Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Cardiologiche, Toraciche e Vascolari

SCUOLA DI DOTTORATO DI RICERCA IN

SCIENZE MEDICHE, CLINICHE E SPERIMENTALI

INDIRIZZO: "SCIENZE CARDIOVASCOLARI"

XXVIII CICLO

Arrhythmogenic Cardiomyopathy: identification of novel genes encoding for intercalated disc proteins by next generation sequencing

Direttore della Scuola : Ch.mo Prof. Gaetano Thiene

Coordinatore d'indirizzo: Ch.mo Prof. Gaetano Thiene

Supervisore: Ch.ma Prof.ssa Alessandra Rampazzo

Dottoranda: Dott.ssa Giulia Poloni

31 gennaio 2016

INDEX

ABSTRACT

Introduction. Arrhythmogenic cardiomyopathy (ACM) is a predominantly genetically determined disease characterized by fibrofatty replacement which leads to right ventricular failure, arrhythmias, and sudden cardiac death (SCD). ACM is inherited as an autosomal dominant trait with incomplete penetrance. Advances in genetic technology have revealed substantial genetic heterogeneity: at least 15 independent loci and 13 disease genes have now been identified associated with the disease, with the large involvement of the genes encoding for desmosomal and *area composita* proteins. Since causative mutations in ACM genes have been detected in about 60% of probands, additional and still unknown disease-genes could be involved.

Aim of the study. This study aimed at detecting causative mutations underlying the disease expression in a cohort of 59 Italian unrelated index cases through a new targeted next generation sequencing (NGS) approach. Moreover, the identification of novel disease loci and genes in four families with recurrence of ACM was attempted by integrating different genetic approaches.

Methods. Three different NGS approaches have been applied: targeted gene panels (TGP), whole exome sequencing (WES), and whole genome sequencing (WGS). Two custom targeted gene panels, including 56 genes associated with different cardiomyopathies or 69 known and candidate ACM genes have been used to screen both familial and sporadic cases. Genetic analysis was extended to available family members to evaluate the segregation of each mutation identified in the index case. The whole exome of 11 subjects belonging to 4 families was sequenced. In one of these families (Family#6) also WGS in 3 subjects in addition to multipoint linkage analysis was performed.

Results. Targeted gene panels resulted a valuable tool for mutation screening in patients affected with ACM. At least a mutation was found in 15 out of 19 probands screened with the 'Cardiomyopathies gene panel'. Moreover, by using the 'ACM known and candidate genes panel', a mutation in a candidate gene has been identified in 40.6% of probands negative for mutations in known-disease genes. Two novel missense variants in *TJP1* gene (p.R265W, p.Y669C), encoding for protein ZO-1 and two in *CDH2* gene (p.E493G, p.V491G), encoding for N-cadherin resulted of particular interest, other than a stop mutation in *TP63* gene (p.R266*). Both *TJP1* mutations affected highly conserved amino acid residues and *in silico* analysis of p.Y669C mutation, which segregates within the proband's family, showed its predicted damaging effect into the protein structure.

WES analysis allowed to identify a pathogenic mutation in *DSP* gene (p.Q1297*) not previously detected by dHPLC in Family#3, and two putative *TTN* mutations

(p.R32573C and p.L32198M) that segregate in Family#4 and Family#5. In Family#5, a rare stop mutation in a candidate gene (*CMYA5*, p.K3597*) was also identified.

In Family #6, a multi-step approach was applied. The availability of many affected and unaffected family members allowed to perform linkage analysis in an 'affected only' approach. The analysis pointed out the presence of two regions with positive pLOD score values on chromosome 19p13.3 and 11q21. The first corresponded to a 7cM region and was shared by all the affected subjects, but one. The 3cM region on chromosome 11q21 segregated in all the affected subjects exception made for two, who carried a *PKP2* splice site variant (c.2578-3 T>C). WES performed in 4 patients of the family failed to identified a possible shared pathogenic mutation neither inside the critical regions, nor in the whole exome. These results were confirmed with WGS performed in 2 affected and a healthy individual. Moreover WGS pointed out the presence of a huge amount of complex variants located in repetitive intronic or intragenic regions.

Discussion. The identification of disease-causing mutations facilitates timely diagnosis, allows the prevention of complications and determines the potential risk in close relatives of a proband. For hereditary cardiomyopathies, as well as for most mendelian diseases, in the last ten years NGS technology improved this process allowing to parallel sequence a large amount of genes in a time and cost-efficient manner. Since mutations in the most index cases sequenced by using a targeted gene panel have been detected, it resulted a valid approach not only for genetic testing but also to identify putative novel disease genes. In the present study, novel mutations in genes encoding for intercalated disc proteins (*TJP1*, *CDH2*) have been identified, confirming the idea that ACM has to be considered a 'junctional disease' rather than only a 'desmosomal disease'.

Despite TGP remains the most commonly used approach for hereditary cardiomyopathies, WES, applied in familial ACM cases, allowed to identify novel variants in candidate genes that would be not detected with TGP and could have a possible pathogenic role or a modifier effect in the phenotypic expression. Finally, even though WGS didn't allow to make a genotype-phenotype correlation in Family #6, the huge amount of data produced represent permanent data that could be re-analysed in the future with new insights and the discovery of novel disease genes.

RIASSUNTO

Introduzione. La cardiomiopatia aritmogena (ACM) è una patologia ereditaria del muscolo cardiaco, caratterizzata da una progressiva sostituzione adiposa o fibroadiposa a carico prevalentemente del miocardio del ventricolo destro. Dal punto di vista clinico, è una patologia eterogenea con ampia variabilità clinica inter- ed intra- familiare; presenta infatti sia forme completamente asintomatiche sia forme molto gravi con rischio di morte improvvisa. Questa patologia, geneticamente eterogenea, è trasmessa come carattere autosomico dominante a penetranza incompleta ed espressività variabile. Ad oggi 15 loci e 13 geni sono stati associati alla malattia, di cui gran parte codificano per proteine desmosomali e proteine della cosiddetta *area composita* dei dischi intercalari. Poiché sono state identificate mutazioni causative in geni noti sono nel 60% dei casi, altri geni, non ancora identificati, potrebbero essere coinvolti nella comparsa del fenotipo patologico.

Scopo della ricerca. Nello studio descritto nella presente tesi, il DNA di 59 casi indice è stato analizzato attraverso due diversi pannelli di geni tramite sequenziamento di nuova generazione (NGS). Inoltre, in quattro famiglie con ricorrenza di casi di ACM, in cui non sono state identificate mutazioni nei geni desmosomali, sono state integrate diverse tecniche allo scopo di identificare nuovi loci e geni malattia.

Metodi. Nell'ambito del sequenziamento di nuova generazione (NGS) tre diversi approcci sono stati utilizzati: il sequenziamento di pannelli di geni target (TGP), il sequenziamento dell'intero esoma (WES) e il sequenziamento dell'intero genoma. Due sono i pannelli di geni considerati, uno comprendente 56 geni associati a diverse cardiomiopatie, l'altro comprendente 69 geni, tra cui i 13 geni associati alla cariomiopatia aritmogena e 56 geni candidati. L'analisi genetica è stata poi estesa ai familiari dei probandi, ove disponibili, per valutare la segregazione delle mutazioni identificate. L'intero esoma è stato poi sequenziato in 11 soggetti appartenenti a 4 diverse famiglie. Infine, in una di queste famiglie (Famiglia#6) è stata eseguita un'analisi di linkage ed è stato sequenziato l'intero genoma di tre soggetti.

Risultati. L'utilizzo dei pannelli di geni target si è rivelato una buona strategia per lo screening di mutazioni in pazienti affetti da ACM: almeno una mutazione è stata trovata in 15 probandi su un totale di 19 analizzati con il pannello di geni associati a diverse cardiomiopatie; inoltre, è stata identificata una mutazione in uno dei geni candidati nel 40.6% dei probandi sequenziati con il secondo pannello e risultati negativi per mutazioni nei geni noti. Tra queste mutazioni, risultano di particolare interesse 2 nuove mutazioni missenso localizzate nel gene *TJP1* (p.R265W,

p.Y669C), 2 mutazioni nel gene che codifica per l'N-caderina (*CDH2*, p.E493G, p.V491G) e una mutazione di stop nel gene *TP63* (p.R266*). Entrambe le mutazioni di *TJP1* riguardano aminoacidi altamente conservati, inoltre un'analisi *in silico* degli effetti della mutazione p.Y669C, che segrega all'interno della famiglia, evidenzia un riarrangiamento della struttura proteica della proteina che riporta la mutazione rispetto alla proteina wild-type.

L'analisi dell'esoma nella Famiglia#3 ha permesso di identificare una mutazione patogena di stop nel gene *DSP*, che non era stato possibile individuare al precedente screening tramite dHPLC. Inoltre, tale approccio ha permesso di identificare due putative mutazioni nel gene *TTN* (p.R32573C and p.L32198M) che segregano nelle Famiglie #4 e #5, rispettivamente. Nella Famiglia #5 inoltre è stata identificata una rara mutazione di stop in un gene candidato (*CMYA5*, p.K3597*). Per lo studio della Famiglia#6, invece, sono stati utilizzati diversi approcci. Data la disponibilità di molti soggetti affetti e sani appartenenti alla famiglia, in seguito alla genotipizzazione degli stessi, è stata eseguita un'analisi di linkage. L'analisi ha evidenziato la presenza di due loci che riportano valori di lod score positivi a livello del cromosoma 19p13.3 e del cromosoma 11q21. Il primo locus corrisponde ad una regione di 7 cM ed è condiviso da tutti i soggetti affetti della famiglia, eccetto uno. La regione di 3cM nel cromosoma 11q21 invece segrega in tutti i soggetti affetti ad eccezione di due, che portano una mutazione in un sito di splicing del gene *PKP2* (c.2578-3 T>C). Il sequenziamento dell'esoma in 4 soggetti affetti della famiglia non ha permesso di identificare nuove mutazioni condivise né all'interno delle due regioni critiche, né all'interno dell'intero esoma. Tali risultati sono stati confermati dal sequenziamento del genoma effettuato in due affetti e un soggetto sano della famiglia. Il sequenziamento del genoma ha inoltre messo in luce la presenza di un'enorme quantità di varianti complesse localizzate in regioni introniche o intrageniche.

Discussione. L'identificazione di mutazioni causative nella cardiomiopatia aritmogena facilita la diagnosi tempestiva, permette di prevenire eventuali complicazioni e determina il rischio di sviluppare la malattia nei familiari di un soggetto affetto. Per le cardiomiopatie ereditarie, come per la maggior parte delle malattie mendeliane, negli ultimi 10 anni le tecniche NGS hanno apportato grossi miglioramenti nel processo di identificazione di mutazioni, permettendo di sequenziare una grande quantità di geni parallelamente, in modo più rapido e meno costoso. In questo studio, dal momento che sono state identificate mutazioni nella maggior parte di probandi analizzati, l'utilizzo di pannelli di geni target si è dimostrato un valido approccio non solo per un test genetico ma anche per l'identificazione di nuovi geni malattia. Sono state infatti identificate nuove mutazioni in geni che codificano per proteine dei dischi intercalari dei cardiomiociti,

confermando l'idea che l'ACM deve essere considerata una 'malattia delle giunzioni' piuttosto che una 'malattia dei desmosomi'.
Nonostante l'utilizzo di pannelli di geni target rimanga l'approccio più comunemente usato nella ricerca di mutazioni nelle cardiomiopatie ereditarie, il sequenziamento dell'intero esoma, applicato in questo studio solo a casi familiari, ha permesso di identificare varianti in geni candidati non inclusi nei pannelli di geni e che potrebbero avere un ruolo nell'espressione del fenotipo patologico. Infine, nonostante il sequenziamento del genoma nella Famiglia#6 non abbia permesso al momento di stabilire una correlazione tra genotipo e fenotipo al momento, la mole di dati prodotti potrà essere rianalizzata in futuro alla luce di nuovi geni annotati e nuovi geni malattia identificati.

1. INTRODUCTION

Cardiomyopathies are a heterogeneous group of myocardial disorders characterized by structural and functional abnormalities of the heart muscle, in the absence of coronary artery disease, hypertension, valvular disease and congenital heart disease sufficient to cause these anomalies (European Society of Cardiology, Maron et al., 2006). Cardiomyopathies were defined as primary if the heart is the sole or predominantly involved organ and secondary if those diseases in which myocardial dysfunction is part of a systemic disorder (Elliot et al., 2008). Most of the forms recognized as primary cardiomiopathies have a genetic origin and a familial presentation. Beyond channelopathies, the main described phenotypes are hypertrophic, restrictive, dilated and arrhythmogenic cardiomyopathies and isolated left ventricular non-compaction.

Arrhythmogenic cardiomyopathy (ACM) is a heritable form of cardiomyopathy characterized by fibrofatty replacement which leads to right ventricular failure, arrhythmias, and sudden cardiac death (SCD) (Marcus et al., 1982, Nava et al., 1988; Thiene et al., 1988).

1.1 History and designation of the disease

The first historical description of ACM date three centuries ago (1736), when Giovanni Maria Lancisi reported the recurrence of the disease in four generations of a family in the book entitled "*De Motu Cordis et Aneurysmatibus".* However, the first comprehensive clinical description of the disease was reported in 1982 by Marcus and colleagues. They described 24 adults cases showing ventricular tachycardia of left bundle branch block (LBBB) morphology (Marcus et al., 1982). Two years later, the electrocardiographic features of the disease, including the characteristic epsilon wave were described (Fontaine et al., 1984). In 1988, Nava and colleagues reported that ACM is manifested at a young age with a polymorphic clinical and arrhythmic profile and the evidence of an autosomal dominant transmission pattern (Nava et al., 1988). Moreover, in the same year, Thiene and colleagues described several cases of sudden death in young people (≤35 years), who showed typical electrical ACM features, like inverted T waves in the right precordial leads and ventricular arrhythmias with LBBB, together with fibrofatty replacement in the right ventricular free wall (Thiene et al., 1988). Since then, ACM is consider one of the most important causes of sudden death among young people and athletes.

In the beginning the term related to the disease was 'arrhythmogenic right ventricular dysplasia (ARVD)' for the histologic features of the myocardium and the

belief that ACM is a congenital heart defect. The name was then replaced with 'arrhythmogenic right ventricular cardiomyopathy (ARVC)' due to acquired nature of myocardial atrophy, occurring after birth as a consequence of cumulative myocyte loss. Moreover, since the left ventricle of the myocardium seems to be involved in more than half of the cases (Corrado et al., 1997) and may precede the onset of significant right ventricular dysfunction, the adoption of the broader term 'arrhythmogenic cardiomyopathy (ACM)' is supported (Sen-Chowdhry et al., 2007).

1.2 Epidemiology

The estimated prevalence of ACM in the general population ranges from 1 in 2000 to 1 in 5000 and it accounts for approximately 30% of cases of sudden cardiac death (SCD) in young and athletes (Nava et al., 1988). At least 50% are familial cases and it is typically transmitted with and autosomal dominant pattern of inheritance characterized by incomplete penetrance and variable expressivity. For these different manifestations, it should be consider an underestimation of the disease prevalence. Moreover, ACM affects men more frequent than woman, with an approximate ratio of 3:1, but it does not appear a priori to harbor adverse effects on long-term survival. Even if the cause remains unclear, important factors such as sexual hormones and physical activity could play a role (Bauce et al., 2008).

1.3 Clinical manifestations

ACM usually arises from the second to the forth decade of life (Nava et al., 1988; Daliento et al., 1995), first symptoms during early childhood or beyond the age of 60 years are in fact unusual.

Traditionally the course of the disease is divided in four main phases based on the long-term follow-up of clinical data (Thiene et al., 1990): (i) an early subclinical phase in which functional and structural abnormalities are rare or absent and sudden cardiac death can be the first manifestation of the disease; (ii) an overt 'electrical' phase with right ventricular arrhythmias and both functional and structural abnormalities. This second phase is characterized by palpitations, syncope, and symptomatic ventricular arrhythmias of left bundle branch block (LBBB) morphology.(iii) A right ventricle (RV) dysfunctional phase with an extensive ad severe involvement of the RV in the absence of left ventricle (LV) dysfunction; (iv) an end-stage phase characterized by a biventricular failure mimicking dilated cardiomyopathy resulting in a more difficult differential diagnosis and sometimes requiring cardiac transplantation. By this stage, patients could

have endocardial mural thrombosis, and the risk of pulmonary and systemic thromboembolism become higher.

In the last twenty years several studies demonstrated also the early involvement of the LV changing the conventional timing of the disease (Pinamonti et al., 1995; Lobo et al., 1999; Sen-Chowdhry et al., 2007) and proposing a general classification in 3 categories based on the patterns of the disease expression: classical pattern associated with right ventricular phenotype, left ventricular and biventricular forms. The three patterns are independent and coexist in the ACM population even within the same family.

Although sudden cardiac death can be the first and final manifestation of ACM, the most common clinical presentation of ACM is symptomatic ventricular tachycardia (VT) with LBBB. Ventricular arrhythmias span from premature ventricular complexes to sustained VT or ventricular fibrillation (VF) leading to cardiac arrest (Basso et al., 2009). Associated symptoms range from palpitations and paroxysmal tachycardia to dizziness, syncope, and sudden cardiac arrest. Results of exercise stress testing and Holter monitoring revealed frequent premature ventricular beats, as well as effort-dependent ventricular couplets and unsustained VT (Corrado et al., 1990).

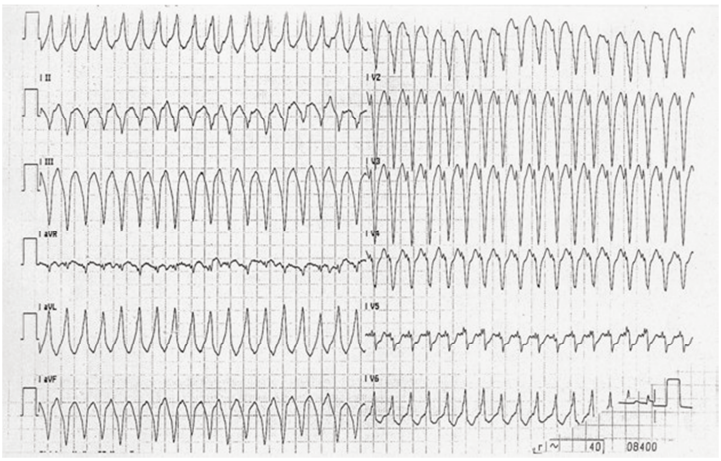


Figure 1.1. 12 lead ECG recording of VT with left bundle branch block (LBBB) morphology. (Thiene et al., 2007).

Furthermore, the presence of the disease is strongly suggested by distinctive ECG findings of inverted T waves in right precordial leads (V1 to V4) which reflect an inferior or superior QRS axis, depending on the origin of the arrhythmias (Figure 1.1). Negative T waves in V1-V3 are observed in the majority of patients with some differences due to the RV dysfunction and the heterogeneous nature of ACM (Nasir et al., 2004; Dalal et al., 2005). Abnormalities of the QRS complex in the right precordial leads, which consist in the QRS duration >100 ms and

differences in QRS duration between right and left precordial leads >25 are typical ECG findings in ACM patients. Moreover, epsilon waves, described in ECG from patients (Fontaine et al., 1984), may represent delayed activation of the surviving myocytes interspersed with fatty and fibrous tissue in the RV. The same phenomenon is seen with the presence of late potentials in the signal average ECG (SAECG) (Figure 1.2).

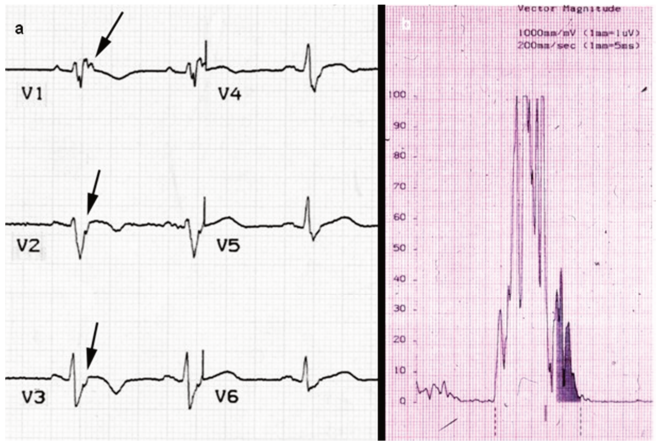


Figure 1.2. ECG recording: (a) post-excitation epsilon waves (arrows) in right precordial leads; (b) positive late potentials at signal-averaged electrocardiography (SAECG) (Thiene et al., 2007).

The absence of a gold standard diagnostic test leads to the use of different non-invasive techniques to determine ACM diagnosis, such as echocardiography, radionucleotide angiography and magnetic resonance imaging (MRI). Echocardiography is a good tool to assess RV size and function both in index cases and family members, because its availability, easy of performance and interpretation. Morphologic abnormalities consisting of wall motion abnormalities, trabecular derangement, and sacculations, as well as global RV dysfunction, can be detected by using this test (Yoerger et al., 2005). Among the imaging methods, MRI allows both qualitative and quantitative analysis of RV function with a high sensitivity and specificity (Tandri et al., 2008). MRI is able to detect intramyocardial fat infiltrations and RV dysfunction also in early and concealed cases of ACM in which other techniques can fail (Figure 1.3). Moreover, MRI with gadolinium enhancement can detect fibrosis in both RV and LV myocardium, a finding that can precede functional abnormalities (Tandri et al., 2005).

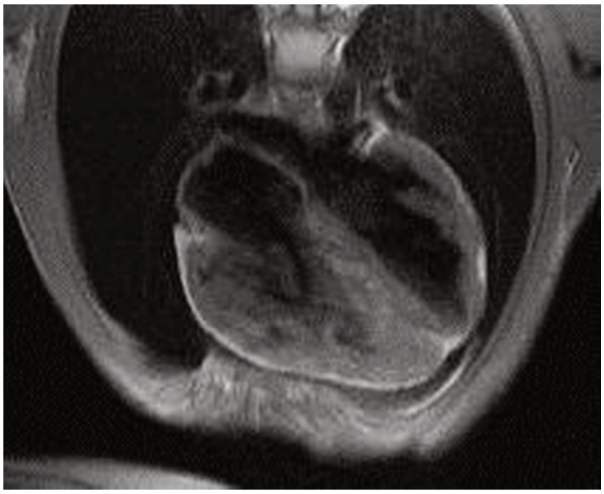


Figure 1.3. ACM fatty variant on cardiac MRI. Note the typical biventricular fatty replacement (*bright white* signal intensity) of the entire RV wall and of the LV lateral wall. From Perazzolo Marra et al., 2015.

Despite its invasiveness, angiography remains a powerful method to diagnose ACM. On angiography, RV end diastolic volume is increased and ejection fraction is reduced with widespread RV impairment (Nava et al., 1988). Moreover, three-dimensional electroanatomical mapping can reveal low voltage areas corresponding to fibrofatty myocardial replacement, thus differentiating ACM from diseases that mimic ACM such as inflammatory cardiomyopathy and idiopathic right ventricular outflow tract tachycardia (Corrado et al., 2005).

1.4 Pathological profile

The typical pathological profile consists of the progressive replacement of the myocardium by fatty or fibro-fatty tissue, starting from epicardium or midmyocardium and then extending to become transmural. This infiltration may be diffuse or regional; typically located at the angles of the so-called “triangle of dysplasia”, including the inferior, apical, and infundibular walls (Figure 1.4).

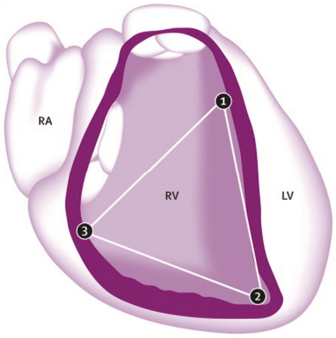


Figure 1.4. “Triangle of dysplasia”. RA=right atrium. RV=right ventricle. LV=left ventricle. Modified from Basso et al., 2009.

Only in the more severe forms of ARVC all the three areas are involved, while usually only one or two regions are affected in the minor cases (Thiene et al., 2007). The ventricular septum seems to be involved in about 20% of the cases and LV in nearly half (Basso and Thiene, 2007). Parietal thinning, as well as endocardial thickening, is found in areas of transmural infiltration and aneurismal dilation providing a substrate for life-threatening arrhythmias (Romero et al., 2013). The fibro-fatty infiltration in fact interferes with electrical impulse conduction accounting for delay (late potentials, epsilon wave, parietal bundle branch block) and onset of re-entrant phenomena.

Based on the prevalent type of tissue accompanying the myocytes loss, two histopathological variants have been defined, namely fatty or fibro-fatty ACM (Basso et al., 1996). The latter is characterized by severe atrophy and scarring of the myocardium, RV wall thickness may be reduced, and the wall becomes translucent and parchment-like, whereas when fatty replacement occurs, it is mainly localized in the anterior wall and infundibulum (Thiene et al., 1997). However, in these cases, a small amount of fibrous tissue is also present. Myocyte death, as well as myocardial substitution with fatty or fibro-fitty tissue are common features of both variants, indicating an injury and repair process (Basso and Thiene, 2007) (Figure 1.5).

