an electron microscopy investigation on endomyocardial biopsies*. European Heart Journal 27: 1847–1854

Basso C., Corrado D., Marcus F. I., Nava A., Thiene G. 2009. *Arrhythmogenic Right Ventricular Cardiomyopathy*. Lancet 373: 1289–1300

Bauce B., Basso C., Rampazzo A., Beffagna G., Daliento L., Frigo G., Malacrida S., Settimo L., Danieli G. A., Thiene G., Nava A. 2005. *Clinical profile of four families with arrhythmogenic right ventricular cardiomyopathy caused by dominant desmoplakin mutations*. European Heart Journal 26: 1666–1675

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

Beffagna G., Occhi G., Nava A., Vitiello L., Ditadi A., Basso C., Bauce B., Carraro G., Thiene G., Towbin J. A., Danieli G. A., Rampazzo A. 2005. *Regulatory mutations in transforming growth factor-h3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1*. Cardiovascular Research 65 366– 373

Bolling M. C., Veenstra M. J., Jonkman M. F., Diercks G. F. H., Curry C. J., Fisher J., Pas H. H., Bruckner A. L. 2010. *Lethal acantholytic epidermolysis bullosa due to a novel homozygous deletion in DSP: expanding the phenotype and implications for desmoplakin function in skin and heart*. British Journal of Dermatology 1-7

Bradshaw D., Groenewald P., Laubscher R., Nannan N., Nojilana B., Norman R., Pieterse D., Schneider M. 2003. *Initial Burden of Disease Estimates for South Africa, 2000.* Cape Town: South African Medical Research Council, 2003.

Burkett E. L., Hershberger R. E. 2005. *Clinical and Genetic Issues in Familial Dilated Cardiomyopathy*. Journal of the American College of Cardiology 45:7 969–981

Cartegni L., Wang J., Zhu Z., Zhang M. Q., Krainer A. R. 2003. *ESEfinder: a web resource to identify exonic splicing enhancers.* Nucleic Acids Research 31:13 3568–3571

Corrado D, Basso C, Schiavon M, Thiene G. 1998. *Screening for hypertrophic cardiomyopathy in young athletes.* New England Journal of Medicine 339:364–369.

Corrado D., Thiene G. 2006. *Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia: Clinical Impact of Genetic Studies*. Circulation 113; 1634-1637

Corrado D., Basso C., Thiene G. 2009. *Arrhythmogenic right Ventricular Cardiomyopathy: An update.* Heart 95: 766-773

Dalal D., Tandri H., Judge D. P., Amat N., Macedo R., Jain R., Tichnell C., Daly A., James C., Russell S. D., Abraham T., Bluemke D. A., Calkins H. 2009. *Morphologic Variants of Familial Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy: A Genetics–Magnetic Resonance Imaging Correlation Study*. Journal of the American College of Cardiology 53:15 1289–99

de Bortoli M., Beffagna G., Bauce B., Lorenzon A., Smaniotto G., Rigato I., Calore M., Mura I. E. A. L., Basso C., Thiene G., Lanfranchi G., Danieli G. A., Nava A., Rampazzo A. 2010. *The p.A897KfsX4 frameshift variation in desmocollin-2 is not a causative mutation in arrhythmogenic right ventricular cardiomyopathy*. European Journal of Human Genetics 1-7

den Haan A. D., Tan B. Y., Zikusoka M. N., Llado L. I., Jain R., Daly A., Tichnell C., James C., Amat-Alarcon N., Abraham T., Russell S. D., Bluemke D. A., Calkins H., Dalal D., Judge D. P. 2009. *Comprehensive Desmosome Mutation Analysis in North Americans With Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy.* Circulation: Cardiovascular Genetics 2; 428-435

Dogan R. I., Getoor L., Wilbur W. J., Mount S. M. 2007. *SplicePort - An interactive splice-site analysis tool*. Nucleic Acids Research 35, Web Server issue W285–W291

Elliott P., Andersson B., Arbustini E., Bilinska Z., Cecchi,F., Charron P., Dubourg O., Kühl U., Maisch B., McKenna W. J., Monserrat L., Pankuweit S., Rapezzi C., Seferovic P., Tavazzi L., Keren A. 2008. *Classification of the cardiomyopathies: A position statement from the european society of cardiology working group on myocardial and pericardial diseases*. European Heart Journal 29: 270–276

Fourlas C. A., Trikas A. G., Stefanadis C. I. 2004. *Familial Dilated Cardiomyopathy: A Genetic Enigma*. Hellenic Journal of Cardiology 45: 42-47

