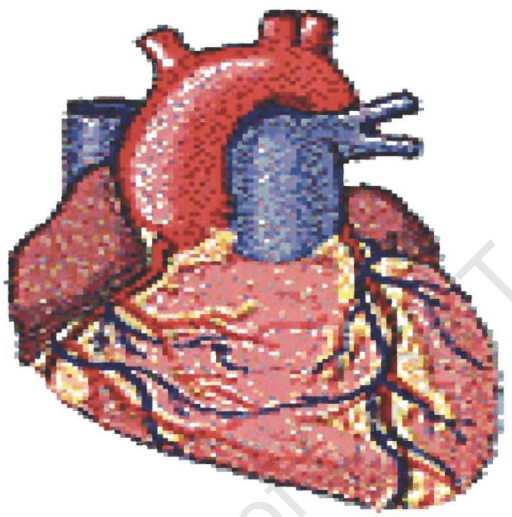
The prevalence of Plakophilin-2 gene (PKP-2) mutations in South African patients with Cardiomyopathy



By

Mzwandile Mbele

(MBLMZW001)

MSc (MED) Human Genetics

Faculty of Health Science

University Of Cape Town

15 August 2008

Supervisor: Dr Gasna Shaboodien, Professor Hanlie Moolman- Smook and Professor Bongani Mayosi

TABLE OF CONTENTS

DECLARATION

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

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 15 August 2008

Prevalence of Plakophilin 2 gene (PKP-2) Mutations in South African patients with Cardiomyopathy.

Mzwandile Mbele

The Cardiac Clinic and Cardiovascular Genetics Laboratory,

ABSTRACT

Genetic mutations in the plakophilin-2 (*PKP2*) gene cause 25-50% of cases of inherited arrhythmogenic right ventricular cardiomyopathy (ARVC), a disease that extends to the left ventricle of the heart in nearly half of cases. The prevalence of *PKP2* gene mutations in South African patients with ARVC and left ventricular forms of cardiomyopathy (such as dilated and restrictive cardiomyopathies) is not known. The aim of the study was to determine the prevalence of *PKP2* gene mutations in patients with ARVC, dilated cardiomyopathy (DCM) and restrictive cardiomyopathy (RCM) patients.

Fourty-five DNA samples from unrelated probands with ARVC (n= 25), DCM (n=19) and RCM (n=1) patients were screened for variants in the *PKP2* gene using the denaturing high-performance liquid chromatography and sequencing. Population frequencies of novel PKP2 variants were determined in control individuals by SNaPshot and restriction enzyme digestions. Haplotype analysis was performed to establish if a founder effect existed in individuals harbouring the same mutation. The T2540C (L-P) was identified in only DCM patients was found to be a rare familial polymorphism.

We found six common polymorphisms (four novel) and three disease-causing mutations in our ARVC cohort. The mutations were (1) a C1132T transition in exon 4 (2) a reported insertion/deletion in exon11 (2197-2202delCACACCinsG) and (3) a reported intronic splice-site mutation in intron 11 (IVS2146-1G-C). The C1132T mutation results in a novel premature stop codon (Q378X) in two affected brothers in family1. The insertion/deletion mutation causes a premature truncation of PKP2 and was found in two affected siblings of family19. The intronic splice-site mutation is known to activate cryptic slice acceptor sites in either intron 12 or exon 13 and was found in an isolated proband. A novel fourth variant, the C1162T (R388W), occurred in three ARVC probands (of families 5, 12 and 19) with the same genetic background.

This variant was not found in >400 control chromosomes but was found in the unaffected mother of family19. The two affected siblings in family19 co-inherited the rare C1162T variant and the insertion/deletion mutation. The severe phenotype observed in these children suggests a disease-modifying effect for this novel variant. The possible disease-causing or disease-modifying role of the C1162T variant in ARVC was supported by the severe phenotype observed in the individuals who co-inherited the 2197-2202delCACACCinsG insertion/deletion mutation.

In our cohort, we have found the prevalence rate of *PKP2* gene mutations to be 12%, which is much lower than the reported global frequencies. Additionally, we postulate that the C1162T variant may have a functional effect on PKP2 that confers a gene-dose effect on the phenotype. We also postulate that this variant, which occurred in probands with the same genetic background, may represent a founder effect.

ACKNOWLEDGEMENTS

I am grateful to the **patients and family members** who participated in this study.

I would like to thank my supervisor, **Professor Bongani Mayosi,** for his guidance, insight and enthusiasm throughout this project. His patience and input in the writing of this dissertation has been very important.

I would also like to thank my co-supervisors, **Professor Hanlie Moolman- Smook and Dr Gasnat Shaboodien**, for his guidance and input in this study, if it was not because of them was going to be impossible for me.

To all the members of **Division of Human Genetics**, you made my life very easy. A big THANK-YOU to **George Rebello**, **Alvera Vorster and Lisa Roberts** for all the assistance and advices.

I am grateful to the **Department of Medicine**, the **University of Cape Town**, and Medical Research Council for their financial assistance.

I thank **my parents** for their support and encouragement throughout my studies.

List of figures

Page

Page

List of tables Page

List of Abbreviations and symbols

ACM = Arrhythmogenic Left Ventricular Cardiomyopathy

AD = Autosomal Dominant

ALVC = Arrhythmogenic Left Ventricular Cardiomyopathy

APS = ammonium persulphate

AR = Autosomal Recessive Anterior Polar Cataract

ARVC = Arrhythmogenic Right Ventricular Cardiomyopathy

BDGP = Berkeley Drosophila Genome Project

BLAST = Basic Local Alignment Search Tool

bp = Base pairs

Ca^{2+} = Calcium

DCM = Dilated cardiomyopathy

dH_20 = Distilled water

DHPLC = Denaturing high-performance liquid chromatography

ddNTP = dideoxy-nucleotide triphosphate

Del = Deletion

DNA = Deoxyribonucleic acid

dNTPs = Deoxytrinucleotide phosphate

DSC = Desmocollin

DSG2 = Desmoglein-2

DSP = Desmoplakin

ECG = Electrocardiogram

F = Forward primer

H_20 = Water

HCl = Hydrochloric acid

HCM = Hypertrophic cardiomyopathy

°C = degrees Celsius

F = Forward primer

FDC = Familial dilated cardiomyopathy

RCM = Restrictive cardiomyopathy

Ins = Insertion

IVS = Intervening sequence

JUP = Plakoglobin

Kcal/mol = Kilocalories per mole

KCl = Potassium chloride

LVNC = Left Ventricular Non-Compaction

$MgCl_2$ = Magnesium chloride

MIM = Mendelian Inheritance in Man

µg = Microgram

µl = Microliter

µM = MicroMolar

MIM = Mendelian Inheritance in Man

MRC = Medical Research Council

MRI = Magnetic resonance image

NCBI = National Center for Biotechnology Information

PAGE = Polyacrylamide gel electrophoresis

PCR = Polymerase chain reaction

PKP2 = Plakophilin-2

PM = Plasma membrane

Q = Glutamine

R = Reverse primer

R = Arginine

RFLP = Restriction Fragment Length Polymorphism

R = Reverse primer

RCM = Restrictive cardiomyopathy

RV = Right ventricle

RyR2 = Cardiac ryanodine receptor

SAP = Shrimp Alkaline Phosphatase

SNP = Single nucleotide polymorphism

STR = Short tandem repeat

SV = Sequence variants

Ta = Annealing temperature

Tm = Melting temperature

TBE buffer = Tris Borate EDTA buffer

TEMED = N'N'N'N'-tetramethylethylenediamine

TGFβ3 = Transforming growth factor-β3

UCT = University of Cape Town

UTR = Untranslated region

V = Volts

USA = United State of America

UV = Ultra violet

w/v = Weight per volume

-/+ = up/downstream of the respective exon

% = percent

Chapter 1: The prevalence of plakophilin-2 gene mutations in South African patients with cardiomyopathy

1.1 Structure of the heart

The wall of the cardiac chambers is composed of three layers: the epicardium, myocardium and endocardium which are surrounded by the pericardium (figure 1). The cells of the myocardium are striated and contain actin and myosin filaments that are arranged in the form of sarcomere, the contractile unit of the myocardial cell.

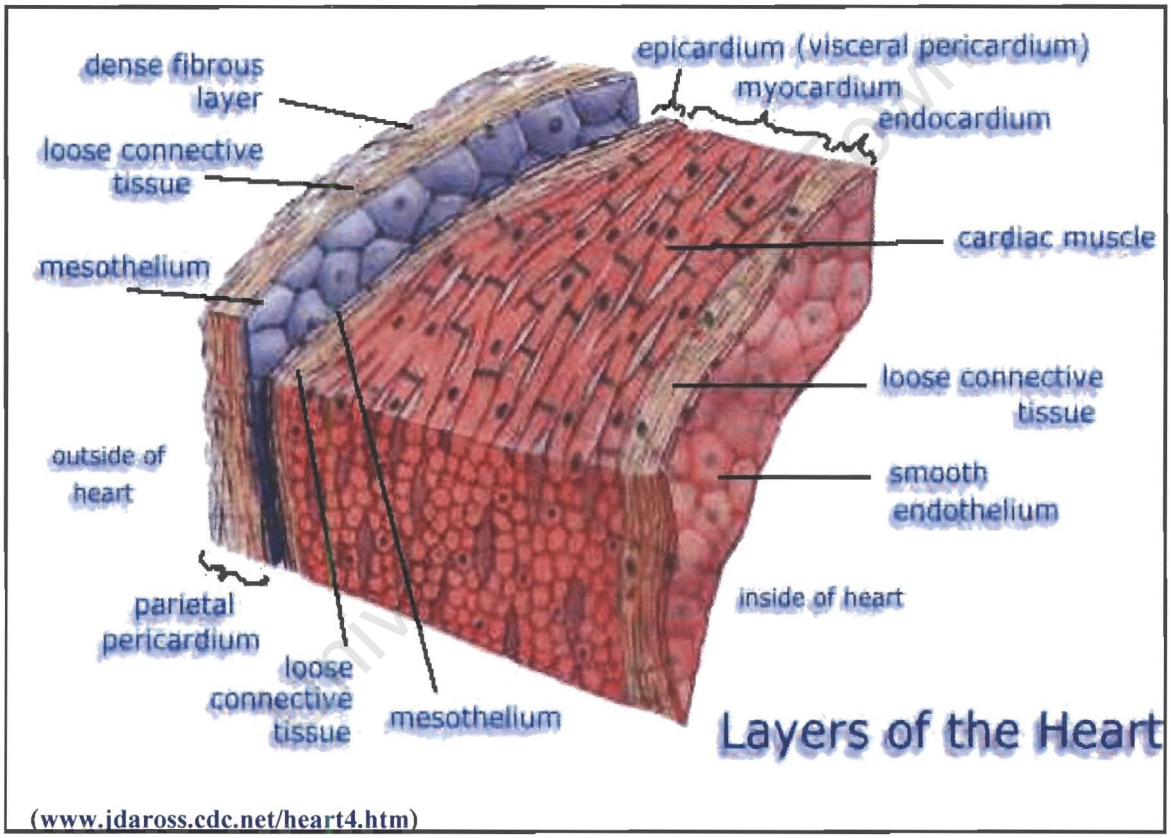


Figure 1: A section of the heart wall from the right ventricle

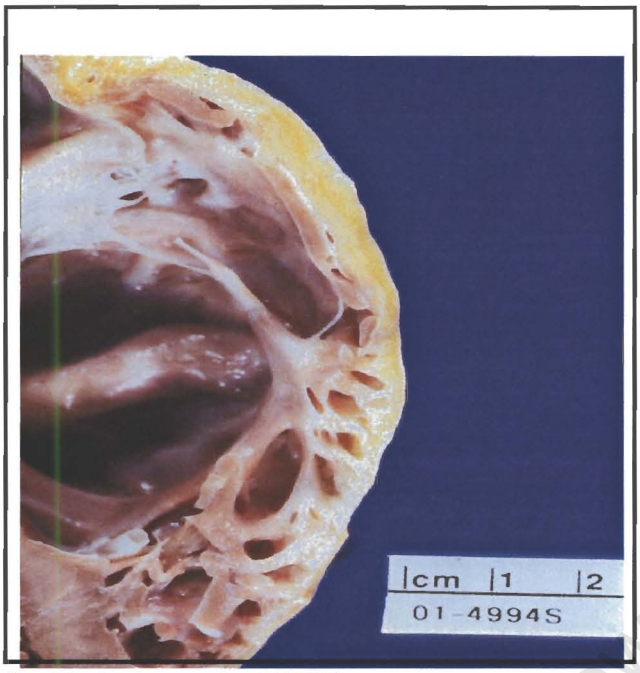
Contractile force generation of sarcomere and its transmission to the extracellular matrix are the fundamental functions of the myocardium. Disease affecting the myocardium leads to cardiac remodelling by hypertrophy or dilation or both (Morita *et al.* 2005). Cardiac dysfunction may lead to arrhythmia, heart failure, and death. Heart disease may arise from diseases of the coronary arteries (coronary artery

disease), myocardium (cardiomyopathy), heart valves (valvular heart disease), pericardium or systemic circulation (e.g., hypertension) (Morita *et al.* 2005). This study will focus on the genetic causes of cardiomyopathy.

1.2. Classification and epidemiology of cardiomyopathy

Cardiomyopathy is defined as a myocardial disorder in which the heart muscle is structurally and functionally abnormal, in the absence of coronary artery disease, hypertension, valvular disease, pericardial disease and congenital heart disease sufficient to cause the observed myocardial abnormality (Elliott *et al.* 2008). Cardiomyopathy is classified into five types depending on the structural and functional changes in the heart, i.e., dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), and the unclassified group. A diagnosis of familial cardiomyopathy is made when the disease is found in at least two first-degree relatives (Mestroni *et al.* 1999). This study will focus mainly on two types of cardiomyopathy: ARVC and DCM.

ARVC is characterized by the loss of myocardial cells in the right ventricle which are replaced by fatty or fibrous tissue (figure 2) (Michael *et al.* 2004).



(health.yahoo.com/.../healthwise/popup/hw141771)

Figure 2: A post mortem section of the right ventricle showing the infiltration of the right ventricular myocardium by fatty tissue in a case of ARVC.

ARVC is associated with re-entrant tachyarrhythmias of right ventricular origin and the risk of sudden death in young people and athletes. In ARVC, the disease affects the free wall of the right ventricle with sparing of the septum and progresses from epicardial to the endocardial surface of the myocardium. The fibro-fatty infiltration is said to affect mainly the apical, subtricuspid and infundibular regions of the right ventricle (Michael *et al.* 2004). There is increasing recognistion that the disease process also affects the left ventricle in a significant proportion of cases of ARVC (Gerull *et al.* 2004; Norman *et al.* 2005). The previous reports considered ARVC as the abnormality of development of the right ventricular myocardium (thus the previous term 'dysplasia') but the recent information suggests that ARVC results from progressive, non-ischaemic loss of myocardium with the replacement by adipose and fibrous tissue (thus the term 'cardiomyopathy is preferred) (Michael *et al.* 2004).

The aetiology of ARVC is not fully understood. The disease is familial in 50 -70 % of affected individuals and sporadic in the rest of cases. It is thought that the disease results from an inherited or acquired insult that causes cardiac cell apoptosis and replacement by fatty tissue or fibrosis. The prevalence and incidence of the disease in South Africa is unknown, but the estimation in European populations is that it

accounts for 3-5% of unexplained sudden death in people under the age of 65 years (Moolman-Smook *et al.* 2003).

DCM is characterized by dilatation of all cardiac chambers and reduction of the ventricular systolic function (figure 3).

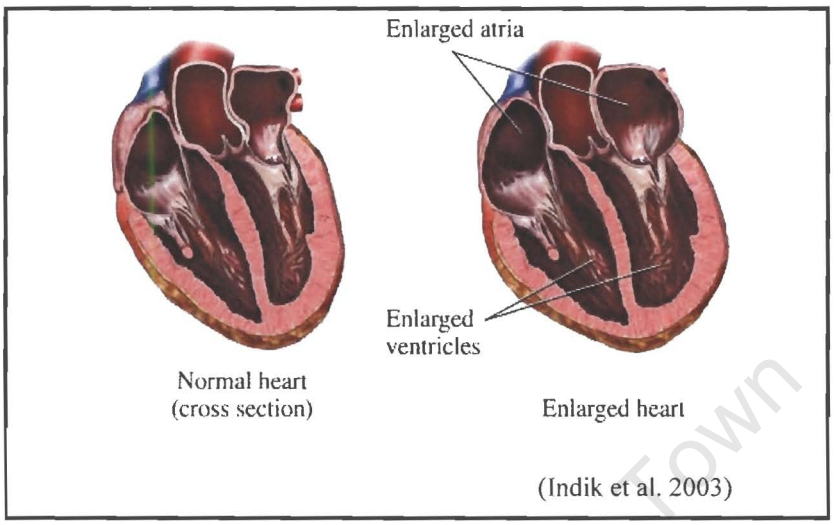


Figure 3: Comparison between normal heart and a heart with DCM

DCM in most cases presents with progressive heart failure. Thrombosis and arrhythmias may occur at any stage and are not uncommon in this type of disease (Richardson et al. 1996). DCM may be idiopathic, familial, viral, auto-immune (Caforio *et al.* 1994, Kühl *et al.* 1996), alcoholic/toxic or associated with recognized cardiovascular disease in which myocardial dysfunction is not explained by the abnormal loading conditions or extent of ischemic heart disease (Richardson *et al.* 1996). DCM has a prevalence of 1/2,500 individuals and the incidence of 7/100,000 per year (Codd *et al.* 1989).

1.3. Clinical diagnosis of ARVC and DCM

In 1994 McKenna and his colleagues proposed the diagnostic criteria for ARVC (McKenna *et al* 1994). In 2003 Ahmad came with a recent version of diagnostic criteria (table1). Both diagnoses are based on the presence of major and minor criteria that encompass electrocardiographic disturbances, structural and functional abnormalities of the right ventricle, histological changes and familial factors (Michael

et al. 2004). The diagnosis of ARVC is made in the presence of two major criteria, one major and two minor criteria or four minor criteria (McKenna *et al.* 1994).