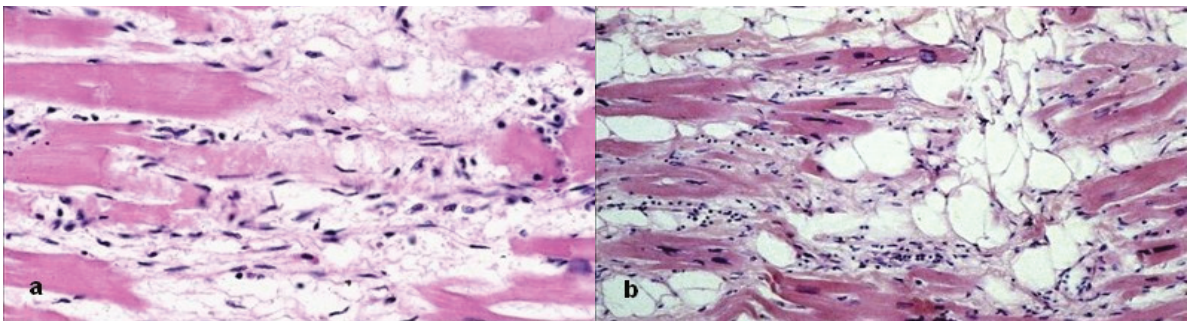


Figure 1.5. Ongoing myocyte death (a) and early fibro-fatty replacement infiltration (b). From Thiene et al., 2007.

Some particular features appear to be related with a specific fatty or fibro-fatty pattern. Fibrofatty infiltration occurs during adolescence, and not after birth as in Uhl's anomaly, a congenital heart defect in which the right ventricular myocardium does not develop during embryonic life (Uhl 1952). In the classical form of the disease the RV is earlier affected but the LV is involved in more than half of the cases (Figure 1.6).

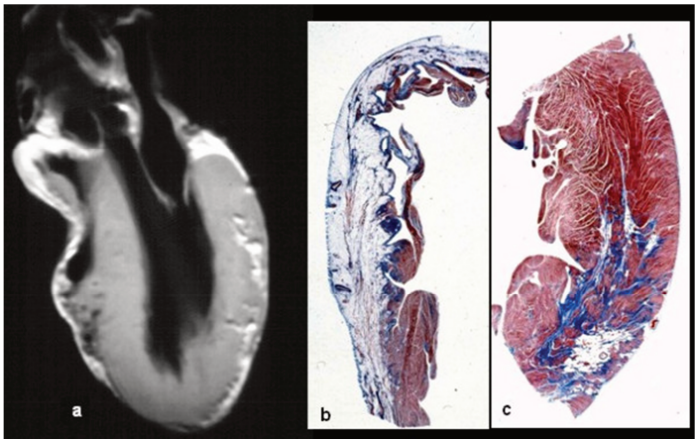


Figure 1.6. Note the biventricular involvement at MRI (a), with transmural fibro-fatty replacement in the RV free wall (b) and focal subepicardial in the LV free wall (c). From Thiene et al., 2007.

Usually in these cases the infiltration is confined to the posterolateral subepicardium and as it is observed in older patients with heavier hearts, it might occur at a later time. It is also well-known that fatty tissue is present in a large amount of normal heart, increasing with age and bodyweight and mainly localized in the subepicardial region. Subsequently the fatty infiltration alone in the right ventricle is not a sufficient hallmark of ACM, presence of fibro-fatty replacement and myocyte degenerative changes are essential to provide a clear-cut diagnosis (Basso and Thiene, 2005).
Furthermore, fibro-fatty pattern shows a high incidence of RV aneurysms and focal myocarditis, consisting of patchy inflammatory infiltrates of lymphocytes associated with myocardial necrosis (Thiene et al., 1991; Basso et al., 1996). Whether the inflammatory response is a reaction to myocyte death or the consequence of infective or immune mechanisms remains not clear. Cardiotropic viruses have been detected in the myocardium of some ARVC patients indicating a possible rule in the disease pathogenesis (Calabrese et al., 2006).

1.5 Diagnosis and Task Force Criteria

In the early stages of ACM the diagnosis can also be difficult because of the non-specific nature of the disease and the broad spectrum of the phenotypic expressions. Dilated cardiomyopathy, Brugada syndrome and myocarditis can mimic ACM and differential diagnosis can be very difficult in absence of classic ACM hallmarks, such as RV aneurysms.
As previously mentioned, there is no gold standard diagnostic test for ACM, and the diagnosis relies on the combination of structural, functional, and electrocardiographic abnormalities caused by the underlying histological changes (McKenna et al., 1994). In 1994, the first International Task Force Criteria were

introduced as common guidelines and abnormalities were divided in major and minor criteria within six categories. A modification of these original criteria was introduced by Marcus and colleagues in 2010, in order to increase the diagnostic sensitivity including quantitative parameters particularly for the imaging studies (Marcus et al., 2010) (Table 1.1).

Original task force criteria	Revised task force criteria
I. Global or regional dysfunction and structural alterations*	
Major	
	By 2D echo:
• Severe dilatation and reduction of RV ejection fraction with no (or only mild) LV impairment • Localized RV aneurysms (akinetic or dyskinetic areas with diastolic bulging) • Severe segmental dilatation of the RV	• Regional RV akinesia, dyskinesia, or aneurysm • *and* 1 of the following (end diastole): — PLAX RVOT ≥32 mm (corrected for body size [PLAX/BSA] ≥19 mm/m^2) — PSAX RVOT ≥36 mm (corrected for body size [PSAX/BSA] ≥21 mm/m^2) — *or* fractional area change ≤33%
	By MRI:
	• Regional RV akinesia or dyskinesia or dyssynchronous RV contraction • *and* 1 of the following: — Ratio of RV end-diastolic volume to BSA ≥110 mL/m^2 (male) or ≥100 mL/m^2 (female) — *or* RV ejection fraction ≤40%
	By RV angiography:
	• Regional RV akinesia, dyskinesia, or aneurysm
Minor	
	By 2D echo:
• Mild global RV dilatation and/or ejection fraction reduction with normal LV • Mild segmental dilatation of the RV • Regional RV hypokinesia	• Regional RV akinesia or dyskinesia • *and* 1 of the following (end diastole): — PLAX RVOT ≥29 to <32 mm (corrected for body size [PLAX/BSA] ≥16 to <19 mm/m^2) — PSAX RVOT ≥32 to <36 mm (corrected for body size [PSAX/BSA] ≥18 to <21 mm/m^2) — *or* fractional area change >33% to ≤40%
	By MRI:
	• Regional RV akinesia or dyskinesia or dyssynchronous RV contraction • *and* 1 of the following: — Ratio of RV end-diastolic volume to BSA ≥100 to <110 mL/m^2 (male) or ≥90 to <100 mL/m^2 (female) — *or* RV ejection fraction >40% to ≤45%
II. Tissue characterization of wall	
Major	
• Fibrofatty replacement of myocardium on endomyocardial biopsy	• Residual myocytes <60% by morphometric analysis (or <50% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy
Minor	• Residual myocytes 60% to 75% by morphometric analysis (or 50% to 65% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy
III. Repolarization abnormalities	
Major	• Inverted T waves in right precordial leads (V_1, V_2, and V_3) or beyond in individuals >14 years of age (in the absence of complete right bundle-branch block QRS ≥120 ms)
Minor	
• Inverted T waves in right precordial leads (V_2 and V_3) (people age >12 years, in absence of right bundle-branch block)	• Inverted T waves in leads V_1 and V_2 in individuals >14 years of age (in the absence of complete right bundle-branch block) or in V_4, V_5, or V_6 • Inverted T waves in leads V_1, V_2, V_3, and V_4 in individuals >14 years of age in the presence of complete right bundle-branch block

continued

Original task force criteria	Revised task force criteria
IV. Depolarization/conduction abnormalities	
Major	
• Epsilon waves or localized prolongation (>110 ms) of the QRS complex in right precordial leads (V_1 to V_3)	• Epsilon wave (reproducible low-amplitude signals between end of QRS complex to onset of the T wave) in the right precordial leads (V_1 to V_3)
Minor	
• Late potentials (SAECG)	• Late potentials by SAECG in ≥1 of 3 parameters in the absence of a QRS duration of ≥110 ms on the standard ECG • Filtered QRS duration (fQRS) ≥114 ms • Duration of terminal QRS <40 μV (low-amplitude signal duration) ≥38 ms • Root-mean-square voltage of terminal 40 ms ≤20 μV • Terminal activation duration of QRS ≥55 ms measured from the nadir of the S wave to the end of the QRS, including R′, in V_1, V_2, or V_3, in the absence of complete right bundle-branch block
V. Arrhythmias	
Major	
	• Nonsustained or sustained ventricular tachycardia of left bundle-branch morphology with superior axis (negative or indeterminate QRS in leads II, III, and aVF and positive in lead aVL)
Minor	
• Left bundle-branch block-type ventricular tachycardia (sustained and nonsustained) (ECG, Holter, exercise) • Frequent ventricular extrasystoles (>1000 per 24 hours) (Holter)	• Nonsustained or sustained ventricular tachycardia of RV outflow configuration, left bundle-branch block morphology with inferior axis (positive QRS in leads II, III, and aVF and negative in lead aVL) or of unknown axis • > 500 ventricular extrasystoles per 24 hours (Holter)
VI. Family history	
Major	
• Familial disease confirmed at necropsy or surgery	• ARVC/D confirmed in a first-degree relative who meets current Task Force criteria • ARVC/D confirmed pathologically at autopsy or surgery in a first-degree relative • Identification of a pathogenic mutation† categorized as associated or probably associated with ARVC/D in the patient under evaluation
Minor	
• Family history of premature sudden death (<35 years of age) due to suspected ARVC/D • Familial history (clinical diagnosis based on present criteria)	• History of ARVC/D in a first-degree relative in whom it is not possible or practical to determine whether the family member meets current Task Force criteria • Premature sudden death (<35 years of age) due to suspected ARVC/D in a first-degree relative • ARVC/D confirmed pathologically or by current Task Force Criteria in second-degree relative

Table 1.1. International Task Force ACM diagnostic criteria. Original criteria are listed in the left column, the revised criteria in the right column. Modified from Marcus et al., 2010.

In a first study of Cox and co-workers applying these new criteria to patients suspected of ACM, 64% of probable ACM patients and 11% of family members were additionally diagnosed, suggesting that revised Task Force Criteria could have a major impact in increasing diagnostic yield (Cox et al., 2010). The diagnosis of ACM is achieved in the presence of two major or one major plus two minor criteria or four minor criteria taken from different groups (Table 1.1). Moreover, a borderline condition is determined by the presence of one major and one minor, or three minor criteria from different groups, while a possible condition is held in the presence of one major or two minor criteria from different categories. The main application of these criteria remains in establishing a diagnosis of ACM in probands. Modifications of the Task Force guidelines have been proposed to

allow detection of ACM in first-degree relatives with mild or incomplete expression of the disease, during familial screening of index cases with confirmed ACM (Hamid et al., 2002). In these familial cases, diagnosis is made considering electrical, functional and structural aspects (Table 1.2).

ARVC in first-degree relatives plus one of the following:	
I. ECG	T-wave inversion in right precordial leads (V2 and V3)
II. SAECG	Late potentials seen on signal-averaged ECG
III. Arrhythmia	LBBB type VT on ECG, Holter monitoring or during exercise testing
	Extrasystoles >200 in 24 hours
IV. Structural or functional abnormality of the RV	Mild global RV dilatation and/or ejection fraction reduction with normal LV
	Mild segmental dilatation of the RV
	Regional RV hypokinesia

Table 1.2 Proposed modification of Task Force Criteria for the diagnosis of familial ACM. From Hamid et al., 2002.

1.6 Risk stratification and management

Prevention of sudden death is the most important management strategy of ACM. Cardiac arrest, syncope, young age, malignant family history, participation in competitive sports, VT, severe right ventricular dysfunction, left ventricular involvement are considered as risk factors for ACM, even though the prognostic value of these single or combined risk factors has not been prospectively assessed. Turrini and colleagues in 2001 reported that QRS dispersion (≥40 ms) was the strongest independent predictor of sudden death in ACM and that risk stratification in these patients was refined by syncope, negative T-waves beyond V1 and QT dispersion >65 ms (Turrini et al., 2001). In a study of Hulot et al. a risk stratification scheme has been developed. Patients with clinical signs of right ventricular failure, left ventricular dysfunction and/or history of ventricular tachycardia are at high risk of cardiac events, while subjects without VT are at very low risk (Hulot et al., 2004). In case of high risk, patients should undergo a pharmacological therapy, beyond avoid competitive sport activities. Anti-arrhythmic therapy, consisting of beta-blockers or class III antiarrhythmic drugs (sotalol and amiodarone) can be useful for controlling arrhythmia but they are not an equivalent alternative to implantable cardioverter defibrillator (ICD) therapy in patients with high risk of sudden death. Accepted indications for ICD implantation include secondary prevention after prior cardiac arrest and VT with hemodynamic compromise (Azaouagh et al., 2011). ICD implantation is imperative if an aborted sudden death is occurred but also indicated in case of sustained VT and/or syncope in presence of risk factors. It provides a life-saving protection by effectively terminating life-threatening ventricular arrhythmias with an estimated mortality reduction of 24–35% at 3-years follow-up (Corrado et al., 2003; Wichter

et al., 2004). Catheter ablation is indicated in subjects with ACM that show monomorphic and well-tolerated VT with localized forms of the disease and drug-refractory or incessant VT or frequent ICD discharges. According to Zou and collegues catheter ablation would not be a long-term curative procedure (Zou et al., 2004). At last, only in the most severe cases of ACM with malignant intractable recurrent arrhythmias or refractory congestive heart failure, cardiac transplantation should be considered (Lacroix et al., 2005).

1.7 Genetic basis of ACM

ACM is a heritable cardiomyopathy characterized by incomplete penetrance and variable expressivity and usually transmitted as an autosomal dominant trait, even though recessive forms have been reported (Nava et al., 1988; McKoy et al., 2000; Norgett et al., 2000). Since 1994, when the first locus was described, 15 independent loci and 13 disease genes have been identified (Table 1.3). One or more disease-causing mutations are detected in nearly half of the patients and most of them are located in genes encoding for desmosomal proteins at the intercalated discs.

locus	gene	protein	mutation prevalence	mode of inheritance
17q21	*JUP*	plakoglobin	rare	AR/AD
1q42-q43	*RYR2*	ryanodine receptor 2	rare	AD
6p24	*DSP*	desmoplakin	1% - 16%	AD
12p11	*PKP2*	plakophilin-2	7% - 70%	AD
14q23-q24	*TGFB3*	transforming growth factor β3	rare	AD
18q12	*DSG2*	desmoglein-2	5% - 25%	AD
18q12	*DSC2*	desmocollin-2	rare	AD/AR
3p25	*TMEM43*	transmembrane protein 43 (LUMA)	rare	AD
2q35	*DES*	desmin	rare	AD
2q31	*TTN*	titin	rare	AD
1q21.2-q21.3	*LMNA*	lamin A/C	rare	AD
6q22.1	*PLN*	phospholamban	rare	AD
10q21	*CTNNA3*	αT-catenin	rare	AD

Table 1.3. ACM known-genes. AD= autosomal dominant; AR= autosomal recessive.

Desmosomes are important intercellular junctions that maintain mechanical integrity of heart tissue providing strong adhesion between cells. Three separate families of protein assemble to form the desmosomes: cadherins (desmoglein and desmocollin), armadillo proteins (plakoglobin and plakophilin), and plakins (desmoplakin). The genes encoding for these five proteins represent the most mutated genes in ACM.

1.7.1 Junctional components

JUP

The first ACM gene (*JUP*) was identified in a rare cardiocutaneous syndrome, characterized by arrhythmogenic right ventricular cardiomyopathy associated with palmoplantar keratoderma and peculiar woolly hair and showing an autosomal recessive pattern of inheritance (McKoy et al., 2000). This syndrome was called "Naxos syndrome", because of its prevalence exceeds 1 in 1000 in the Greek island of Naxos. Despite the cutaneous phenotype appears early, cardiac abnormalities present later, during adolescence and young adulthood. In 2007, Asimaki and colleagues identified a dominant mutation in plakoglobin gene in a German family with ACM but no cutaneous abnormalities. The mutation was found to affect plakoglobin stability at the junctions (Asimaki et al., 2007).

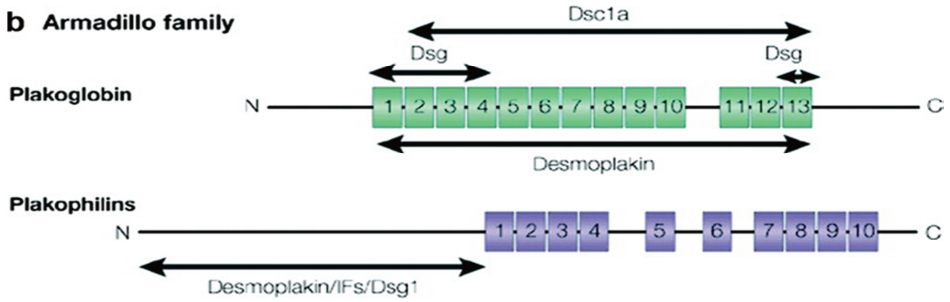


Figure 1.7. Desmosomal armadillo family members: plakoglobin and plakophilins. Structures and interaction with other desmosomal proteins. Modified from Delmar and McKenna, 2010.

Plakoglobin (γ-catenin) together with plakophilin, belongs to the armadillo proteins, which are characterised by the presence of a central domain containing a variable number of imperfect 42 aa repeats. Plakoglobin has 12 *arm* repeats, in which both desmoglein and desmocollin proteins interact with distinct sites through their intracellular domains (Figure 1.7). Plakoglobin also interacts with the N-terminal plakin domain of desmoplakin which complete the link with desmin intermediate filaments (IF) through their C-termini (Garrod and Chidgey, 2008). The desmosome-IF complex is a network that maintains the integrity of tissues and gives mechanical strength to them. Moreover, it seems that C-terminal truncations

of plakoglobin alter desmosome morphology in cultured cells, while deletion of the N-terminus do not have a dramatic effect on the structure of desmosomes in these cells. This suggests that the presence of the C-terminus of plakoglobin limits the size of desmosomes, perhaps through regulating protein-protein interactions required for assembly of the desmosomal plaque (Palka and Green, 1997). In addition to its structural function and due to its high homology with β-catenin, plakoglobin has a role also in intracellular signal transduction. It is now well known that plakoglobin interacts and compete with β-catenin and plays a unique role in Wnt signaling pathway, an ancient and evolutionarily conserved pathway that regulates crucial aspects of cell fate determination, migration, proliferation and apoptosis (Zhurinsky et al., 2000).

The two catenins interact at multiple cellular levels with a net negative effect on the canonical Wnt/β-catenin signaling pathway through T cell/lymphoid-enhancing binding (Tcf/Lef) transcription factors (Ben Ze'ev et al., 1998; Klymkowsky et al., 1999). Garcia-Gras and coworkers speculate that when plakoglobin is free from the desmosomes, it translocates into the nucleus and through competition with β-catenin suppresses signaling through the canonical Wnt/β-catenin–Tcf/Lef pathway (Garcia-Gras et al., 2006). This is the case of ACM patients, when mutations in desmosomal genes compromise the integrity of these complexes.

DSP

Desmoplakin gene (*DSP*) was quickly associated with the classical autosomal dominant form of the disease (Rampazzo et al., 2002) even though recessive mutations in *DSP* have been earlier described in Carvajal syndrome, which consists of palmoplantar keratoderma, woolly hair, and biventricular dilated cardiomyopathy (Norgett et al., 2000). The first described autosomal dominant mutation is thought to disrupt a protein kinase C (PKC) phosphorylation site which is involved in plakoglobin binding and in clustering of desmosomal cadherin-plakoglobin complexes. Other *DSP* mutations, both recessive and dominant, as well as different types of mutation in different domains of the protein and associated or not with cutaneous phenotype have been identified (Alcalai et al., 2003; Jonkman et al., 2005; Whittock et al., 1999).

Desmoplakin is an important constituent of the desmosomal plaque. It interacts with desmin intermediate filaments in cardiomyocytes, anchoring them to the plasma membrane and forming an essential scaffold for tissue integrity. Desmoplakin's primary structure is divided in three parts: globular head or plakin domains, tail domains and a coiled-coil rod region. Alternative splicing of the gene generates two isoforms of desmoplakin (DPI and DPII), which differ only in the

length of the central rod domain. DPI is the only one expressed in cardiomyocytes (Angst et al., 1990) (Figure 1.8).

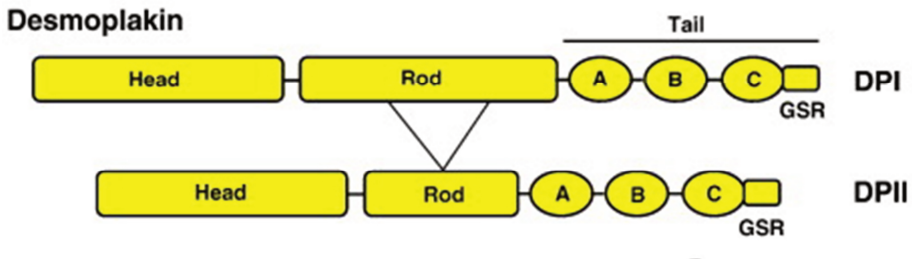


Figure 1.8. Structure of desmoplakin. The two desmoplakin isoforms are shown. A, B and C are plakin repeat domains. Modified from Garrod and Chidgey, 2008.

PKP2

In 2004, Gerull and collegues screened a cohort of 120 ACM patients for all the 14 exons of *PKP2* gene and identified 25 different heterozygous mutations in 32 of them (Gerull et al., 2004). *PKP2* gene encodes plakophilin-2, an essential protein of the cardiac desmosome which belong to the armadillo protein family (Figure 1.7). Three plakophilins are known (PKP1-3) and show complex tissue specific patterns of expression. Plakophilin-2 is the only one isoform expressed in cardiomyocytes (Hatzfeld, 2006). Plakophilins are located in the outer dense plaque of desmosomes linking desmosomal cadherins with desmoplakin and the intermediate filament system, but it is also found in the nucleus, where they may have a role in transcriptional regulation (Mertens et al., 2001). Gerull et al. speculate that lack of plakophilin-2 or the presence of mutant plakophilin-2 into cardiac desmosomes impairs cell-cell contacts and, as a consequence, might disrupt adjacent cardiomyocytes, especially in response to mechanical stress or stretch (Gerull et al., 2004). In addition, a possible indirect role in Wnt signaling has been proposed following the evidence that PKP2 interacts with β-catenin through the N-terminal domain and expression of PKP2 in SW480 cells up-regulates β-catenin/Tcf signalling in a β-catenin dependent fashion (Chen et al., 2002).

DSG2

DSG2 together with *DSP* and *PKP2*, is one of the three most mutated genes, the so-called “three big genes”, with a mutation prevalence that span from 5% to 25% in different cohorts (Gerull et al., 2004; Dalal et al., 2006; van Tintelen et al., 2006; Lahtinen et al., 2008; den Haan et al., 2009; Fressart et al., 2010). In 2006, Pilichou and colleagues studied a cohort of 80 unrelated ACM cases, in which, for the first time, they detect heterozygous mutations exactly in 10% of probands

(Pilichou et al., 2006). In an independent study, Awad and collegues described four probands with *DSG2* mutations and, among them one presented compound-heterozygous mutations (Awad et al., 2006).

Desmoglein-2 is one of the four desmogleins found in humans and this form is expressed in all tissues containing desmosomes. Together with desmocollin-2, desmoglein-2 belongs to the cadherin family and, through its cytoplasmic domains, mediates intercellular adhesion at the cardiac desmosomal plaque. Within the cell, both desmocollin and desmoglein have an intracellular cadherin-like sequence (ICS) domain which provide binding sites for other desmosomal constituents such as the armadillo proteins (Garrod and Chidgey, 2008). The cadherins are a superfamily of calcium-dependent adhesion molecules, in fact both DSCs and DSGs extracellular domains consist of four cadherin homology repeats (EC1-4) with a calcium binding motif and a membrane proximal extracellular anchor domain (Dusek et al., 2007) (Figure 1.9).