Fowler S. J., Napolitano C., Priori S. G. 2010. *When is genetic testing useful inpatients suspected to have inherited cardiac arrhythmias?* Current Opinion in Cardiology 25:37–45

Fox S. I. 2006. *Human Physiology, Ninth Edition* © The McGraw-Hill Companies, Inc, New York

Fressart V., Duthoit G., Donal E., Probst V., Deharo J.-C., Chevalier P., Klug D., Dubourg O., Delacretaz E., Cosnay P., Scanu P., Extramiana F., Keller D., Hidden-Lucet F., Simon F., Bessirard V., Roux-Buisson N., Hebert J.-L., Azarine A., Casset-Senon D., Rouzet F., Lecarpentier Y., Fontaine G., Coirault C., Frank R., Hainque B., Charron P. 2010. *Desmosomal gene analysis in arrhythmogenic right ventricular dysplasia/cardiomyopathy: spectrum of mutations and clinical impact in practice*. Europace Advance Access Publication doi:10.1093/europace/euq104

Gallicano G. I., Kouklis P., Bauer C., Yin M., Vasioukhin V., Degenstein L., Fuchs E. 1998. *Desmoplakin Is Required Early in Development for Assembly of Desmosomes and Cytoskeletal Linkage*. The Journal of Cell Biology 143 2009-2022

Garret R. H., Grisham C. M. 2005. *Biochemistry (3rd Edition)* Brooks/Cole

Gaziano T. A. 2005. *Cardiovascular Disease in the Developing World and its Cost-Effective Management*. Circulation 112: 3547-3553

Gaziano T. A. 2007. *Reducing The Growing Burden Of Cardiovascular Disease In The Developing World*. Health Affairs 26:1 13–24

Hamid M. S., Norman M., Quraishi A., Firoozi S., Thaman S., Gimeno J. R., Sachdev B., Rowland E., Elliott P. M., McKenna W. J. 2002. *Prospective Evaluation of Relatives for Familial Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia Reveals a Need to Broaden Diagnostic Criteria.* Journal of the American College of Cardiology 40:8 1445–50

Hartwell L. H., Hood L., Goldberg M., Reynolds A. E., Silver L. M., Veres R. C. 2004. *Genetics: From Genes to Genomes (2nd Edition)* McGraw Hill New York

Hebsgaard S.M., Korning P.G., Tolstrup N., Engelbrecht J., Rouze P., Brunak S. 1996. *Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information*. Nucleic Acids Research 24:17, 3439-3452

Hofacker I. L. 2003. *Vienna RNA secondary structure server*. Nucleic Acids Research 31:13 3429–3431

HRM Assay Design and Analysis CorProtocol™ 6000-1-July06. © Corbett Life Sciences 2006

Hughes S.E., McKenna W. J. 2005. *New Insights into the pathology of Inherited Cardiomyopathies*. Heart 2005 91: 257–264

Jonkman M. F., Pasmooij A. M. G., Pasmans S. G. M. A., van den Berg M. P., ter Horst H. J., Timmer A., Pa H. H. 2005. *Loss of Desmoplakin Tail Causes Lethal Acantholytic Epidermolysis Bullosa.* American Journal of Human Genetics 77:653–660

Khogali S. S., Mayosi B. M., Beattie J. M., McKenna W. J., Watkins H., Poulton J. 2001. *A common mitochondrial DNA variant associated with susceptibility to dilated cardiomyopathy in two different populations*. Lancet 357: 1265-7

Kiès P., Bootsma M., Bax J., Schalij M. J., van der Wall E. E. 2006. *Arrhythmogenic right ventricular dysplasia/cardiomyopathy: Screening, diagnosis, and treatment.* Heart Rhythm 3: 225–234

Kolar A.J.O., Milroy C.M., Day P.F., Suvarna S.K. 2008. *Dilated cardiomyopathy and sudden death in a teenager with palmar-plantar keratosis (occult Carvajal syndrome)* Journal of Forensic and Legal Medicine 15 185-188

Lai Cheong J. E., Wessagowit V., McGrath J. A. 2005. *Molecular abnormalities of the desmosomal protein desmoplakin in human disease*. Clinical and Experimental Dermatology 30: 261–266

Luckey J. A., Drossman H., Kostichka A. J., Mead D. A., D'Cunha J., Norris T.B., Smith L. M. 1990. *High speed DNA sequencing by capillary electrophoresis*. Nucleic Acids Research 18:15 4417-4421

Luk A., Ahn E., Soor G. S., Butany J. 2009. *Dilated Cardiomyopathy: A Review*. Journal of Clinical Pathology 62: 219-225