The diagnosis of DCM is normally made on the basis of dilated cardiac chambers with depressed systolic function by non-invasive cardiac imaging, particularly through two-dimensional echocardiography, in the absence of other diseases such as hypertension, valvular heart disease and coronary artery disease (Schönberger and Seidman, 2001; Elliot *et al.* 2008).

Table 1: Criteria for diagnosis of arrhythmogenic right ventricular cardiomyopathy

Criterion	Description
I	Global and/ or regional dysfunction and structural alterations Major: Severe reduction and dilatation of right ventricular ejection fraction with no left ventricular impairment. Localized right ventricular aneurisms. Severe segmental dilatation of the right ventricle. Minor: Mild global right ventricular dilatation and/ or reduced ejection fraction with normal left ventricle. Mild segmental dilatation of right ventricle. Regional right ventricular hypokinesia.
II	Tissue walls Major: Fibrofatty replacement of heart muscles cells on endomyocardial biopsy Repolarization abnormalities.
III	Repolarization abnormalities Minor: Inverted T waves in right praecordial leads (V2 and V3) (individuals aged > 2yrs, in absence of right BBB)
IV	Depolarization/ conduction abnormalities Major: Epsilon waves or localized prolongation (>110ms) of the QRS complex in right praecordial leads (V2-V3) Minor: late potentials (signal averaged ECG)
V	Arrhythmias Minor: Left BBB type ventricular tachycardia (sustained and nonsustained) (ECG, holter, exercise testing) Frequent ventricular extrasystoles (>1000/24h) (holter)
VI	Family history Major: Familial disease confirmed at necropsy or surgery Minor: Familial history of premature sudden death (<35yrs) due to suspected right ventricular dysplasia Familial history (clinical diagnosis based on present criteria)

BBB= bundle branch block, ECG = electrocardiogram

(Ahmad 2003)

1.4. Molecular genetics of ARVC and DCM

Familial cases of ARVC, which account for 30-50% of all cases of ARVC, exhibit autosomal dominant and autosomal recessive patterns of inheritance (McKoy *et al.* 2000, Gerull *et al.* 2004; Awad *et al.* 2006a; Awad *et al.* 2008) but a recent publication mention that familial cases of ARVC account for 25% and prevalence of PKP-2 gene mutations is 50 -70 % (van Tintelen *et al* 2006). Thirteen chromosomal loci have been reported to date for ARVC. Mutations in the plakoglobin gene may the cause autosomal recessive Naxos disease, which manifests as ARVC with hair and skin abnormalities (McKoy *et al.* 2000), or as autosomal dominant disease (Awad *et al.* 2008). Similarly, mutations in plakophilin-2 and desmoplakin genes may cause either autosomal dominant or autosomal recessive disease. The four desmosomal protein genes that are implicated in ARVC, i.e., plakoglobin (JUP), desmoplakin (DSM), plakophilin-2 (PKP-2), desmoglein (DSG2) (Pilichou *et al.* 2006 Awad *et al* 2006b), and desmocollin (DSC2) (Heuser et al 2006) are listed in Table 2. Mutations in desmoplakin gene may be associated with a phenotype that is indistinguishable for DCM, in that there is involvement of both the right and left ventricles (Awad *et. al.*, 2008).

Table 2: The cell adhesion genes that are involved in ARVC

Type of ARVC	*Reference*	*MIM*	*Chromosome*	*Mode of inheritance*	*Gene*	*Reference*
ARVC-1	(Rampazzo *et al.* 1994)	107970	14q23-q24	AD	Transforming Growth Factor-β3 (*TGFβ3*)	(Beffagna *et al.* 2005) (Nattel S. and Schott J. 2005)
ARVC-2	(Rampazzo *et al.* 1995)	600996	1q42-q43	AD	Cardiac Ryanodine Receptor (*RyR2*)	(Tiso *et al.* 2001)
ARVC-3	(Severini *et al.* 1996)	602086	14q12-q22	AD	unknown	-
ARVC-4	(Rampazzo *et al.* 1997)	602087	2q32.1-q32.3	AD	unknown	-
ARVC-5	(Ahmad *et al.* 1998)	604400	3p23	AD	unknown	-
ARVC-6	(Li *et al.* 2000) (Matolweni *et al.* 2006)	604401	10p12-p14	AD	unknown	-
ARVC-7	(Melberg *et al.* 1999)	609160	10q22.3	AD	unknown	-
ARVC-8	(Rampazzo *et al.* 2002)	607450	6p24	AD/AR	Desmoplakin (*DSP*)	(Rampazzo *et al.* 2002)
ARVC-9	(Gerull *et al.* 2004)	609040	12p11	AD	Plakophilin-2 (*PKP2*)	(Gerull *et al.* 2004)
ARVC-10	(Pilichou *et al.* 2006)	Nm	18q12.1	AD	Desmoglein-2 (*DSG2*)	(Pilichou *et al.* 2006) (Awad *et al* 2006b)
Naxos	(Coonar *et al.* 1998)	601214	17q21	AR	Plakoglobin (*JUP*)	(McKoy *et al.* 2000)
ARVC/APC	(Frances *et al.* 1997)	115650	14q24-q terminal	AR	unknown	-

Familial DCM, which is found in 20-35% of all cases of DCM, is associated with various modes of inheritance including X-linked, autosomal dominant, autosomal recessive and mitochondrial transmission (Fourlas *et al.* 2004). There are at least twenty five different genetic loci that are associated with familial DCM and some families are not linked to any of the reported loci, which suggest that other chromosomal loci and genes are involved (Moolman *et al.* 2003; Ahmad *et al.* 2005). This underlines the extreme genetic heterogeneity of familial DCM.

1.5 Hypothesis

It is of interest that mutations in the genes coding for desmosomal proteins may be associated with a phenotype that is characteristic of both ARVC and DCM (Kolar *et al* 2008). For example, mutations in the desmoplakin gene may be associated with ARVC or DCM with palmar plantar keratosis (the Carvajal syndrome) (Norgett *et al.* 2000; Nattel et al. 2005). It is established that plakophilin-2 mutations may be the most common genetic mutation in ARVC, accounting for at least 25% of genotype positive cases (Awad *et al.* 2008). We hypothesized that plakophilin-2 mutations may account for ARVC and familial DCM cases in South African patients.

1.6 Plakophilin-2 gene (PKP-2)

Plakophilin-2 gene encodes for plakophilin-2 protein, a component of the desmosome. This gene, which consists of 14 exons, has been mapped to chromosome 12p11 (Gerull *et al.* 2004). Plakophilin-2 protein has 881 amino acids which make up the three domains of the protein: the N-terminal domain, nine armadillo repeat units domain and C- terminal domain. Figure 4 is an illustration of PKP-2 gene and its relationship with the PKP-2 domains.

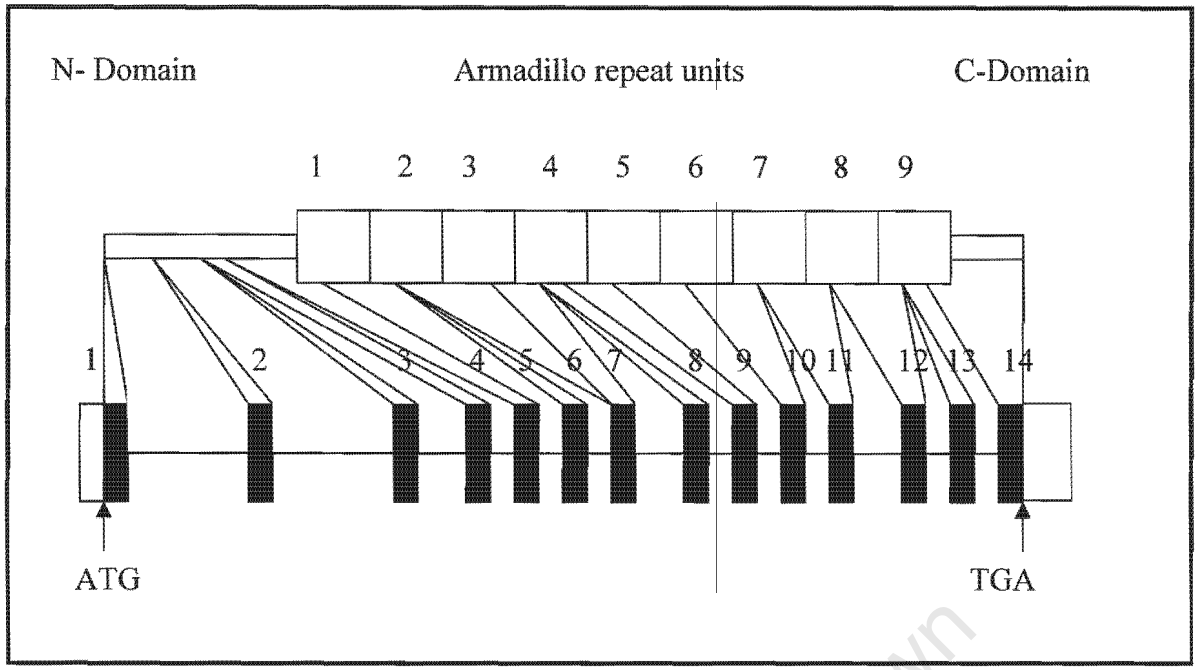


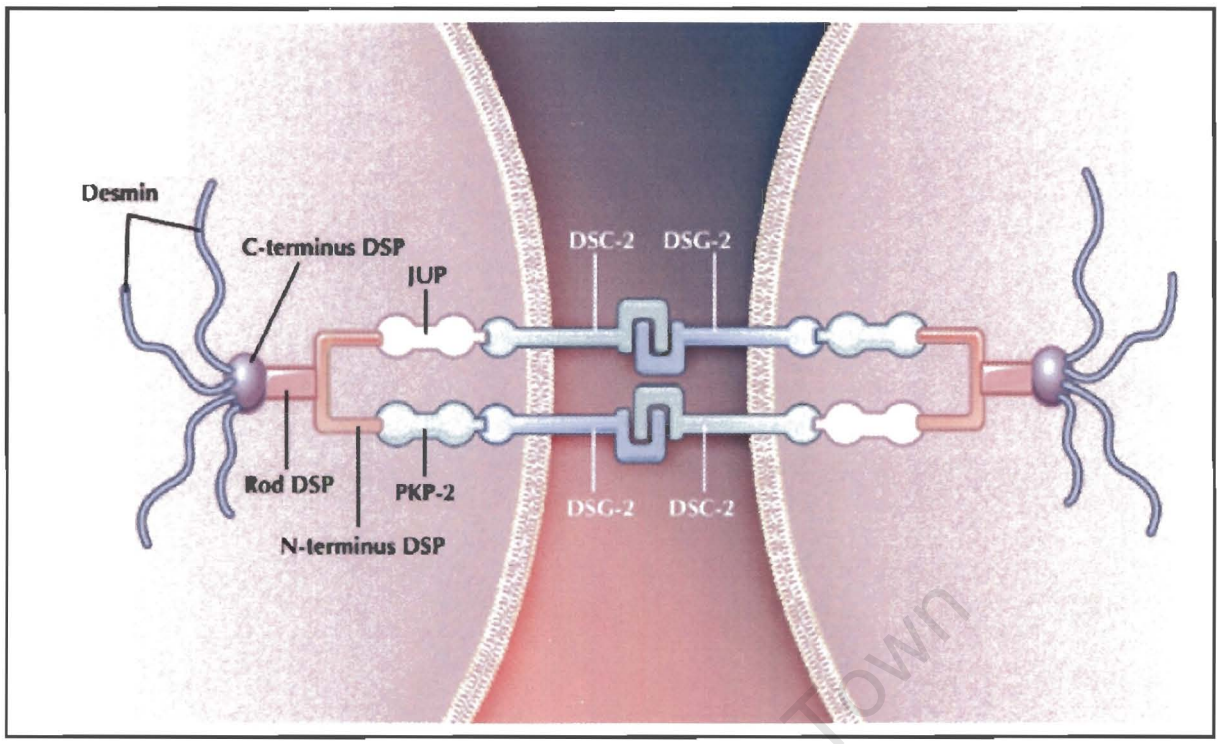
Figure 4: PKP-2 exons with its arm repeat units

The exons in this figure are represented by vertical boxes whereas the introns are represented by horizontal lines. The lines linking the exons to the domains indicate the areas of the protein and their coding exons.

Mutations in PKP-2 gene cause about 25% of ARVC cases in Europe and North America (Gerull *et al.* 2004; Awad *et al.* 2008). No disease-causing PKP-2 mutations have been reported in other forms of cardiomyopathy. Gerull and colleagues postulated that mutations in plakophilin-2 gene of individuals with ARVC cause apoptosis of myocytes and that defect led to the pathological changes that were represented by a wall thinning and aneurysm formation (Gerull *et al.* 2004, Michael *et al.* 2004). Mutations identified in this gene including insertion/deletions, missense, nonsense and splice site mutations were all heterozygous, underlying the autosomal dominant mode of inheritance of most patients. In 2006, Awad and colleagues identified an autosomal recessive homozygous mutation which, although predicted to be translationally silent, actually results in a cryptic splice site causing a 7 base pair deletion in about 80% of the transcripts at 3' untranslated region (Awad *et al.* 2006).

1.7 Plakophilin-2 protein

Plakophilins are armadillo repeat proteins that contribute to the formation of desmosomes. Desmosomes are complex multiprotein structures of the cell membrane and provide structural and functional integrity to adjacent cells, e.g., epithelial cells and cardiomyocytes. A simplified structure of desmosomes is represented by figure 5. the desmosomal cadherins, desmoglein (DSG)-2 and desmocollin (DSC)-2, comprise the transmembrane component of the desmosomal complex. Their extracellular domains interface directly with their counterparts in the neighbouring cells. The intracellular portions of the desmosomal cadherins interact with proteins of the armadillo family, i.e., PKP-2 and junctional plakoglobin (JUP), which in turn bind to the N-terminal domain of desmoplakin (DSP). At its C-terminal, DSP anchors desmin intermediate filaments (i.e., the cytoskeleton) to the cell surface.



DSP- Desmoplakin; DSC-2 – Desmocollin-2;JUP- Junctional plakoglobin;DSG-2 –Desmoglobin-2; PKP-2 – Plakophilin-2 (Sen-Choudry S *et al.* 2007)

Figure 5: Schematic of desmosomes structure

Desmosomes serve as anchoring sites for rope-like intermediate filaments that form a structure of great tensile strength (Alberts *et al.* 2002). Desmosomal proteins also play a role in cell signalling (Gerull *et al.* 2004). Desmosomes are abundant in tissues that experience mechanical stress, such as the heart and the skin. This proposal is supported by the existence of diseases in which tissue integrity is destroyed by gene defects or by autoimmune antibodies targeting desmosomes (Green *et al.* 2000).

1.8 Mechanism of disease due to desmosomal protein defects

There are several hypothesis that have been proposed to account for the cardiomyopathies that are associated with defects in desmosomal proteins. The desmosomal model of ARVC proposes that the compromise of the desmosome leads to detachment and death of affected cells of the heart. Myocardial regenerative capacity is limited, resulting in fibro-fatty replacement of affected myocardium (Awad *et. al.*, 2008). The 'adipogenesis' model of ARVC proposes that Wnt/β-catenin

signalling pathway, which normally inhibits adipogenesis by preventing mesodermal precursors from differentiating into adipocytes, is dysregulated in the disease. Disruption of desmoplakin in desmosomal disease allows plakoglobin, a γ-catenin, to translocate to the nucleus and suppress Wnt/β-catenin signalling. Suppression of the Wnt/β-catenin sugnaling pathway could promote the differentiation of adipose tissue in the myocardium of patients with ARVC. There is also evidence to show gap junction remodelling (through the reduction in the number of gap junctions) and abnormal calcium haemostasis in animal models of ARVC (Awad *et al.* 2008). Thus, loss of cell adhesion, transcriptional dysregulation of adipogenesis, gap junction remodelling, and dysregulation of calcium haemostasis may lead to the cell death, and adipose and fibrous tissue replacement which are the pathological hallmark of ARVC.

1.9 Aims and objective

The aim of this study was to determine the prevalence of plakophilin-2 gene mutations in South African patients with ARVC and familial DCM. The specific objective of this project was to perform mutation screening of all PKP-2 exons in 26 cases of ARVC, and 19 cases of familial DCM; one case of familial RCM was also included (i.e. 46 cases of cardiomyopathy were studied). These cases were available for testing at the time of commencement of this study in 2006.

Chapter 2: Materials and Methods

2.1. Study Cases

DNA samples were available for 46 probands with cardiomyopathy which were classified into ARVC (n=26), familial DCM (n=19) and familial RCM (n=1). These patients were recruited at the Cardiac Clinic, Groote Schuur Hospital, Cape Town (Tables 3, 4 and 5). All probands were selected from the registry due to the availability of the DNA.