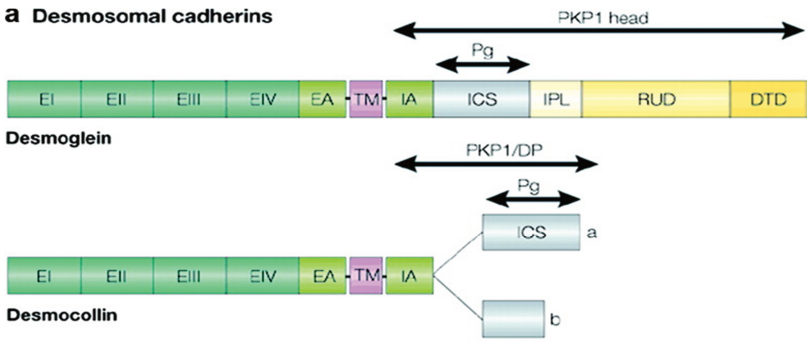


Figure 1.9. Structure of desmosomal cadherins: desmoglein and desmocollin. Modified from Delmar and McKenna, 2010.

All the 4 isoforms of desmoglein and the 3 of desmocollin are encoded by separate genes that are clustered together on opposite sides of a central region on chromosome 18q12.1, suggesting a shared control by a long-range master regulatory region.

DSC2

Belonging to the cadherin family together with *DSG2*, *DSC2* presents similar structure and function at the intercalated discs and shortly became a good candidate gene for ACM. In 2006, Syrris et al. identified 4 mutations in *DSC2*: 2 heterozygous variants, a deletion (1430delC) and an insertion (2687_2688insGA). Both result in frameshift and premature truncation of the DSC2 protein (Syrris et al., 2006). In some individuals with *DSC2* mutations a premature and prominent left ventricular dysfunction was reported suggesting the involvement of both ventricles in the disease manifestation (Sen-Chowdhry et al., 2007). In another

study Beffagna et al. reported that two heterozygous mutations in the N-terminal domain of desmocollin-2 affect the normal cellular localization of the protein (Beffagna et al., 2007). Autosomal recessive ACM forms associated with homozygous *DSC2* mutations were also described in combination with or without cutaneous manifestations (Simpson et al., 2009; Lorenzon et al., 2015). Hair or skin abnormalities have been associated only with homozygous mutations in *DSC2*, probably because in these tissue a gene dosage effect may exceed the ability of other DSC isoforms to compensate (Delmar and McKenna, 2010).

TMEM43

In 2008 Merner and co-workers identified a common region on chromosome 3p (locus ARVD5) in fifteen unrelated ACM families from Newfoundland. They found a missense mutation (S358L) in a highly conserved transmembrane domain of *TMEM43* gene (transmembrane protein-43). They concluded that ACM at locus ARVD5 is a lethal, fully penetrant, sex-influenced morbid disorder (Merner et al., 2008). The same mutation was reported also in a Danish families by Christensen et al. who identified another novel variant of unknown significance (c.705+7G>A) in one patient. In this study immunostaining experiments showed that the signal level of TMEM43 was reduced in cardiomyocytes from all the three *TMEM43*-positive patients in addition to the reduction of plakoglobin (Christensen et al., 2011).

TMEM43 gene encodes for LUMA, which was initially considered to be a tetraspan transmembrane protein of the nuclear envelope. In a recent study, Franke and colleagues discovered that LUMA is a component of *zonula adherens* and *punctum adherens* plaques of diverse epithelia and epithelial cell cultures and is also located in the *area composita* at myocardiac intercalated discs (IDs). These findings allow the general conclusion that LUMA should be considered as an important component of (or associated with) cardiac IDs (Franke et al., 2014).

Interestingly, two different missense mutations in *TMEM43* gene were identified in patients affected by Emery-Dreifuss syndrome characterized by muscular dystrophy, joint contractures, and cardiomyopathy with conduction defects (Liang et al., 2011). This finding indicates that mutations located in different protein domains of transmembrane protein-43 may result in the expression of distinct phenotypes.

CTNNA3

Recently van Hengel and colleagues discovered a novel disease gene, *CTNNA3* gene, located on chromosome 10q21 and considered the fifth largest gene in the human genome (van Hengel et al., 2013). *CTNNA3* gene encodes for αT-catenin,

a member of α-catenin proteins, which plays key functional roles in cadherin-catenin cell-cell adhesion complexes and binds plakophilin-2 in the intercalated discs (Goossens et al., 2007). A heterozygous missense mutation (V94D) and a small deletion (del765L) were identified in two probands, both of them located in fundamental domains of αT-catenin.

In order to characterize the functional effects of the *CTNNA3* mutations, yeast two-hybrid assays and cell transfection studies have been performed. The interaction between the p.V94D mutant protein and β-catenin was affected, whereas the p.del765L mutant protein showed a much stronger dimerization potential compared to wild-type protein (van Hengel et al., 2013). Moreover, in HEK293T cells, endogenous β-catenin was co-immunoprecipitated with exogenous GFP–αT-catenin fusion protein, resulting in a diminished interaction between p.V94D proteins and full-length β-catenin. In contrast to p.V94D mutant protein, the stronger homodimerization potential of the p.del765L mutant protein could lead to a dominant negative or dominant positive effect and the formation of aggresomes in transfected HEK293T cells.

1.7.2 Non-junctional components

RYR2

The first identified ACM gene in a dominant form was cardiac ryanodine receptor-2 involved in ARVD2 (Tiso et al., 2001). *RYR2* gene is one of the largest human genes (105 exons) and all the mutations described involve highly conserved amino acids in three critical domains: N-terminal domain, central and C-terminal domains. In myocardial cells, the RYR2 protein, a homo-tetrameric structure activated by Ca^{2+}, induces the release of calcium from the sarcoplasmic reticulum into the cytosol (Missiaen et al., 2000; Stokes and Wagenknecht, 2000).

Mutations affecting *RYR2* gene are responsible of a form of ACM characterized by a less pronounced fibro-fatty substitution of the myocardial tissue and ventricular polymorphic arrhythmias, induced by effort. *RYR2* mutations might be either gain and loss of function, thus suggesting a heterogeneity in functional consequences. Moreover, loss-of-function mutations have been also associated with catecholaminergic polymorphic ventricular tachycardia (CPVT), characterized by stress-induced, bidirectional ventricular tachycardia that may lead to cardiac arrest and sudden death in the absence of structural heart abnormalities (Priori et al., 2001). Whether CPVT and ARVD2 are different diseases or different degrees of phenotypic expression of the same disease remains unsettled (d'Amati et al., 2005).

TGFβ3

TGFβ3 is a member of the transforming growth factor superfamily, a large group of regulatory proteins playing a pivotal role in development and tissue homeostasis. They inhibit proliferation in many cell types and induce apoptosis in epithelial cells. Two single nucleotide substitutions have been detected in 5'UTR and 3'UTR of *TGFβ3* gene (c.-36G>A, c.1723C>T) in a large ACM family and in one proband respectively (Beffagna et al., 2005). Endomyocardial biopsy in the two probands carrying the UTR-*TGFβ3* mutations revealed extensive replacement-type fibrosis and fatty infiltration, supporting the idea that TGFβ stimulates mesenchymal cells to proliferate and to produce extracellular matrix.

Moreover, it has been demonstrated that TGFβs modulate expression of genes encoding desmosomal proteins in different cell types. Cardiac fibroblasts incubated in the presence or in the absence of exogenous TGFβs show increased expression of different genes, including plakoglobin (Kapoun et al., 2004). In another study, Yoshida et al. reported that TGFβ1 exposure of cultured airway epithelial cells increases the content of both the isoform I and II of desmoplakin (Yoshida et al., 1992). This suggests that TGFβs may play an important role into the regulation of cell–cell junctional complexes and mutations in UTR regions, that cause overexpression of TGFβ3, might affect cellular stability.

DES

Van Tintelen and colleagues in 2009 identified a p.S13F *DES* founder mutation associated with a severe cardiac phenotype with right ventricular predominance (van Tintelen et al., 2009). One year later, Klauke et al. reported a *de novo* missense mutation (N116S) in *DES* gene, in segment 1A of the desmin rod domain. The mutation leads to the aggresome formation in cardiac and skeletal muscle without signs of an overt clinical myopathy (Klauke et al., 2010). Other ACM-related *DES* mutations have been linked to clinically heterogeneous phenotype frequently associated with muscle alterations, the so-called desmin-related myopathy (DRM) (Otten et al., 2010; van Spaendonck-Zwarts et al., 2012).

In the skeletal and cardiac muscle cells, desmin forms a flexible 3-dimensional scaffold around the myofibrillar Z-discs. Desmin represents the cardiac intermediate filaments and connects myofibrils to the plasma membrane, nuclear lamina, mitochondria, and desmosomes. It plays key structural and signaling roles in myocytes and is critical for cytoskeletal organization and maintaining the cardiomyocyte structure (McLendon et al., 2011).

TTN

Titin (*TTN*) is the largest described protein in mammals and constitutes the third most abundant type of filament, after myosin and actinin, both in cardiac and skeletal human muscle (Fürst et al.,1988). TTN is organized into four structurally and functionally distinct regions: the amino-terminal Z-line, the "TTN-anchor"; the I-band, that behaves as "bidirectional springs" regulating the sarcomeric length; an inextensible A-band which binds myosin and myosin-binding protein and is thought to be critical for biomechanical sensing and signaling; and the carboxy-terminal M-line extremity (Figure 1.10).

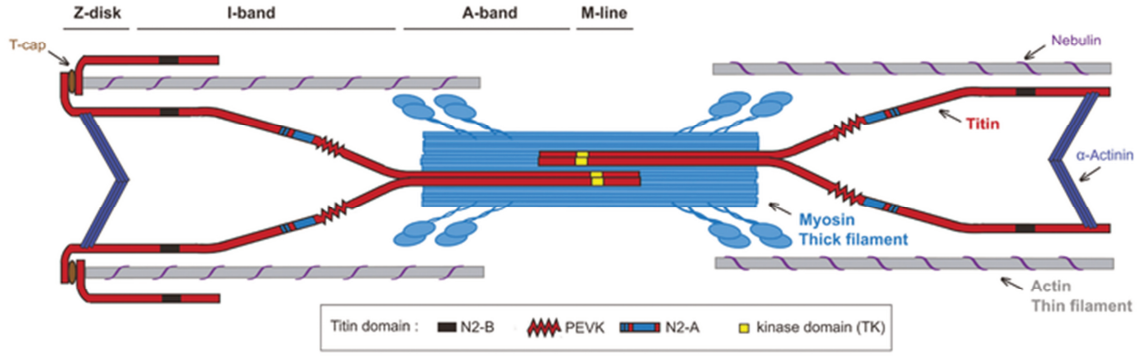


Figure 1.10. *TTN* structure. Schematic representation of 4 *TTN* molecules (red) and their main domains, with their integration in the sarcomere. PEVK= *TTN* region rich in proline (P), glutamate (E), valine (V), and lysine (K). Modified from Chauveau, et al., 2014.

TTN has multiple key roles in all striated muscle cells that span from structural and architectural roles to sensory and regulatory roles (Bennet et al., 2006). *TTN* gene encodes for 364 exons that undergo extensive alternative splicing to produce many isoforms, including two major full-length titin isoforms, N2BA and N2B, expressed both in cardiac and skeletal muscles. The N2BA variant has an additional element in the N2 region (segment N2A), which confers a higher level of elasticity (LeWinter and Granzier, 2010). *TTN* mutations have been associated with different phenotypes: the majority with a purely cardiac phenotype, 35 mutations lead to a purely skeletal muscle phenotype, and nine mutations lead to overlap syndromes (Chauveau et al., 2014). Among the cardiac phenotypes associated with *TTN* mutations, dilated cardiomyopathy is the most widely represented, with 69 mutations identified (Itoh-Satoh et al., 2002; Gerull et al., 2002; 2006). The first description of ACM patients carrying *TTN* mutations was published in a paper including 38 families, in which Taylor and colleagues identified 8 unique *TTN* variants, with a prevalence of the T2896I mutation (Taylor et al., 2011). This missense mutation is absent in all other tested controls and shows a complete segregation within the families, providing a strong genetic evidence that it is linked to the ACM phenotype. A recent study reported that *TTN*

variant carriers display distinct phenotypic characteristics including a greater risk for supraventricular arrhythmias and conduction disease but less severe clinical outcome compared to patients with mutations in desmosomal genes (Brun et al., 2014).

LMNA

Lamin A/C is a nuclear envelope protein located on the nuclear surface of the inner nuclear membrane and it is involved in numerous signaling pathways including adipogenesis.

It has been reported that genetic defects in distinct domains of lamin A/C selectively cause dilated cardiomyopathy with conduction system disease or autosomal dominant Emery–Dreifuss muscular dystrophy (Bonne et al., 1999; Fatkin et al., 1999), other than peripheral neuropathy, lipodystrophies and premature ageing syndromes. More recently, four novel missense variants in *LMNA* gene have been found in patients negative for desmosomal mutations and affected with a severe form of ACM (Quarta et al., 2012). It has been supposed that the effects caused by *LMNA* mutations observed in the laminopathies may arise from varying degrees of impaired nuclear mechanics and transcriptional activation, suggesting an involvement of this protein in cellular mechanotransduction (Lammerding et al., 2004).

PLN

PLN is a Ca^{2+}-ATPase regulator, and together with other sarcoplasmatic Ca^{2+}-cycling proteins is a key regulator of cardiac contractility. *PLN*-R14del mutation was first identified in a large family with hereditary cardiomyopathy by the group of Haghighi, who performed functional studies using a transgenic mouse model overexpressing this mutation (Haghighi et al., 2006). These mice showed depressed cardiac function, histopathological abnormalities (fibrosis), and premature death. More recently, the same specific mutation was identified both in ACM and DCM cases, supporting the concept of 'arrhythmogenic cardiomyopathy' as an entity encompassing ACM, including left-dominant arrhythmogenic cardiomyopathy, and arrhythmogenic forms of DCM (van der Zwaag et al., 2012).

1.8 From a 'desmosomal disease' to a 'disease of the intercalated discs'

Cardiomyocytes are connected each other through complex entities, called intercalated discs (IDs), that provide mechanical and electrochemical coupling between cells. The original description of the IDs in vertebrates identified three separate structures: gap junctions (GJs), adherens junctions (AJs, also called *fascia adherens* in cardiac tissue), and desmosomes (Forbes and Sperelakis,

1985). Gap junctions represent the electric components of IDs, maintaining the electrical and chemical coupling by direct diffusion of small molecules and ions between cardiac myocytes (Saffiz, 2005).

GJ channel is formed by two hemichannels, called connexons, resulting from assembly of six connexin (Cx) molecules that span across the extracellular space, forming a permeable pore (Delmar and Sorgen, 2009). In the myocardium the most abundant connexin isoform is connexin-43.

Conversely, desmosomes and AJ are highly specialized anchoring structures, fundamental for the maintenance of the adhesion and integrity of a tissue such as the myocardium during exposure of mechanical stress. The molecular structure of the two components is similar, both of them are composed of intercellular adhesion molecules that bind in the extracellular space the adhesion proteins extending from the adjacent cells. Intracellularly, there are various adaptor proteins linked to the cytoskeleton or involved in signaling pathways (Rampazzo et al., 2014). For many years AJs and desmosomes have been considered as different entities. Different studies demonstrated that in the mature mammalian IDs the colocalization of different junctional proteins in both the structures lead to the idea that a unique and promiscuous complex exists, the so-called *area composita* (Borrmann et al., 2006; Pieperhoff and Franke, 2008) (Figure 1.11).

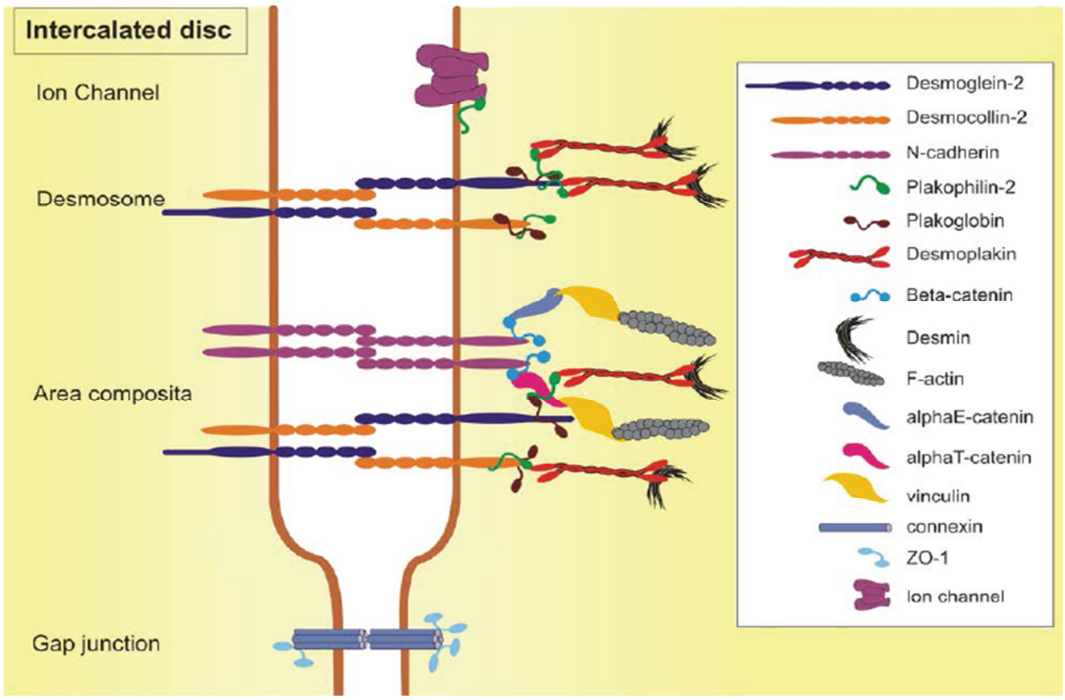


Figure 1.11. Molecular organization of the IDs connecting cardiomyocytes. From Rampazzo et al., 2014.

Franke and colleagues proposed a novel molecular organization of the IDs in the heart, in which desmosomes and AJs proteins participate to the formation of the

area composita, and it develops rather early in embryonic cardiac differentiation (Franke et al., 2006). These findings, together with the recent involvement of mutated αT-catenin, a protein found in the *area composita* (van Hengel et al., 2013), suggest that ACM may be considered a disease of the IDs, rather than a purely desmosomal disease.

1.9 Mutation frequency in ACM genes

Genetic screening of the most mutated genes in different laboratories worldwide allowed to define the mutational spectrum of the disease-genes. As previously mentioned, most of mutations are located in one of the 3 big genes: *PKP2*, *DSP*, and *DSG2*, but respective frequencies can vary a lot with patients' geographic origins. Both the groups of Dalal and van Tintelen reported a frequency of 43% of mutations in *PKP2* gene, with a peak of 70% of the index Dutch patients with affected family members (Dalal et al., 2006; van Tintelen et al., 2006), whereas in other studies these frequencies are largely reduced to around 10% (Pilichou et al., 2006; Sen-Chowdhry et al., 2007; Christensen et al., 2010). The varying mutation prevalence of *PKP2* gene in different cohorts could be explained with a founder effect in certain populations, differences in patient selection, and geographical variations in genetic and nongenetic factors. In any case, the mutational trend is almost conserved and *PKP2* remains the most mutated gene in ACM patients.
The first described ACM gene, *DSP*, shows a mutation rate that span from 1% to 16%, while heterozygous mutations in *DSG2* have been identified in patients with ACM with a frequency range of 5% to 25% in different cohorts (Rampazzo et al., 2014). Overall mutation screening of all the 13 known ACM genes can detect causative mutations in in ≈60% of probands (Groeneweg et al., 2015). Moreover, it is well known that in a significant proportion of cases, concomitant causes such as either a "second hit" in the same gene (compound heterozygosity) or in a different gene (digenic heterozygosity) is required for the overt clinical phenotype and it is often associated with a more severe phenotype in terms of LV and RV alterations (Bauce et al., 2010; Xu et al., 2010, Bhonsale et al., 2015).

1.10 Pathogenetic mechanisms of ACM

Significant progresses have been made in the genetics of ACM, which have led to identification of several causative mutations in junctional and non-junctional genes. Despite these findings, a consensus on the etiopathogenesis of ACM remains elusive, and the molecular processes underlying cardiomyocyte injury, fibro-fatty tissue repair, ventricular remodeling and arrhythmias are not completely clear.

At the beginning an inflammatory and an apoptotic theory (Valente et al., 1998) have been proposed based on histopathological findings. Inflammatory infiltrates are common in ventricular myocardium of ACM patients (Basso et al., 1996) and, they probably play a major part in triggering life-threatening arrhythmias (Thiene et al., 1991; Pinamonti et al., 1996). Therefore, cardiotropic viruses such as enteroviruses, cytomegalovirus, parvovirus, hepatitis C virus, and adenovirus, have been identified in some ACM patients (Bowles et al., 2002; Calabrese et al., 2006). Nevertheless, this theory cannot explain all the ACM cases and the most reliable theory remains the desmosomal dysfunction.

A desmosomal gene mutation may lead to unstable mechanical linkage between cells and disassociation at the intercalated discs which, in turn, can lead to myocyte injury, cell death either by apoptosis or necrosis, tissue remodeling and a clinical picture characterized by contractile dysfunction and cardiomyopathy.

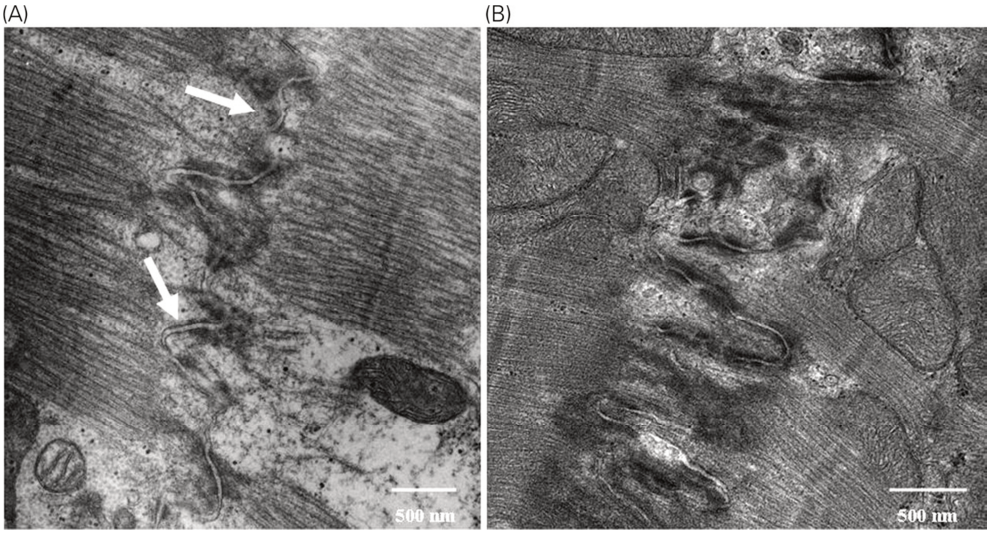


Figure 1.12. Paleness of desmosome internal plaques (arrows) in ACM (*A*) vs. control (*B*). From Basso et al., 2006.

An electron microscopy and morphometric analysis in a series of ACM patients provided evidence of highly convoluted nuclei, increased lipid droplets, decreased desmosome numbers and intercalated disc remodeling (Basso et al., 2006, Figure 1.12).

In 2009, Asimaki's group demonstrated a significant reduction in immunoreactive signal levels for plakoglobin and variable patterns of altered distribution for the other desmosomal proteins at the intercalated discs (Asimaki et al., 2009; Figure 1.13). The molecular mechanism responsible for this reduction remains unclear but it seems a consistent feature in patients with ACM and is not seen in other forms of heart-muscle disease.

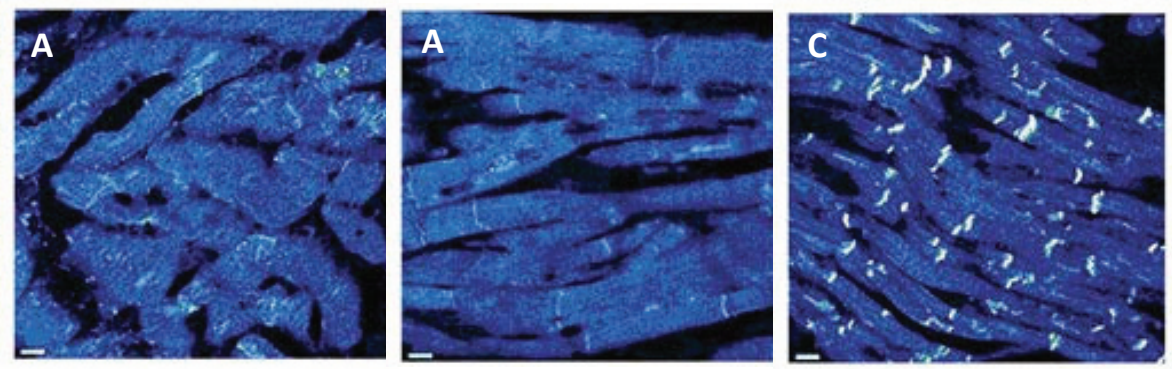


Figure 1.13. Immunofluorescence signal of plakoglobin of left ventricular myocardium from two subjects with ACM (A1 and A2) and a control subject (C). Modified from Asimaki et al., 2009.