Marcus F. I., Nava A., Thiene G. 2007. *Arrhythmogenic RV Cardiomyopathy/Dysplasia: Recent Advances*. © Springer Publishers. Pages 7-14

Marcus F. I., McKenna W. J. , Sherrill D., Basso C., Bauce B., Bluemke D. A., Calkins H., Corrado D., Cox M. G. P. J., Daubert J. P., Fontaine G., Gear K., Hauer R., Nava A., Picard M. H., Protonotarios N., Saffitz J. E., Sanborn D. M. Y., Steinberg J. S., Tandri H., Thiene G., Towbin J. A., Tsatsopoulou A., Wichter T., Zareba W. 2010. *Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia - Proposed Modification of the Task Force Criteria.* Circulation 121;1533-1541

Mathe E., Olivier M., Kat S., Ishioka C., Hainaut P., Tavtigian S. V. 2006. *Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods*. Nucleic Acids Research 34:5 1317–1325

Matolweni L. O., Bardien S., Rebello G., Oppon E., Munclinger M., Ramesar R., Watkins H., Mayosi B. M. 2006. *Arrhythmogenic right ventricular dysplasia/cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes.* BMC Medical Genetics 7:29

McGuffin L. J., Bryson K., Jones D. T. 2000. *The PSIPRED protein structure prediction server.* Bioinformatics Applications Note 16:4 404-405

McKenna W. J., Thiene G., Nava A., Fontaliran F, Blomstrom-Lundqvist C., Fontaine G., Camerini F. 1994. *Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy*. British Heart Journal 71:215-218

McKoy G., Protonotarios N., Crosby A., Tsatsopoulou A., Anastasakis A., Coonar A., Norman M., Baboonian C., Jeffery S., McKenna W. J. 2000. *Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease)*. Lancet 355: 2119–24

Merner N. D., Hodgkinson K. A., Haywood A. F. M., Connors S., French V. M., Drenckhahn J.-D., Kupprion C., Ramadanova K., Thierfelder L., McKenna W., Gallagher B., Morris-Larkin L., Bassett A. S., Parfrey P. S., Young T.-Y. 2008. *Arrhythmogenic Right Ventricular Cardiomyopathy Type 5 Is a Fully Penetrant, Lethal Arrhythmic Disorder Caused by a Missense Mutation in the* TMEM43 *Gene*. The American Journal of Human Genetics 82, 809–821

Moolman-Smook J. C., Mayosi B. M., Brink P. A., Corfield V. A. 2003. *Molecular genetics of cardiomyopathy: changing times, shifting paradigms*. Cardiovascular Journal of South Africa 14:3 145-155

Ng P. C., Henikoff S. 2003. *SIFT: predicting amino acid changes that affect protein function*. Nucleic Acids Research 31:13 3812–3814

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

Norgett E. E., Lucke T. W., Bowers B., Munro C. S., Leigh I. M. , Kelsell D. P. 2006. *Early Death from Cardiomyopathy in a Family with Autosomal Dominant Striate Palmoplantar Keratoderma and Woolly Hair Associated with a Novel Insertion Mutation in Desmoplakin*. Journal of Investigative Dermatology 126: 1651–1654

Norman M., Simpson M., Mogensen J., Shaw A., Hughes S., Syrris P., Sen-Chowdhry S., Rowland E., Crosby A., McKenna W. J. 2005. *Novel Mutation in Desmoplakin Causes Arrhythmogenic Left Ventricular Cardiomyopathy*. Circulation 112: 636-642

Opie L. H., Mayosi B. M. 2005. *Cardiovascular Disease in Sub-Saharan Africa.* Circulation 112: 3536-3540

Parmley J. L., Hurst L. D. 2007. *How do synonymous mutations affect fitness?* BioEssays 29:515–519

Ramensky V. Bork P., Sunyaev S. 2002. *Human non-synonymous SNPs: server and survey*. Nucleic Acids Research 30:17 3812–3814

Rampazzo A., Nava A., Malacrida S., Beffagna G., Bauce B., Rossi V., Zimbello R., Simionati B., Basso C., Thiene G., Towbin J. A., Danieli G. A. 2002. *Mutation in Human Desmoplakin Domain Binding to Plakoglobin Causes a Dominant Form of Arrhythmogenic Right Ventricular Cardiomyopathy.* American Journal of Human Genetics 71: 1200–1206

Schönberger J., Seidman C. E. 2001. *Many Roads Lead to a Broken Heart: The Genetics of Dilated Cardiomyopathy.* American Journal of Human Genetics 69:249–260