Table 3: The age, race and gender of ARVC probands

DNA sample	*Age*	*Race*	*Gender*
ACM 1.2	20	Caucasian	Male
ACM 2.4	23	Caucasian	Male
ACM 3.1	40	Caucasian	Female
ACM 4.1	26	Mixed	Female
ACM 5.1	44	Caucasian	Female
ACM 6.1	52	Mixed	Male
ACM 8.3	14	Caucasian	Female
ACM 9.1	27	Caucasian	Male
ACM 10.1	39	Mixed	Male
ACM 11.2	15	Mixed	Male
ACM 12.1	50	Caucasian	Male
ACM 13.1	18	Caucasian	Male
ACM 14.1	45	Mixed	Male
ACM 15.1	27	Caucasian	Male
ACM 16.1	35	Caucasian	Male
ACM 17.1	22	Black	Male
ACM 18.1	21	Caucasian	Male
ACM 19.2	21	Caucasian	Female
ACM 20.1	53	Caucasian	Female
ACM 21.1	51	Caucasian	Male
ACM 22.1	38	Mixed	Male
ACM 23.1	41	Black	Male
ACM 24.1	62	Caucasian	Female
ACM 25.1	31	Caucasian	Male
ACM 26.1	54	Caucasian	Male
ACM 27.1	54	Caucasian	Male

Table 4: The age, race and gender of DCM probands

DCM sample	*Age*	*Race*	*Gender*
DCM 3.5	33	Mixed	Male
DCM 5.1	37	Caucasian	Male
DCM 9.1	35	Caucasian	Female
DCM 10.3	55	Indian	Male
DCM 11.1	38	Caucasian	Male
DCM 24.1	40	Black	Male
DCM 91.1	49	Mixed	Male
DCM 126.1	55	Black	Female
DCM 128.1	42	Black	Female
DCM 136.1	55	Mixed	Female
DCM 175.2	29	Black	Male
FDC 1.1	21	Black	Male
FDC 9.1	21	Mixed	Male
DCM 219.6	14	Mixed	Male
DCM 226.1	14	Black	Male
DCM 236.2	34	Mixed	Male
DCM 240.1	48	Black	Female
DCM 241.1	55	Black	Male
DCM 242.2	-	Black	Female

Table 5: The age, race and gender of the RCM proband

RCM sample	*Age*	*Race*	*Gender*
RCM 1.1	30	Indian	Male

The diagnosis of cardiomyopathy was made according to standard criteria (McKenna *et al.* 1994; Richardson *et al.* 1996; Elliott *et al.* 2008). The DNA was extracted by a technologist in the Division of Human Genetics at the University of Cape Town (UCT) where the PureGene™ DNA isolation kit (Gentra system) was used according to the instructions of the manufacturer. The patients gave written informed consent for the study of the genetic causes of cardiomyopathy (see Appendix 1). This project forms part of the Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa, which was approved by the Research Ethics Committee of the Faculty of Health Sciences of the University of Cape Town (REC REF 047/2003).

2.2. DNA Amplification by Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a rapid *in vitro* method for amplifying defined target DNA sequences present within a source of DNA (Strachan & Read 2001). The technique uses oligonucleotide primers to anneal to specific DNA template sequences, which flank the genomic region of interest. A DNA polymerase enzyme is used to extend the sequence from the primers across the target region in many cycles and each cycle consists of denaturation, annealing and elongation in order to produce numerous copies of the target sequence. We screened the 14 exons of the plakophilin-2 (PKP2) gene for mutations that cause cardiomyopathy in the South African patients. We used the primer sequences and PCR conditions that had been published by Syrris *et al* (2006) (for exons 1, 2, 3, 4, 5, 6, 7, 8) and Gerull *et al.* 2004 (for exons 9, 10, 11, 13, 14). All the primers were ordered from Inqaba Biotechnology except for exon 12 which was designed by the author and ordered from the UCT supplier (Table 6).

Table 6: Primer sequence used and PCR product sizes for each exon

Primer	*PCR products Size (bp)*	*Primer sequence (5'–3')*	*Annealing Temperature (°C)*
1F	414	ACT CGA GCG GGG CGG GGC TCG C	68
1R		ACT CCC AGC ACG CGG GGT GAG	
2F	332	TAC TTG TTC TTG GCC TTC ATT AC	55
2R		ACT TGG GAA AAG TAA ACA CTC	
3F1	423	AGTCCTCAGCAAAGTTGAAATTTG	53
3R1		GTCAAAAACGGTGTCGCTAACAGA	
3F2	433	CACATACCACAGACAGTACCAG	58
3R2		TGTGCTGGCAATGACTGAACTG	
4F	291	AGT ATT CGC TGA GTC GTC TCT	55
4R		GCA AAG TCA CCA TAA TAG AAG	
5F	413	GAA AGG TTA TAG TCA GCA TCA G	55
5R		CAT CAA TCA TTT GCT CCA GGA	
6F	258	TTG CTG TGT TCA TAA AGG AGC C	55
6R		ATT ACA GGC GCA GAC CAC GAC A	
7F	305	TCC AGC GGT CAT TTT GGT CCC A	55
7R		TGA CTT CCT TGG GGC TAC CTA A	
8F	250	CAA AGA CCT GTT GGA TAC ACA	55
8R		CCA AGC GGC TAT CTT AAG AAT	
9F	493	TAC TCA TTG CAT TTC CCC CAG	57
9R		TCC TCA CTG GTA AAT GAG GG	
10F	458	CAG TAT TTC TGG TCT CCT GG	55
10R		TGA CTT GAC TTG TCA GTC AAG CAG CCT GAC	
11F	417	ACA TCT TCA ACC TCT GGT AAT CTA CAG	58
11R		AGA AAG CCT GTT TGT GAT ATC TGG TGG CAC	
12F	402	CAA CAG AGC AAG ATT CCG	58
12R		GTT GTG AAT GTG CGG TCA TG	
13F	539	AGT TGA GGA GCG AAG GGG ACC AC	57
14R		TTC TGG ATT CAG GGG ACC AC	

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

2.2.1 Optimisation of the PCR

The integrity of the DNA was identified by electrophoresing 2µl of the sample on an agarose gel before diluting to a concentration of 100ng/µl. If a DNA sample was degraded a different storage sample which showed no degradation was diluted. PCR optimization was first performed on five control samples (DNA samples from the

individuals that had no history of cardiac disease). The published annealing temperatures were used in the PCR of exons 1, 2, 3, 4, 5, 6, 7, 8 9, 11 and 13 plus 14 (which were amplified together). The annealing temperature for amplifying exons 10 and 12 had to be changed for successful amplification. To identify the best annealing temperature for exons 10 and 12, the gradient function of the PX2 Thermal Cycler machine was used. PCR products were electrophoresed on the gel. The temperature that gave best amplification was the one selected as the annealing temperature for exons 10 and 12. The annealing temperatures are presented in Table 6.

2.2.2 PCR amplification of DNA samples from cardiomyopathy patients

PCR amplification was performed in a total volume of 25µl, which contained 100ng of genomic DNA, 1.5µM (1X) Go Taq ™ reaction Buffer that contains 1.5µM $MgCl_2$, 200µM dNTPs, 20 µM of each of forward (F) and reversed (R) primers and 0.5 Units (U) of GoTaq polymerase. The cycling conditions used during PCR are presented in Table 7.

Table 7: PCR cycling conditions

Temperature (°C)	*Time*	*Number of Cycles*
94	5 minutes	1
94	30 seconds	30
Annealing temperature*	30 seconds	
72	30 seconds	
72	7 minutes	1

*Annealing temperatures are presented in Table 3

PCR was performed in either the GeneAmp® System 9700 thermal cycler, PX 0.2 Thermal cycler (Thermo Electron Corporation), or the PX 2 Thermal cycler (Thermo Electron Corporation). A negative control (whereby dH_2O was used instead of DNA) was included in all PCR experiments. This was done to identify possible contamination that could interfere within the reactants.

Contamination was encountered when amplifying exons 13 together with 14. The PCR was therefore set up in the Laminar Flow Hood. A different set of pipettes was used and the table top of the hood and different pipettes were first cleaned with chlorine solution.

Agarose was used to visualize the PCR products. It is a standard method for separation of PCR products. The agarose is cross-linked to form pores through which DNA particles migrate at different speeds depending on size of the fragment. The PCR products migrate towards the cathode (positive terminal) in an electric field due to the negative charge of DNA. The velocity at which molecules move through these pores depends on the size of the pore; which is related to the percentage of the agarose in the gel. To visualize the PCR products in the agarose gel, ethidium bromide was used to stain the PCR products. Ethidium bromide intercalates between the base pairs of the DNA strands and is visualised under ultraviolet (UV) light. The loading of the DNA and the monitoring of electrophoretic progression were assisted with the loading buffer, 3µl of 1X loading dye (Appendix 2). PCR products together with a GeneRuler^{TM} 100bp DNA ladder Plus, that gave the sizes of the PCR products, which were electrophoresed through a 1% agarose gel in 1X TBE buffer (Appendices 2 and 3).

2.3 Mutation Screening by Denaturing High-performance Liquid Chromatography

The principle behind denaturing high-performance liquid chromatography (dHPLC) is to separate heteroduplex from homoduplex DNA fragments by ion-pair reverse-phase liquid chromatography (Frueh and Noyer-Weidner, 2003). A heteroduplex is a double-stranded DNA complex in which the two strands do not have the perfect base complementarity and homoduplex is a double stranded DNA with perfect base complementarity (figure 6).

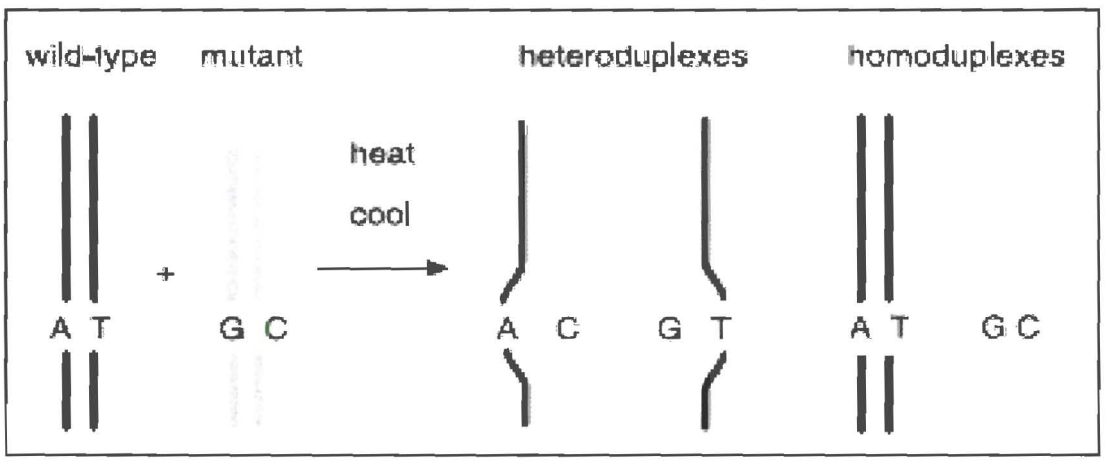


Figure 6: Illustrations of the formation of both the heteroduplexes and the homoduplexes in the process of denaturing high-performance liquid chromatography. (Frueh and Noyer-Weidner, 2003)

The method detects all types of nucleotide mismatches regardless of the position of the mismatch within the fragment. The WAVE® system of dHPLC operates in three different modes: partial denaturation, full denaturation and non-denaturation (Xiao *et al.* 2001). This study utilised the partial denaturation mode, as it is applicable to mutation detection. The separation chemistry in partial denaturation is based on primer sequence, the size and the melting temperature (Tm) of the PCR products. Once the PCR products are loaded onto the autosampler, they are passed through the separation cartridge by the mobile phase.

2.3.1 WAVE® method creation

DHPLC-based WAVE® system has the Wavemaker software programme; the programme was used to create methods for mutation detection for each PCR amplicon. Each method was designed with different temperatures to allow elution of the partially denatured PCR product through the separation cartridge.

2.3.2 Heteroduplexing of the PCR products

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

After the denaturation the mixture was allowed to cool slowly at room temperature for 45 minutes (figure 6).

2.4 Identification of PKP-2 gene Mutations by Dideoxy-cycle sequencing in Probands with Cardiomyopathy

All DNA samples that showed variation on WAVE® machine were sequenced to identify the cause of the abnormal pattern. PCR amplifications were performed on the selected individuals with variations. PCR products were run on gel matrix and were excised and processed according to the instructions and reagent of the Wizard Purification System Kit Protocol (Wizard® SV. Gel and PCR clean-up system). The DNA was recovered using 30µl of dH_2O (Sabax). Five µl of PCR product was run on agarose gel to visualize the concentration of the PCR product electrophoresed. The purified PCR product was sequenced according to instructions given in the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, CA, and USA). The amount of the PCR product used in the DNA Terminator Cycle Sequencing depended on the concentration of the PCR product after purification. The brighter the bands the smaller the volume of the respective PCR product used in sequencing reactions. The products were sequenced in both forward (F) and reversed (R) directions; this was done to confirm the potential changes. The parameters of the half reactions were as follows: a minimum of 0.8µM of each F and R primers, 5X Dilution Buffer, 5X Termination Mix, amount of DNA product and dH_2O to make up the total volume of 20µl. The conditions applied in this type of sequencing are presented in Table 8.

Table 8: Dideoxy-cycle sequencing conditions

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

2.4.1 Purification of cycle sequencing product

Purification of the cycle sequencing products was performed to remove unincorporated dideoxy-nucleotide triphosphate (ddNTP) terminators. These molecules could interfere with further processing steps. CENTRI-SEP Columns

(Princeton Separations, Inc. Adelphia, NJ) were used with 7% Sephadex ™ G-50 (Amersham Pharmacia Biotech, AB, Uppsala, Sweden) to purify the cycle sequencing products. This method was performed according to the manufacture's protocol. All sequencing was performed on the ABI Prism® 3100 Genetic Analyzer (Applied Biosytems).

2.4.2 Sequence analysis

The sequencing electropherograms from the ABI Prism ® 3100 were analyzed using the BioEdit programme. The reverse and the forward electropherograms were aligned with the wildtype sequences from NCBI and analysed for changes. Splice site prediction by the neural network programme on the Berkeley Drosophila Genome Project (BDGP) website (www.fruitfly.org/seqtools/splice.htm/) was used to predict whether intronic variants could be disease causing.

The frequency of novel mutations detected was examined in normal controls using two methods. The first method was by SNaPshot® multiplex technology (see section 2.5 below), and the second method was by Restriction Fragment Length Polymorphism (RFLP) in the case of variants that created restriction sites for enzyme digestion (see section 2.6 below).

2.5 Screening of DNA from normal control for polymorphisms by SNaPshot® multiplex technology

SNaPshot® multiplex technology (Applied Biosystems) uses a single-base extension method to genotype single nucleotide polymorphism (SNP). This method enables robust multiplex SNP interrogation of PCR products. SNaPshot® technology was used in this study to genotype population controls for the new variants that were discovered in the mutation screening study of the probands with cardiomyopathy.

A single primer was designed next to the variation, either as a forward or a reverse primer. These primers were analyzed using an oligo analyzer programme to identify any primer hairpins and dimers. Then primers were aligned against different sequences with variations using BioEdit programme to see if there are any dimers or primers that bind to each other and produce the variations that may produce false positive results. Then these primers were aligned on the AutoDimers programme to see if there are any primers producing the same variations that might give false

positive results. Tails were introduced to these primers, with different length so that they do not give peaks that overlap to each other. The primers are shown in a Table 9.

Table 9: SNaPshot® primer sequence, size and colour of expected peaks

Primer sequence	*Size(bp)*	*Colour expected*
Exon 4.1 >T32822T Reverse GATGACACTCACCCTCTTCC	20	Blue- Green
Exon 10 >G80401A Reverse aaCTGAATTGAATGTAGGTATGTCTA	26	Red- Black
Exon 11 >C99354T Forward ccacgtaatcCTGCAGAATGAAATTGGTGAGT	32	Black- Red
Exon 9 >C79290G Reverse atttgatttgAACTTAAAATAGTGCCGATATATACCAA	38	Blue- Red
Exon 13.2 >G109196A Forward tgtgtctcggacctggcagtcatgCTTGACCCTGGGAAGAAATC	44	Blue- Green
Exon 4.2 >C32792T Forward cacccttcctagtccggattcccacgacccCTGCTGCAGCTACTTTCATA	50	Black- Red
Exon13.2 >T109076C Forward gaaaaaataatacgccgaacgcatactccacgtgccTTCCGTCCTTCTGTATTCTC	56	Red- Black

In this study, mutation screening of the PKP-2 gene in patients with cardiomyopathy revealed seven novel variants, whose frequency in the general population was verified by genotyping by the SNaPshot® technique in 241 normal controls (i.e., 482 chromosomes) of different ethnic groups (86 blacks, 80 Caucasians, and 75 people of mixed ancestry).

Seven DNA samples that had been found to have each particular SNP for each particular exon (i.e., positive controls) were selected to optimize the SNaPshot® technique. PCR for these DNA samples was performed and run on an agarose gel; only reactions with bright bands on agarose were accepted for the next step of the experiment. These PCR products were cleaned using Exonuclease1 (EXO1) and Shrimp Alkaline Phosphatase (SAP). A quantity of 0.1µl of EXO1 and 1µl SAP were added to the PCR product, and, incubated in a 37°C water bath for an hour then transferred to 75°C water bath for 15 minutes to deactivate oligonucleotides and dNTPs; EXO1degrades excess single stranded primer oligonucleotide from a reaction

mixture containing double stranded extension products and SAP removes all unincorporated dNTPs.

When the cleanup step was complete, the SNaPshot® reaction was performed as follows:

- 1µl PCR product was added to each representative tube according to the volume in Table 10.
- Added 1µl SNaPshot® primer to each represented tube.

Table 10: Multiplex mix reaction

Reagents	*Amount added*
Reaction Mix	2.5µl
dH2O	5µl

Cycling conditions for this technique are presented in Table 11.

Table 11: Conditions of multiplex reactions in PCR machine

Temperature (°C)	*Time*	*Number of Cycles*
96	10 seconds	25
50	5 seconds	
60	30 seconds	

A negative control was also included whereby, instead of a cleaned PCR product, dH2O was used to control for contamination. When all these cycles were done, an additional microlitre of SAP was added and incubated at 37°C water bath for an hour and 75°C water bath for 15 minutes. Then (Table 12) 1µl of each product was mixed with 9µl of formamide and 0.2µl LIZ standard 120 in an ABI3100 machine plate and denatured at 98°C for two minutes.

Table 12: Reagents added in an ABI 3100 plate with a SNaPshot® multiplex mix to run

Reagents	*Amount added*
High dye Formamide	9µl
Liz standard 120	0.2µl

Thereafter, the reactions were kept on ice while preparing to transfer to ABI3100 machine for analysis.

The SNaPshot® results were analyzed using a Gene Mapper programme.

2.6 Screening of DNA from Normal Controls for Polymorphisms by Restriction Enzyme Digests Analysis

Restriction enzymes are enzymes isolated from bacteria that recognize specific sequences in DNA and then cut the DNA to produce fragments, called restriction fragments. Restriction enzymes play a very important role in the construction of recombinant DNA molecules, as is done in gene cloning experiments. Another application of restriction enzymes is to map the locations of restriction sites in DNA (Williams *et al* 1996). Restriction enzymes cut the DNA in different ways. Some enzymes cut in the middle of the recognition sequence, resulting in a blunt end. Other enzymes cleave in a staggered fashion, resulting in DNA products that have short single-stranded overhangs at each end. These are called cohesive ends, as these single-stranded overhangs could come together again through complementary base-pairing.