Desmosomal dysfunction can also lead to gap junction remodeling because their strong associations with these structures and this could contribute to disease development (Delmar and McKenna, 2010). It does appear that reduced junctional expression of connexin-43 (Cx43), the major ventricular gap junction protein, occurs early in ACM, during the concealed phase (Kaplan et al., 2004; Asimaki et al., 2009). A significant decrease of Cx43, was observed also in N-cadherin–depleted hearts of knockout mouse model (Kostetskii et al., 2005), demonstrating the hierarchical relationship of the structural components of the intercalated disc in the working myocardium. Moreover, Sato and colleagues demonstrated that loss of desmosomal integrity can affect sodium channel function, with deleterious consequences to the electrical stability of the heart (Sato et al., 2009, 2011). These findings were also supported by another study, in which a PKP2 haploinsufficiency mouse model shows I_{Na} deficit that could contribute to generation and/or maintenance of arrhythmias in murine hearts (Cerrone et al., 2012). The remodeling of gap junctions in ACM patients occurs diffusely in regions of the heart that show no apparent structural or functional alterations, suggesting that this process could cause conduction abnormalities that promote arrhythmias in ACM (Asimaki et al., 2009). Taken together, these findings demonstrate that desmosomes, gap junctions, and sodium channels act as a functional triad in which changes in the composition of one constituent can affect the function and integrity of the others.

Together with the disruption of mechanical cell-cell junctions, various signaling pathways have been proposed to be involved in ACM pathogenesis, highlighting the dual function of junctional proteins as both components of intercellular adhesion structures and transcriptional regulators (Calore et al., 2015).

The first described pathway involved in ACM pathogenesis was Wnt/β-catenin pathway, which is known to regulate adipogenesis, fibrogenesis and apoptosis (Ross et al., 2000) and is primarily involved in differentiation of cardiac progenitor

cells and embryonic development of the right ventricle and its outflow tract (Ai et al., 2007). Garcia-Gras and colleagues in 2006 demonstrated that suppression of desmoplakin expression leads to nuclear localization of plakoglobin and a 2-fold reduction in canonical Wnt/β-catenin signaling through Tcf/Lef1 transcription factors. This event leads to enhanced fibrogenesis, adipogenesis, myocyte apoptosis, and accumulation of fat droplets (Garcia-Gras et al., 2006). The involvement of Wnt-signaling in ACM pathogenesis was also found in a plakoglobin knockout mouse model, in which ablation of plakoglobin causes an increase of the dephosphorilated and active β-catenin form, which can lead to cardiac hypertrophy (Li J et al., 2011). In contrast, another group reported the main involvement of TGFβ-mediated signaling in *JUP* mutant cardiomyocytes at the early stage of cardiomyopathy, excluding an alteration of Wnt/β-catenin signaling (Li D et al., 2011). TGFβ signaling regulates myocyte cell death in addition to its important influence on cardiac fibrosis and hypertrophy.

More recently, it has been postulated that an emerging signaling pathway, called Hippo/YAP pathway, seems to cross-talk with the canonical Wnt pathway and may have a role in its regulation and, in turn, in the pathogenesis of the disease. As β-catenin for Wnt signaling, neurofibromin 2 (NF2) is an upstream molecule for the Hippo pathway and it is localized at IDs. Perturbed molecular changes at the IDs could affect both Wnt and Hippo pathways and, the activation of the latter may cause a suppression of the canonical Wnt pathway. After IDs perturbation in fact, the Hippo kinases downstream of NF2 are cascade-phoshorylated, which in turn, cause YAP phosphorylation (p-YAP) and its cytoplasmic retention (Chen et al. 2014). p-YAP sequesters β-catenin and plakoglobin, resulting in a reduction of transcriptional activities in both pathways, followed by increased adipogenesis in ACM hearts (Figure 1.14).

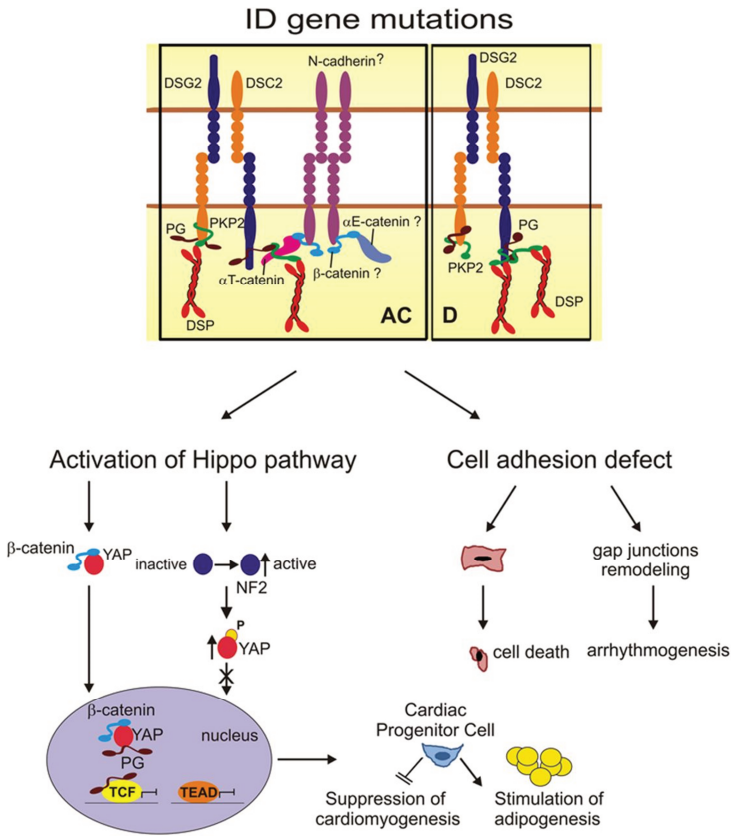


Figure 1.14. Model of the molecular pathogenesis of ACM. Crosstalk between Hippo and Wnt/ β-catenin pathways after mutations occurring on ID proteins. From Rampazzo et al., 2014.

1.11 Animal models for ACM

1.11.1 Spontaneous animal models

ACM, as other forms of cardiovascular diseases, occurs commonly in pets, particularly in domestic cats and dogs. These are spontaneous animal models for ACM because they closely resemble the clinical and pathological features of human disease.

Fox and co-workers, for the first time, described a spontaneously occurring ACM in 12 domestic cats, all of them showing pronounced RV lesions including marked myocardial injury (myocyte death and atrophy) and repair (fibrous and/or fatty replacement) (Fox et al., 2000).

Similar to feline models, boxer dogs with ventricular tachycardia and fatty or fibrofatty replacement were described as a spontaneous model for ACM (Basso et al., 2004). The precise inheritance pattern in these cases is unsolved but transmission from one generation to another suggest the likelihood of dominant inheritance which, in turn, was reported by Meurs and colleagues, and consistent with that of the human disease (Meurs et al., 2007).

Moreover a line of mice with inherited right ventricular dysplasia caused by a mutation of the gene laminin receptor 1 (*LAMR1*) is reported (Asano et al., 2004),

despite the fact that mutations in *LAMR1* have never been identified in humans with ACM.

These spontaneous animal models are considered a potentially important investigative tool that could enhance our understanding of the complex clinical and pathophysiological mechanisms responsible for sudden death and disease progression.

1.11.2 Genetically-engineered models

Plakophilin-2 mutant murine models

The first mouse model with a disruption of the desmosomal protein plakophilin-2 was created by the group of Grossmann in 2004. In the absence of plakophilin 2 ($PKP2^{-/-}$), mutant mice exhibited lethal alterations in heart morphogenesis and stability at mid-gestation (E10.5–E11) (Grossmann et al., 2004). In addition, disarrayed cytoskeleton and reduced trabeculations within the ventricles and atrial wall thinning was observed. Moreover, desmoplakin did not colocalize with junctional proteins at the intercalated disk but was dispersed over the cytoplasm, forming granular aggregates. In contrast to the myocardium, embryonic epithelia showed normal junctions. More recently, PKP2-heterozygous-null (PKP2-Hz) mutant mice have been created and studied (Cerrone et al., 2012). Ultrastructural but not histological or gross anatomical differences in PKP2-Hz hearts compared with wild-type (WT) were reported. On the other hand, decreased amplitude and a shift in gating and kinetics of the sodium current (I_{Na}) were observed in PKP2-Hz hearts as previously reported (Sato et al., 2009).

Desmoplakin mutant murine models

Desmoplakin targeted deletion mice $DSP^{-/-}$ die at embryonic stage due to desmosome instability and loss of tissue integrity (Gallicano et al., 1998; Garcia-Gras et al., 2006). Partial rescue of desmoplakin expression in extra embryonic tissues promoted survival of embryos not beyond gastrulation, with major defects in heart muscle, neuroepithelium and skin epithelium (Gallicano et al., 2001). As these animals die before birth, they cannot be consider a proper model for ACM. Heterozygous $DSP^{+/-}$ mice developed normally but exhibited excess adipocytes and fibrosis in the myocardium, thin ventricular walls, increased myocyte apoptosis, cardiac dysfunction, and ventricular arrhythmias (Garcia-Gras et al., 2006). In another study, $DSP^{+/-}$ mice showed normal echocardiograms but delayed conduction and inducible ventricular tachycardia associated with a delocalization and reduced expression of Cx43 (Gomes et al., 2012). Furthermore, transgenic (Tg) mice with cardiac-restricted overexpression of the C-terminal DSP mutant (R2834H) demonstrated histological evidence of cardiomyocyte apoptosis, cardiac

fibrosis, and lipid accumulation, along with ventricular enlargement and cardiac dysfunction of both right and left ventricles, confirmed by necroscopy (Yang et al., 2006). Mutant mice also displayed interruption of DSP-desmin interaction and marked ultrastructural changes of the IDs. Recently, the same model was studied by Martherus and colleagues, who confirmed that Tg mice exhibit RV dilation and wall thinning compared to non-transgenic, when subjected to exercise. In these conditions AKT1 and GSK3-β signaling is perturbed in Tg mice, suggesting that endurance exercise causes progression of the ACM phenotype in these models (Martherus et al., 2016).
Taken together, these results underline the role of desmoplakin in mantaining desmosomal stability and heart tissue integrity.

Plakoglobin mutant murine models

In 1996, targeted deletion of plakoglobin was described by two independent groups (Bierkamp et al., 1996; Ruiz et al., 1996).
Homozygous $JUP^{-/-}$ mutant mice die between days 12-16 of embryogenesis due to severe defects in heart function; in addition, one of these models showed a blistering skin phenotype, with epidermolytic hyperkeratosis (Bierkamp et al., 1996). Heterozygous $JUP^{+/-}$ mice appeared healthy and fertile. Ten-months-old animals had perturbed RV function and spontaneous ventricular ectopy, in the absence of left ventricular involvement. Histology did not identify right ventricular abnormalities and gap junctions localized correctly at the IDs (Kirchhof et al., 2006). Two transgenic mouse models overexpressing respectively wild-type (PG^{WT}) or truncated plakoglobin (PG^{TR}) were created (Lombardi and Marian, 2011). Both of them exhibited fibro-adiposis, cardiac dysfunction, and premature death; immunofluorescence experiments showed nuclear localization of PG^{WT} and PG^{TR}. A reduced binding with desmosome proteins *DSP* and *DSG2* was uniquely found in PG^{TR} mice.
More recently, double knockout mice for both plakoglobin and β-catenin (DKO) were generated, to determine whether β-catenin is required to maintain cardiac conduction in the absence of plakoglobin (Swope et al., 2012). These mice showed reduced gap junctions at the IDs, fibrous tissue replacement, and conduction abnormalities resulting in high percentage of SCD.

Desmoglein-2 mutant murine models

Transgenic mice with cardiac overexpression of the human homologue mutation p.N266S were generated (Tg-NS) (Pilichou et al., 2009). Contrary to homozygous $DSG2^{-/-}$ described earlier (Eshkind et al., 2002), Tg-NS did not manifest embryonic lethality, but they generally recapitulated human ACM phenotype, including

sudden death at young age, spontaneous ventricular arrhythmias, cardiac dysfunction, and biventricular dilatation and aneurysms.
In another study, an homozygous DSG2 mutant mouse lacking two extracellular adhesion domains was considered and showed myocardial fibrosis and calcification (Krusche et al., 2011). These data suggest that desmoglein-2 is important in embryonic viability and loss of this protein compromises cell adhesion, but new findings demonstrated that the presence of DSG2 is not essential for late heart morphogenesis and for cardiac contractility to support postnatal life (Kant et al., 2016).

Ryanodine receptor-2 murine models

Knock-in mice with the human disease-associated *RYR2* mutation R176Q ($RYR2^{R176Q/+}$) were generated by Kannankeril and colleagues. Histological analysis revealed no evidence of fibrofatty infiltration or structural abnormalities, but it was reported that right ventricular end diastolic volume was decreased in $RYR2^{R176Q/+}$ mice compared with controls. This model was inconclusive regarding the phenotype but demonstrated that R176Q mutation in *RYR2* predisposes the heart to catecholamine-induced oscillatory calcium-release events that trigger a calcium-dependent ventricular arrhythmia, consistent with a CPVT phenotype (Kannankeril et al., 2006).

Zebrafish models

A zebrafish model system was used by Heuser and coworkers to investigate the physiological effects of a heterozygous splice-acceptor-site mutation in intron 5 (c.631-2A>G) of *DSC2* gene, which causes a creation of a premature stop codon (Heuser et al., 2006). Morpholino-induced knockdown of *DSC2* gene resulted in a cardiac phenotype with dose-dependent bradycardia, chamber dilatation, abnormal cardiac contractility, and progressive pericardial oedema. Co-injection with wild-type human DSC2 mRNA, but not with mutant mRNA, rescued the morphant phenotype revealing the paramount role of DSC2 in the establishment of the normal myocardial structure and function in early cardiac morphogenesis.
Knockdown of plakoglobin in zebrafish was also generated and resulted in decreased heart size, reduced heartbeat, cardiac oedema, valvular dysfunction leading to intrachamber reflux and a twisted tail (Martin et al., 2009). Moreover, morphant embryos showed an increased expression level of Wnt target genes, indicating that Wnt signaling was activated in absence of plakoglobin and β-catenin cannot compensate this loss in desmosomes.
Most recently, a zebrafish ACM model with the cardiomyocyte-specific expression of the human c.2057del2 *JUP* mutation was generated (Asimaki et al., 2014). The animals exhibited reductions in I_{Na} and I_{K1} current densities, heart enlargement

and peripheral oedema. Electron microscopy revealed also interruptions in cell-cell boundaries and structural disarray in mutant fishes. High-throughput drug screening has identified SB216763 as a molecule able to prevent or reverse the disease phenotype in these models, activating the canonical Wnt signaling pathway. The efficacy of SB216763 has been supported by additional experiments performed in two different cell lines: neonatal rat ventricular myocytes with the same *JUP* mutation and induced pluripotent stem cells from two ACM probands. Treatment with SB216763 restored the subcellular distribution of plakoglobin, connexin-43 and Nav1.5 in cardiomyocytes and of SAP97, a protein known to mediate the forward trafficking of Nav1.5 and Kir2.1, in both cell types. These findings were consistent with previous demonstrations of the involvement and dysregulation of Wnt/β-catenin signaling pathway in ACM pathogenesis.

1.12 Induced Pluripotent Stem Cells (iPSCs)

In the last few years, a new powerful tool to study ACM mechanism has been introduced and is represented by patient iPSC-derived cardiomyocytes (iPSC-CMs).

Up to now, all the published studies on iPSCs and ACM examined the effect of different *PKP2* mutations on patients derived iPSCs.

Kim and colleagues generated cardiomyocytes from fibroblasts of two patients, who carry a homozygous p.G828G and a heterozygous p.K672RfsX12 PKP2 mutations respectively (Kim et al., 2013). Mutant iPSC-CMs demonstrated abnormal plakoglobin nuclear translocation and decreased β-catenin activity in cardiogenic conditions, but it was not sufficient to reproduce the pathological phenotype. The induction of an adult-like metabolism by culturing beating mutant embryoid bodies in a lipogenic milieu co-activated master peroxisome proliferator-activated receptor (PPAR)-α-dependent metabolism and the PPAR-γ pathway and recapitulated the classical ACM features consisting in an exaggerated lipogenesis and apoptosis. In another study, Ma et al. reprogrammed iPSC from dermal fibroblasts of an ACM patient, harboring an heterozygous PKP2 mutation (p.L614P) (Ma et al., 2013). Gene expression levels of *PKP2* and plakoglobin in these mutant cells were significantly lower compared with cardiomyocytes from control iPSCs, whereas no significant differences in the expression of desmoplakin, N-cadherin, and connexin 43 between the two groups was observed. Moreover, ACM–iPSC cardiomyocytes appeared larger and, darker lipid droplets were found. Clusters of lipid droplets were identified also in the ACM-iPSC cardiomyocytes from two other patients with two different frameshift *PKP2* mutations (Caspi et al., 2013). Immunostaining experiments revealed reduced

densities of PKP2, the associated desmosomal protein plakoglobin, and the gap-junction protein connexin-43. Transmission electron microscopy identified widened and distorted desmosomes in the ACM-iPSC cardiomyocytes and it seemed to be correlated with the upregulation of pro-adipogenic PPAR-γ pathway. The effect of the adipogenic stimuli on the ACM-iPSC cardiomyocytes could be prevented by activation of the canonical Wnt pathway in these cells by a specific inhibitor of GSK-3β, suggesting the possible role of the canonical Wnt/β-catenin pathway in ACM pathogenesis.

Despite the unique advantages of the iPSCs approach for modeling ACM, such as to overcome limitations arising from species differences in electrophysiological properties, this method still possesses some inherent limitations. It is unable to study pathological phenomena at the whole-organ and systemic levels and, up to now, only PKP2 mutant iPSC cardiomyocytes have been generated.

2. AIM OF THE STUDY

Sudden death, ventricular arrhythmia and heart failure are common features in arrhythmogenic cardiomyopathy (ACM), an inheritable heart muscle disease, characterized by clinical and genetic heterogeneity. The typical pathological profile associated with the disease consists of the progressive replacement of the myocardium by fatty or fibro-fatty tissue. So far, 13 disease-genes have been identified, responsible for around 60% of all ACM cases, resulting in about 40% of patients genotype-negative with a strong clinical diagnosis.

The role of the genetic analysis appears fundamental, not only to confirm a diagnosis and determine the potential risk in close relatives but also to research the genetic causes underlying the phenotypic expression. To this purpose, the introduction of the next generation sequencing (NGS) techniques made this process faster and cheaper, allowing to parallel sequence a large amount of genes.

The objective of this study is twofold: to set up a custom targeted gene panel performing a comprehensive mutation screening in a cohort of affected subjects; and to identify putative novel ACM genes in both sporadic and familial cases by using three different NGS approaches (targeted gene panels, whole exome sequencing, whole genome sequencing).

Identification of novel disease-genes is of paramount importance to understand the molecular pathogenesis of ACM, as well as to increase the power of genetic screening and develop successful targeted therapies.

3. MATERIAL AND METHODS

3.1 Clinical evaluation

The study population consisted of 63 unrelated index cases of Italian descent with a diagnosis of ACM. Moreover, a total of 70 family members were subsequently clinically evaluated in the University of Padua Medical School-Azienda Ospedaliera.

After written informed consent, blood for DNA extraction was obtained from all participating individuals, according to the pertinent Italian legislation. Clinical evaluation consisted of a detailed personal/family history, physical examination, 12-lead electrocardiogram (ECG), signal-averaged ECG (SAECG), 24-hour Holter ECG, and 2-dimensional echocardiography (2D-echo). Invasive tests including angiography and right ventricular endomyocardial biopsy were performed in particular cases, when deemed necessary for the diagnosis. Clinical diagnosis of ACM was made according to the major and minor criteria established by the International Task Force of the European Society of Cardiology/International Society and Federation of Cardiology (Marcus et al., 2010).

3.2 Salting-out DNA extraction

Blood samples were stored at -20°C in EDTA vacutainer tubes and genomic DNA was extracted from peripheral blood leukocytes using a modification of the salting-out procedure (Miller et al., 1988) at the Human Genetic Laboratory of University of Padua. To facilitate haemolysis of Red Blood Cells it is recommended to store a fresh sample for few hours in a freezer. After thawing, N-N solution (0,9%, NaCl 0,1% Nonidet) was added to the blood and transferred into a Falcon tube, to a final volume of 40 mL. Two sequential sample centrifugations for 35 minutes at 6000 rpm at 4°C followed. After centrifugation, supernatants were discarded and pellets were resuspended in 4 mL of TEN solution (TrisHCl 10mM; EDTA 2mM, pH 8; NaCl 400 mM) and well vortexed. After adding 300 µl of 20% SDS, the tubes were incubated at 80°C for 3 hours under mixing, to denature cellular proteins. After the incubation, 1 ml of saturated NaCl solution was added to the samples to induce protein precipitation and, after tubes spinning for 10 minutes at 6000 rpm at room temperature (RT), supernatant was poured off into a fresh tube, using a sterile Pasteur pipette. An isovolume of chloroform was added and the tubes were inverted for 10-15 times and centrifuged for 10 minutes at 6000 rpm at RT. After centrifugation, three different phases appeared: the upper containing nucleic acids, the second composed of proteins and cellular debris, and the third, consisting of chloroform. The upper aqueous phase was transferred into a clean

and sterile centrifuge tube (15 ml) and an equal volume of isopropanol was added; DNA was precipitated by gentle swirling of the tube and observed visually as a white thread. Samples were then centrifuged for 10 min at 6000 rpm at RT and supernatant was discarded. Pellets were washed twice with 1-2 ml of 70% ethanol and centrifuged for 10 minutes at 6000 rpm at RT. After the last centrifugation, pellets were dried from the excess of ethanol leaving the tubes open for 1 hour at RT or in a heater for less time, till the pellets appeared transparent. Finally, the dried pellets were resuspended in 300-500 µl TE buffer.

3.3 DNA quantification

DNA concentration was determined using NanoDrop ND-1000 UV-Vis Spectrophotometer (CELBIO). This instrument can measure DNA concentrations starting from 1 µL samples, also for highly concentrated samples without dilution, with high accuracy and reproducibility. Samples readings at 260 nm and at 280 nm provide directly DNA and proteins concentration in the sample, respectively. The ratio of these two values of absorbance was used to assess the purity of DNA, checking the presence of proteins or other contaminants that absorb near 280 nm. Pure DNA preparations have ratio values between 1.8 and 2.0. A ratio inferior than 1.8 indicates that there may be proteins and/or ultraviolet (UV) absorbers in the sample, thus it is recommended to re-precipitate the DNA, whereas a ratio higher than 2.0 indicates that samples may be contaminated with chloroform or phenol. Moreover, the spectrophotometric analysis provided the absorbance ratio at 260 and 230 nm, as secondary measure of DNA purity. Expected 260/230 values are commonly in the range of 2.0-2.2. If the ratio is appreciably lower than expected, it may indicate the presence of contaminants which absorb at 230 nm, such as phenolate ion, thiocyanates, and other organic compounds.
The quantified DNA samples were diluted in sterile bidistilled water to a final concentration of 50 ng/µl and used for subsequent polymerase chain reactions (PCR).

3.4 Genotyping and Linkage Analysis

Linkage analysis was the primary tool used for the genetic mapping of Mendelian and complex traits with familial aggregation. It is still used to determine if there is significant evidence for co-segregation (linkage) of alleles at a marker locus and alleles at a hypothetical disease locus. The rate at which two alleles are transmitted together to the offspring is related to the probability of a crossing-over event between the loci and hence to their distance along the chromosome.

In the parametric linkage analysis the LOD score (logarithm of odds) value, is calculated to assesses the probability that a given pedigree, where the disease and the marker are co-segregating, is due to the existence of linkage or to chance (Morton, 1955). By convention, a LOD score greater than 3.0 is considered evidence for linkage, as it indicates 1000 to 1 odds that the linkage being observed did not occur by chance; conversely linkage hypothesis can be rejected if the LOD score is less than -2.0. Values of LOD score between -2 and +3 are inconclusive.

Individual genotyping for 33 subjects belonging to the same family (Family #6) was carried out using the High-density SNP marker set Infinium HumanCNV370-Duo BeadChip (Illumina, San Diego, CA, USA), which includes 370,404 SNP markers with average spacing of 7.9 Kb. SNPs with high heterozygosity and uniform genome-wide distribution were selected to improve the probability to detect recombination events.

For each family member, 200 ng of genomic DNA (gDNA) was analyzed with the SNP arrays on Illumina Iscan System according to the manufacturer's protocol (Figure 3.1).