Sen-Chowdhry S., Syrris P., McKenna W. J. 2007. *Role of Genetic Analysis in the Management of Patients With Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy*. Journal of the American College of Cardiology 50:19 1813-1821

Sen-Chowdhry S., Syrris P., Prasad S. K., Hughes S. E., Merrifield R., Ward D., Pennell D. J, McKenna W. J. 2008. *Left-Dominant Arrhythmogenic Cardiomyopathy: An Under-Recognized Clinical Entity.* Journal of the American College of Cardiology 52:25 2175–87

Sen-Chowdhry S., Morgan R. D., Chambers J. C., McKenna W. J. 2010. *Arrhythmogenic Cardiomyopathy: Etiology, Diagnosis and Treatment*. Annual Review of Medicine 61:233–53

Shapiro M. B., Senapathy P. 1987. *RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression*. Nucleic Acids Research 15:17 7155-7174

Sliwa K., Damasceno A., Mayosi B. M. 2005. *Epidemiology and Etiology of Cardiomyopathy in Africa.* Circulation 112: 3577-3583

Stokes D. L. 2007. *Desmosomes from a structural perspective*. Current Opinion in Cell Biology 19: 565–571

Tanaka A., Lai-Cheong J. E., Café M. E. M., Gontijo B., Salomão P. R., Pereira L., McGrath J. A. 2008. *Novel truncating mutations in PKP1 and DSP cause similar skin phenotypes in two Brazilian families*. British Journal of Dermatology 160, pp692–697

Tavtigian S. V., Deffenbaugh A. M., Yin L., Judkins T., Scholl T., Samollow P. B., de Silva D., Zharkikh A., Thomas A. 2006. *Comprehensive statistical study of 452* BRCA1 *missense substitutions with classification of eight recurrent substitutions as neutral.* Journal of Medical Genetics 43:295–305

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

Uzumcu A., Norgett E. E., Dindar A., Uyguner O., Nisli K., Kayserili H., Sahin S. E., Dupont E., Severs N. J., Leigh I. M., Yuksel-Apak M., Kelsell D. P., Wollnik B. 2006. *Loss of desmoplakin isoform I causes early onset cardiomyopathy and heart failure in a Naxos-like syndrome*. Journal of Medical Genetics 43 e05 Available at: http://www.jmedgenet.com/cgi/content/full/43/2/e5

van der Zwaag P. A., Jongbloed J. D. H., van den Berg M. P., van der Smagt J. J., Jongbloed R., Bikker H., Hofstra R. M. W., van Tintelen J. P. 2009. *A Genetic Variants Database for Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy.* Human Mutation 30:9 1278–1283

Watkins D., Hendricks N., Shaboodien G., Mbele M., Parker M., Vezi B. Z., Latib A., Chin A., Little F., Badri M., Moolman-Smook J. C., Okreglicki A., Mayosi B. M. 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: S10–S17

Wexler R., Elton T., Pleister A., Feldman D. 2009. *Cardiomyopathy: An Overview*. American Family Physician 79:9 778-784

Whittock N. V., Wan H., Morley S. M., Garzon M. C., Kristal L., Hyde P., McLean W. H. I., Pulkkinen L., Uitto J., Christiano A. M., Eady R. A. J., McGrath J. . 2002. *Compound Heterozygosity for Non-Sense and Mis-Sense Mutations in Desmoplakin Underlies Skin Fragility/Woolly Hair Syndrome.* Journal of Investigative Dermatology 118:2 232-238

Xu T., Yang Z., Vatta M., Rampazzo A., Beffagna G., Pillichou K., Scherer S. E., Saffitz J., Kravitz J., Zareba W., Danieli G. A., Lorenzon A., Nava A., Bauce B., Thiene G., Basso C., Calkins H., Gear K., Marcus F., Towbin J. A. 2010. *Compound and Digenic Heterozygosity Contributes to Arrhythmogenic Right Ventricular Cardiomyopathy*. Journal of the American College of Cardiology 55:6 587–97

Yang Z., Bowles N. E., Scherer S. E., Taylor M. D., Kearney D. L., Ge S., Nadvoretskiy V. V., DeFreitas G., Carabello B., Brandon L. I., Godsel L. M., Green K. J., Saffitz J. E., Li H., Danieli G. A., Calkins H., Marcus F., Towbin J. A. 2006. *Desmosomal Dysfunction due to Mutations in Desmoplakin Causes Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy*. Circulation Research 99: 646-655