Restriction enzyme digestion was used to obtain the frequency of variants in normal controls for haplotype analysis. To design restriction enzymes at specific restriction sites the Webcutter 2.0 programme was used (http://rna.lundberg.gu.se/cutter2/.) for the restriction site of 2196-2202delCACACCinsG and an enzyme Sda I was identified. This restriction enzyme recognizes 5'...C C T G C A G G...3' sequence and it cuts between A and G bases. This programme did not work for T1097C and IVS11-1G-C restriction sites and another programme was used, a zeon.well.ok.ac.uk/. This programme has two boxes for exon sequences, one for wild type sequence and

another for mutant sequence, once you submit these sequences you get a number of enzymes for that specific restriction site and one enzyme must be selected. In exon 4 for T1097C restriction site Aci I restriction enzyme was selected. This enzyme recognizes 5'...C C G C...3' sequence and it cuts between C and C bases in the 5 prime and G and C in the 3 prime. In exon 11 for IVS11-1G-C restriction site Nsp I was selected and in this restriction site all restriction enzymes were cutting more than once. It recognizes 5'...Pu C A T G Py...3' sequence.

A protocol from the manufacturer instructions on the use of restriction enzymes was followed:

- Nuclease-free water 16µl
- Restriction enzyme buffer 2µl
- Restriction enzyme (10units/µl) 1µl

A buffer for Aci I is 10X Buffer O (appendix 2), Nsp I is 10X Tango Buffer (appendix 2) and Sda I is 10X Buffer Sda I (appendix 2). A reaction mix for each exon was prepared and 10µl of PCR product was added in each respectively tube. All restriction digests were incubated at 37°C overnight. Exon 11 restriction enzyme digests with 2196-2202delCACACCinsG restriction site was run in agarose gel to visualize digested bands. Exons 4 and 11 respectively were run in polyacrylamide gel because the bands sizes were too close to each other and acrylamide is very sensitive in small and close PCR products.

2.6.1 Polyacrylamide gel analysis to analyse restriction fragment length polymorphisms

Polyacrylamide gel analysis was used to run and visualise the DNA fragments that were subjected to enzyme digestion. Polyacrylamide gel was set between two glass plates (11.8cm x 22cm). One of the plates contained well plates made of Dymo tape. The other plate is referred to as the back plate. Both glass plates were cleaned thoroughly with ethanol three times to remove any excess water. The well plate was cleaned with acetone to prevent the gel from adhering to this plate. A mixture of plate glue and 10% acetic acid was wiped vigorously over the back plate so that the gel adheres to this plate. The plates were clamped together (cleaned sides facing inwards)

with a 0.5cm wide spacer in between. The plates were set upright with the wells at the bottom.

A polyacrylamide gel was prepared by mixing 20ml 40% polyacrylamide solution with 200µl 10% ammonium persulphate (APS) and 20µl N'N'N'N'-tetramethylethylenediamine (TEMED). APS and TEMED cause polymerisation of the polyacrylamide solution, therefore this mixture was quickly poured between the plates using a 10 ml syringe. The gel was allowed to polymerise for approximately an hour before use. The clamps were removed, the plates separated and the back plate (containing the gel) was put into position on the Multiphor II Electrophoresis System. Six filter paper strips (5cm x 28cm) were soaked in 1x TBE (Appendix 2). Three strips were placed on both the top and bottom of the gel. A volume of 5µl PCR product was mixed with 3µl agarose loading dye (Appendix 2) and loaded into the well. The PCR products were electrophoresed through the gel at approximately 355V for one hour alongside a 100 base pair ladder. The gel was first rinsed with distilled water then stained in silver staining solution 1 (Appendix 2) for 10 minutes. The gel was then washed in dH_20 to remove any excess silver nitrate before being counter stained in silver staining solution 2 (Appendix 2) for 5-10 minutes depending on the appearance of the bands.

2.7 Haplotyping by Microsatellite (Short Tandem Repeat) Analysis

Microsatellites are simple repeated sequences of nucleic acids and they are also called short tandem repeat (STR). In most cases they are usually two to six base pair repeats, for example $(CA)_n$, $(GT)_n$, and $(CAG)_n$. The $(CA)_n$ is the most common repeat sequence. Their length is determined by the copies of the repeat unit. Microsatellites are used as highly polymorphic genetic markers which are not only used in the construction of genomic linkage map, gene location and cloning but also used in the linkage analysis and gene diagnosis of genetic disease.

Three STR markers were used in this study; D12S1692 is a published marker in position 7275538- 727786 of the PKP-2 gene. Primers for this marker were published. The other two markers, the D12S2415474 marker at position 2415474 to 241529 and D12S825316 marker at position 825316 to 825348 were selected from the genomic sequence. The genomic sequence used for this study was a reverse complement so the

position of the two selected markers is not the same as the normal sequence. Primers were designed for the new markers. Tails of different sizes that are made of random base pairs were added to the primers so that they do not overlap on a master mix. All these primers were labelled with Hex or Fam flourescence dyes depending on the size of the primer. These primers were ordered from University of Cape Town.

A cohort of normal controls representing the Black, Caucasian and mixed ancestry communities of South Africa was amplified for all three markers. These controls were obtained from unaffected members of DCM families, normal controls from Division of Human Genetics and from the University of Stellenbosch. The microsatellite results were analysed using a Gene Mapper technology.

Chapter 3: Results

3.1 PCR Amplification of DNA samples from affected patients

PCR product bands of the expected size and the appropriate brightness were obtained when the PCR product was electrophoresed on the agarose gel (Fig. 7).

Figure 7: Illustration of PCR amplification of exon 10 of PKP-2 gene in 11 control individuals. The expected product size of 458 bp is run on an agarose gel, and corresponds to the correct size according to the DNA ladder.

All exons were amplified successfully using the published and redesigned (i.e., exon 12) primers.

3.2 Mutation screening using WAVE® technology

An abnormal WAVE® pattern/variation in a DNA sequence of an individual sample is present when a chromatogram presents with a different peak profile to the normal (wildtype) (figure 8) peak profile. The DNA of 46 individuals in our cohort (ARVC=26; DCM=19; RCM=1) was screened for variations on the WAVE® system. Abnormal WAVE profiles were found in 19 probands: ARVC (n=12), familial DCM (n=6), and RCM (n=1) which resulted in a total of 66 profile variations in exons 4, 9, 10, 11, 13 and 14. The distribution of WAVE® profiles were as follows: three in exon 4, two in exon 9, 16 in exon 10, 13 in exon 11 and 32 in exons 13+14. As quite a number of the chromatogram profiles appeared to be duplications it was decided to sequence the chromatograms presenting with different peak profiles. This resulted in a decrease in the number of samples to be sequenced; from 66 to ten.

Figure 8 illustrates a normal sequence represented by DNA sample RPD 16.9 (population control), which was run concurrently with DNA sample of ARVC proband ACM 1.2 on the WAVE system.

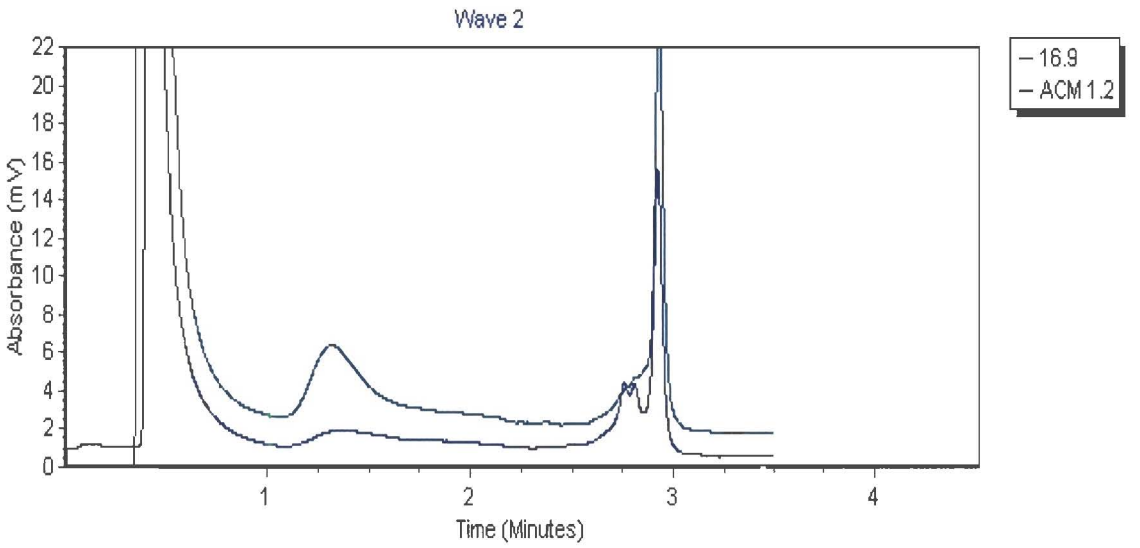


Figure 8 shows a three peak electropherographic profile of ACM 1.2 produced by a WAVE run. The profile of ACM 1.2 is significantly different from the wildtype profile (i.e., three peaks in ACM 1.2 versus one peak in wild-type 16.9 at the elution time of 3 minutes).

The DNA samples from ARVC probands ACM 5.1, 8.3, 12.1, and 19.2 also had a similar chromatographic profile to ACM 1.2. These abnormal profiles do not necessarily indicate the same genetic mutation in PKP2.

3.3 DNA sequencing analysis

The abnormal WAVE® electropherograms of the 19 probands (ten profiles) investigated were sequenced and the results summarised in Table 13. Our study has identified three known variants (two disease-causing mutations and one polymorphism) and seven novel variants (one disease causing mutation, six polymorphisms) in the PKP2 gene. These results are discussed in more detail below.

Table 13: Sequencing results of individuals with abnormal WAVE electropherograms

Region	*No*	*Nucleotide change*	*Amico acid change*	*Disease-causing/ Polymorphism*	*Patient ID*	*Reference*
Known variants						
Exonic	4	**T1094C**	M-T	Polymorphism	ACM 5.1, ACM 12.1, ACM 19.2	rs1046116
Intronic	10	**IVS2146-1G-C**	Intronic splice product	Disease-causing	ACM 22.1	Gerull *et al.*2004
Exonic	11	**2197-2202del CACACCinsG**	-Stop	Disease-causing	ACM 19.2	Dalal *et al* 2005
Novel variants						
Exonic	4	**C1132T**	Q- stop	Disease-causing	ACM 1.2	-
Exonic	4	**C1162T**	R-W	Disease-modifier	ACM 5.1,ACM 12.1, ACM 19.2	-
Intronic	9	**IVS9+83C-G**	-	Polymorphism	ACM 4.1	-
Intronic	10	**IVS10+44G- A**	-	Polymorphism	ACM9.1; ACM 20.1,DCM 10.3, DCM 226.1, FDC 1.1 & RCM 1.1	-
Intronic	11	**IVS11+7C- T**	-	Polymorphism	ACM 15.1 & DCM 219.6	-
Intronic	13+14	**IVS13+83G-A**	-	Polymorphism	ACM 6.1, ACM 8.3, ACM 18.1 & DCM 3.5	-
Exonic	13	**T2540C**	L-P	Rare polymorphism	DCM 3.5	-

ACM, arrhythmogenic right ventricular cardiomyopathy; DCM, dilated cardiomyopathy; FDC, familial dilated cardiomyopathy; RCM, restrictive cardiomyopathy; IVS, intervening sequence (intron); -/+, up/downstream of the respective exon; del, deletion; ins, insertion; -: substitution, M: methionine, Q: Glutamine, R: Arginine, T: threonine,W: Tryptophan, Novel variants are presented in bold lettering.

3.3.1 Identification of three known variants in the PKP-2 gene

Sequencing of the abnormal WAVE® electropherograms identified three known variants (refer to Table 13) in the PKP2 gene. The T1094C polymorphism in exon 4, the IVS2146-1G-C disease-causing mutation in intron 10 (Gerull *et al.* 2004) and the 2197-2202 del CACACC/insG disease-causing mutation in exon 11 (Dalal *et al.* 2005).

The T1094C polymorphism was detected in exon 4 of *PKP-2* and changed the amino acid from a Methionione to a Threonine. Three probands (ACM 5.1; ACM 12.1; ACM19.2) with ARVC was found to have this polymorphism. The mutation in intron 10, IVS2146-1G-C, was found to produce a mutant splice product and was identified in only one proband (ACM 22.1) with ARVC; no additional family members were available for testing. The 2197-2202 del CACACC/insG disease-causing mutation in exon 11 was found in one ARVC proband (ACM 19.2) and produces a stop codon that results in truncation of the PKP-2 protein. The discovery of this mutation led to screening of the whole family (ACM19). Our data indicated that the father and the two affected children were found to have this mutation (figure 9). Both the affected siblings and the father had earlier been diagnosed as having clinical ARVC.

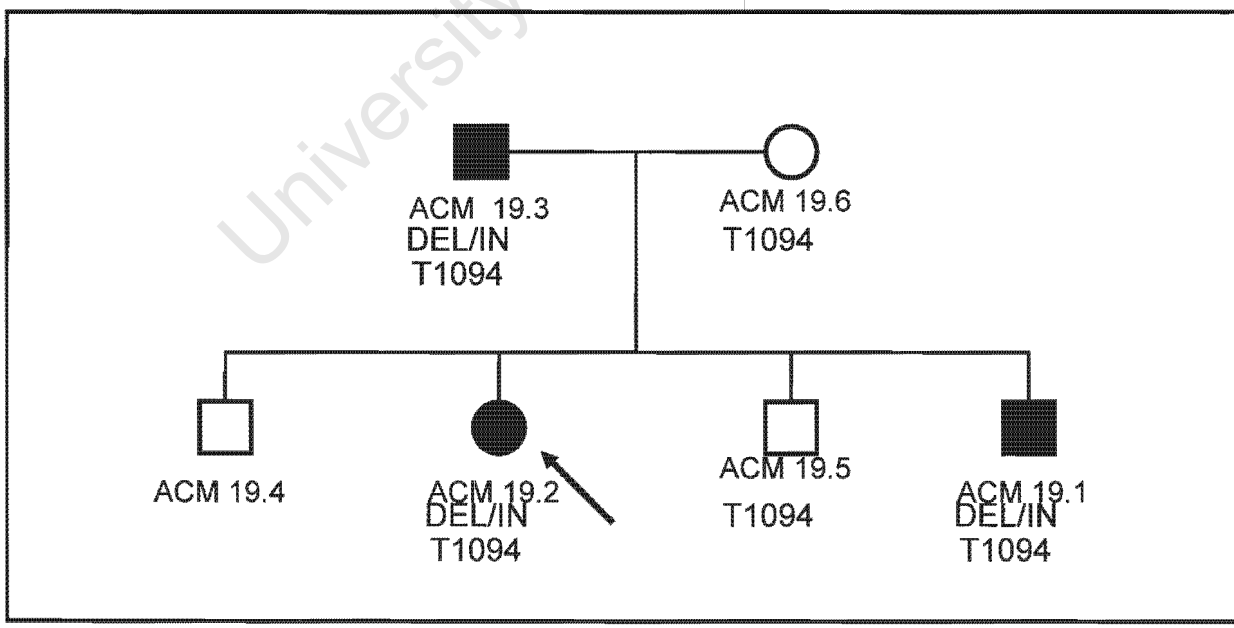


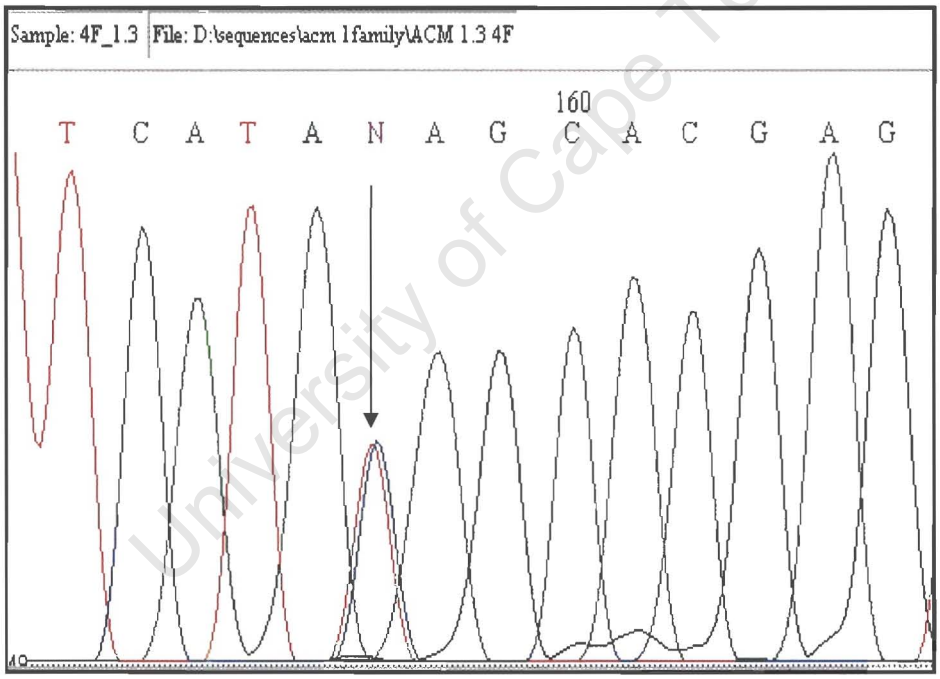
Figure 9: The pedigree for ARVC family 19 showing segregation of the T1094C variant and the 2197-2202 ins/del in the family members

3.3.2 Discovery of seven novel variants in PKP-2

Sequencing of the abnormal WAVE® electropherograms identified seven novel variants: one disease causing mutation and six polymorphisms in *PKP-2*. All of the sequencing results were heterozygous substitutions in the exonic or intronic regions of this gene. These findings are reported in relation to the affected regions as follows:

In exon 4, we detected a C1132T polymporphism in PKP-*2*. The mutation alters a CAG codon to a TAG codon; the TAG codon codes for a stop in the translation of the protein. This disease causing mutation was found to occur in only one ARVC proband (ACM 1.2). When the Homo sapiens protein sequences was aligned and compared with other modern animals it was found to be a highly conserved (Figure 10B).