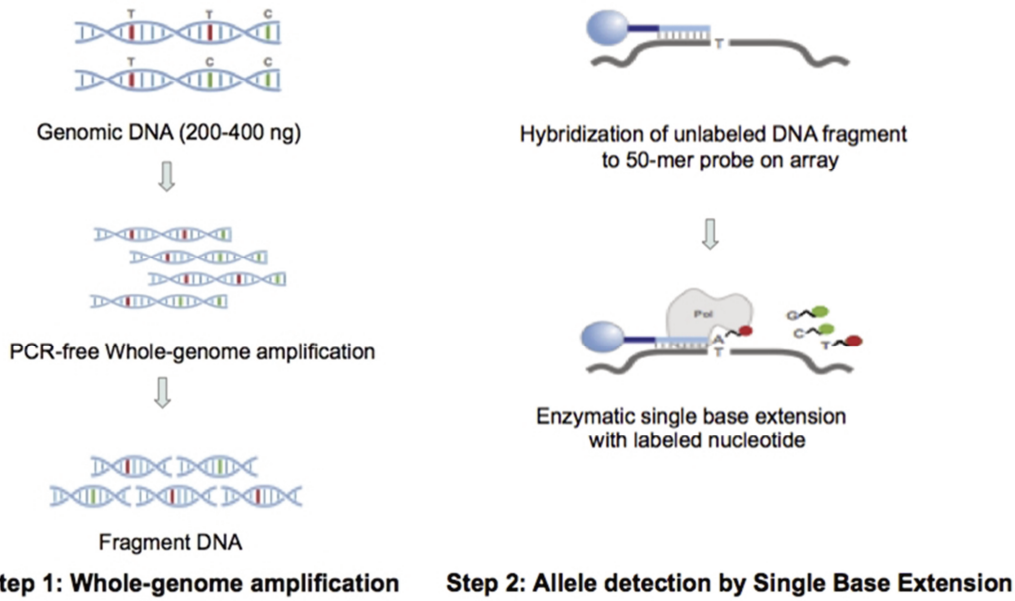


Figure 3.1. Illumina whole-genome genotyping on DNA arrays (www.illumina.com). **Step 1.** Whole-genome amplification reaction amplifies the input gDNA by over 1000-2000X. This genome representation is fragmented to approximately 300 to 600 bp, precipitated, resuspended, and hybridized onto a BeadChip. **Step 2.** After capture of the target loci, the SNPs are scored using a single-base extension (SBE) reaction. SBE employs a single probe sequence 50 bp long (attached to the bead) designed to hybridize immediately adjacent to the SNP query site. The arrayed SNP locus-specific primers are extended with a single hapten-labeled dideoxynucleotide. Two different colored labeled terminators are used to discriminate the alleles. Finally, the haptens are detected in a multilayer immunohistochemical sandwich assay and genotype calls are generated by Illumina's software (Gunderson et al., 2005).

Genotyping raw data were processed by Plink software v1.07 (Purcell et al., 2007) to obtain a single file (*.snp) including all genotyping data of each family member, with markers sorted according to their physical map position. Quality control of genotypes was carried out with the program PLINK and SNPs not fulfilling the following criteria were excluded from analysis: minor allele frequency <1%; call rate <95% (defined as the percentage of successful genotype calls among subjects); Hardy-Weinberg equilibrium $p<0.001$.
Moreover, SNPs in the same physical position or mapped on multiple loci were not included in the analysis.

Linkage analysis was conducted by GeneHunter software version 2.1 (Kruglyak et al., 1996) through the graphical user interface EasyLinkage Plus v5.08 (Lindner and Hoffmann, 2005), using a parametric multipoint approach.
An autosomal dominant inheritance with 0.0001 disease allele frequency and 0.02 phenocopy rate was assumed, as well as equal allele frequencies for the markers tested. In order to avoid decrease of LOD score values due to low penetrance, an ‘‘affected-only’’ approach was used, which classifies all individuals with no clinical manifestations as “unknown’’. For all regions showing positive LOD score values, haplotypes were constructed using Viterbi algorithm implemented in GeneHunter software and displayed by HaploPainter program (Thiele and Nurnberg, 2005). Haplotypes were subsequently checked on the basis of the minimal number of recombinations and both upper and lower borders of each critical interval were refined by GeneHunter, based on recombination events observed in affected individuals.

Next Generation Sequencing Technology

Background

First described by Sanger in 1977, capillary electrophoresis (CE)-based sequencing of DNA allowed to sequence the full genome of any species in a reliable, reproducible manner (Sanger et al., 1977). After the first automated sequencer (AB370), the Genome Analyzer emerged in 2005 really revolutionized the sequencing capability, sequencing up to 1 Gb per run, thus introducing the “next generation” in genomic science. From that point forward, the data output of next generation sequencing (NGS) has outpaced Moore’s law (Figure 3.2).

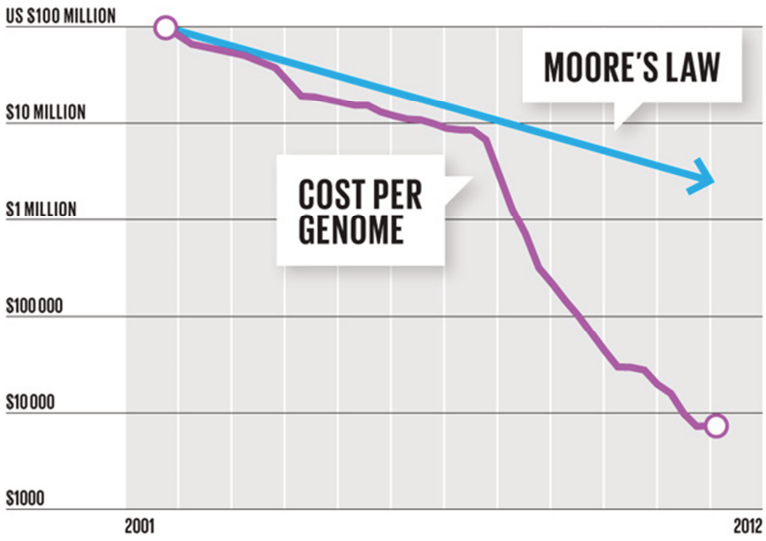


Figure 3.2. Sequencing costs per basepair: after the introduction of NGS technology the cost are falling faster that Moore's law.

The concept behind NGS technology is similar to dideoxy-terminal method: DNA polymerase catalyzes the incorporation of fluorescently labeled dNTPs into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, at the point of incorporation, the nucleotides are identified by fluorophore excitation. In contrast to Sanger sequencing, NGS extends this process across millions of fragments in a massively parallel fashion.

Different NGS approaches share general processing steps, based on the Illumina sequencing by synthesis (SBS) chemistry. Illumina NGS workflows include 4 basic step (Figure 3.3):

A. Library preparation: DNA is randomly fragmented and adapters are ligated at 5' and 3' ends. Adapter-ligated fragments are then PCR amplified and gel purified.
B. Cluster generation: the library is loaded into a glass slide flow cell that displays oligonucleotides complementary to Illumina adapter sequences. Each fragment is then amplified into distinct, clonal clusters through bridge-PCR (Figure 3.5). The clusters templates are ready for sequencing.
C. Sequencing: in Illumina SBS technology, all 4 nucleotides are added simultaneously to the flow cell channels, along with DNA polymerase, for incorporation into the oligo-primed cluster fragments. The nucleotides carry a base unique fluorescent label and the 3'-OH group is chemically blocked, so that each incorporation is a unique event and natural competition minimizes incorporation bias. Subsequently, the 3' blocking group is chemically removed to prepare each strand for the next incorporation. The

cycle is repeated, one base at a time, generating a series of images each representing a single base extension at a specific cluster.

D. Data analysis: The newly identified sequence reads are then aligned to the human reference genome. Following alignment, many variations such as single nucleotide variants (SNVs) or insertions-deletions (InDels) are called.

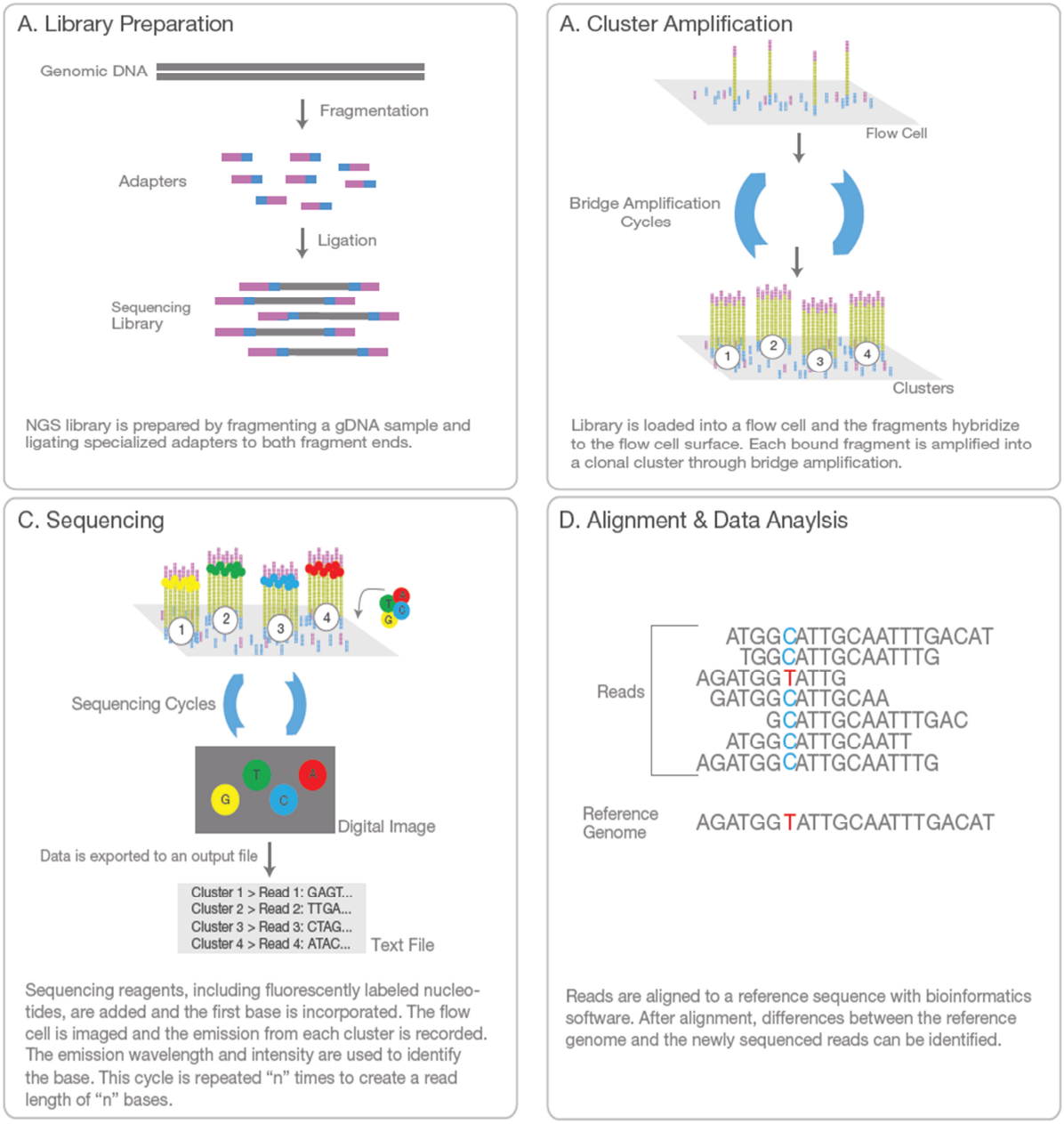


Figure 3.3. Illumina SBS chemistry workflow.
Adapted from http://www.illumina.com/technology/next-generation-sequencing.html.

3.5 Targeted gene panels

The study involved a cohort of 59 unrelated index cases with diagnosis of ACM. In the majority of them (50 out of 59) the presence of mutations in the three most

mutated ACM genes (*DSP*, *PKP2* and *DSG2*) was excluded by previous mutation screening with direct Sanger sequencing and dHPLC analysis. Only 9 samples sequenced with the 'Cardiomyopathies gene panel' (described below) were not previously investigated. Moreover, DNA from 4 patients were analyzed with both panels.

Two different Targeted NGS gene panels have been generated in collaboration with BMR Genomics, a spin-off of University of Padua.
The first panel, named 'Cardiomyopathies gene panel' comprises 56 genes (443 kb), associated with different cardiomyopathies (Table 1, Appendix A). Specific exonic regions of 19 ACM probands were captured and sequenced on a MiSeq sequencer which provides a mean coverage depth of 200X.
On the second panel, conventionally called 'ACM known and candidate genes panel', 44 samples were sequenced by using a NextSeq500 platform. The exonic and intronic flanking regions of 69 genes were sequenced at a mean coverage depth of 267X. The genes included in this panel have been selected because their previous association with ACM phenotype or as candidate genes. A gene is considered 'candidate' if encodes for a protein which functions and/or localization can be associated with the disease pathogenic mechanism (Table 2, Appendix A).
In both panel strategies the preliminary enrichment step has been performed with Haloplex Target Enrichment System as described below.

a. Samples precipitation

Prior to capture, DNA samples must be resuspended in nuclease-free water to a final concentration of 200 ng/µl and a minimum quantity of 4 µg. Since all the samples have been already diluted in TE buffer after DNA extraction (paragraph 3.2), they must be precipitated.
A volume of sodium acetate 3M (pH=5.2) equal to 1/10 of the total DNA volume in addition to a 2.5X volume of absolute ethanol were added to the samples. After mixing, the samples were incubated overnight at -20°C and then centrifuged for one hour at 4 °C. Supernatants were discarded and two washes with cold 70% ethanol followed. After a short centrifugation (5-10 min), supernatants were eliminated and pellets were dried at RT or in a heater (37°C). Finally, the dried pellets were resuspended in a volume of bidistilled water to reach the theoretical requested concentration (200 ng/µl).

b. Haloplex Target Enrichment System

Haloplex workflow is optimized for a total of 225 ng of genomic DNA (gDNA) and comprises the following steps (Figure 3.4).

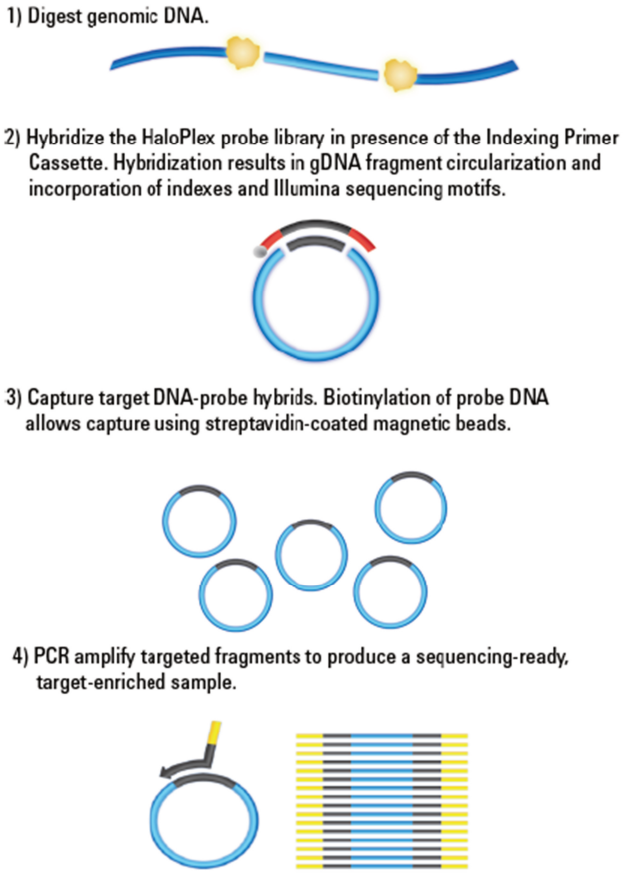


Figure 3.4. Haloplex workflow. From HaloPlex Target Enrichment System Automation Protocol for Illumina Sequencing (http://www.chem.agilent.com/library/usermanuals/Public/G9900-90020.pdf).

1. Digest genomic DNA with restriction enzymes. In this step, gDNA samples are digested by 16 different restriction enzymes to create a library. The gDNA is digested in 8 different restriction reactions, each containing 2 enzymes.
2. Hybridize digested DNA to HaloPlex probes for target enrichment and sample indexing. In this step, the restriction fragments are hybridized to the HaloPlex probe capture library. HaloPlex probes are designed to hybridize selectively to fragments originating from target regions of the genome and to direct circularization of the targeted DNA fragments. During the hybridization process, Illumina sequencing motifs including index sequences are incorporated into the targeted fragments.
3. Capture, amplify and purify the target DNA. The circularized target DNA-HaloPlex probe hybrids, containing biotin, are captured on streptavidin beads. After capture, DNA ligase is added to seal nicks, then target DNA is eluted and PCR-amplified. The amplified target DNA samples are purified using AMPure XP beads.

4. Validate enrichment and quantify enriched target DNA. Prior to sample pooling and sequencing sample preparation, enriched target DNA in each library sample is validated and quantified by microfluidics analysis using the 2100 Bioanalyzer. Enriched library samples may also be qualitatively analyzed using gel electrophoresis. Each amplicon in the prepared library includes 50 to 500 bp of target DNA insert and 125 bp of sequencing motifs.
5. Pool samples with different indexes for multiplexed sequencing. Bioanalyzer-measured concentration of 175-625 bp products in each sample are used to pool equimolar amounts of differentially indexed samples in order to optimize the use of sequencing capacity. The final enrichment pool is ready for direct sequencing using Illumina paired-end primers and chemistry on the Illumina MiSeq or NextSeq500.

c. Sequencing

Sequencing of the enriched libraries was done on MiSeq or NextSeq500 platform (Illumina) which produce 150 bp paired-end (PE) reads. PE sequencing involves sequencing both ends of the DNA fragments, and aligning the forward and reverse reads as read pairs. As previously described Illumina technology relies on solid-phase bridge PCR (Adessi et al., 2000; Fedurco et al., 2006) to amplify each DNA template (Figure 3.5) and reversible terminator chemistry to sequence the amplified libraries.

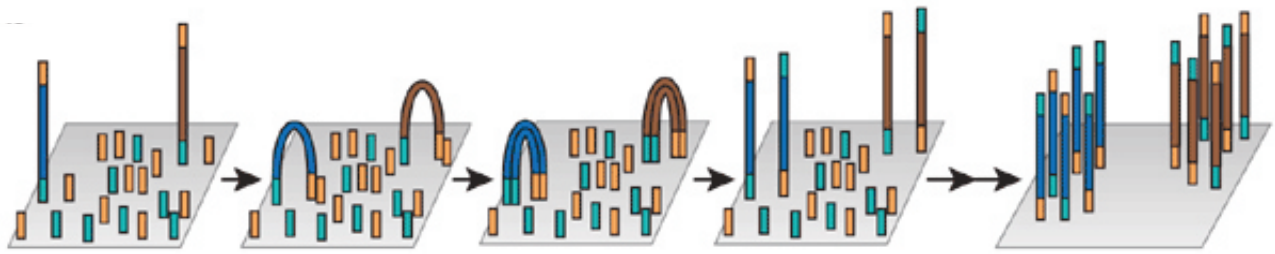


Figure 3.5. Bridge-PCR. The flow cell surface is coated with two different populations of oligonucleotides 5'-attached by a flexible linker and complementary to the adapters of the library DNA fragments. Single-stranded library molecules are annealed to the immobilized oligonucleotides and a new strand is synthesized from the original template which is then removed by denaturation. The copied strand bends over and the adapter sequence at the 3' end hybridizes to another complementary oligonucleotide, forming a bridge and allowing the synthesis of a second covalently bound strand. At the conclusion of the PCR, each clonal cluster contains about 1,000 copies of a single member of the template library. Modified from Shendure and Ji, 2008.

d. Bioinformatic analysis

Sequence reads were mapped to the UCSC reference human genome (hg 19, NCBI build 37.1) (http://genome.ucsc.edu) and *Binary Alignment Map* files (.bam) have been generated. These files can be loaded to genomic public browsers as

Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA) and represent the bases for the variant calling process. Different variant callers such as VarScan and FreeBayes were used to detect any difference between the sequencing product and the reference genome, generating a *Variant Call Format* (.vcf) file. The annotation step performed by BMR Genomics was done using ANNOVAR (Wang et al., 2010; http://www.openbioinformatics.org/annovar/), which integrates several information to the .vcf file about the coordinates, the frequencies and the predictions of the identified variants. To facilitate variants annotation and data analysis, BMR Genomics created also an on-line tool, called *SNP-shot* (paragraph 3.11).

e. Mutation validation

Direct Sanger sequencing of PCR products from genomic DNA was performed to confirm the presence of the detected variants and to study the segregation within the family members, where possible.

All PCR primers were designed by Primer3 web version 4.0.0 software (http://primer3.ut.ee/), based on the sequence data in GenBank. PCR reactions were carried out using a standard or touch-down protocol.

3.6 DNA amplification by PCR

All the uncovered exons of the known-disease genes in the 19 probands screened with the 'Cardiomyopathies gene panel' other than exons containing a detected variant were amplified by Polymerase Chain Reaction (PCR) assay and sequenced by direct Sanger sequencing.

PCR reactions were carried out in a total volume of 25 µl containing 50 ng of DNA, 10 pmoles of each primer (10 pmol/µl), 100 µM deoxinucleotide triphosphate (dNTPs 1mM, Invitrogen), 1X PCR Buffer (Promega), 5U/µl of GoTaq DNA polymerase (Promega), and bidistilled water to reach the final volume.

DNA amplification was performed in a Peltier PTC-200 thermal cycler (MJ Research) using standard or touch-down protocols.

After activation of the enzyme at 95°C, the cycling conditions (denaturation at 95°C for 30 sec, annealing at the working temperature for 30 sec, extension at 72°C for 30 sec) were repeated 35 times in a standard PCR protocol. For a touch-down (TD) protocol the thermal cycler was programmed to perform runs in which the annealing temperature is lowered incrementally during the PCR cycling from an initial value above the expected Tm to a value below the Tm. TD-PCR should increase the specificity of a PCR reaction, discouraging the formation of spurious products.

An aliquot of PCR product was loaded onto a 2% w/v agarose gel stained by Gel Red (Biotinum), in order to check the amplification reaction. Amplicons were then visualized under ultraviolet trans-illumination and identified by comparing with a known molecular-weight marker.

3.7 Sanger sequencing

Before the sequencing step, PCR amplicons had to be purified to eliminate possible dNTPs and primers not used in the amplification that could interfere with the sequencing reaction. Purification was performed with a mix of two enzymes (ExoSAP, Affymetrix): SAP, an alkaline phosphatase, that removes exceeding nucleotides, and EXOI, an exonuclease that removes single strand filaments and primers. A volume of 2.5 µl of ExoSAP solution was add to 5 µl of PCR product. The mixture was then incubated in a thermal cycler at 37°C for 15 minutes and then at 80°C for other 15 minutes to inactivate the enzymes. The purified PCR products together with 1 µl of primer (forward or reverse) were sequenced at the BMR Genomics (Padua, Italy) by ABI Prism 3730XL DNA sequencer (Applied Biosystems), with the BIG DYE dideoxy-terminator chemistry (Applied Biosystems). Sequencing results were visualized as electropherograms and analyzed by using Seqman II software (DNASTAR). It allows to align and compare the obtained sequence with the consensus sequence and eventually detect single nucleotide variants (SNVs), small insertions and deletions (InDels).

3.8 *In silico* analysis of *TJP1* gene variants

Two different mutations (p.R265W, p.Y669C) affecting the human tight junction protein ZO-1 (*TJP1*) were identified in two probands with the 'ACM known and candidate genes panel'. To predict the effect of these variations, a conservation analysis was performed in collaboration with Dr. Giovanni Minervini and Prof. Silvio Tosatto, Department of Biomedical Sciences, University of Padua.

The *TJP1* sequence (accession code: Q07157) was downloaded from the Uniprot database. Homologous sequences were retrieved from the OMA browser database using standard parameters and then realigned with Clustal Omega using a gap opening penalty 3.0 and gap extension penalty 1.8. The final alignment has been refined using Jalview (i.e. removing macroscopic misalignment and used in the subsequent analysis). Search into PFAM database was used to identify conserved domains, whereas search into MobiDB was used to investigate the presence of intrinsically disordered regions. For the second mutation (p.Y669C), which affects the guanylate kinase-like domain of *TJP1*, a modular dynamics simulation was performed. The mutation was modeled with Modeller using the wild

type crystal structure of SH3-Guanylate kinase core domain (PDB code 3LH5) as template. The resulting model was used to perform molecular dynamics simulation in order to highlight the structural impact of amino acid substitution. The overall fold stability was studied by performing 40 ns of molecular dynamics (MD) simulation with Gromacs 4.6 using the CHARMM 27 force field. Finally, networks of interacting amino acids affected by mutations were calculated with RING.