Yu C.-C., Yu C.-H., Hsueh C.-H., Yang C.-T., Juang J. M., Hwang, Lin J.-J., Lai L.-P. 2008. *Arrhythmogenic Right Ventricular Dysplasia: Clinical Characteristics and Identification of Novel Desmosome Gene Mutations*. Journal of the Formos Medical Association 107(7):548–558

Zhou X., Stuart A., Dettin L. E., Rodriguez G., Hoel B., Gallicano G. I. 2004. *Desmoplakin is required for microvascular tube formation inc culture.* Journal of Cell Science 117, 3129-3140

Zuker M. 2003. *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research 31:13 3429–3431

ELECTRONIC RESOURCES

1. Structure of the heart image: Last Date accessed: 10 August 2010

   http://www.ohiohealth.com/bodymayo.cfm?id=6&action=thumbnail&image=/images/image_popup/ch7_theheart.jpg&caption=A normal heart has two upper and two lower chambers. The upper chambers, the right and left atria, receive incoming blood. The lower chambers, the more muscular right and left ventricles, pump blood out of your heart. The heart valves, which keep blood flowing in the right direction, are gates at the chamber openings. &title=Atrial septal defect (ASD)

2. Layers of the heart image: Last Date accessed: 10 August 2010

   http://www.ohiohealth.com/bodymayo.cfm?id=6&action=thumbnail&image=/images/image_popup/r7_heartmuscle.jpg&caption=Your%20heart%20wall%20has%20several%20layers.%20An%20inner%20layer%20(endocardium),%20a%20middle%20layer%20of%20muscle%20(myocardium)%20and%20an%20outer%20layer%20(epidcardium).%20A%20thin,%20saclike%20membrane%20(pericardium)%20surrounds%20the%20heart.%20&title=Cardiomyopathy

3. Arrhythmogenic right ventricular cardiomyopathy heart cross-section image:

   Last Date accessed: 10 August 2010

   http://www.ottawaheart.ca/images/v3i2_-_heart_tissue_ENG.jpg

4. Dilated cardiomyopathy image:

   Last Date accessed: 10 August 2010

   http://www.ohiohealth.com/bodymayo.cfm?id=6&action=thumbnail&image=/images/image_popup/mcdc7_dilatedcardiomyopathy.jpg&caption=Compared with a healthy heart, dilated cardiomyopathy causes the chambers of the heart to enlarge, which can lead to heart failure if left untreated. &title=Dilated cardiomyopathy

5. Wikipedia: High Resolution Melt:

   Last Date accessed: 10 August 2010
   http://en.wikipedia.org/wiki/High_Resolution_Melt

6. Gene2Promoter:

   Last Date accessed: 10 August 2010

   http://www.genomatix.de/online_help/help_eldorado/Gene2Promoter_Intro.html
   (Note: Access to this tool is restricted, and registration is required to gain access)

APPENDICES

APPENDIX 1

ARVC STUDY PATIENT CONSENT FORM

UNIVERSITY OF CAPE TOWN

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

Date ___/___/_____
DD MM YEAR

Informed Consent Form

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

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

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

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

REQUEST FOR MOLECULAR STUDIES (DNA)

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

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

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

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname: _____ First Name(s): _____

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

Medical Aid: _____ Medical Aid No: _____

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

Number of children: _____

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

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

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

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

Reason for Referral (Clinical diagnosis):

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

Arrhythmogenic Right Ventricular Cardiomyopathy ☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

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

Other: _____

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

If Yes, where: _____

For Laboratory use only:

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

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

CONSENT FOR DNA ANALYSIS AND STORAGE

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

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

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

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

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

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

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

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

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

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

APPENDIX 2

SOLUTIONS

Loading Buffer

1 ml 5X GoTaq® Flexi Buffer (Promega)

5 µl GelRed™ Nucleic Acid Gel Stain (Biotium)

1X TBE Buffer

10.8 g Tris (Promega)

5.5 g Boric Acid (Riedel-de Haën)

7.4 g EDTA (Promega)

Make up to 1 litre with dH_2O

1.5% Agarose Gel

1.5 g Agarose (Roche)

Volume made up by 100 ml 1X TBE Buffer

3% Agarose Gel

3 g Agarose (Roche)

Volume made up by 100 ml 1X TBE Buffer

APPENDIX 3

100 bp DNA LADDER (PROMEGA)