A



B

```
H. sapiens_   PSRISAAATFIQHECFQKSEARKRV
C. lupus_     PPRISAAATFIQHECFQKSEARKRV
M. musculus_  VSKISAAATFIQHESFQKSEARKRV
G. gallus_    TPRILAAVTFIQHECFQKAEARRKV
```

Figure 10: In (A) an electropherograph of ACM 1.2 showing variant C1132T and (B) the protein sequence alignment with modern animals.

The discovery of this mutation led to screening of the whole family (ACM1) consisting of a mother and two affected sons with ARVC. Our data indicated that the unaffected mother of the proband did not carry the variation while the affected brother (ACM1.3) of the proband did carry the mutation (figure 11).

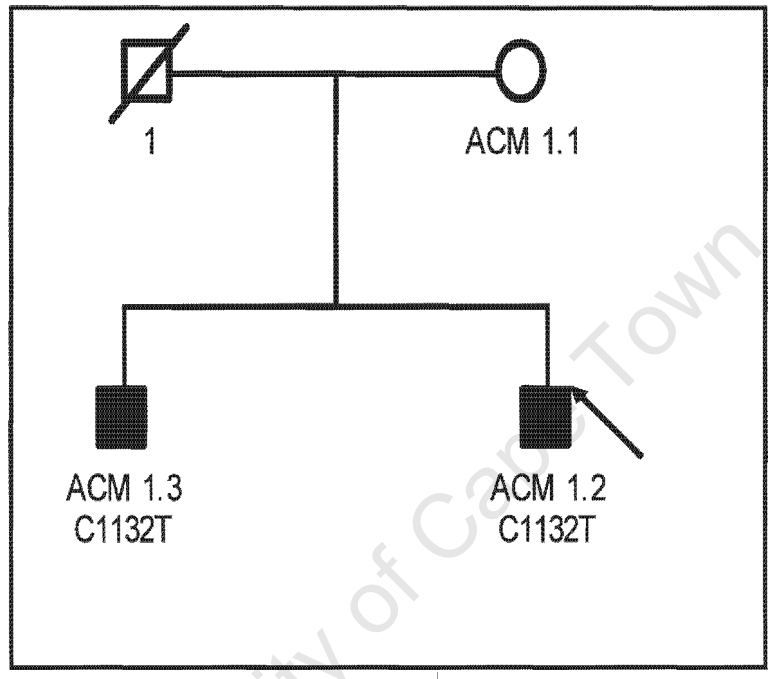
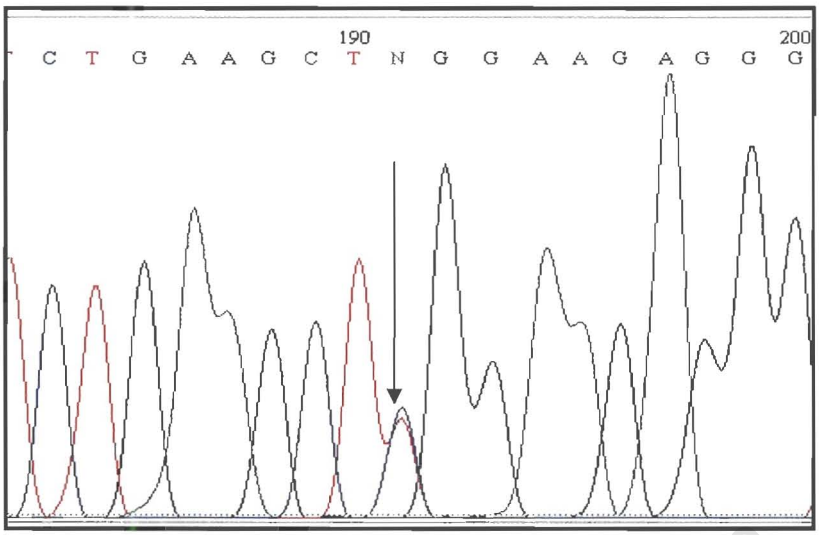


Figure 11: Pedigree for family 1 showing variant C1132T and members with ARVC

Another variation found in exon 4 was the C1162T variant (Figure 13A). This variation changed the amino acid from an Arginine to a Tryptophan and was found to occur in three ARVC probands (ACM5.1, ACM12.1; ACM 19.2). Analysis of our data indicated that the C1162T variation was found to occur in conjunction with the published exonic T1094C polymorphism in cases ACM 5.1, ACM 12.1 and ACM 19.2 (Figure 13).

A



B

```
H. sapiens_ TFIQHECFQKSEARKRVNQL
C. lupus_   TFIQHECFQKSEARKRVNQL
M. musculus_TFIQHESFQKSEARKRVNQL
G. gallus_  TFIQHECFQKAEARRKVFSL
```

Figure 12: In (A) an electrophorogram of ACM19.2 showing the C1162T variant and (B) sequence alignment with other modern animals

When the amino acid sequence from Homo sapiens was aligned with other modern animals it was found that the position of the C1162T is highly conserved through evolution (fig. 12B). Due to the highly conserved nature of this polymorphism we screened the family members of these three probands. It was however found that no extended families for ACM5.1 and ACM12.1 were available for screening. The C1162T variant was however detected in the extended family of proband ACM19.2 (figure 13).

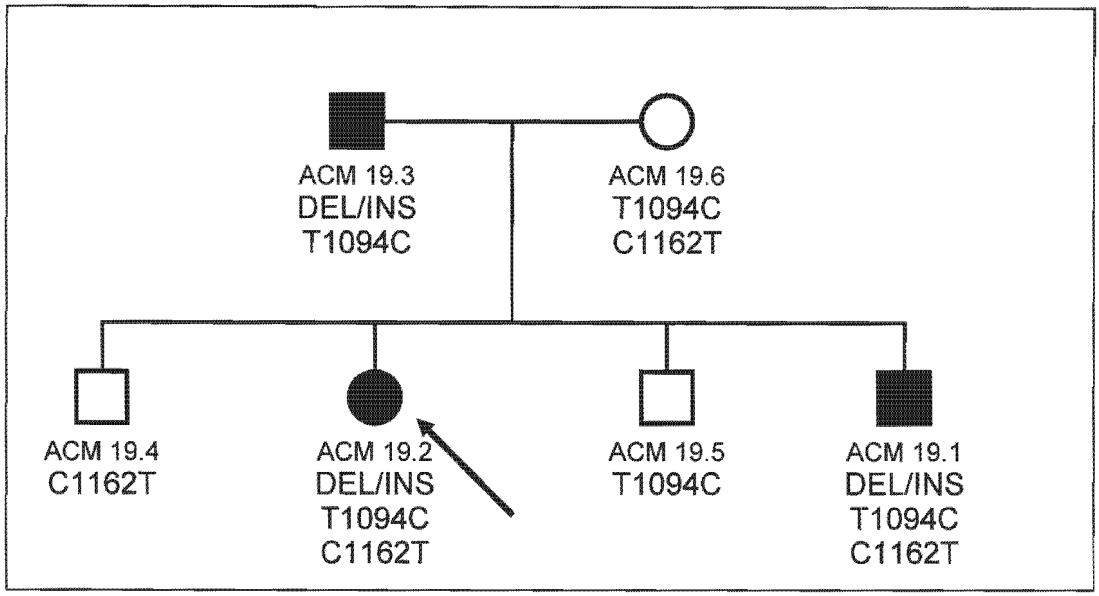


Figure13: The pedigree for ARVC family 19 showing segregation of the T1094C variant, the 2197-2202 ins/del and the C1162T variant

It was elected to use proband ACM 19.2 and do a family study for the recurrent T1094C and C1162T variants to establish whether they occurred in cis or in trans (i.e., whether they recurred on the same haplotype background or not). Figure 13 shows the transmission of the T1094C and C1162T from ACM19.6 (mother) to the different offspring, suggesting that the variants are in trans (i.e., they are on different chromosomes).

The phenotypic expression of ARVC in family ACM19 is of great interest. The proband (ACM19.2) presented with severe childhood onset ARVC and she required heart transplantation in her adolescence. Similarly, her brother (ACM19.1) presented with severe ARVC in his teens and required a heart transplant by his mid-teens. By contrast, the parents were not symptomatic with the father only having minor criteria for ARVC (i.e., ECG T wave inversion in V1-3, and mild wall motion abnormality of the right ventricle on cardiac MRI). The clinical genetics of ARVC in this family suggests that the rare C1162T mutation that is carried by the apparently unaffected mother (ACM19.6) may be acting as a modifier of the phenotype in the two affected offspring. The functional significance of the C1162T variant requires further study.

In exon 9, a novel intronic variation (IVS9+83C-G) was only identified in ACM 4.1. We screened for this variant in the background population; the results are reported in section 3.4.1

In exon 10, a novel intronic variation IVS10+44G-A, was identified in six probands individuals (i.e., ACM 9.1, ACM 20.1, DCM 10.3, FDC 1.1, DCM 226.1 and RCM 1.1). We screened for this variant in the background population; the results are reported in section 3.4.1

In exon 11, a novel intronic variation, IVS11+7C- T, was only identified in DCM 219.6. We screened for this variant in the background population; the results are reported in section 3.4.1

In exon 13+14, a novel intronic variant (IVS13+83G-A) was identified in ACM 6.1, ACM 8.3, and ACM 18.1. We screened for this variant in the background population; the results are reported in section 3.4.1

An exonic variation (T2540C) where the amino acid changed from a Leucine to Proline was only identified in proband DCM 3.5. Sequence alignments to other modern animals revealed that this region was found to be highly conserved. The variant was then screened in the family members of DCM3.5 in order to establish whether the T2504C variant segregated with disease. The T2540C was then identified in seven other family members (figure 14). However, the variant does not segregate with DCM in this family, raising doubt about its causal role in DCM.

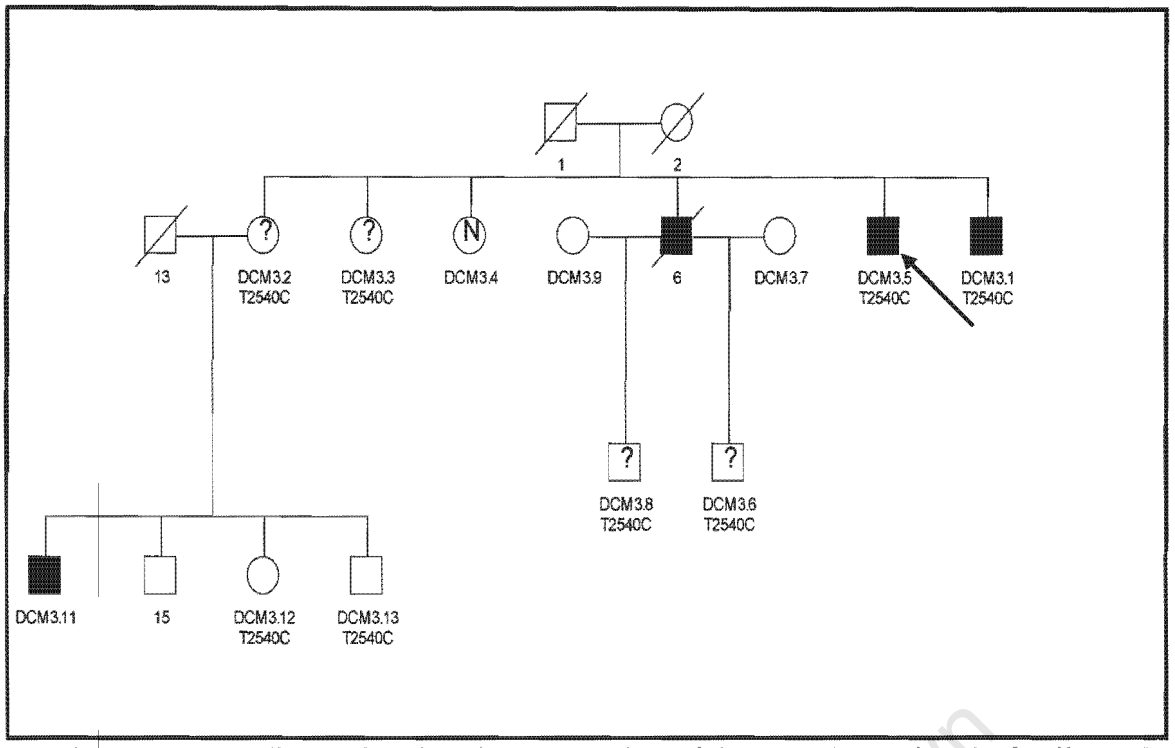


Figure 14: A pedigree showing the segregation of the T2540C variant in family DCM 3. Note the individual DCM 3.11, who is affected with DCM, does not carry the T2540C variant, raising doubt about the segregation of the variant with phenotype.

The 10 variants (three known and seven novel) detected in *PKP-2* were superimposed onto a cartoon drawing of the PKP-2 gene in an attempt to provide a global overview of the nature and position of the changes detected (figure 15).

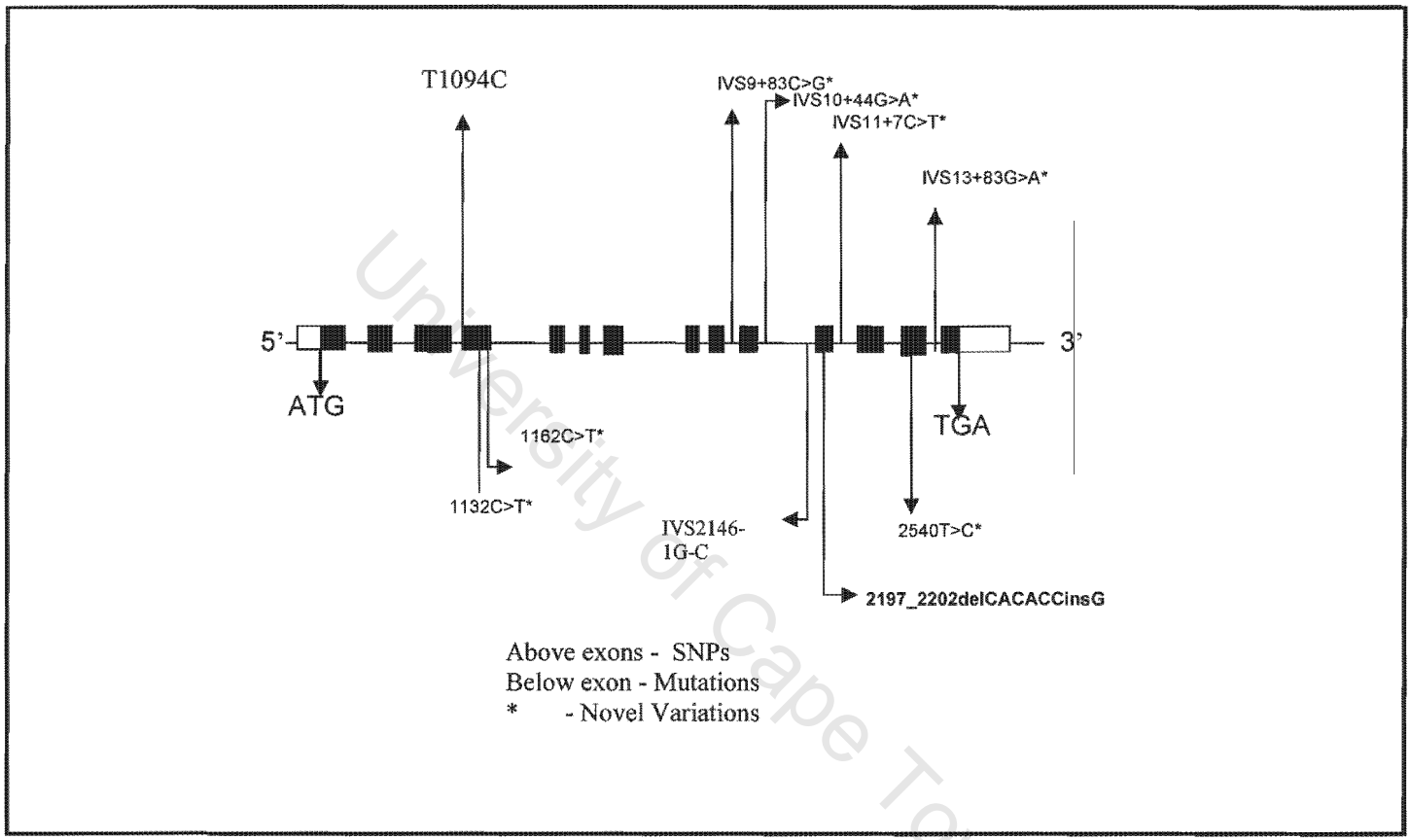


Figure 15: A schematic diagram showing the location of all variants identified in *PKP-2* exons in ARVC and DCM patients

3.4 SNaPshot results

SNaPshot was used to genotype the background population (normal control DNA) in order to determine the frequency of the seven new variants that were detected in this study. SNaPshot results were obtained through electropherograms with either homozygous or heterozygous peaks (figure 16). Heterozygous peaks confirmed a polymorphism/SNP, while homozygous peaks represented the wildtype (normal). A negative control was included to eliminate the possibility of contaminants during the experiment.