3.9 Whole Exome Sequencing (WES)

Whole exome sequencing was performed in 11 total affected subjects at the Beijing Genomics Institute (BGI, http://www.genomics.cn/en/index): 4 subjects of Family #6 (IV-5, IV-8, IV- 9, IV-12), 3 subjects of Family #5 (I-1, II-2, II-4), and 2 subjects belonging to Family #3 (II-4, II-7) and Family #4 (II-4, III-5), respectively.

a. SureSelect Target Enrichment and sequencing

For each sample, individual library preparations, hybridizations, and captures were performed, following the manufacturer's protocols.

The qualified gDNA sample was randomly fragmented by Covaris and paired-end sequencing adapters were ligated to both ends of 150-200 bp DNA fragments. The adapter-ligated templates were purified by the Agencourt AMPure SPRI beads (Beckman Coulter Genomics) and fragments with insert size about 176 bp were excised. Extracted DNA was amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the SureSelect Biotinylated RNA Library (BAITS) for enrichment with the Agilent SureSelect All Exon Kit (Figure 3.6). Hybridized fragments were bound to the strepavidin coated magnetic beads and, after washing non-hybridized fragments and RNA digestion, captured products were amplified by PCR. Captured LM-PCR products were subjected to Agilent 2100 Bioanalyzer to estimate the magnitude of enrichment. Each captured library was then loaded on Illumina Hiseq2000 platform, and high-throughput sequencing were performed as described in paragraph 3.5 c.

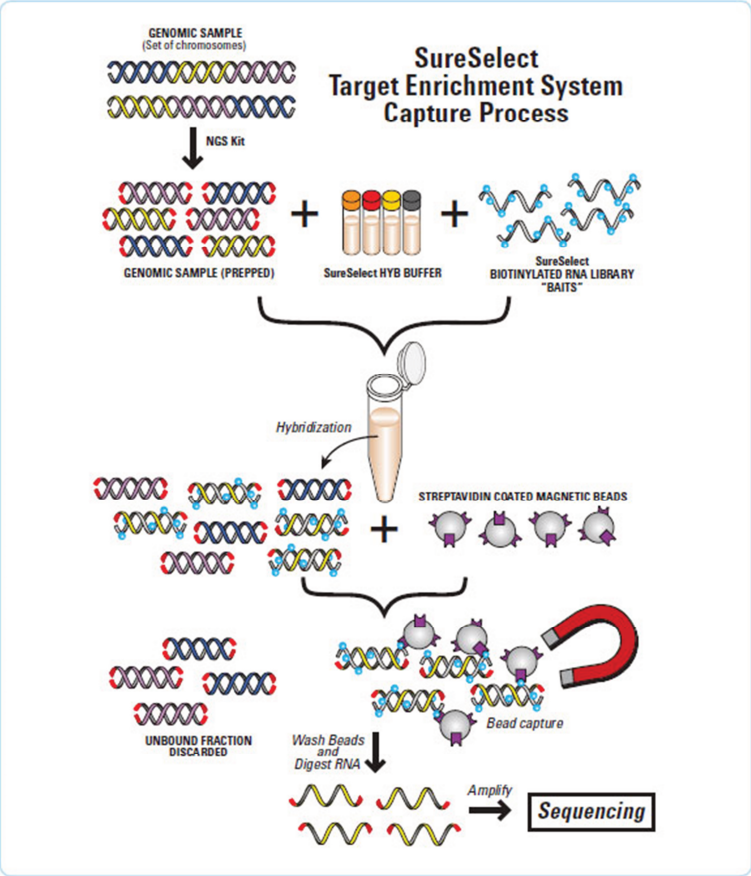


Figure 3.6. SureSelect Target Enrichment workflow.

1. Starting from gDNA, a shearing step produces small fragments.
2. Prepare library with sequencer specific adaptors and indexes.
3. Hybridize sample with biotinylated RNA library baits (120mer).
4. Select targeted regions using magnetic streptavidin beads.
5. Amplify and load on the sequencer.

Sequencing-derived raw image files were processed by Illumina base-calling Software 1.7 for base-calling with default parameters and the sequence data of each individual was generated as paired-end reads, which was defined as "raw data" and stored in *FASTQ* format.

b. Bioinformatic analysis

Starting from the raw data produced by the Illumina sequencer, the “clean data” were generated through a filtering step. Data filtering included: i) Removing reads containing sequencing adapter; ii) Removing reads whose low-quality base ratio (base quality less than or equal to 5) is more than 50%; iii) Removing reads whose unknown base ('N' base) ratio is more than 10%. All “clean data” of each sample were mapped to the human reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA V0.7.12) software. Considering that the same DNA molecules can be sequenced several times during the sequencing process and that the resulting duplicate reads are not informative, Picard tools (v1.118) was used to mark and remove these duplicates. Local realignment around InDels and base quality score recalibration were performed using Genome Analysis Toolkit (GATK). The sequencing depth and coverage for each individual were calculated based on the alignments. In WES, the variant calling is restricted to the exonic

target regions and their flanking regions (extending 200bp towards both sides of each target region). The HaplotypeCaller of GATK(v3.3.0) was used to call both SNVs and InDels simultaneously via local *de novo* assembly of haplotypes. The raw variation set containing all potential variants was outputted into a .vcf file and, after that, an hard-filtering method was applied to get high-confident variant calls.
The AnnoDB tool developed by BGI was applied to perform: i) Gene-based annotation: identify whether SNVs or InDels cause protein coding changes and which amino acid is affected; ii) filter-based annotation: identify variants that are reported in SNPs databases (1000 Genomes Project database, ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release; dbSNP141, http://www.ncbi.nlm.nih.gov/snp) or identify the pathogenicity score given to the subset of coding non-synonymous variants by Sorting Intolerant From Tolerant (SIFT, http://sift.jcvi.org/) or Polymorphism Phenotyping v2 (PolyPhen2**,** http://genetics.bwh.harvard.edu/pph2/).
Each putative pathogenic variant identified was validated with Sanger sequencing and the analysis of the co-segregation with ACM phenotype within the families was carried on, as described in paragraphs 3.6 and 3.7.

3.10 Whole Genome Sequencing (WGS)

Unlike focused approaches such as targeted sequencing or WES, which analyze a limited portion of the genome, WGS delivers a comprehensive view of the 3.2 billion bases of the human genome.

The whole genome of subjects IV-3, IV-10 and IV-12 belonging to Family #6 was sequenced at Novogene (http://en.novogene.com/). A total amount of 1 µg DNA per sample was used to libraries preparation. Sequencing libraries were generated using Truseq Nano DNA HT Sample preparation Kit (Illumina) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. The genomic DNA were then polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real time PCR (Figure 3.7).

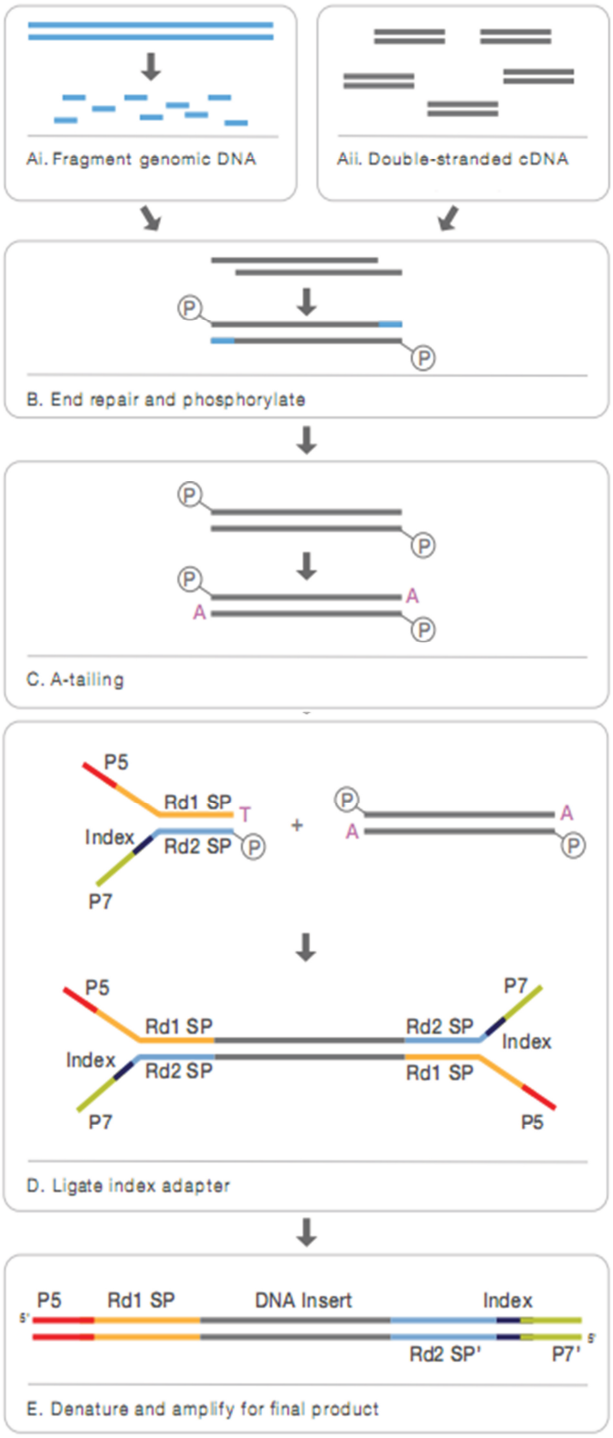


Figure 3.7. WGS workflow. Modified from preliminary report from Novogene.

After cluster generation with Hiseq X HD PE Cluster Kit (Illumina) according to the manufacturer's protocols, the library preparations were sequenced on the HiSeq X Ten Illumina platform and paired-end reads were generated. Raw data obtained from sequencing contains adapter contamination and low-quality reads, that were removed. Two parameters were considered to establish the sequencing quality: the error rate (e) and the base quality Phred score expressed as Qphred. They are linked by the following equation: Qphred= -10log10(e)

The sequencing error rate was obtained from the base calling process and the corresponding relation is listed as below:

Phred score	sequencing error rate	correct sequencing rate	Q-sorce
10	01/10	90%	Q10
20	1/100	99%	Q20
30	1/1000	99.9%	Q30
40	1/10000	99.99%	Q40

Table 3.1. Relation between Qphred and sequencing error rate.

Error rate is increasing with sequencing reads extension due to the consumption of chemical reagents and is higher in the first six bases of a read. Generally, each bases sequencing error rate should be smaller than 1% and most base quality is required to be greater than Q20.

Reads alignment and duplicated reads removal have been performed as described for WES. To detect genomic variations different software were used: SAMtools 1.0 (http://samtools.sourceforge.net/) to identify SNVs and InDels, BreakDancer 1.4.4 (Chen et al., 2009; http://breakdancer.sourceforge.net/) for structural variants, and control-FREEC v6.7 (http://bioinfo-out.curie.fr/projects/freec/) for copy number variations (CNVs). SNVs and InDels were then annotated by using ANNOVAR.

Moreover, exomes were extracted from the genomes and re-annotated using the in-house pipeline at the Human Genetics Department of the Radboud University medical center (Nijmegen) in collaboration with Dr. Christian Gilissen and Prof. Joris Veltman. Data from about 5000 WES projects were integrated and used to filter for the allele frequency.

3.11 Bioinformatic tools

UCSC Human Genome Browser (http://genome.ucsc.edu)

It's a browser that contains the reference sequence and working draft assemblies for a large collection of genomes. It comprises many useful tools and links for genomic analysis. The Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. Gene Sorter shows expression, homology and other informations on groups of genes. Moreover, graphical representation of the fragments making up a region of draft genome can be displayed, showing the relative size and overlaps of each fragment and also whether any gap between fragments are bridged by mRNAs or paired BAC end sequences. There is an

increasing amount of data becoming available from large-scale gene expression studies. More recently genomes from many different species are available on the browser.

Primer3 (http://primer3.ut.ee)
Primer3 is a free online tool to design and analyze primers for PCR and real time PCR experiments developed and supported from WI/MIT. Primer3 can also select single primers for sequencing reactions and can design oligonucleotide hybridization probes.

SeqManII (DNASTAR)
SeqManII is a software used to assemble from two to tens of thousands of electropherograms into contigs. Prior to assembly, SeqManII evaluates the quality of the underlying trace data to eliminate contaminating and poor quality data from the projects, thus generating the most accurate consensus sequence possible. This software is useful to detect nucleotides that differ from the consensus sequence and that could be a pathogenic mutation.

IGV (Integrative Genomic Browser)
The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of many different types of large genomic datasets. It is freely available for download from http://www.broadinstitute.org/igv. IGV includes a large number of specialized features for exploring next generation sequencing read alignments (.bam files), including features for sequencing coverage and variant visualization (.vcf files) (Figure 3.8).

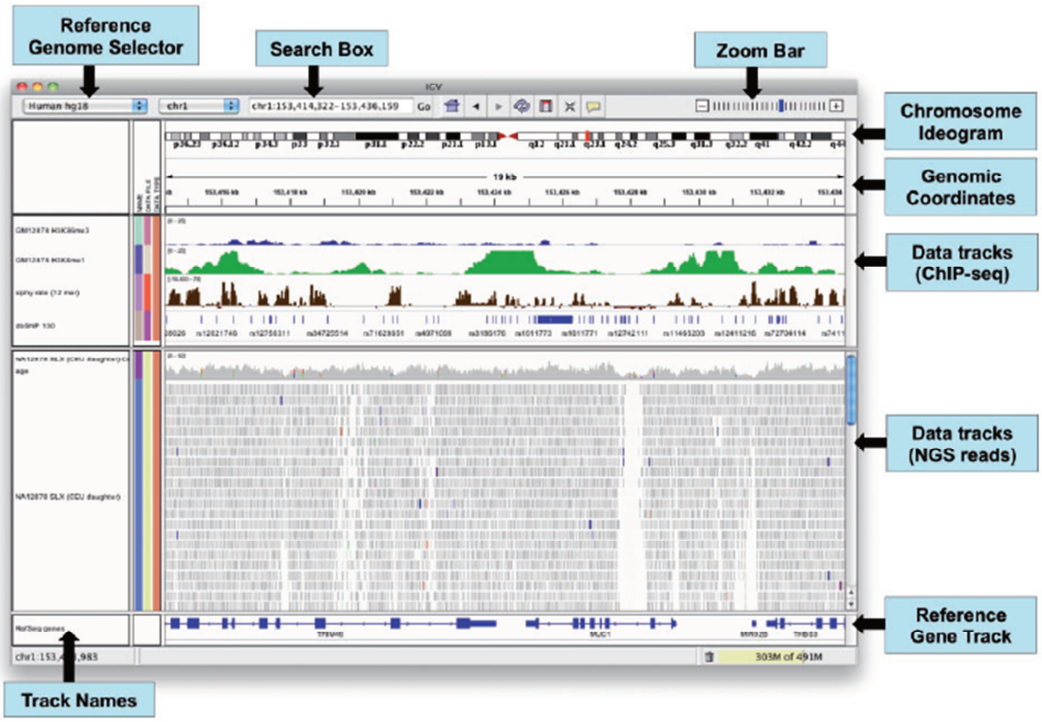


Figure 3.8. IGV application window. From Thorvaldsdóttir et al., 2013.

***SNP-shot* tool** (http://www.bmr-genomics.it/SNP-Shot)

It is an on-line tool, created by BMR Genomics to facilitate the analysis of the next generation sequencing data. *SNP-shot* integrates several information such as the genomic position of a variant, the resulting effect on the protein, the allele frequencies (1000 Genomes, EVS, ExAC), the predicted mutation effects (PolyPhen2, SIFT) and, when possible, the association with a particular phenotype. It considers all the mutations previously reported in a Locus Specific Database (LSDB) such as HGMD, ClinVar, ARVD/C Genetic Variants Database and LOVD. It allows to select and analyze all the variants in a subgroup of genes and make a comparison with two samples.

Variation public databases

1000 GENOMES (http://www.1000genomes.org)

1000 Genomes Project started in 2008 and was the first project to sequence the genomes of a large number of people, to provide genetic variants present in the population studied. The genomes of about 2500 unidentified people from about 25 populations around the world have been sequenced using NGS technologies. Data obtained are available to the worldwide scientific community through freely accessible public databases.

dbSNP (Single Nucleotide Polymorphism Database)
(http://www.ncbi.nlm.nih.gov/projects/SNP)
The Single Nucleotide Polymorphism Database (dbSNP) is a free public archive for genetic variations within and across different species developed and hosted by the NCBI. Other than single nucleotide polymorphisms (SNPs), it contains a range of molecular variations such as short deletion and insertion polymorphisms (InDels). In February 2010, dbSNP contained more than 184 million submissions representing more than 64 million distinct variants for 55 organisms. Each submitted variation can be identified with an unique identification number and a reference SNP ID number ("rs#"), that gives the basic information.

EVS (Exome Variant Server) (http://evs.gs.washington.edu/EVS)
The NHLBI Exome Sequencing Project (ESP) is focused on understanding the contribution of rare genetic variation to heart, lung and blood disorders through the sequencing of well-phenotyped populations. Variant count data are available on the Exome Variant Server (EVS), which currently contains NGS data on 6503 individuals, and allele frequencies are provided for African-Americans and European-Americans.

ExAC (Exome Aggregation Consortium) (http://exac.broadinstitute.org)
The data set provided by the Exome Aggregation Consortium (ExAC) span 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. All the sequencing data divided in specific population groups (Latino, European, European (Finnish), East Asian, South Asian, African and 'others') are available and freely downloaded for the benefit of the wider biomedical community.

ARVD/C Genetic Variants Database (http://www.arvcdatabase.info)
It contains information from clinical research and other types of data on variants in genes causing Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy (ARVD/C). When the project started, in 2009, it contained 481 variants in eight genes. A total of 144 variants were considered pathogenic, 73 were unknown/unclassified, and 264 had no known pathogenicity (van der Zwaag et al., 2009). The database was converted into the Leiden Open Variation Database (LOVD) format, a gene-centered collection of DNA variations.

Genetic diseases databases

OMIM (Online Mendelian Inheritance in Man) (http://www.ncbi.nlm.nih.gov/OMIM)
It is an online catalog of Human Genes and Genetic Disorders, focusing on the relationship between phenotype and genotype. It is freely available and daily

updated and contains information on all known mendelian disorders and over 15,000 genes. This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. In 1995, it was developed for the World Wide Web by NCBI, the National Center for Biotechnology Information.

ClinVar (http://www.ncbi.nlm.nih.gov/clinvar)
The National Center for Biotechnology Information has also a public archive of reports of the relationships among human variations and phenotypes, named ClinVar. ClinVar is freely accessible and processes submissions reporting variants found in patient samples, assertions made regarding their clinical significance, information about the submitter, and other supporting data. A major goal is to support computational (re)evaluation, both of genotypes and assertions, and to enable the ongoing evolution and development of knowledge regarding variations and associated phenotypes. Novel variants submitted to ClinVar are in turn submitted to dbSNP or dbVar, as appropriate, for accessioning.

HGMD (The Human Gene Mutation Database)
(http://www.hgmd.cf.ac.uk/ac/index.php)
The Human Gene Mutation Database (HGMD) represents an attempt to collate known (published) gene variants responsible for human inherited disease. Other than various types of mutations associated with human inherited disease, HGMD contains disease-associated/functional polymorphisms reported in the literature. It may also include additional reports for certain mutations if these reports serve to enhance the original entry (e.g. functional studies). There are two versions of the database: a public freely-available and less up-to date HGMD and a weekly-update version (HGMD Professional), for which a license is required.

Mutation prediction tools

PolyPhen2 (Polymorphism Phenotyping v2)
(http://genetics.bwh.harvard.edu/pph2)
It's a software tool which predicts a possible impact of amino acid substitutions on the structure and function of human proteins using straightforward physical and evolutionary comparative considerations. PolyPhen2 makes functional predictions based on a number of features comprising the sequence, phylogenic and structural information characterizing the substitution. This analysis classifies mutations as benign and possibly or probably damaging.

SIFT (Sorting Intolerant From Tolerant) (http://sift.jcvi.org)

SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and whether an amino acid substitution affects protein function. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST. SIFT can be applied to naturally occurring nonsynonymous polymorphisms or laboratory-induced missense mutations.

Mutation Taster (http://www.mutationtaster.org)

It is a free, web-based application for rapid evaluation of the disease-causing potential of DNA sequence alterations. Mutation Taster integrates information from different biomedical databases and uses established analysis tools such as 1000 Genomes database, ExAC browser and protein prediction tools. Analyses comprise evolutionary conservation, splice-site changes, loss of protein features and changes that might affect the amount of mRNA. Test results are then evaluated by a naive Bayes classifier2, which predicts the disease potential. A typical query is completed in less than 0.3 seconds.

PhyloP (phylogenetic P-values)

PhyloP is a program freely available (Pollard et al., 2010) that compares probability of observed substitutions under hypothesis of neutral evolutionary rate. The scores generated reflect either conservation (positive scores) among 44 vertebrate species or selection (negative scores). The scores generated range from -7 to +7, and a position is considered highly conserved for phyloP scores ≥2.5.

RVIS (Residual Variation Intolerance Score) (http://genic-intolerance.org)

RVIS is a gene-based score intended to help in the interpretation of human sequence data. The intolerance score in its current form is based upon allele frequency data as represented in whole exome sequence data from the NHLBI-ESP6500 data set. The score is designed to rank genes in terms of whether they have more or less common functional genetic variation relative to the genome wide expectation given the amount of apparently neutral variation the gene has. A gene with a positive score has more common functional variation, and a gene with a negative score has less and is referred to as "intolerant".

4. RESULTS

4.1 Targeted gene panels: 'Cardiomyopathies gene panel'

4.1.1 Coverage

Target regions of 19 DNA from ACM probands have been captured with HaloPlex kit (Agilent) and sequenced using 150 bp paired-end reads on a MiSeq sequencer (Illumina). This kit is able to capture target genomes with size <500 kb and consists of 20.000 probes, providing a mean theoretical coverage of 200X. This value does not correspond to the effective coverage and therefore it is necessary to establish a cutoff value, below which the base called in that position is not informative. There is no ideal coverage threshold, but 20X has been adopted by many research groups as a good cutoff point.

Overall, coverage >20X was obtained for 91.95% of 1.416.998 nucleotides analyzed and for 90.93% of the 13 ACM genes. In particular, *DSG2* and *DSC2* ACM genes resulted to be the most poorly covered genes, with 84.07% and 68.08% of target nucleotides covered >20X, respectively (Table 4.1).

ACM genes	% bases >20X
PKP2	90.51
DSP	97.91
DSG2	84.07
DSC2	68.08
CTNNA3	95.16
DES	99.92
JUP	98.18
LMNA	95.63
TMEM43	93.91
PLN	78.27
RYR2	89.14
TTN	92.51
TGFβ3	98.81
Mean for all ACM genes	90.93

Table 4.1. Percentage of bases with coverage >20X in ACM disease genes.

All the uncovered exons of the most mutated desmosomal genes in ACM (*PKP2*, *DSP*, *DSG2*, *DSC2* and *JUP*) were re-sequenced with traditional Sanger method, but only common polymorphisms have been detected in these regions.

4.1.2 Previously detected variants

Among the 19 probands analyzed by using this panel, 10 were previously investigated for mutations in the 'three big ACM genes' (*PKP2*, *DSP* and *DSG2*)

with Sanger sequencing. Three mutations and a rare variant were identified in four different subjects in addition to 23 total polymorphisms (Table 4.2, Table 4.3).

Gene	Nucleotidic variant	Aminoacidic variant	MAF	#proband
Mutations				
PKP2	c.1216delG	p.V406FfsX13	-	3a
DSP	c.3912G>C	p.E1304D	-	6a
PKP2	c.148_151delCAGA	p.T50SfsX60	-	7a
Rare variant				
DSG2	c.1003A>G	p.T335A	G: 0.02%	12a

Table 4.2. Mutations and a rare variant previously identified with Sanger sequencing. *PKP2* c.148_151delCAGA and *DSP* c.1003A>G variants are classified as pathogenic in ARVC/D Genetic Database.