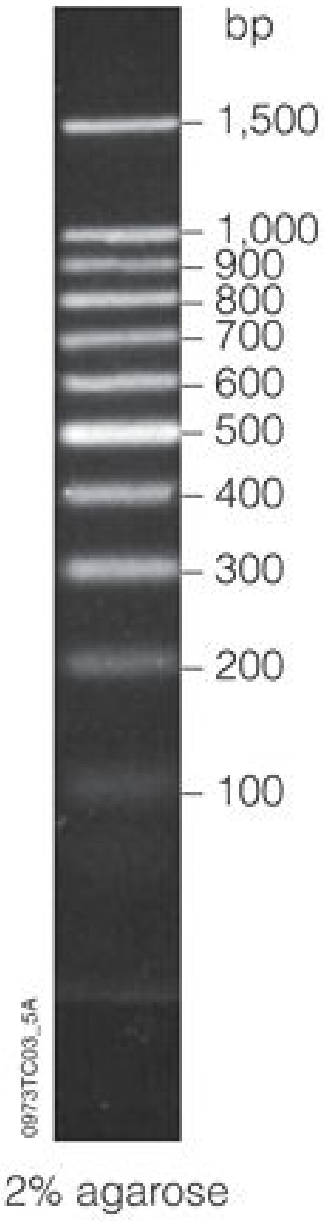


Image available from:
http://www.promega.com/catalog/catalogproducts.aspx?categoryname=productleaf_179

White bands represent DNA fragments of a known range of sizes, as indicated. This is run alongside DNA samples of interest on an agarose gel to determine the size of the fragment.

APPENDIX 4

MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION (MLPA) PROBES

MLPA was performed using the SALSA MLPA Kit P168 ARVC-PKP2. This kit contains 18 probes for *PKP2,* 7 probes for *DSP* and 3 probes each for *JUP*, *TGFβ3* and *RyR2*. The probes in this kit are listed below, with an indication of the regions which these probes cover.

Length (nt)	SALSA MLPA probe	Chromosomal position		
		reference	PKP2	DSP
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA			
92	Synthetic control probe at 2q14			
136	Reference probe 0662-L0158	6p21.3		
142	**PKP2 probe** 5383-L4780		**Promoter**	
148	**PKP2 probe** 5391-L4788		**Exon 8**	
154	**DSP probe** 5417-L4827			**Exon 1**
159	**TGFb3 probe** 5414-L4824	**TGFb3 exon 1**		
172	**PKP2 probe** 5384-L4781		**Exon 1**	
178	**PKP2 probe** 5392-L4789		**Exon 9**	
184	**DSP probe** 5418-L4828			**Exon 5**
190	**JUP probe** 5424-L4834	**JUP exon 4**		
202	**PKP2 probe** 5385-L4782		**Exon 2**	
208	**PKP2 probe** 5393-L4790		**Exon 10**	
213	**DSP probe** 5419–L4829			**Exon 7**
218	Reference probe 2107-L0794	14q24.3		
229	**PKP2 probe** 5934-L5803		**Exon 5**	
238	**PKP2 probe** 5386-L4783		**Exon 3**	
247 †	**PKP2 probe** 5394-L4791		**Exon 11**	
256 *	**DSP probe** 5420-L4830		**Exon 9**	**Exon 9**
265	**TGFb3 probe** 5415-L4825	**TGFb3 exon 6**		
281	**PKP2 probe** 5387-L4784		**Exon 4**	
293	**PKP2 probe** 5395-L4792		**Exon 12**	
302	**RYR2 probe** 5935-L4837	RYR2 exon 3		
310	**DSP probe** 5936-L4831			**Exon 21**
330	**PKP2 probe** 5388-L4785		**Exon 5**	
337	**PKP2 probe** 5396-L4793		**Exon 13**	
346	**DSP probe** 5422-L4832			**Exon 23**
353	**RYR2 probe** 5428-L4838	**RYR2 exon 97**		
374	**PKP2 probe** 5389-L4786		**Exon 6**	
381	**PKP2 probe** 5397-L4794		**Exon 14**	
392 §	**DSP probe** 5423-L4833			**Exon 24**
400 §	**TGFb3 probe** 5416-L5804	**TGFb3 exon 7**		
419	**PKP2 probe** 5390-L4787		**Exon 7**	
427	**PKP2 probe** 5411-L4821		**Exon 4**	
436	**JUP probe** 5426-L4836	**JUP exon 12**		
454	**JUP probe** 5937-L4835	JUP exon 9		
463	**PKP2 probe** 5413-L4823		**Exon 11**	
472	**RYR2 probe** 5429-L4839	**RYR2 exon 105**		

† Less reliable.
* Shoulder peak.
§ At 396 nt an unspecific peak is present.