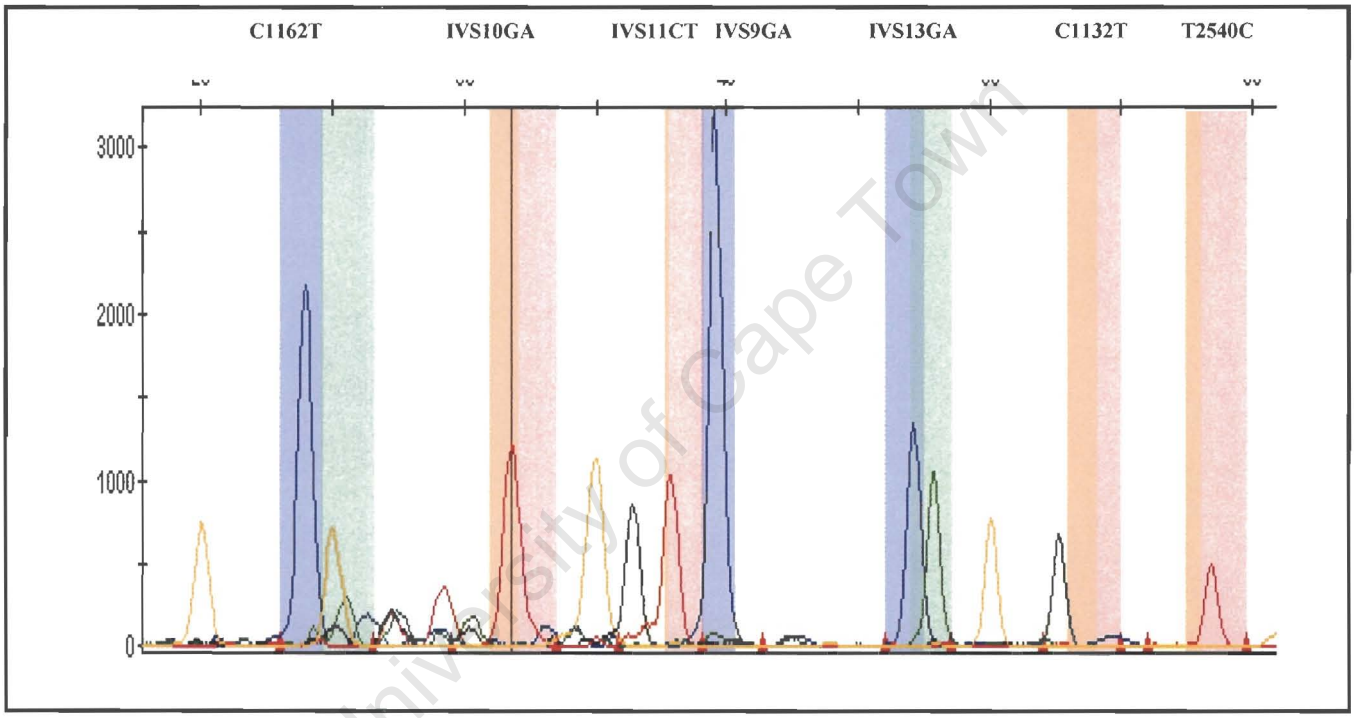


Figure16 : SNaPshot electropherogram of DNA from a control individual showing (1) a single blue peak for C1162T (i.e., homozygous wild type), (2) a single red peak for IVS10+44GA (homozygous wild-type), (3) two peaks for IVS11+7CT (i.e., heterozygote), (4) single blue IVS9+83GA (homozygous wild-type), (5) two peaks for IVS13+83GA (heterozygote), (6) one peak for C1132T (homozygous wild-type), and (7) one peak for T2540C (homozygous wild type).

3.4.1 Frequency of mutations in common population groups

The frequencies of the one known and seven novel variants were screened in three population groups, i.e., mixed ancestry, Blacks and Caucasians in order to establish their frequency in the South African background population. Tables 14 to 16 represents a summary of these frequencies.

Table 14: Frequency of mutations in the Black African population

BLACK AFRICANS

SNP's	*Alleles*	*Chromosome*	*Reference allele*	*Frequency*	*Other allele*	*Frequency*
T1094C	T/C	12	T	71/86(83.0%)	C	15/86(17.0%)
C1132T	C/T	12	C	86/86(100.0%)	T	0/86(0.0%)
C1162T	C/T	12	C	86/86(100.0%)	T	0/86(0.0%)
IVS9+83C-G	C/G	12	C	80/86(73.0%)	G	23/86(27.0%)
IVS10+44G-A	G/A	12	G	40/86(47.0%)	A	46/86(53.0%)
IVS11+7C-T	C/T	12	C	34/86(40.0%)	T	52/86(60.0%)
IVS13+83G-A	G/A	12	G	23/86(27.0%)	A	63/86(76.0%)
T2540C	T/C	12	T	86/86(100.0%)	C	0/86(0.0%)

IVS: intervening sequence (intron),-/+:up/downstream of the respective exon, %: percentage, SNP: single nucleotide polymorphism, ARVC: arrhythmogenic right ventricular cardiomyopathy, DCM: dilated cardiomyopathy, RCM: restrictive cardiomyopathy

Table 15: Frequency of mutations in the Caucasian population

CAUCASIAN

SNP's	*Alleles*	*Chromosome*	*Reference allele*	*Frequency*	*Other allele*	*Frequency*
T1094C	T/C	12	T	48/80(60.0%)	C	32/80(40.0%)
C1132T	C/T	12	C	80/80(100.0%)	T	0/80(0.0%)
C1162T	C/T	12	C	80/80(100.0%)	T	0/80(0.0%)
IVS9+83C-G	C/G	12	C	71/80(89.0%)	G	9/80(11.0%)
IVS10+44G-A	G/A	12	G	61/80(76.0%)	A	19/80(24.0%)
IVS11+7C-T	C/T	12	C	39/80(49.0%)	T	41/80(51.0%)
IVS13+83G-A	G/A	12	G	23/80(29.0%)	A	57/80(71.0%)
T2540C	T/C	12	T	80/80(100.0%)	C	0/80(0.0%)

IVS: intervening sequence (intron),-/+:up/downstream of the respective exon, %: percentage, SNP: single nucleotide polymorphism, ARVC: arrhythmogenic right ventricular cardiomyopathy, DCM: dilated cardiomyopathy, RCM: restrictive cardiomyopathy

Table 16: Frequency of mutations in the Mixed Ancestry population

MIXED ANCESTRY						
SNP's	*Alleles*	*Chromosome*	*Reference allele*	*Frequency*	*Other allele*	*Frequency*
T1094C	T/C	12	T	72/126(57.0%)	C	54/126(43.0%)
C1132T	C/T	12	C	75/75(100.0%)	T	0/75(0.0%)
C1162T	C/T	12	C	75/75(100.0%)	T	0/75(0.0%)
IVS9+83C-G	C/G	12	C	65/75(87.0%)	G	10/75(13.0%)
IVS10+44G-A	G/A	12	G	50/75(67.0%)	A	25/75(33.0%)
IVS11+7C-T	C/T	12	C	64/75(85.0%)	T	11/75(15.0%)
IVS13+83G-A	G/A	12	G	27/75(36.0%)	A	48/75(64.0%)
T2540C	T/C	12	T	75/75(100.0%)	C	0/75(0.0%)

IVS: intervening sequence (intron),-/+:up/downstream of the respective exon, %: percentage, SNP: single nucleotide polymorphism, ARVC: arrhythmogenic right ventricular cardiomyopathy, DCM: dilated cardiomyopathy, RCM: restrictive cardiomyopathy

In exon 4, we confirm the known status of the T1094C as a common polymorphism with the minor allele (C) frequency of 17% in blacks, 40% in Caucasians, and 43% in people of mixed ancestry. W also screened for both the stop codon, C1132T (Q-*(stop)), and the C1162T (R-W) variant but neither was found in 241 controls samples (482 chromosomes). Both variations were rare in all population groups screened.

In intron 9, IVS9+83C-G was identified in the population with the minor allele (G) at a frequency of 27% in blacks, 11% in Caucasians and 13% in people of mixed ancestry.

In intron 10, IVS10+44G-A was identified in the population with the minor allele (A) at a frequency of 53% in the black population, 24% in the Caucasian population and 33% in the mixed ancestry group.

In intron 11, IVS11+7C-T was identified in the population with the minor allele (T) at a frequency of 60% in the black population, 51% of Caucasian population and 15% of people of mixed ancestry.

In intron 13, IVS13+83G-A was found to be a common polymorphism with the minor allele (A) at a frequency of 76% in black controls, 71% in Caucasian controls and 64% in people of mixed ancestry. The T2540C was not identified in any of the controls but was only found in the DCM 3 family.

Table 17: Frequency of novel variants in PKP-2 by cardiomyopathy type

Variant	*Significance*	*ARVC*	*DCM*	*RCM*
C1132T	Disease-causing	1/25(4%)	0/19(0%)	-
C1162T	?Functional variant	3/25(12%)	0/19(0%)	-
IVS9+83C-G	Polymorphism	1/25(4%)	0/19(0%)	-
IVS10+44G-A	Polymorphism	2/25(8%)	4/19(21%)	1/1(100%)
IVS11+7C-T	Polymorphism	1/25(4%)	12/19(63%)	-
T2540C	?Functional variant	0/25(0%)	1/19(5%)	-
IVS13+83G-A	Polymorphism	13/25(52%)	1/19(5%)	-

IVS: intervening sequence (intron),-/+:up/downstream of the respective exon, %: percentage, SNP: single nucleotide polymorphism, ARVC: arrhythmogenic right ventricular cardiomyopathy, DCM: dilated cardiomyopathy, RCM: restrictive cardiomyopathy

The table 17 presents a summary of the 7 novel variants that were discovered in this study by disease type.

C1132T was identified in one ARVC proband (4%) and none of DCM or RCM probands had this variation.

C1162T was identified in three unrelated ARVC probands (12%) but did not appear in DCM or RCM probands. It is therefore possible that this variant may act as a modifier or disease-causing variant with variable penetrance.

IVS9+83C-G was only found in one of the ARVC probands (4%) but was not identified in DCM or RCM individuals. This variant was proven to be a common polymorphism, appearing in >1% of the population.

IVS10+44G-A was identified in two ARVC (8%), four DCM (21%) and one RCM (100%) individuals. This variant was proven to be a common polymorphism, appearing in >1% of the population.

IVS11+7C-T was identified one ARVC proband (4%), 12 DCM individuals (63%) and no RCM cases. This variant was proven to be a common polymorphism, appearing in >1% of the population.

T2540C was identified in one DCM proband (5%); none of the ARVC and RCM presented with this variation. This proved to be a rare variant that does not segregate with DCM in the family involved. The functional significance of this variant is yet to be established.

IVS13+83G-A was identified in 13 ARVC (52%), one DCM individual (5%) and no RCM cases. This variant was proven to be a common polymorphism as it was found in 1% of the population.

3.5 Haplotyping results

The rare C1162T occurred in three unrelated affected probands with ARVC (ACM5.1, ACM12.1, ACM19.2) (Table 13), but also in two individuals in family ACM19 without overt ARVC (see ACM19.4 and 19.6 in figure 13). There was evidence from family ACM19 (Figure 13) that, in view of the severe phenotype in the two affected offspring, the variant may be contributing to disease either on basis of compound heterozygosity (i.e., disease-causing with variable penetrance) or as a modifier gene (i.e., functional polymorphism that does not cause disease in the absence of another functional gene defect). We tested the hypothesis that the C1162T may have arisen on the same chromosome (i.e., founder effect) by constructing haplotypes in the three 'unrelated' probands who are carrying the C1162T variant.

The typing of three STR markers, D12S825316, D12S2451474 and D12S1692, together with the 10 variants in *PKP-2*, shows that the minor T allele of the C1162T variant is likely to occur on the same haplotype background. This rare haplotype background is made up of the following alleles (of the STRs) in the three unrelated individuals (ACM19.6, ACM5.1 and ACM12.1) : 1-2-1 (D12S825316, D12S2451474 and D12S1692) (Figures 17 and 18) and appears to occur together with the rare C1162T variant.

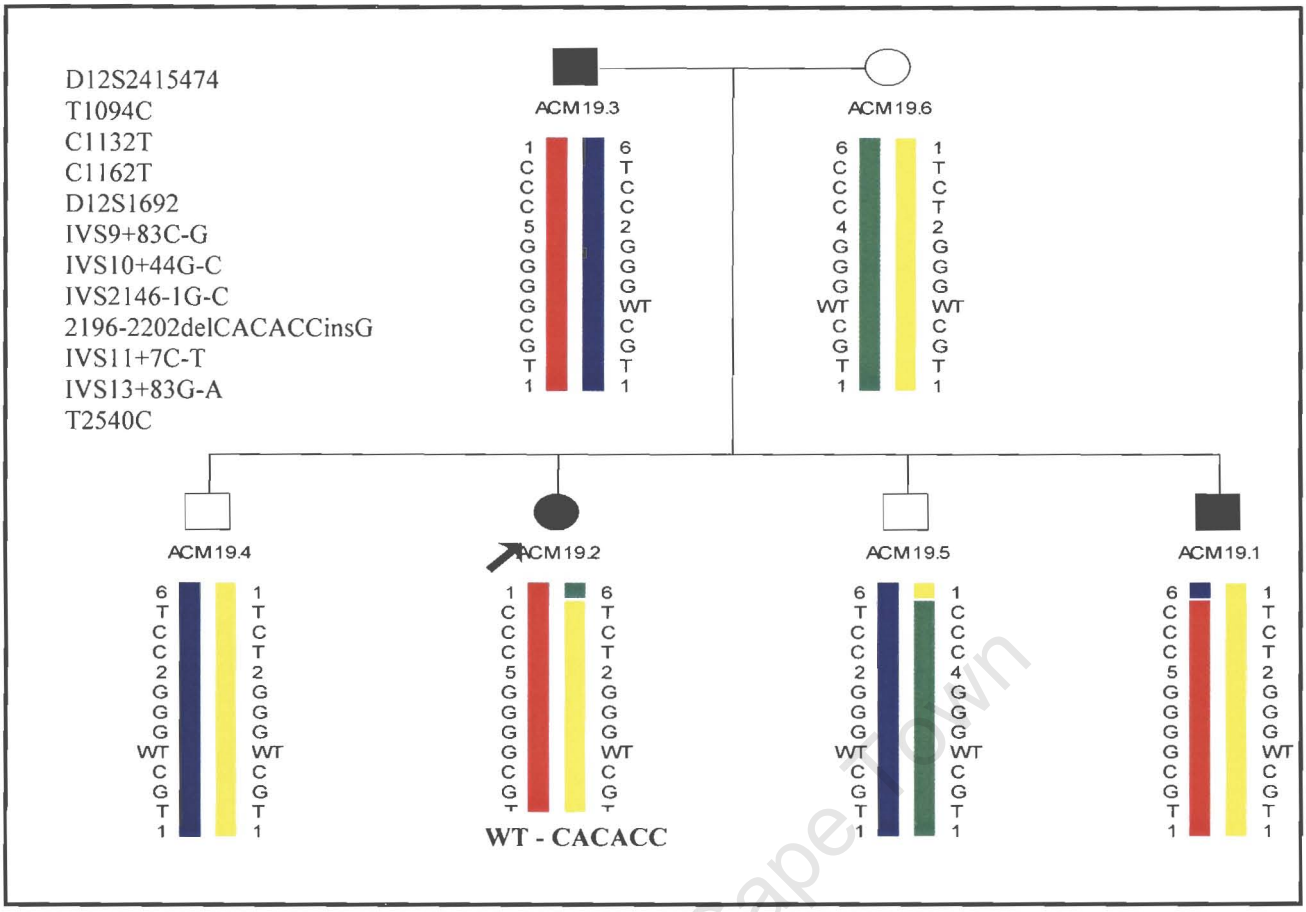


Figure 17: Haplotype results for ACM 19 family

Genotyping of ACM5.1 and ACM12.1, who also carry the C1162T variant, suggest that the T allele may be inherited on the 1-2-1 STR haplotype background (Figure 19).

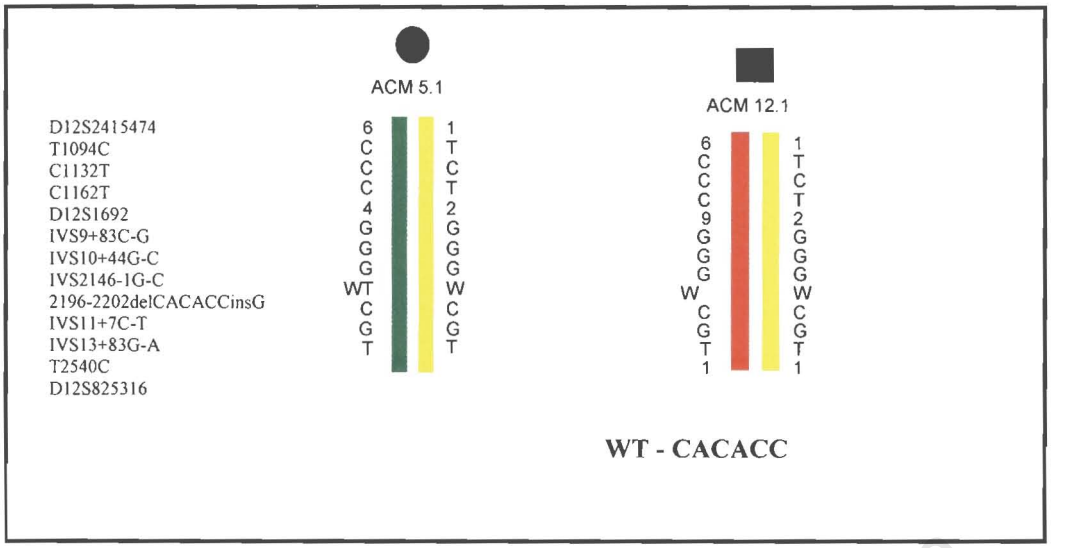


Figure 18: Haplotype results for ACM 5.1 and ACM 12.1

The absence of family information for individuals ACM5.1 and ACM12.1, however, makes it difficult to be certain of the assignment of phase in these individuals. Nevertheless, the apparent occurrence of the C1162T variant on the same haplotype background in three unrelated affected individuals (ACM5.1 and ACM or potentially affected (ACM19.6) with ARVC raises the possibility of a founder effect, should the C1162T variant be shown to be disease causing in functional studies.

Haplotyping of the PKP-2 gene was also carried out in the DCM family with the T2540C to confirm that the affected individual who did not carry this rare variant (i.e., DCM 3.11) was related to the family (Figure 19), but did not inherit the haplotype carrying the rare variant from his mother (DCM 3.2). The lack of segregation of the T2540C with DCM in this pedigree makes it unlikely to be the disease-causing mutation for DCM.