Gene	SNP ID	Nucleotidic variant	Aminoacidic variant	MAF	#proband
Polymorphisms					
DSG2	rs2230233*	c.861T>C	p.N287N	C: 47.34%	3a, 4a, 14a
	rs1042769	c.2505A>G	p.T835T	G: 19.75%	3a, 4a
	rs1791235	c.3321C>T	p.V1107V	T: 45.71%	3a, 4a, 14a
	rs2230234	c.877A>G	p.I293V	G: 3.23%	6a
	rs3737378	c.828+16C>A	-	A: 21.15%	14a
	rs62095194	c.1014+27G>A	-	A: 21.63%	14a
	rs79241126	c.2137G>A	p.E713K	A: 2.6%	14a
	rs2278792	c.2318G>A	p.R773K	A: 24%	14a
PKP2	rs10772008	c.2145+45A>G	-	G: 40.83%	5a, 6a, 14a
	rs77991187	c.2489+73T>A	-	NA	5a
	rs6488091	c.2490-56C>A	-	A: 45.03%	5a
	rs6488092	c.2490-104C>G	-	G: 46.69%	5a
	rs7956824	c.2578-69G>A	-	A: 45.05%	5a
	rs61927769	c.2489+72G>A	-	A: 24.94%	14a
	rs71447623	c.2489+73delT	-	–: 24.94%	14a, 5a
DSP	rs10484326	c.274-31T>C	-	C: 15.62%	6a
	rs2806234	c.741G>T	p.A247A	T: 0%	6a
	rs2076296	c.1141-44T>C	-	C: 22.72%	6a
	rs2076304	c.2091G>A	p.G697G	A: 22.56%	6a
	rs2064217	c.2862C>T	p.C954C	T: 27.52%	6a
	rs2064218	c.2877+79A>C	-	C: 30.61%	6a
	rs2076300	c.7122C>T	p.T2374T	T: 21.17%	6a
	rs2744380	c.8472C>G	p.G2824G	G: 29.05%	6a

Table 4.3. Polymorphisms previously detected with Sanger sequencing. Asterisk indicates the polymorphism not detected with NGS analysis. MAF: Minor Frequency Allele from dbSNP database; NA= data not available.

Targeted NGS approach allowed to identify 96% of the variants previously detected by Sanger: all the 3 mutations, the rare variant and 22 out of 23

polymorphisms. The SNP c.861T>C in *DSG2* gene (rs2230233) has been detected in probands n° 4a and 14a but not in patient n° 3a. Looking at the aligned reads with IGV software, it appeared evident that the region where the SNP is located is poorly covered (coverage=4X) for this sample.

4.1.3 Detection of variants with Targeted 'Cardiomyopathies gene panel'

All the variants detected by at least one variant caller were annotated and analyzed by using *SNP-Shot* tool, which allowed to filter the variants for different parameters, such as the frequency in the population, the genomic region involved, and the type of variant (SNV synonymous, missense, non-sense or InDel). All the missense/non-sense variants with a MAF ≤1% and located both in ACM known-disease genes (with exclusion of *TTN* gene) and in genes associated with other cardiomyopathies were considered.

For each annotated variant the sequencing alignments were analyzed with IGV to determine: i) the real coverage at this position; ii) the balance between the reference and the alternative alleles; iii) the position of the variant in the read. These information are useful to recognize false positives from real variants.

Overall, applying the filters described above, 34 variants were identified, including the 3 mutations and the rare variant previously detected. Among the 30 novel variants, 28 were subsequently confirmed with Sanger sequencing (Table 4.4) resulting in two false positives detected in patient n°1a in *MYH7* gene (c.3382G>A, p.A1128T) and in patient n°14a in gene encoding for *DMD* (c.1633A>G, p.R545G). In both cases the two variants were called only by one variant caller and the alleles resulted unbalanced.

#proband	gene	variant	frequences	predictions
1a	*ANK2*	c.2075T>G p.I692S		PolyPhen2: p.damaging SIFT: deleterious
2a	-	-		
3a	*PKP2*	c.1216delG p.V406FfsX13		
4a	*MYBPC3*	rs373164247 c.148A>G p.S50G	EVS:0.0077% ExAC:0.0038%	PolyPhen2: benign SIFT: tolerated
5a	*JUP*	c.1355C>T p.T452M	ExAC:0.0074%	PolyPhen2: p.damaging SIFT: deleterious
	DSC2	c.577_624del48 p.G193_F208del16		
6a	*DSP*	c.3912G>C p.E1304D		
7a	*PKP2*	c.148_151delCAGA p.T50SfsX60		

#proband	gene	variant	frequences	predictions
8a	*DES*	rs62636490 c.322G>A p.E108K		PolyPhen2: p.damaging SIFT: deleterious
	DSC2	rs61731920 c.2616C>T p.C872C	dbSNP:0.16% EVS:0.23% ExAC:0.06%	
	GPD1L	rs72552293 c.370A>G p.I124V	dbSNP:0.08% EVS:0.1% ExAC:0.2%	
	SCN5A	rs41276525 c.647C>T p.S216L	dbSNP:0.08% EVS:0.09% ExAC:0.1%	
9a	*ABCC9*	c.1274T>C p.M425T	ExAC:0.0008%	PolyPhen2: possibly damaging SIFT: deleterious
	SCN1B	c.794G>A p.R265Q		PolyPhen2: - SIFT:deleterious
	ANK2	rs66785829 c.4619T>A p.V1540D	dbSNP:0.2% EVS:0.23% ExAC:0.27%	
10a	*KCND3*	rs142744204 c.641A>G p.K214R	dbSNP:0.02% EVS:0.03% ExAC:0.023%	
11a	*PKP2*	c.2043delT p.I681MfsX3		
	VLC	rs146278697 c.1237G>A p.A413T	dbSNP:0.02% EVS:0.023% ExAC:0.024%	
	CACNA2D1	rs35131433 c.3134A>C p.D1045A	dbSNP:0.18% EVS:0.5% ExAC:0.28%	
12a	*DSG2*	rs191564916 c.1003A>G p.T335A	dbSNP:0.02% EVS:0. 4% ExAC:0.055%	
13a	*TNNI3*	c.232C>T p.T78S		PolyPhen2: benign SIFT: deleterious
	DSP	rs121912998 c.88G>A p.V30M	dbSNP:0.22% EVS:0.12% ExAC:0.3%	
14a	*DSP*	c.944G>C p.R315P		PolyPhen2: p.damaging SIFT: deleterious
	JUP	rs143043662 c.1942G>A p.V648I	dbSNP:0.26% EVS:0.67% ExAC:0.7%	
15a	*PKP2*	c.1643delG p.G548VfsX15	-	
	DES	c.1282T>G p.F428V	ExAc:0.0008%	PolyPhen2: benign SIFT: tolerated
16a	*TMEM43*	rs145619906 c.424G>A p.E142K	dbSNP:0.08% EVS:0.13% ExAC:0.048%	
17a	*MYL2*	c.34G>T p.G12C	ExAC:0.0008%	PolyPhen2:unknown SIFT:tolerated

#proband	gene	variant	frequences	predictions
18a	*LMNA*	c.1835C>T p.S612F		PolyPhen2:probably damaging SIFT:deleterious
	PSEN2	rs143501870 c.149A>G p.Q50R	EVS:0.0077%	PolyPhen2: benign SIFT: tolerated
	TPM1	rs397516389 c.775G>A p.E259K		PolyPhen2: benign SIFT: tolerated
19a	*PKP2*	c.39C>G p.I13M		PolyPhen2: p.damaging SIFT: deleterious
	DSP	c.1067C>T p.T356M		PolyPhen2: p.damaging SIFT:tolerated

Table 4.4. Variants detected with Targeted NGS and confirmed with Sanger sequencing. The *SNP-Shot* annotation includes frequency data from dbSNP, Exome Variant Server (EVS) and ExAC Browser and pathogenicity predictions from PolyPhen2 and SIFT.

Among the 32 variants reported, 5 were deletions, 4 of which affecting *PKP2* gene and a homozygous deletion of 48 bp in *DSC2* gene identified in patient n° 8a (Figure 4.1).

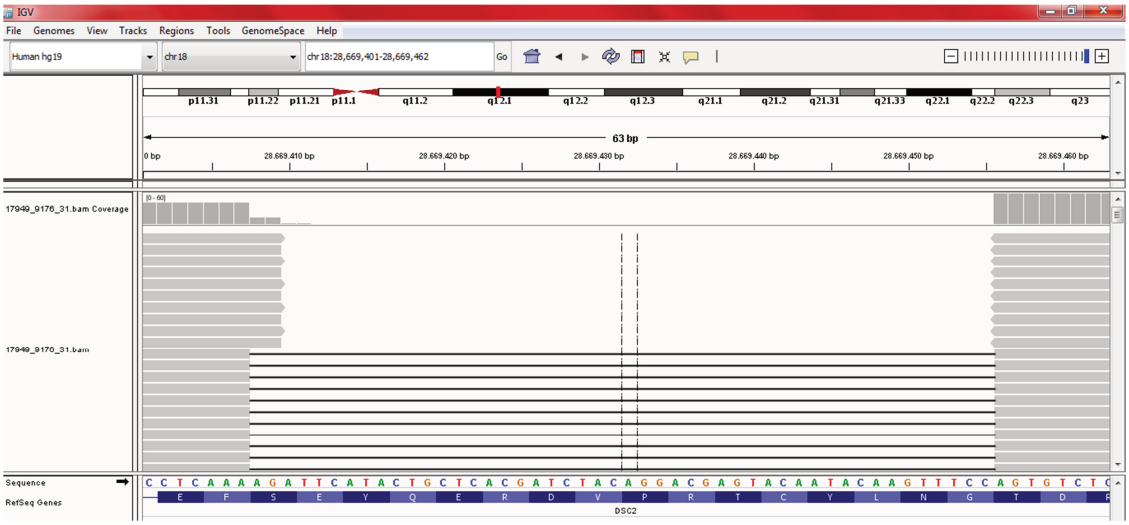


Figure 4.1 . IGV view: sequencing alignments of DSC exon 5 for patient n°8.

Taking into account that the allele frequency of ACM in the general population span from 0.01% and 0.025%, a genetic variant should be considered as a 'mutation' if its minor allele frequency (MAF) is ≤ 0.01%, as a 'rare variant' if reported with 0.01%≤ MAF≤ 0.5% and as a 'low frequency variant' with 0.5%≤ MAF≤ 1%.

Considering this classification, 15 out of 19 probands analysed reported one or more mutations in a known-ACM gene (patients n° 3a, 5a, 6a, 7a, 8a, 11a, 14a, 15a, 18a, 19a) or in a gene associated with other cardiomyopathies (patients n°

1a, 4a, 9a, 13a, 17a). Several rare or low frequency variants were also detected in all the samples (Table 4.4). Among these, the rare variant in *DSP* gene (p.V30M) found in patient n° 13a and the missense variant in *GPD1L* gene (p.I124V) in patient n° 8a are reported in ClinVar database and associated with ACM and Brugada syndrome, respectively. Moreover, the missense variant in *ANK2* gene (p.V1540D) identified in patient n° 9a has been associated with cardiac arrhythmia in HGMD database.
In some familial cases, described below, it was possible to study the segregation of the detected variants within the families.

Proband #14a

The patient resulted to carry 2 variants in 2 ACM genes: a novel missense variant in *DSP* gene (p.R315P), predicted damaging by the two prediction tools, and a low frequency variant (ExAC MAF: 0.7%) in *JUP* gene (p.V648I) classified as "unknown variant" in the ARVD/C Genetic Variants database. The presence of both variants was assessed in subjects II-2, III-4 and III-5 (Figure 4.2).

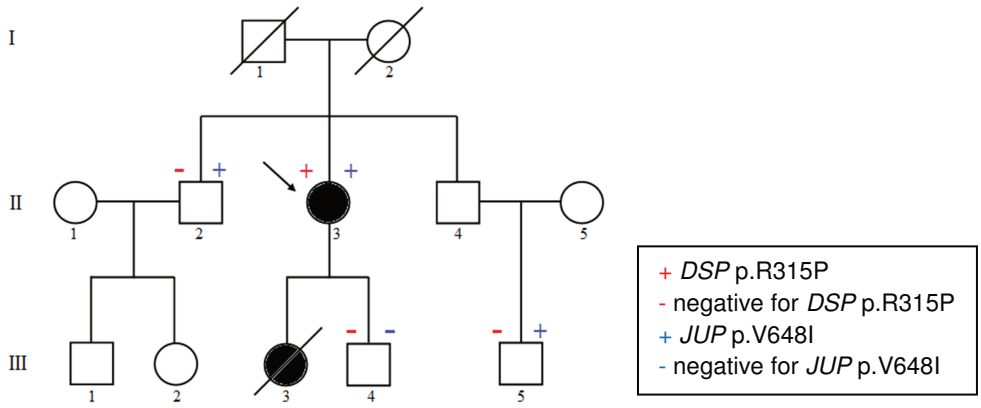


Figure 4.2. Family's pedigree of patient n°14a (pointed with the arrow).

None of the healthy family members present the *DSP* mutation, whereas two of them (II-2, III-5) reported the variant in *JUP* gene.

Proband #15a

The patient reported a frameshift *PKP2* mutation (p.G548VfsX15) and a missense mutation in *DES* gene (p.F428V), predicted to be tolerated by SIFT and PolyPhen2. Genomic DNA of 4 family members (III-2, III-3, III-4, III-5) was available and analyzed for the two variants identified in the proband (II-3) (Figure 4.3).

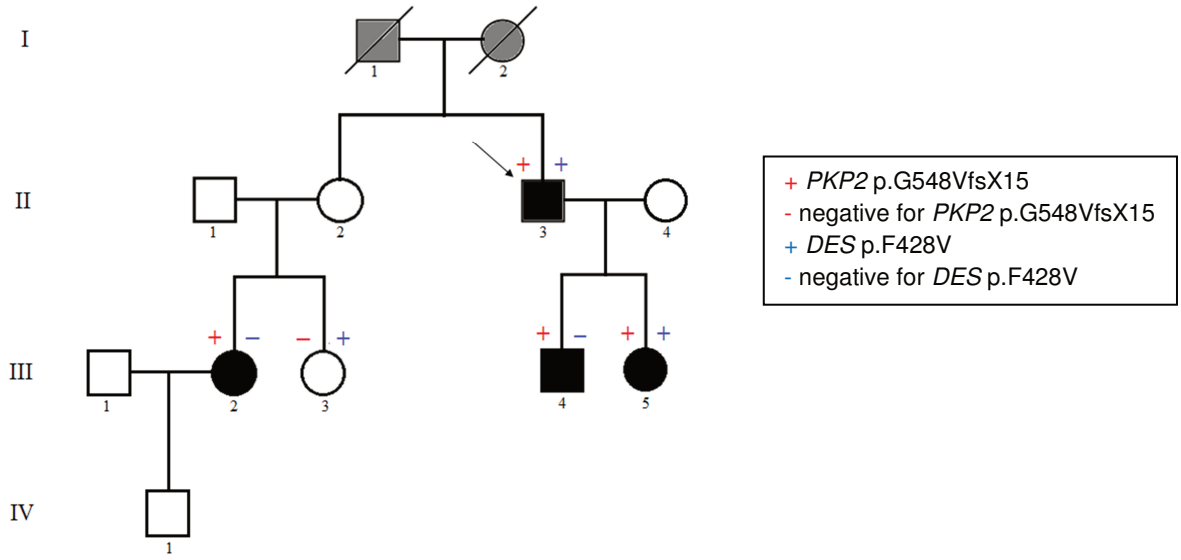


Figure 4.3. Family's pedigree of patient n°15a (pointed with the arrow).

All the affected family members resulted positive for *PKP2* frameshift mutation, which is not present in the DNA of the healthy subject III-3. The missense *DES* variant is present in 2 out of 4 individuals (III-3 and III-5) beyond the proband.

4.1.4 *TTN* variants

First analysis of the sequencing data excluded all the variants found in *TTN* gene, because of the size of this gene and the difficulty to interpret the frequent missense variants identified also in individuals from control populations. However, among the huge amount of *TTN* variants identified in our cohort of patients, all the variants called in 10 or more samples were ignored as possible sequencing errors, whereas the others were divided in two groups: the 'private variants', present in only one proband, and the 'common variants' present at least in 2 samples. Excluding the synonymous *TTN* variants, 32 'private variants' (Table 1, Appendix C) have been found, resulting in a mean of 1.68 *TTN* missense variants for each subject. The only one mutation reported at least in one database is the c.30274C>T (p.H10092Y) variant (rs72650011) identified in proband n° 15a, which is present both in ClinVar and HGMD databases and classified as 'pathogenic' in ARVD/C Genetic Variants database. Moreover, other 8 missense variants present in more probands have been found (Table 4.5). For all these variants the frequency data in the general population is not reported in the databases.

TTN variant	frequences	predictions	#probands
c.85136C>A p.A28379E	-	PolyPhen2: p.damaging SIFT: deleterious	2a, 3a
c.47860G>T p.A15954S	-	PolyPhen2: p.damaging SIFT: deleterious	3a, 5a, 6a, 8a, 12a, 14a, 16a
c.84562G>A p.G28188S	-	PolyPhen2: p.damaging SIFT: deleterious	8a, 9a, 12a
rs55634791 c.82691C>T p.A27564V	-	PolyPhen2: possibly damaging SIFT: deleterious	9a, 11a, 16a, 18a
rs201675016 c.53062A>G p.K17688E	-	PolyPhen2: benign SIFT: deleterious	9a, 11a
c.10831C>A p.Q3611K	-	PolyPhen2: benign SIFT: deleterious	9a, 10a, 11a, 12a, 16a, 17a, 18a, 19a
c.55784C>T p.T18595I	-	PolyPhen2: p.damaging SIFT: deleterious	10a, 19a
c.49814T>G p.V16605G	-	PolyPhen2: p.damaging SIFT: deleterious	14a, 15a, 16a

Table 4.5. *TTN* 'common variants' detected in 19 ACM probands.

4.2 Targeted gene panels: 'ACM known and candidate genes panel'

4.2.1 Coverage

Four subjects previously analyzed with the 'Cardiomyopathies gene panel' (2a, 4a, 16a, 17a) together with other 40 ACM probands were screened using a panel that includes a total of 69 genes, 13 known-disease genes and 56 ACM candidate genes. All the patients included in this study were diagnosed affected based on the Revised Task Force Criteria (Marcus et al., 2010) and resulted negative for mutation screening of the 'three big ACM genes' (*PKP2*, *DSP* and *DSG2*).

Overall, coverage >20X was obtained for 92.45% of sequenced nucleotides and for 91% considering only the 13 ACM genes (Table 4.6). On average, 445 variants per sample have been detected, of them 409 are SNPs and 36 InDels.

ACM genes	% bases >20X
PKP2	89.1
DSP	96.8
DSG2	90.6
DSC2	84.3
CTNNA3	95.2
DES	97.7
JUP	96.8
LMNA	95.3
TMEM43	94.2
PLN	76.6
RYR2	89.5
TTN	82.6
TGFβ3	94.8
Mean for all ACM genes	91.0

Table 4.6. Percentage of bases with coverage >20X in ACM disease genes.

4.2.2 Detection of variants with Targeted 'ACM known and candidate genes panel'

After the variant calling process performed with FreeBayes and VarScan, the annotation step was carried out by using *SNP-Shot* tool, as previously described. *SNP-Shot* interface reported also a specific parameter called 'class', indicating the presence and the classification of the variant in mutation databases (ClinVar, HGMD). All the missense/non-sense variants with a MAF ≤1% were considered, located both in ACM known or candidate genes (with exclusion of *TTN* gene).
The number of probands carrying mutations (MAF ≤0.01%) in known or candidate genes is listed in Table 4.7.

		TOT n° of probands	N° probands with mutations in candidate genes
N° probands carrying mutations in known-disease genes	In ACM genes	6 (13.6%)	3 (50%) out of 6
	In other cardiomyopathies genes	6 (13.6%)	2 (33.3%) out of 6
	none	32 (72.7%)	13 (40.6%) out of 32
	TOT	44	18 (40.9%) out of 44

Table 4.7. N° of probands in whom mutations have been identified in known disease genes divided in ACM known genes and in genes associated with other cardiomyopathies (on the left) and in candidate genes (column on the right).

a. Mutations in ACM genes

In 6 out of 44 probands (Table 4.7), one or more mutations (MAF ≤0.01%) in ACM-known disease genes have been identified and confirmed with Sanger sequencing (Table 4.8).

#proband	gene	variant	frequences	predictions	class
6b	***DSP***	**rs121912997 c.3799C>T p.R1267***			I
	PKP2	**rs146102241 c.1759G>A p.V587I**		PolyPhen2: p.damaging SIFT: tolerated	I
	MYZAP	rs141271435 c.209G>C p.G70A	ExAC: 0.008%	PolyPhen2: p.damaging SIFT: tolerated	
	FAT1	rs180820128 c.2563G>A p.G855R	EVS:0.2% ExAC:0.2%		
	XIRP1	rs199632343 c.2066C>T p.P689L	ExAC: 0.04%		
12b	***DSP***	**c.1267-2A>G**			
	TJP1	rs2229517 c.1412A>G p.N471S	dbSNP:0.7% EVS:1% ExAC:0.9%		
	KCNJ11	c.820G>A p.D274N	ExAC: 0.006%	PolyPhen2: benign SIFT: deleterious	
	FAT1	rs116784674 c.7700G>A p.R2567H	dbSNP:0.5% EVS:1% ExAC:0.9%		
14b	***PKP2***	**c.1063C>T p.R355***			I
25b	***DSG2***	**c.1557C>A p.D519E**		PolyPhen2: p.damaging SIFT: deleterious	
	DSG2	rs2230232 c.266A>G p.Y89C	dbSNP:0.2% ExAC:0.1%		IX
	DSG2	rs191143292 c.473T>G p.V158G	dbSNP:0.2% ExAC:0.5%		III
27b	***JUP***	**c.1355C>T p.T452M**	ExAC:0.007%	PolyPhen2: p.damaging SIFT: deleterious	
	DSP	rs41302885 c.6208G>A p.D2070N	dbSNP:0.2% EVS:0.3% ExAC:0.4%		IX
33b	***JUP***	**c.1359G>T p.E453D**	ExAC:0.006%	PolyPhen2: p.damaging SIFT: deleterious	IV
	FAT1	c.8938G>T p.V2980L	ExAc:0.0008%	PolyPhen2: benign SIFT:deleterious	
	AXIN1	rs146947903 c.1485C>G p.D495E	dbSNP:0.4% EVS:0.6% ExAC:0.6%		
	DCHS1	rs117368891 c.1245C>A p.S415R	dbSNP:0.6% EVS:0.8% ExAC:0.6%		

Table 4.8. Probands reporting one or more mutations in ACM-known disease genes. Class I: Pathogenic, variants reported in LSDB (Locus Specific Databases) and/or HGMD professional as pathogenic for a given phenotype; class III: Uncertain Pathogenicity, variants reported in LSDB and for which the evidence for pathogenicity for a given phenotype may be dubious; class IV: Unclassified, variants reported in LSDB as unclassified; class IX: Not Pathogenic; variants reported in LSDB as likely not pathogenic. In bold type are the mutations in the known-disease genes.

For patients n° 6b, 12b, 14b and 25b the mutation screening previously performed with dHPLC or Sanger sequencing failed to identify the stop-mutations in *DSP* gene (p.Arg1267*) and *PKP2* gene (p.Arg355*), the missense mutation in *DSG2* gene (p.Asp519Glu) and the variant affecting a *DSP* splice-site (c.1267-2A>G).

b. Mutations in genes not previously associated with ACM

As observed in the 19 probands screened with the 'Cardiomyopathies gene panel', also in this cohort of ACM patients there is a fraction of subjects (Table 4.7) carrying one or more mutations in genes associated with other cardiomyopathies (Table 4.9). All the putative pathogenic mutations have been subsequently validated by Sanger sequencing.

#proband	gene	variant	frequences	predictions	class
4b	***LMNA***	**rs267607571 c.569G>A p.R190Q**		PolyPhen2: p.damaging SIFT: deleterious	I/IV
	XIRP1	rs146284335 c.5352G>A p.M1784I	dbSNP:0.5% EVS:1% ExAC:0.7%		
11b	***SCN5A***	**c.3763C>A p.L1255M**	ExAC: 0.0008%	PolyPhen2: p.damaging SIFT: deleterious	
	PKP4	rs146222858 c.2764G>A p.G923S	ExAC:0.1%	PolyPhen2: possibly damaging SIFT: tolerated	
	FAT1	rs77280820 c.3446T>C p.M1149T	dbSNP:0.3% EVS:0.5% ExAC:0.3%		
	CTNNA1	rs28363394 c.536C>T p.A179V	EVS:0.2% ExAC:0.1%		
	TJP2	rs140340673 c.985G>A p.V329I	ExAC: 0.008%	PolyPhen2: benign SIFT: tolerated	
15b	***TNNC1***	**rs267607124 c.435C>A p.D145E**	ExAC: 0.02%	PolyPhen2: possibly damaging SIFT: deleterious	I
32b	***TNNC1***	**c.208G>A p.G70S**	ExAC:0.0016%	PolyPhen2: p.damaging SIFT: deleterious	
	XIRP1	rs146284335 c.5352G>A p.M1784I	dbSNP:0.5% EVS:1% ExAC:0.7%		
36b	***TPM1***	**c.803A>G p.L268R**		PolyPhen2: benign SIFT: tolerated	
	DCHS1	rs377088444 c.1669C>T p.P557S	ExAC:0.0024%	PolyPhen2: benign SIFT: tolerated	
	TJP1	rs201943761 c.3505G>T p.G1169C	EVS:0.1% ExAC:0.1%		
	FAT1	c.13621G>A p.A4541T		PolyPhen2: benign SIFT: tolerated	

#proband	gene	variant	frequences	predictions	class
44b(17a)	***MYL2***	**c.34G>T p.G12C**	ExAC:0.0008%	PolyPhen2:unknown SIFT:tolerated	

Table 4.9. Probands reporting one or more mutations in genes not previously associated with ACM. Class I: Pathogenic, variants reported in LSDB and/or HGMD professional as pathogenic for a given phenotype; class IV: Unclassified, variants reported in LSDB as unclassified. In bold type are the mutations in genes associated with other diseases.