APPENDIX 5

INTERPRETATION OF HIGH RESOLUTION MELT GRAPHS

High resolution melt (HRM) analysis was used in this study for the screening of the ARVC and DCM cohorts for mutations. HRM runs were analysed by comparing the melting curves from HRM experiments of control samples to the melting curves of diseased samples from our cohort, and determining whether there was a difference in melting behaviour between the diseased and control sample amplicons. By reference to the melting curves plotted by the RotorGene 6000 software, it is possible to distinguish wildtype/control samples from mutation homozygotes and heterozygotes. In addition to this, the RotorGene software can be used to plot a derivative graph that indicates the rate of change of fluorescence over increasing temperature, which can be used in conjunction with the original melting curve to detect changes in amplicon melting behaviour.

Figure 1 shows results from HRM experiments for DNA sequence variants identified in this study. Figure 1(A) shows the HRM curve for an amplicon with a single melting domain, indicating the wildtype and mutant samples, which are clearly distinguishable by their melting behaviour. Figure 1(B) indicates the derivative graph of the HRM curve in Figure 1(A), which can be used in conjunction with the HRM curve to detect changes in amplicon melting behaviour. Figure 1(C) indicates the HRM curve for an amplicon with 2 melting domains, showing wildtype and mutant samples, while figure 1(D) shows the derivative graph for this HRM curve. As shown, the 2 melting domains are clearly distinguishable with HRM. As shown in Figures 1(E) and (F), it is also possible to distinguish 3 melting domains in an amplicon by means of HRM.

Figure 2(A) shows another melting curve of an amplicon with 2 melting domains, with a close-up view of the melting domain with inter-sample changes in Figure 2(B). As shown, wildtype samples can be distinguished from the heterozygote and homozygote mutant samples with the melt curve (Figure 2(A) & (B)) as well as the derivative graph (Figure 2(C)).