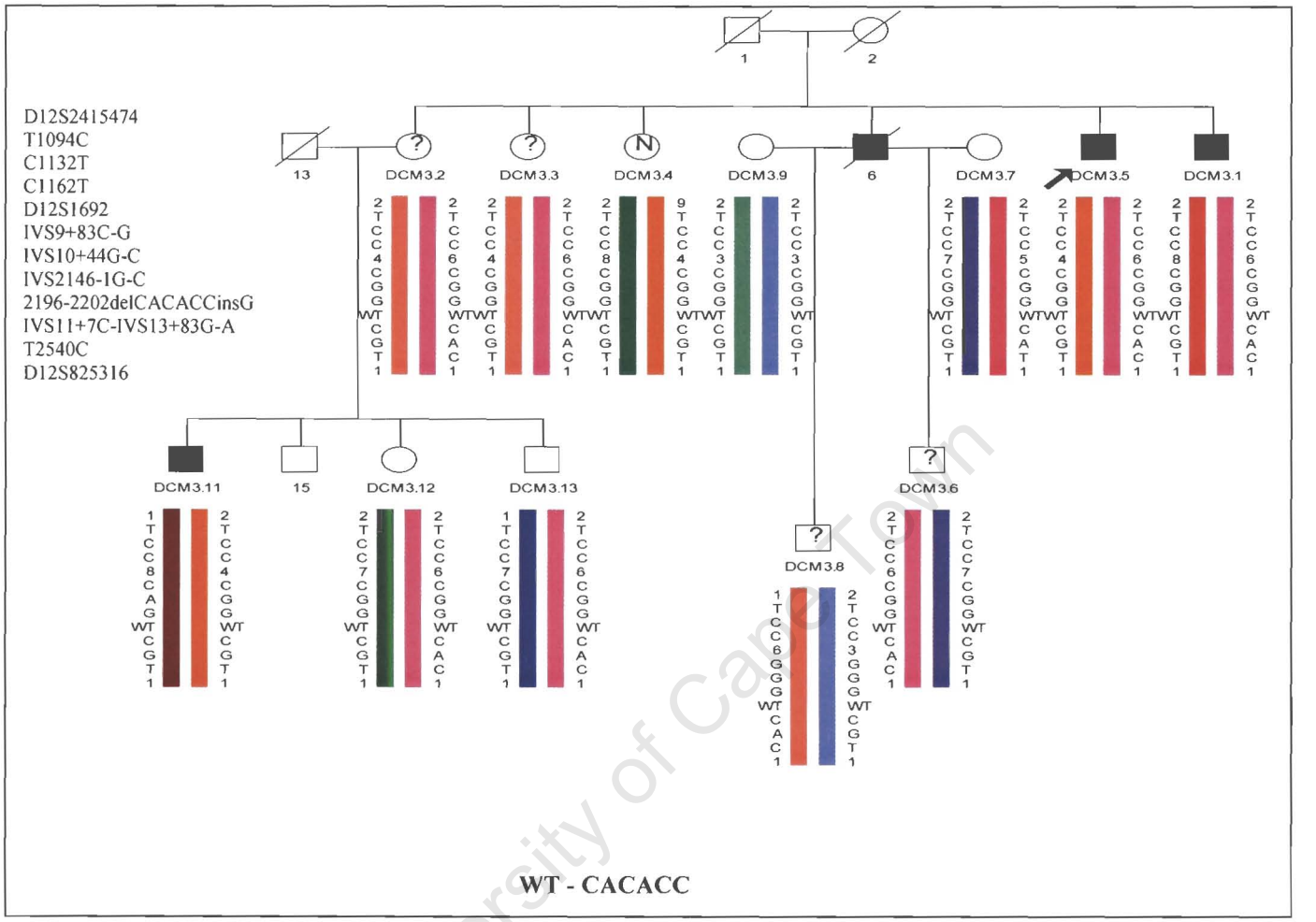


Figure 19: The haplotype results of family DCM 3

Chapter 4: Discussion and Conclusions

We have conducted, to the best of my knowledge, the first study of the prevalence of the Plakophilin-2 gene mutations in patients with cardiomyopathy in Africa. The mutation screening of the PKP-2 gene in 45 unrelated individuals yielded 10 variants, seven of which are novel. The prevalence of definite disease-causing mutations in the PKP-2 gene in patients with ARVC was 12% (3/25 individuals). This prevalence rate is lower than the average of 50- 70% that has been found in other studies (van Tintelen *et al 2006,* Awad et al. 2008). This might be caused by the small number of ARVC probands we had at that time. However, a rare variant in exon 4, the C1162T variant, was found in three unrelated ARVC probands but was not present in over 400 chromosomes in the general population. The functional significance of this variant is unknown because it has been observed in two relatives on an affected ARVC member who do not have the phenotype of ARVC. If this variant is shown in functional studies to be contributing to disease, with variable penetrance, then the prevalence of PKP-2 gene mutations could be as high as 20% in this study. All six intronic polymorphisms identified in this study were common and was found not to be significant; the four novel polymorphism will be recorded in the SNP database. Common polymorphisms were identified in all cohorts but only one novel polymorphism was found in the DCM cohort. It is known though, that DCM is a heterogeneous disease. This topic has been reviewed extensively by Seidman but is not addressed further in this thesis (Seidman *et al* 2006).

4.1 The important PKP-2 variants that were observed in this study will be discussed on an individual basis as follows:

4.1.1 Exon 4, C1132T, Q- stop codon

A novel C1132T single nucleotide change in exon 4 which is predicted to cause premature termination of translation of the *PKP2* protein in a family with two affected brothers was discovered. The stop codon introduced by this variant terminates the protein at position 378 (appendix 4), resulting in the loss of 503 amino acids, and termination of the protein in arm repeat unit one. The multiple alignment sequence of

this variant was compared to three modern species to determine the conservation of this amino acid among the species, rat, mouse and dog. The multiple sequence alignment showed that this amino acid was indeed conserved among species. This variant was not found after screening over 400 chromosomes in the general population. The C1132T mutation is considered to be disease-causing on the basis of (1) the predicted introduction of the stop codon, (2) the segregation with disease in the family, (3) the absence from the general population, and (4) the conservation of the region through evolution, indicating an essential role in the function of the PKP-2 gene.

4.1.2 Exon 4, C1162T, R-W (Arginine to tryptophan change)

In the same exon, a second variation was found in two unrelated probands and a family with early-onset ARVC. This variation is predicted to alter amino acid Arginine to Tryptophan in PKP-2 (appendix 4). Arginine is a basic amino acid with a positively charged side chain whilst tryptophan is a non-polar aromatic amino acid. The replacement of polar with a non-polar amino acid may be predicted to have a deleterious effect on the structure of *PKP-2*. Furthermore, multiple sequence alignment of the human protein with other species shows that this region is highly conserved through evolution. These lines of evidence, together with the fact that the variant was not found in over 400 normal chromosomes, suggest that the variant may be disease-causing. The appearance of this mutation on the same haplotype background in three unrelated individuals raises the possibility of a founder effect.

The study of family ACM 19, however, shows that this variant is seen in individuals who are not overtly affected with ARVC. It is therefore not possible to conclude, on the basis of the available data, on whether this variant is disease-causing or not. There is a need for functional studies of this mutation to establish its pathogenetic role in ARVC.

The significance of the C1162T variant has wider importance in view of the recurrent nature of mutation in patients with ARVC in this cohort. Haplotype analysis revealed that probands ACM 5.1, ACM 12.1 and ACM 19.6 may be related, raising the possibility of having descended from a common ancestor or founder. All three individuals carrying the recurrent variant are of Caucasian ancestry, which lends further support to the hypothesis of a founder effect. If the C1162T mutation proves to

be a disease-causing mutation, this will be another example of a founder effect in ARVC (Awad *et al.* 2008).

4.1.3 Exon 11, 2196-2202 del CACACC/ ins G mutation

The exonic deletion/insertion mutation, 2197-2202delCACACCinsG, is located in exon 11. This deletion/insertion mutation is predicted to cause a premature termination of PKP-2 protein in patients with ARVC (Gerull *et al.* 2004). This termination might change the folding or conformation of the arm repeat units and shorten the protein by 148 amino acids in arm repeat unit number eight (appendix 4). In family ACM 19, it was found that the mutation segregates with ARVC. The clinical genetics of ARVC in family ACM19 appear to exhibit the phenomenon of anticipation, in which the disease becomes progressively severe in subsequent generations. The father (ACM19.3) has no symptoms but mild ECG changes (T wave inversion in V1-3) and mild right ventricular motion abnormality on cardiac MRI. However, the two affected offspring have a severe phenotype that manifested in childhood. The mechanism for the severe phenotype in the offspring may be related to the inheritance of both the deletion/insertion mutation from the father and the rare C1162T variant of unknown functional significance from the mother. The mechanisms of disease in this family remain to be established by means of functional studies of the C1162T variant.

4.1.4 Exon 13, T2540C, (Leucine to Proline change)

A rare exonic variation was identified in exon 13 in a proband with familial DCM (DCM3.5). It was found that it alters a Leucine to Proline which belongs to the nonpolar group of amino acids (appendix 4). This variant is rare (i.e., not found in over 400 chromosomes in the general population), is in a conserved region of the protein, but did not segregate with DCM in the family. The impact of this variant on the function and structure of PKP-2 protein remains to be established.

4.2 Implications for clinical practice

We have identified disease-causing mutations in the PKP-2 gene in three families with ARVC. Presymptomatic screening is now available to first degree relatives of the affected individuals. It is well established that endurance exercise is an important

determinant of the clinical manifestations of ARVC. Furthermore, 50% of patients with ARVC present for the first time with unexpected sudden death. Therefore, there is a role for presymptomatic diagnosis in order to identify unaffected individuals who do not need screening, and to counsel affected individuals against activities that increase the risk of sudden death, such as competitive sports.

4.3 Implications for research

The functional significance of the rare C1162T variant in exon four require further study to establish its role in disease-causation or disease-modification, which would resolve the question of the presence of a founder effect for ARVC in South Africa. These studies will involve (1) the analysis of endomyocardial biopsy samples of the right ventricle from members of family ACM 19 for the expression of desmosomal proteins in the tissue, (2) yeast-two-hybrid studies of the interaction of mutant proteins with related proteins in the desmosome, and (3) the examination of the expression of mutant proteins in cell lines, such as the human embryonic kidney cells Asimaki *et al.* 2007).

All four novel sequence variants identified in this study need to be submitted to the SNP database.

4.4 Conclusion

We have found an initial prevalence of 12% for definite disease causing mutations for ARVC, which is much lower than the 50- 70% frequency found in other studies and this lower frequency may be caused by the small number of ARVC probands. We postulate that the rare variant found (i.e., the C1162T variant) may have a functional effect on *PKP2* that results in a compound heterozygous effect, and that there is a gene-dose effect on the phenotype. We also postulate that the recurrent C1162T variant, which occurred in 12% of the probands on a similar haplotype background, may represent a founder effect. In this study we identified six PKP-2 gene polymorphisms, of which two was shown to be specific for a particular disease eg T1094C in ARVC disease and T2540C (familial polymorphism) in DCM. The other four polymorphism were common in both diseases.

References

1. Ahmad F, Seidman JG, Seidman CE. (2005). **The Genetic Basis for Cardiac Remodeling**. *Annual Review of Genomics and Human Genetics.* 6(1):185-216.

2. Ahmad F, Li D, Karibe A, Gonzalez O, Tapscott T, Hill R, Weilbaecher D, Blackie P, Furey M, Gardner M, Bachinski LL, Roberts R. (1998). **Localization of a gene responsible for arrhythmogenic right ventricular dysplasia to chromosome 3p23.** *Circulation.* 98(25):2791-5.

3. Alberts B., Johnson A., Lewis J., Raff M., Roberts K. and Walter P. (2002). **Molecular biology of the cell.** *Fourth edition. Published by Garland Science.* 1066-1079.

4. Awad MM, Dalal D, Tichnell C, James C, Tucker A, Abraham T, Spevak PJ, Calkins H, Judge DP. (2006a). **Recessive arrhythmogenic right ventricular dysplasia due to novel cryptic splice mutation in PKP2**. *Hum Mutat.* 27(11):1157.

5. Awad MM, Dalal D, Cho E, Amat-Alarcon N, James C, Tichnell C, Tucker A, Russell SD, Bluemke DA, Dietz HC, Calkins H, Judge DP. (2006b). **DSG2 mutations contribute to arrhythmogenic right ventricular dysplasia/cardiomyopathy.** *Am J Hum Genet.* 79(1):136-42.

6. Awad MM, Calkins H, Judge DP. (2008). **Mechanisms of Disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy**. *Nat Clin Pract Cardiovasc Med.* 5(5):258-267.

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

8. Codd MB, Sugrue D., Gersh BJ, Melton LJ. (1989). Epidemiology **of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted Country, 1975-1984**. Circulation. 80(3):564-72.

9. Coonar AS, Protonotarios N, Tsatsopoulou A, Needham EW, Houlston RS, Cliff S, Otter MI, Murday VA, Mattu RK, McKenna WJ. (1998). **Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21.** *Circulation.* 97(20):2049-58.

10. Dalal D, Nasir K, Bomma C, Prakasa K, Tandri H, Piccini J, Roguin A, Tichnell C, James C, Russell SD, Judge DP, Abraham T, Spevak PJ, Bluemke DA, Calkins H. (2005). **"Arrhythmogenic Right Ventricular Dysplasia: A United States Experience."** *Circulation.* 112(25):3823-32.

11. Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kühl U, Maisch B, McKenna WJ, 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**. *Eur Heart J.* 29(2):270-6.

12. Frances R, Rodriguez Benitez AM, Cohen DR (1997) **Arrhythmogenic right ventricular dysplasia and anterior polar cataract**. *Am J Med Genet.* 73(2):125-6.

13. Frueh and Noyer-Weidner. (2003). **The use of denaturing High-Performance Liquid Chromatography for the analysis of genetic variations: Impact for diagnostics and phamacogenetics**. *Clin Chem Lab Med.* 41(4): 452-461

14. Fourlas CA, Trikas AGand Stefanadis CI. (2004). **Familial Dilated Cardiomyopathy a genetic enigma**. *Hellenic J. Cardiol.* (45): 42-45.

15. Gerull B, Heueser A, Wichter T, Paul M, Basson CT., McDermott DA., Lerman BB., Markowitz SM., Ellinor PT., MacRae CA, Peters S, Grossmann KS, Michely B, Sasse-Klaassen S, Birchmeier W, Dietz R, Breihardt G, Schulze-Bahr E and Thierfelder L. (2004). **Mutation in the desmosomal protein plakophilin-2 is common in arrhythmogenic right ventricular cardiomyopathy**. *Nat Genet.* 36(11):1162-4.

16. Green KJ and Gaudry CA. (2000). **Are desmosomes more than tethers for intermediate filaments? Nat Rev Mol Cell Biol**. 1(3):208-16. Review.

17. Heuser A, Plovie ER, Ellinor PT, Grossmann KS, Shin JT, Wichter T, Basson CT, Lerman BB, Sasse-Klaassen S, Thierfelder L, MacRae CA, Gerull B. (2006). **Mutant Desmocollin-2 Causes Arrhythmogenic Right Ventricular Cardiomyopathy**. *Am J Hum Genet*. 79(6):1081-8.

18. Indik JH. and Marcus FI. (2003). **Arrhythmogenic right ventricular cardiomyopathy/Dysplsia.** *Indian Pacing Electrophysiol J.* 3(3):148-56.

19. Kolar AJO, Milroy CM, Day PF and suvarna K. (2008). **Dilated cardiomyopathy and sudden death in a teenager with palmar-plantar keratosis (occult Carvajal syndrome).** *Forensic and legal medicine J.* 15: 185-188.

20. Li D, Ahmad F, Gardner MJ, Weilbaecher D, Hill R, Karibe A, Gonzalez O, Tapscott T, Sharratt GP, Bachinski LL, Roberts R. (2000). **The locus of a novel gene responsible for arrhythmogenic right-ventricular dysplasia characterized by early onset and high penetrance maps to chromosome 10p12-p14**. *Am J Hum Genet.* 66(1):148-56.

21. Matolweni LO, Bardien S, Rebello G, Oppon E, Munclinger MJ, Ramesar R and Mayosi BM.(2006). **Arrhythmogenic right ventricular dysplasia/cardiomyopathy type 6 (ARVC6): corroboration of locus assignment, narrowing of the critical region and mutation screening of three candidate genes**. *BMC Medical Genetics*. 7:29

22. Melberg A, Oldfors A, Blomstrom-Lundqvist C, Stalberg E, Carlsson B, Larrson E, Lidell C, Eeg-Olofsson KE, Wikstrom G, Henriksson G, Dahl N. (1999). **Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy linked to chromosome 10q**. *Ann Neurol.* 46(5):684-92.

23. McKenna WJ, Thiene G, Nava A, Fontaliran F, Blomstrom-Lundqvist C, Fontaine G and Camerini F. (1994). **Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology**. *Br Heart J.* 71(3):215-8.

24. McKoy G, Protonotarios N, Crosby A, Tsatsopoulou A, Anastasakis A, Coonar A, Norman M, Baboonian C, Jeffrey S and McKenna WJ. (2000). **Identification of a deletion in Arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease).** *Lancet.* 355(9221):2119-24.

25. Mestroni L, Maisch B, McKenna WJ, Schwartz K, Charron P, Rocco C, Tesson F, Richter A, Wilke A, Komajda M. (1999). **Guidelines for the study of familial dilated cardiomyopathies**. *Eur Heart J.* 20:93-102.

26. Michael KA, Latib MA, Matolweni LO and Mayosi BM. (2004). **Arrhythmogenic right ventricular cardiomyopathy: pathogenesis, diagonosis and management**. *S.A. heart*. 13- 22.

27. Morita H, Seidman J, and Seidman C E. (2005). **Genetic causes of human heart failure**. *J Clin Invest*. 115(3):518-26. Review.

28. Moolman-Smook JC, Mayosi BM, Brink PA and Corfield VA. (2003). **Molecular genetics of cardiomyopathy: shifting paradigms**. *Cardiovasc J S Afr.* 14(3):145-55. Review.

29. Nattel S. and Schott J. (2005). **Arrhythmogenic right ventricular dysplasia type 1 and mutations in transforming growth factor β3 gene regulatory regions: a breakthrough?** *Cardiovasc Res.* 65(2):302-4.

30. Norgett EE., Hatsell, SJ, Carvajal-Huerta, L., Cabezas, JC, Common J, Purkis, PE, Whittock, N, Leigh IM, Stevens HP, and Kelsell, DP. (2000). **Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma**. *Hum Mol Genet.* 9(18):2761-6.

31. Norman M, Simpson M, Mogensen J, Shaw A, Hughes S, Syrris P, Crossby A. and McKenna WJ. (2005). **Novel mutation in desmoplakin causes arrhythmogenic left ventricular cardiomyopathy**. *Circulation.* 112(5):636-42.

32. Kühl U, Noutsias M, Seeberg B and Schulthesis HP. (1996). **Immunological analysis for a chronic intramyocardial inflammatory processs in dilated cardiomyopathy**. *Heart.* 75(3):295-300.

33. Pilichou K, Nava A, Basso C, Beffagna G, Bauce B, Lorenzon A, Frigo G, Vettori A, Valente M, Towbin J, Thiene G, Danieli GA, Rampazzo A. (2006). **Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy**. *Circulation.* 113(9):1171-9.

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

35. Rampazzo A, Nava A, Danieli GA, Buja G, Daliento L, Fasoli G, Scognamiglio R, Corrado D, Thiene G. (1994). **The gene for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 14q23-q24**. *Hum Mol Genet*. 3(6):959-62.

36. Rampazzo A, Nava A, Malacrida S, Beffagna G, Bauce B, Rossi V, Zimbello R, Simionati B, Basso C, Thiene G, Towbin JA, Danieli GA. (2002). **Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy.** *Am J Hum Genet.* 71(5):1200-6.

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

38. Rampazzo A, Nava A, Miorin M, Fonderico P, Pope B, Tiso N, Livolsi B, Zimbello R, Thiene G, Danieli GA. (1997). **ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy, maps to chromosome 2 long arm**. *Genomics*. 15; 45(2):259-63.

39. Richardson P, McKenna W, Bristow M, Maisch B, Mautner B, O'Connel J, Olsen E, Thiene G, Goodwin J, Gyarfas I, Martin I And Nordet P. (1996). **Report of the 1995 World Health Organisation/ International Society and Federation of Cardiology task Force on the Definition and Classification of Cardiomyopathies.** *Circulation.* 1; 93(5):841-2.

40. Tiso N, Stephan DA, Nava A, Bagattin A, Devaney JM, Stanchi F, Larderet G, Brahmbhatt B, Brown K, Bauce B, Muriago M, Basso C, Thiene G, Danieli GA, Rampazzo A. (2001). **Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2).** *Hum Mol Genet.* 1; 10(3):189-94.

41. Schönberger J. and Seidman. (2001). **Many roads lead to a broken heart: The genetics of dilated cardiomyopathy.** *Am J Hum Genet.* 69(2):249-60. Review.

42. Sen-Chowdhry S, Syrris P, McKenna WJ. (2007). **Role of genetic analysis in the management of patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy.** *J Am Coll Cardiol.* 6; 50(19):1813-21.

43. Severini GM, Krajinovic M, Pinamonti B, Sinagra G, Fioretti P, Brunazzi MC, Falaschi A, Camerini F, Giacca M, Mestroni L. (1996). **A new locus for arrhythmogenic right ventricular dysplasia on the long arm of chromosome 14.** *Genomics.* 15; 31(2):193-200.

44. Strachan T. and Read A.P. (1996). **Human Molecular genetics.** *Third ed. John Willey and Sons publications.* 119-135

45. Syrris P., Ward D., Asimaki A., San-Chowdhry S., Abrahim H.Y., Evans A., Hitomi N., Norman M Pantazis A., Shaw A. L., Elliot P.M. and McKenna W.J. (2006). **Clinical expression of plakophilin-2 mutations in familial arrhythmogenic right ventricular cardiomyopathy.** *Circulation.* 24; 113(3):356-64.

46. van Tintelen JP, Entius MM, BhuiyanzaZA, Jongbloed R, Wiefeld AC, Wilde AA, van der Smagt J, Boven LB, manners MM, van Langen IM, Hofstra RM, Otterspoor LC, Doevendans PA, Rodriguez LM van Gelder IC, hauer RN. (2006). **Plakophilin-2 mutations are major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy.** *Circulation.* 113:1650-1658.

47. Xiao W. and Oefner P.J. (2001) **Denaturing High-Performance Liquid Chromatography.** *Hum Mutat.* 17(6):439-74. Review.

48. Williams R., Kline M. and Smith R. (1996) **BSA and Restriction Enzyme digestion.** *Promega notes magazine.* 59:46

Appendices

Appendix 1: An example of a consent form

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 and Buffers

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

216g (0.89M) Tris (ICN Biomedicals, Inc.)
110g (0.89M) Boric Acid (ICN Biomedicals, Inc.)

14.8g (0.5M) EDTA, pH 8.0 (Merck)

Make up to 2L with dH_2O

Agarose Loading Dye solution

0.125g (0.25%) Bromophenol blue

20g (40%) Sucrose

Make up to 50ml with dH_2O and ensure the pH is 8 or more basic

1X Buffer O solution (for 100% Aci I digestion)

50mM Tris-HCL (pH 7.5)

10mM $MgCl_2$

100mM NaCl

0.1mg/ml BSA

1X Buffer Tango solution (for 100% Nsp I digestion)

33mM Tris- acetate (pH 7.9)

10mM magnesium acetate

66mM potassium acetate

0.1mg/ml BSA

1X Buffer Sda I solution (for 100% Sda I digestion)

37mM Tris- acetate (pH 7.0)

15mM magnesium acetate

150mM potassium acetate

0.1mg/ml BSA

Silver staining solutions

Solution 1:

1g Silver Nitrate

Make up to 1L with dH_2O

Solution 2:

15g NaOH pellets

800ml dH_2O

Mix on magnetic stirrer

Add 10ml of 15% formaldehyde

Make up to 1L

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

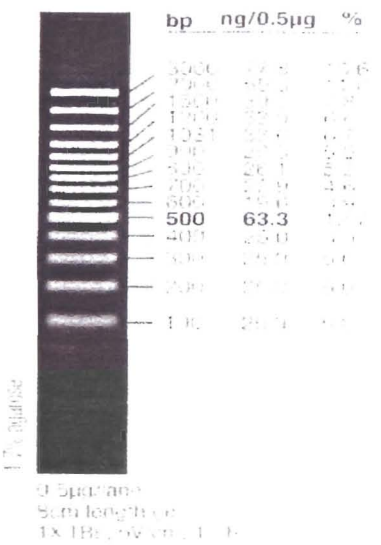
216g (0.89M) Tris (ICN Biomedicals, Inc.)
110g (0.89M) Boric Acid (ICN Biomedicals, Inc.)

14.8g (0.5M) EDTA, pH 8.0 (Merck)

Make up to 2L with dH_2O

Appendix 3: GeneRuler^{TM} 100bp DNA Ladder Plus

GeneRuler™ 100bp DNA Ladder Plus



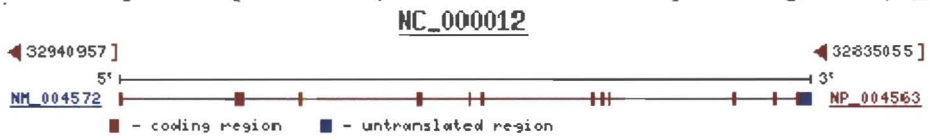
Appendix 4: mRNA sequence showing all exonic variants that are possible disease causing

PKP2 plakophilin 2 [*Homo sapiens*]

GeneID: 5318 Locus tag: HGNC:9024; MIM: 602861 updated 14-Apr-2005

Official Symbol: PKP2 **and Name:** plakophilin 2

Transcripts and products: (shown on reverse complement genome) RefSeq below



Genomic context: chromosome: 12; **Maps:** 12p11

The Sequence Manipulation Suite: Show Translation

Results for 2646 residue sequence "NM_004572" starting "atggcagccc".

```
     1 M  A  A  P  G  A  P  A  E  Y  G  Y  I  R  T  V  L  G  Q  Q
     1 atggcagcccccggcgccccagctgagtacggctacatccggaccgtcctgggccagcag
    21 I  L  G  Q  L  D  S  S  S  L  A  L  P  S  E  A  K  L  K  L
    61 atcctgggacaactggacagctccagcctggcgctgccctccgaggccaagctgaagctg
    41 A  G  S  S  G  R  G  G  Q  T  V  K  S  L  R  I  Q  E  Q  V
   121 gcggggagcagcggccgcggcggccagacagtcaagagcctgcggatccaggagcaggtg
    61 Q  Q  T  L  A  R  K  G  R  S  S  V  G  N  G  N  L  H  R  T
   181 cagcagaccctcgcccggaagggccgcagctccgtgggcaacggaaatcttcaccgaacc
    81 S  S  V  P  E  Y  V  Y  N  L  H  L  V  E  N  D  F  V  G  G
   241 agcagtgttcctgagtatgtctacaacctacacttggttgaaaatgattttgttggaggc
   101 R  S  P  V  P  K  T  Y  D  M  L  K  A  G  T  T  A  T  Y  E
   301 cgttcccctgttcctaaaacctatgacatgctaaaggctggcacaactgccacttatgaa
   121 G  R  W  G  R  G  T  A  Q  Y  S  S  Q  K  S  V  E  E  R  S
   361 ggtcgctggggaagaggaacagcacagtacagctcccagaagtccgtggaagaaaggtcc
   141 L  R  H  P  L  R  R  L  E  I  S  P  D  S  S  P  E  R  A  H
   421 ttgaggcatcctctgaggagactggagatttctcctgacagcagcccggagagggctcac
   161 Y  T  H  S  D  Y  Q  Y  S  Q  R  S  Q  A  G  H  T  L  H  H
   481 tacacgcacagcgattaccagtacagccagagaagccaggctgggcacaccctgcaccac
   181 Q  E  S  R  R  A  A  L  L  V  P  P  R  Y  A  R  S  E  I  V
   541 caagaaagcaggcgggccgccctcctagtgccaccgagatatgctcgttccgagatcgtg
   201 G  V  S  R  A  G  T  T  S  R  Q  R  H  F  D  T  Y  H  R  Q
   601 ggggtcagccgtgctggcaccacaagcaggcagcgccactttgacacataccacagacag
   221 Y  Q  H  G  S  V  S  D  T  V  F  D  S  I  P  A  N  P  A  L
   661 taccagcatggctctgttagcgacaccgtttttgacagcatccctgccaacccggccctg
   241 L  T  Y  P  R  P  G  T  S  R  S  M  G  N  L  L  E  K  E  N
   721 ctcacgtaccccaggccagggaccagccgcagcatgggcaacctcttggagaaggagaac
   261 Y  L  T  A  G  L  T  V  G  Q  V  R  P  L  V  P  L  Q  P  V
   781 tacctgacggcagggctcactgtcgggcaggtcaggccgctggtgcccctgcagcccgtc
   281 T  Q  N  R  A  S  R  S  S  W  H  Q  S  S  F  H  S  T  R  T
   841 actcagaacagggcttccaggtcctcctggcatcagagctccttccacagcacccgcacg
   301 L  R  E  A  G  P  S  V  A  V  D  S  S  G  R  R  A  H  L  T
   901 ctgagggaagctgggcccagtgtcgccgtggattccagcgggaggagagcgcacttgact
```

```
  321 V  G  Q  A  A  A  G  G  S  G  N  L  L  T  E  R  S  T  F  T
  961 gtcggccaggcggccgcaggggaagtgggaatctgctcactgagagaagcactttcact
  341 D  S  Q  L  G  N  A  D  M  E  M  T  L  E  R  A  V  S  M  L
 1021 gactcccagctggggaatgcagacatggagatgactctggagcgagcagtgagtatgctc
  361 E  A  D  H  M  P  P  S  R  I  S  A  A  A  T  F  I  Q  H  E
Q Δ to *
 1081 gaggcagaccacatgccgccatccaggatttctgctgcagctactttcatacagcacgag
CAG Δ to TAG
  381 C  F  Q  K  S  E  A  *  K  R  V  N  Q  L  R  G  I  L  K  L
* Δ to *
 1141 tgcttccagaaatctgaagctcggaagagggttaaccagcttcgtggcatcctcaagctt
CGG Δ to TGG
  401 L  Q  L  L  K  V  Q  N  E  D  V  Q  R  A  V  C  G  A  L  R
 1201 ctgcagctcctaaaagttcagaatgaagacgttcagcgagctgtgtgtggggccttgaga
  421 N  L  V  F  E  D  N  D  N  K  L  E  V  A  E  L  N  G  V  P
 1261 aacttagtatttgaagacaatgacaacaaattggaggtggctgaactaaatggggtacct
  441 R  L  L  Q  V  L  K  Q  T  R  D  L  E  T  K  K  Q  I  T  D
 1321 cggctgctccaggtgctgaagcaaaccagagacttggagactaaaaaacaaataacagac
  461 H  T  V  N  L  R  S  R  N  G  W  P  G  A  V  A  H  A  C  N
 1381 catacagtcaatttaagaagtaggaatggctggccgggcgcggtggctcacgcctgtaat
  481 P  S  T  L  G  G  Q  G  G  R  I  T  R  S  G  V  R  D  Q  P
 1441 cccagcactttgggaggccaaggcgggcggatcacgaggtcaggagttcgagaccagcct
  501 D  Q  H  G  L  L  W  N  L  S  S  N  D  K  L  K  N  L  M  I
 1501 gaccaacatggtttgctgtggaatttgtcatctaatgacaaactcaagaatctcatgata
  521 T  E  A  L  L  T  L  T  E  N  I  I  I  P  F  S  G  W  P  E
 1561 acagaagcattgcttacgctgacggagaatatcatcatccccttttctgggtggcctgaa
  541 G  D  Y  P  K  A  N  G  L  L  D  F  D  I  F  Y  N  V  T  G
 1621 ggagactacccaaaagcaaatggtttgctcgattttgacatattctacaacgtcactgga
  561 C  L  R  N  M  S  S  A  G  A  D  G  R  K  A  M  R  R  C  D
 1681 tgcctaagaaacatgagttctgctggcgctgatgggagaaaagcgatgagaagatgtgac
  581 G  L  I  D  S  L  V  H  Y  V  R  G  T  I  A  D  Y  Q  P  D
 1741 ggactcattgactcactggtccattatgtcagaggaaccattgcagattaccagccagat
  601 D  K  A  T  E  N  C  V  C  I  L  H  N  L  S  Y  Q  L  E  A
 1801 gacaaggccacggagaattgtgtgtgcattcttcataacctctcctaccagctggaggca
  621 E  L  P  E  K  Y  S  Q  N  I  Y  I  Q  N  R  N  I  Q  T  D
 1861 gagctcccagagaaatattcccagaatatctatattcaaaaccggaatatccagactgac
  641 N  N  K  S  I  G  C  F  G  S  R  S  R  K  V  K  E  Q  Y  Q
 1921 aacaacaaaagtattggatgttttggcagtcgaagcaggaaagtaaaagagcaataccag
  661 D  V  P  M  P  E  E  K  S  N  P  K  G  V  E  W  L  W  H  S
 1981 gacgtgccgatgccggaggaaaagagcaaccccaagggcgtggagtggctgtggcattcc
  681 I  V  I  R  M  Y  L  S  L  I  A  K  S  V  R  N  Y  T  Q  E
 2041 attgttataaggatgtatctgtccttgatcgccaaaagtgtccgcaactacacacaagaa
  701 A  S  L  G  A  L  Q  N  L  T  A  G  S  G  P  M  P  T  S  V
 2101 gcatccttaggagctctgcagaacctcacggccggaagtggaccaatgccgacatcagtg
  721 A  Q  T  V  V  Q  K  E  S  G  L  Q  H  T  R  K  M  L  H  V
 2161 gctcagacagttgtccagaaggaaagtggcctgcagcacaccgaaagatgctgcatgtt
  741 G  D  P  S  V  K  K  T  A  I  S  L  L  R  N  L  S  R  N  L
 2221 ggtgacccaagtgtgaaaaagacagccatctcgctgctgaggaatctgtcccggaatctt
  761 S  L  Q  N  E  I  A  K  E  T  L  P  D  L  V  S  I  I  P  D
 2281 tctctgcagaatgaaattgccaaagaaactctccctgatttggtttccatcattcctgac
  781 T  V  P  S  T  D  L  L  I  E  T  T  A  S  A  C  Y  T  L  N
 2341 acagtcccgagtactgaccttctcattgaaactacagcctctgcctgttacacattgaac
  801 N  I  I  Q  N  S  Y  Q  N  A  R  D  L  L  N  T  G  G  I  Q
 2401 aacataatccaaaacagttaccagaatgcacgcgaccttctaaacaccgggggcatccag
  821 K  I  M  A  I  S  A  G  D  A  Y  A  S  N  K  A  S  K  A  A
 2461 aaaattatggccattagtgcaggcgatgcctatgcctccaacaaagcaagtaaagctgct
  841 S  V  L  L  Y  S  L  W  A  H  T  E  L  H  H  A  Y  K  K  A
L Δ to L
 2521 tccgtccttctgtattctctgtgggcacacacggaactgcatcatgcctacaagaaggct
CTG Δ to TTG
```

```
 861 Q  F  K  K  T  D  F  V  N  S  R  T  A  K  A  Y  H  S  L  K
2581 cagtttaagaagacagattttgtcaacagccggactgccaaagcctaccactcccttaaa
 881 D  *
2641 gactga
```

Appendix 5: Electronic Resources

1. faculty.plattsburgh.edu/donald.slish/Electrophoresis.html: Electrophoresis.
2. health.yahoo.com/.../healthwise/popup/hw141771: A diagram of a normal and heart with dilated cardiomyopathy.
3. rna.lundberg.gu.se/cutter2/
4. www.nihlbi.nih.gov/health/dci/diseases: The heart failure.
5. www.hoslink.com/heart.htm: Different types of diseases that causes the heart failure.
6. www.ualberta.ca/~stothard/javascript/: DNA analysis (sequence translation).
7. www.fruitfly.org/seq tools/splice.htm/: Intronic analysis.
8. www.jdaross.cwc.net/heart4.htm: Composition of the heart