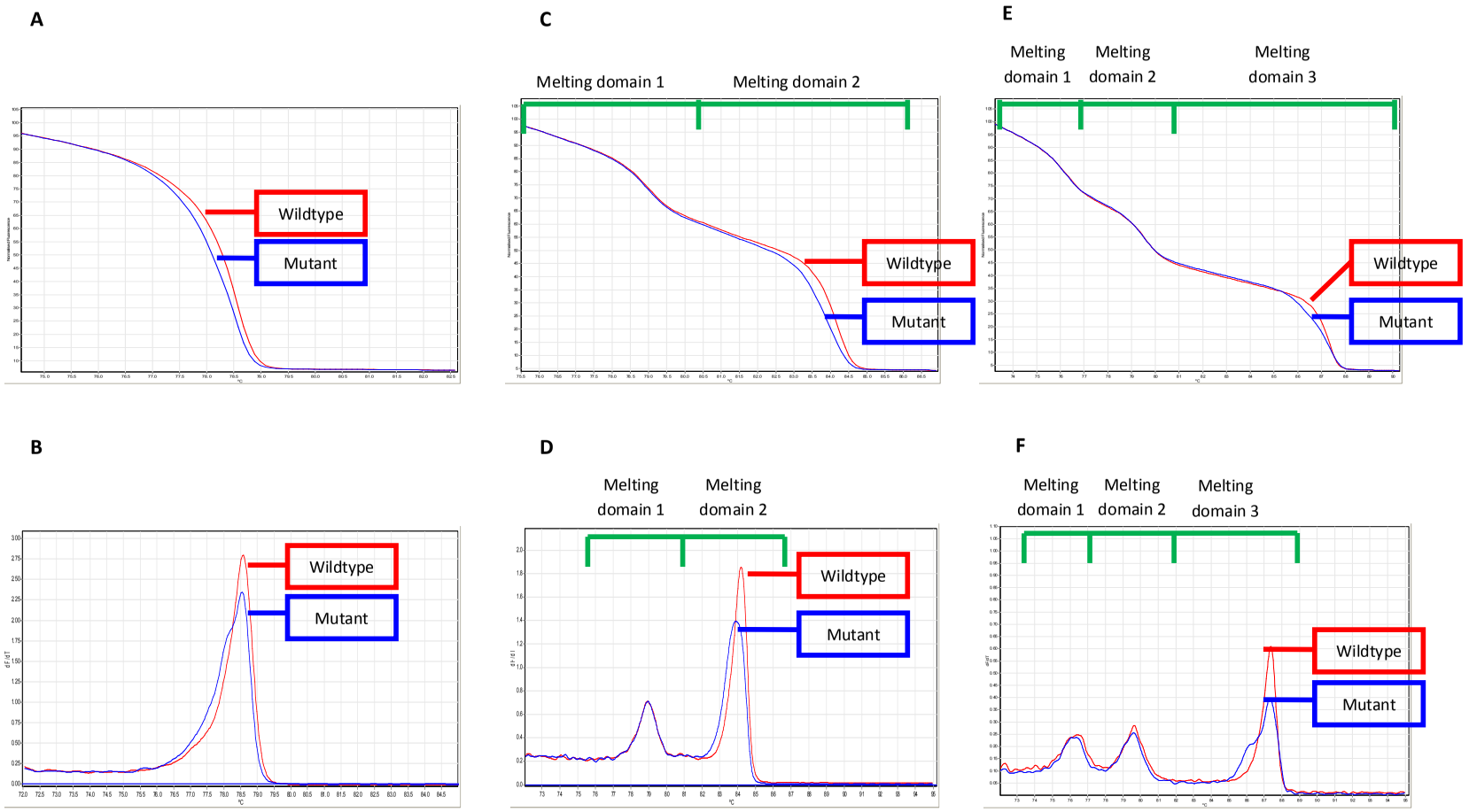


Figure 1: (A) Melt curve of amplicon with 1 melting domain; (B) Derivative graph of melt curve in (A); (C) Melt curve of amplicon with 2 melting domains; (D) Derivative graph of melt curve in (C); (E) Melt curve of amplicon with 3 melting domains; (F) Derivative graph of melt curve in (E)

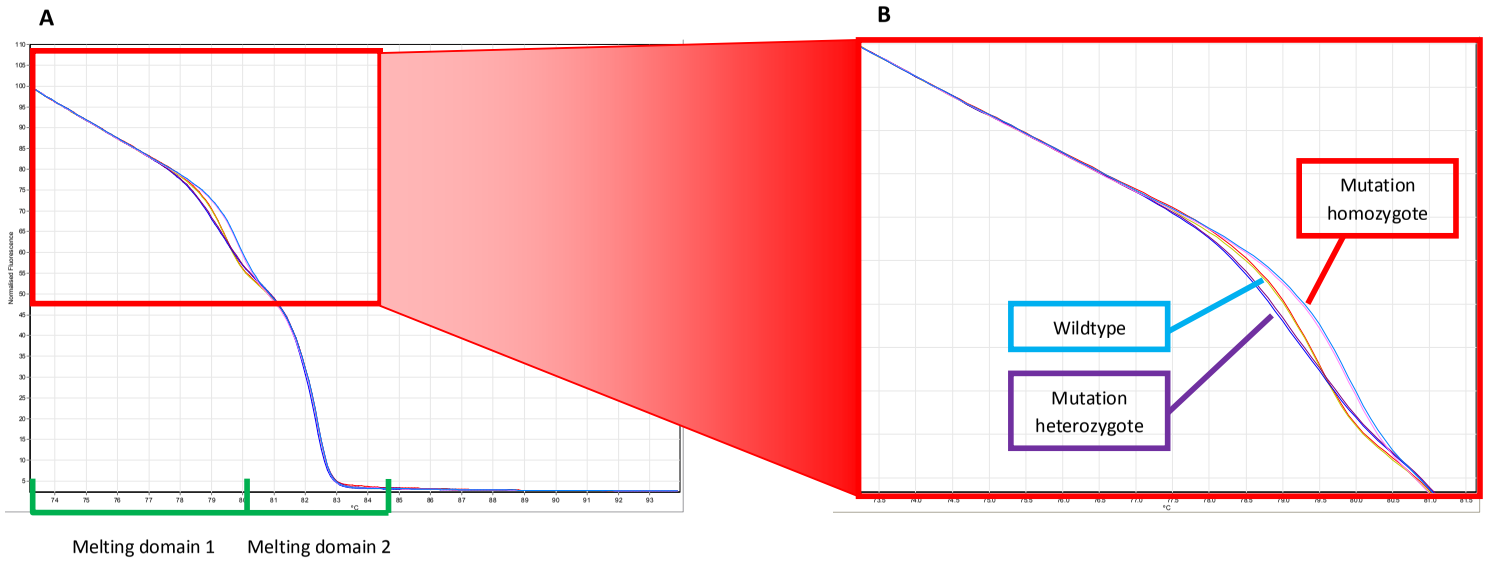


Figure 34: (A) Melt curve of amplicon with 2 melting domains
(B) Close-up view of melting domain showing inter-sample changes
(C) Derivative graph of melt curve in (A)