Despite the fact that *LMNA* is a gene already associated with ACM, in the literature the mutation p.Arg190Gln has been reported in a patient affected with Emery-Dreifuss syndrome (ClinVar). Two probands (15b and 32b) showed missense variants in *TNNC1* gene, and one of these (p.Asp145Glu) has been already identified both in patients affected with HCM and DCM. Patient n° 44b was previously screened with the 'Cardiomyopathies gene panel' (proband n° 17a), but in both cases only p.Gly12Cys mutation in *MYL2*, a gene associated with HCM, has been identified.

c. Mutations in candidate ACM genes

The application of this Targeted gene panel allowed to identify missense and non-sense mutations in candidate genes in 40.6% of the index cases negative for mutations in known-disease genes (Table 4.7). Among the 56 candidate genes included in the panel, 13 reported very rare coding variants (MAF ≤0.01%) and in 4 of them more than one different missense mutation was found (Table 4.10).

#proband	gene	variant	frequences	predictions	class
8b	*TP63*	c.796C>T p.R266*			
	TLN1	c.3707A>G p.Q1236R		PolyPhen2: possibly damaging SIFT: tolerated	
16b	*FAT1*	c.2755G>A p.D919N		PolyPhen2: p.damaging SIFT: deleterious	
17b	*AXIN*	rs140190126 c.1532C>T p.S511L	ExAC:0.007%	PolyPhen2: benign SIFT: tolerated	
19b	*XIRP1*	c.223G>A p.E75K	ExAC: 0.016%		
	CASQ2	c.1147_1149delGAT p.D383del			
24b	*DCHS1*	c.6803T>C p.I2268T	ExAC:0.0016%	PolyPhen2: benign SIFT: tolerated	
26b(2a)	*DCHS1*	c.259G>A p.G87S	ExAC: 0.0008%	PolyPhen2: p.damaging SIFT: tolerated	
30b	*DCHS1*	c.3206T>A p.I1069K	ExAc:0.006%	PolyPhen2: benign SIFT: tolerated	
31b	*AXIN1*	rs140190126 c.1532C>T p.S511L	ExAC:0.007%	PolyPhen2: benign SIFT: tolerated	
35b	*ANK3*	c.2309C>A p.T770K			
37b	*TJP1*	c.793C>T p.R265W	ExAC:0.0008%	PolyPhen2: p.damaging SIFT: deleterious	

#proband	gene	variant	frequences	predictions	class
38b	*KAZN*	c.2242G>T p.E748*			
40b	*CDH2*	c.1478A>G p.E493G		PolyPhen2: p.damaging SIFT: deleterious	
	CDH2	c.1472T>G p.V491G		PolyPhen2: possibly damaging SIFT: deleterious	
	FAT1	rs201352448 c.7138G>A p.V2380I	ExAC:0.003%	PolyPhen2: benign SIFT: tolerated	
41b	*TJP1*	c.2006A>G p.T669C		PolyPhen2: p.damaging	

Table 4.10. Probands reporting one or more mutations in candidate ACM genes.

Other than the patients reported in Table 4.10, who carried only mutations in one or more candidates genes, other probands resulted positive for mutations in these genes together with mutations in ACM known-disease genes (proband n° 6b, 12b and 33b, Table 4.8), or in genes associated with other cardiomyopathies (proband n° 11b and n° 36b, Table 4.9) (Table 4.7).

The most mutated genes resulted *FAT1* and *DCHS1*, both encoding for members of the cadherin superfamily. Mutations in *FAT1* gene were found in 4 subjects (16b, 33b, 36b and 40b), whereas 4 probands showed mutations in *DCHS1* gene (24b, 26b, 30b and 36b). In most of the cases, these missense mutations were predicted 'tolerated' from prediction tools. Moreover many rare or low-frequency variants in these two genes were also found in our cohort of patients.

Among these 13 candidate mutated genes, 2 showed single nucleotide variants which cause the formation of a stop codon. The c.2242G>T (p.E748*) mutation located in gene encoding for kazrin (*KAZN*, proband n° 38b), resulted in a stop-gain variant affecting only one out of 5 isoforms considered for this gene (NM_201628), whereas the same variant is located in an intergenic region for other isoforms. The other stop mutation identified in *TP63* gene (p.R266*) in proband n° 8b will be described below (Family #1).

CDH2 gene

Proband n° 40b showed 2 single nucleotide variants, one 6 bp far from the other in exon 10 of *CDH2* gene (c.1478A>G p.E493G, c.1472T>G p.V491G). Both mutations were validated with Sanger sequencing, which highlighted also the presence of a homozygous SNP (c.1377C>G, rs1628684) with a MAF= 0.18% for the G allele.

CDH2 gene (NM_001792) encodes for a calcium dependent cell-cell adhesion glycoprotein. It belongs to the cadherin superfamily and is composed by five extracellular cadherin repeats, a transmembrane region and a highly conserved cytoplasmic tail. N-cadherin is highly expressed by the developing and mature myocardium, where it is found predominantly in the fascia adherens of the transverse region of intercalated discs (IDs) and in the regions of close lateral

contact between neighboring cardiomyocytes. Both the variants are absent in all SNP databases and affected two highly conserved amino acids (phyloP=4.9 for p.E493G mutation, phyloP=3.3 for p.V491G mutation), localized into the extracellular protein domain. Moreover, both the amino acid substitutions have been predicted to be deleterious by SIFT tool and probably or possibly damaging by PolyPhen2 tool.

In two familial cases the analysis was extended to the available family members.

Family #1

Proband n°8b showed first signs of the disease at the age of 14. He reported infundibular arrhythmias and late potentials. The proband's mother was also diagnosed affected with ACM. The proband underwent to traditional screening for the 5 most mutated genes in ACM, but no putative causative mutation was found.
Analysis of the data obtained by Targeted gene panel highlighted the presence of two novel mutations: a missense variant in *TLN1* gene (c.3707A>G, p.Q1236R) and a stop mutation in *TP63* gene (c.796C>T, p.R266*). The latter was called only by FreeBayes variant caller and not by VarScan, but Sanger sequencing allowed to confirm both variants.
TP63 stop mutation was carried also by the mother (II-2), but was not found in subject (II-3) for whom clinical data are not available. Conversely, the missense mutation in *TLN1* gene was detected only in the proband, but not in subjects II-2 and II-3 (Figure 4.4).

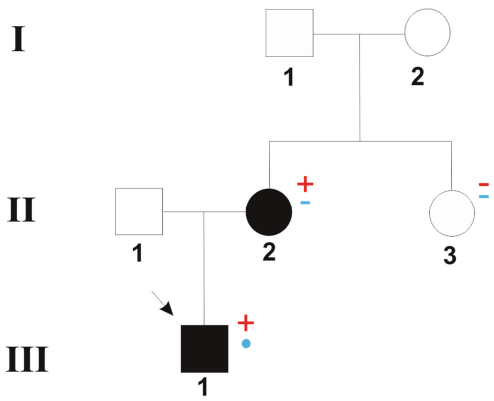


Figure 4.4. Family's pedigree of patient n° 8b (pointed with the arrow). Black and white symbols represent clinically affected and unaffected individuals, respectively.

\+ *TP63* c.796C>T p.R266*
\- negative for *TP63* c.796C>T p.R266*
• *TLN1* c.3707A>G p.Q1236R
\- negative for *TLN1* c.3707A>G p.Q1236R

TP63 gene

This gene encodes a member of the p53 family of transcription factors and acts as a sequence specific DNA binding transcriptional activator or repressor. For this gene several isoforms have been identified. Different isoforms contain a varying set of transactivation and auto-regulating transactivation inhibiting domains thus showing an isoform specific activity. c.796C>T mutation causes the formation of a stop codon located in exon 4 (isoform 1) or exon 6 (isoform 6) encoding a beta strand domain. The gene is widely expressed, notably in heart, kidney, placenta, prostate, skeletal muscle, testis and thymus.

Family #2

Proband n° 41b, came to clinical evaluation for ACM after his brother's sudden death at the age of 35 (II-2, Figure 4.5). Subject II-2 presented negative T waves at ECG and experienced 3 syncopes before death. Proband (II-4) showed a moderate form of ACM, with RV dilatation and initial biventricular involvement. Previous mutation screening of the 5 most mutated genes (*PKP2*, *DSP*, *DSG2*, *DSC2*, *JUP*) didn't reveal the presence of any pathogenic mutation. Genetic analysis through Targeted gene panel identified a novel missense mutation in *TJP1* gene (c.2006A>G, p.Tyr669Cys). Since the DNA of 8 family members was available (I-1, I-2, II-5, II-8, II-9, III-1, III-2, and III-4), the analysis was extended to the whole family (Figure 4.5).

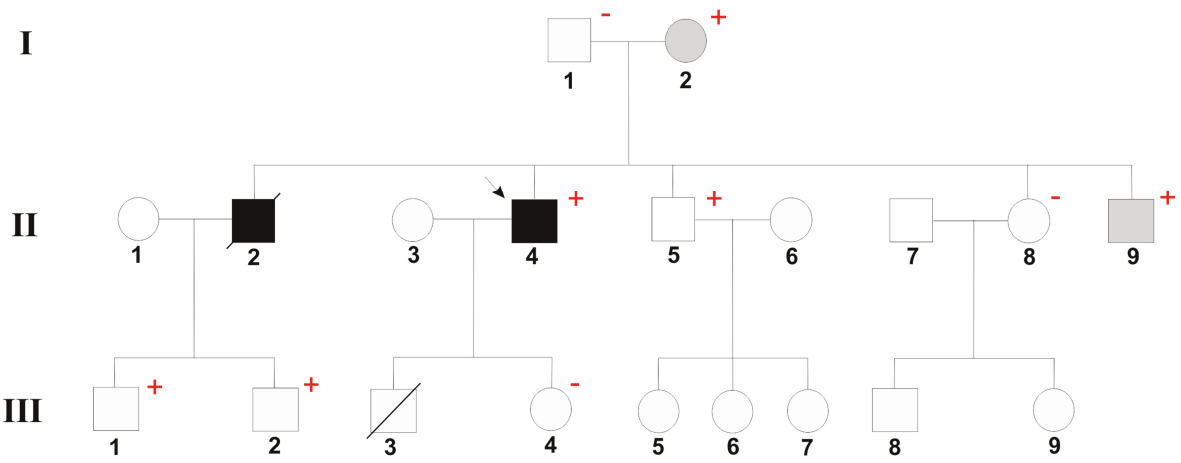


Figure 4.5. Family's pedigree of patients n° 41b (pointed with the arrow). Black and white symbols represent clinically affected and unaffected individuals, respectively; grey symbols denote subjects with minor signs of the disease.

\+ *TJP1* c.2006A>G p.Y669C
\- negative for *TJP1* c.2006A>G p.Y669C

The variant is present in 5 out of 8 family members including the two sons of the dead proband's brother (III-1, III-2). Two other proband's brothers showed the

variant, one (II-9) reporting minor signs of the disease (late potentials) but negative RMN, and subject II-5 with no clinical signs at last evaluation. The mutation is carried by the mother (I-2), who has minor ACM signs, but reporting a family history for sudden cardiac death. Subject III-3 is dead at young age, but not for cardiac causes.
Since the variant is absent in all databases and affects a good candidate gene, which resulted mutated also in proband n° 37b (p.Arg265Trp), the analysis was carried on studying the effect of the amino acid change into the protein structures.

TJP1 **gene**

TJP1 (NM_003257) encodes for tight junction protein 1 (also called zonula occludens 1, ZO-1), that forms regulated intercellular barriers between both epithelial and endothelial cells. Barrier function can be further subdivided into protective barriers and functional barriers serving purposes such as material transport and maintenance of osmotic balance. TJP1 is located on the cytoplasmic membrane surface precisely at sites of cell-cell contact and interacts with different junctional proteins such as connexin-43, myozap, Ankrd2, other than with other tight junction proteins (TJP2/ZO-2 and TJP3/ZO-3). The N-terminal domain may be involved in transducing a signal required for tight junction assembly, while the C-terminal may have specific properties of tight junctions.
Mutation p.Tyr669Cys affects a highly conserved residue at the guanylate kinase-like (GUK) domain of protein ZO-1 (Figure 4.6). The mutation falls at the contact region between strands β2/β3, the two are members of a large five β-sheet which forms the GUK hydrophobic core. A 40 ns full atom molecular dynamics simulation was performed with Gromacs 4.6 using the CHARMM 27 force field, in order to investigate the structural impact of the p.Tyr669Cys mutation. When superimposed, the wild-type and the mutant proteins show a similar organization for the PDZ3 domain, while the GUK domain resulted to adopt a more opened conformation in the mutant protein (Figure 4.7).
Moreover, network of interacting residues generated with RING around 10 Å from position 669 showed that several hydrophobic interactions were lost during the simulation of the mutant protein. In particular, the Tyr669Cys substitution breaks Van der Waals interactions between Tyr669 and Leu66, Ile789, promoting a clear structural rearrangement.

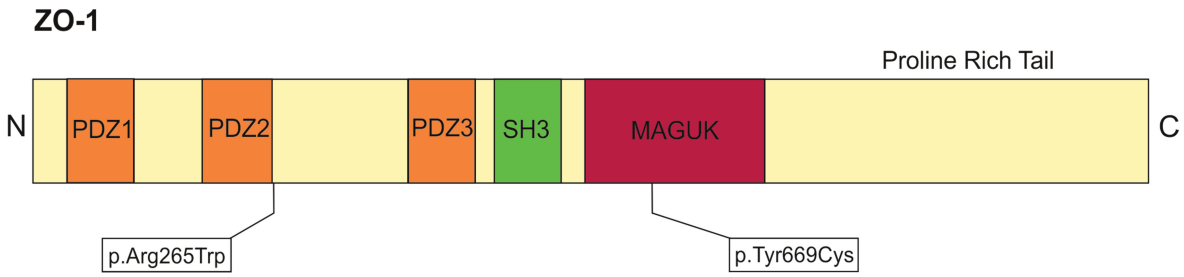


Figure 4.6. Schematic representation of protein ZO-1 and localization of the identified mutations p.Arg265Trp and p.Tyr669Cys.

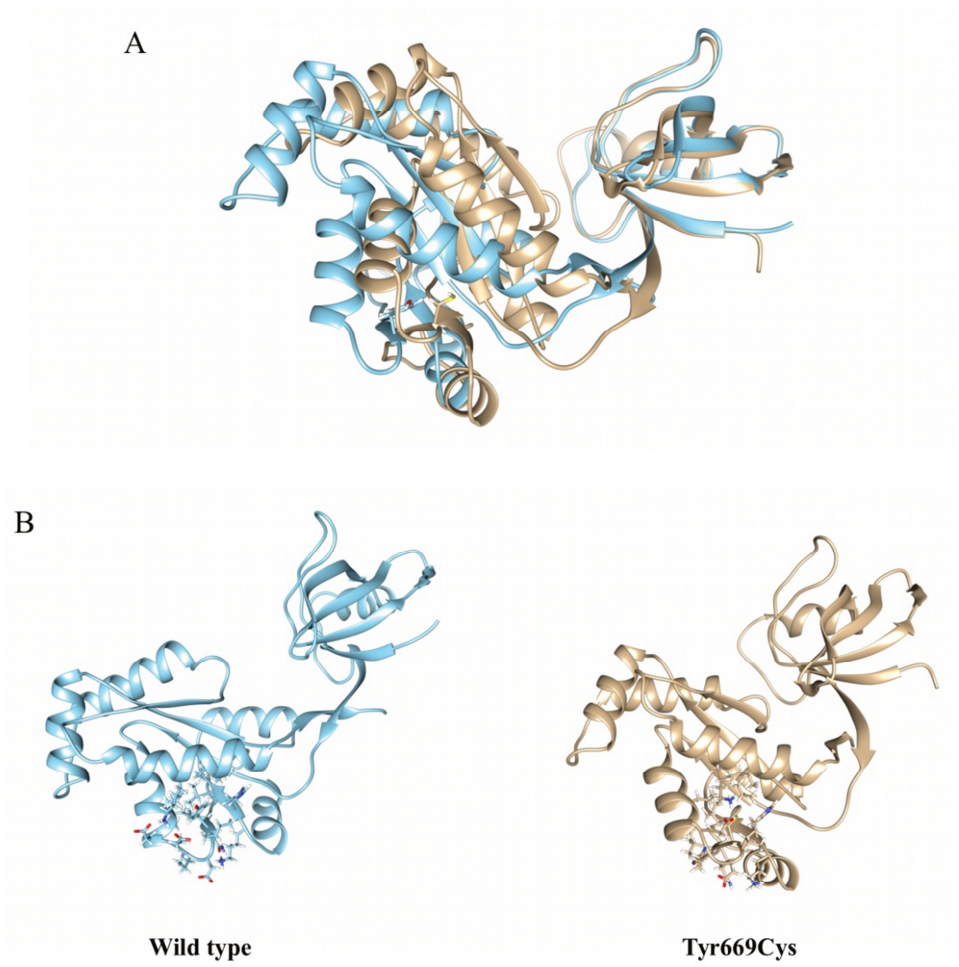


Figure 4.7. Graphical representation of both wild type and mutant protein ZO-1. A) Superimposition of the last frame (40 ns) obtained after two independent runs of molecular dynamics simulation. PDZ3 maintains the same fold, while GUK domain shows relevant variation suggesting a local unfold induced by p.Tyr669Cys substitution. B) Comparative view of the same structures shows a remarkable spatial reorganization of the region surrounding the position 669. Affected residues are presented in stick style. Analysis from Dr. Giovanni Minervini and Prof. Silvio Tosatto.

Mutation p.Arg265Trp (found in proband n° 37b) is localized at the end of the second PDZ domain, in an intrinsically disordered region. To identify the

conservation rate of the amino acid Arg265, 43 orthologous sequences were aligned. Multi sequences alignment revealed that the position is fully conserved, albeit mapping outside the PDZ2 (Figure 4.8). Moreover, a high rate of conservation was observed for the 9 residues immediately after Arg265.

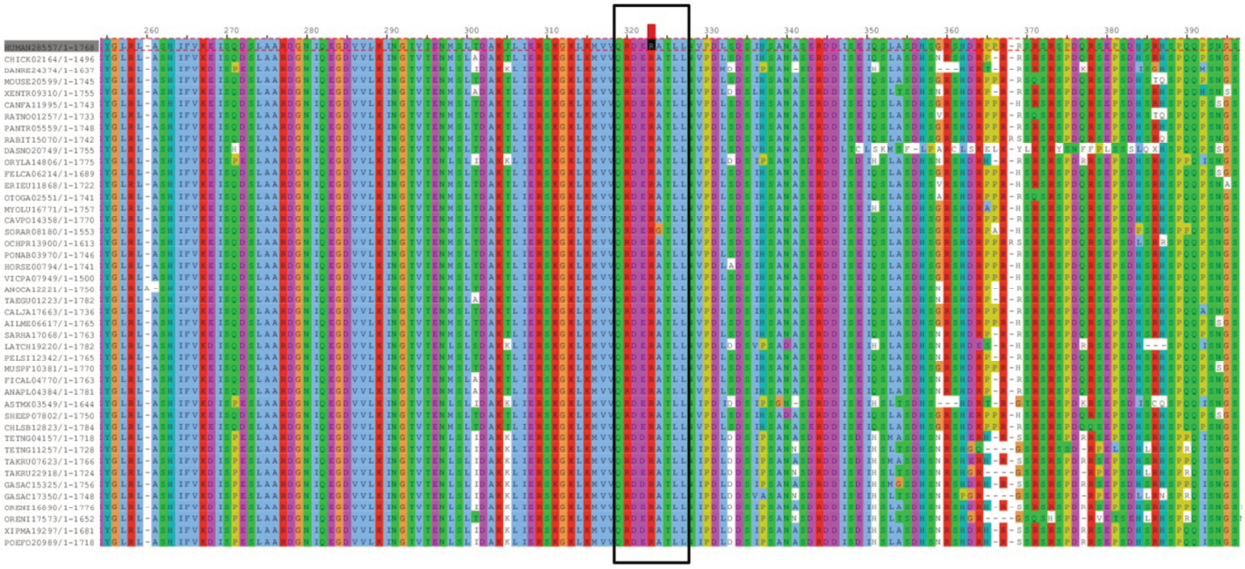


Figure 4.8. Protein sequence alignment of TJP1 PDZ2 domain C-terminal flanking region. The extra boundary conserved region is marked with a black box, while red dot is used to mark residues corresponding to the human Arg265.

a. Rare and low frequency variants

Other than in patients described in the paragraphs above, in additional 17 probands rare or low frequency variants (0.01%≤ MAF≤ 1%) were identified, whereas only in 2 samples (1b and 43b) no interesting variants were found (Table 2, Appendix C).

In our cohort of patients, some of these variants are present in more than one individual, thus resulting in most cases in a MAF ten times higher than the MAF reported by common SNP databases (Table 4.11).

gene	variation	global frequencies	#probands	frequency in 44 subjects
SCN5A	rs45489199 c.6016C>G p.P2006A	EVS:0.1% ExAC:0.1%	2b, 10b, 21b	3.4%
ANK3	rs61845768 c.10055A>G p.E3352G	dbSNP:0.2% EVS:0.4% ExAC:0.4%	3b, 16b, 21b,23b	4.5%
XIRP1	rs146284335 c.5352G>A p.M1784I	dbSNP:0.5% EVS:1% ExAC:0.7%	4b, 32b, 40b, 42b	4.5%
STK3	rs187757501 c.203G>A p.S68N	dbSNP:0.2% EVS:0.4% ExAC:0.3%	5b, 13b	2.3%
ABCC9	rs61688134 c.2200G>A p.V734I	dbSNP:0.6% EVS:0.9% ExAC:0.7%	8b, 16b, 23b	3.4%
TJP1	rs2229517 c.1412A>G p.N471S	dbSNP:0.7% EVS:0.1% ExAC:0.9%	10b, 12b, 21b	3.4%
AXIN1	rs140190126 c.1532C>T p.S511L	ExAC:0.007%	17b, 31b	2.3%
XIRP1	rs61736135 c.1741C>A p.P581T	dbSNP:0.6% EVS:0.9% ExAC:0.7%	28b, 31b	2.3%
FAT1	rs116784674 c.7700G>A p.R2567H	dbSNP:0.5% EVS:1% ExAC:0.9%	2b, 12b	2.3%

Table 4.11. 'Common rare and low-frequency variants' in 44 probands.

4.2.3 *TTN* variants

As described in paragraph 4.1.4, several variants in gene encoding for titin have been detected also in 44 probands sequenced with Targeted 'ACM known and candidate genes panel'.

Excluding the huge amount of synonymous variants, 48 'private variants' (Table 3, Appendix C) have been found in addition to 9 variants present in more than one subject (Table 4.12).

***TTN* variant**	**frequences**	**predictions**	**#probands**
rs55837610 c.71705T>C p.I23902T	EVS:0.3% ExAC:0.2%	SIFT: deleterious	5b, 38b
rs72648972 c.23232C>G p.N7744K	EVS:0.2% ExAC:0.4%	SIFT: deleterious	10b, 22b, 39b
rs72646819 c.55139T>C p.I18380T		SIFT: deleterious	10b, 22b
rs72677231 c.47191C>T p.R15731C	dbSNP:0.2% EVS:0.2% ExAC:0.2%	PolyPhen2: p.damaging SIFT: deleterious	17b, 27b, 33b
rs72677233 c.47315G>A p.R15772Q	dbSNP:0.2% EVS:0.3% ExAC:0.3%	SIFT: deleterious	19b, 25b
rs72648949 c.18776C>G p.T6259S	EVS:0.3% ExAC:0.3%	SIFT: deleterious	19b, 25b
rs72646885 c.69676A>G p.S23226G	dbSNP:0.2% EVS:0.7% ExAC:0.8%	PolyPhen2: p.damaging SIFT: deleterious	22b, 32b, 42b
rs72677237 c.47723G>A p.R15908H	dbSNP:0.2% EVS:0.7% ExAC:0.6%	PolyPhen2: p.damaging SIFT: deleterious	22b, 32b
rs33971253 c.11969C>T p.P3990L	dbSNP:0.6% EVS:0.9% ExAC:0.7%	SIFT: deleterious	20b, 37b

Table 4.12. *TTN* 'common variants' detected in 44 ACM probands.

4.3 Whole exome sequencing

In some particular familial cases, in which traditional mutation screening failed to identify mutations in the 5 most mutated genes in ACM, a different approach was applied. The entire coding part of the genome was captured and sequenced for at least two members of each family.

Family #3

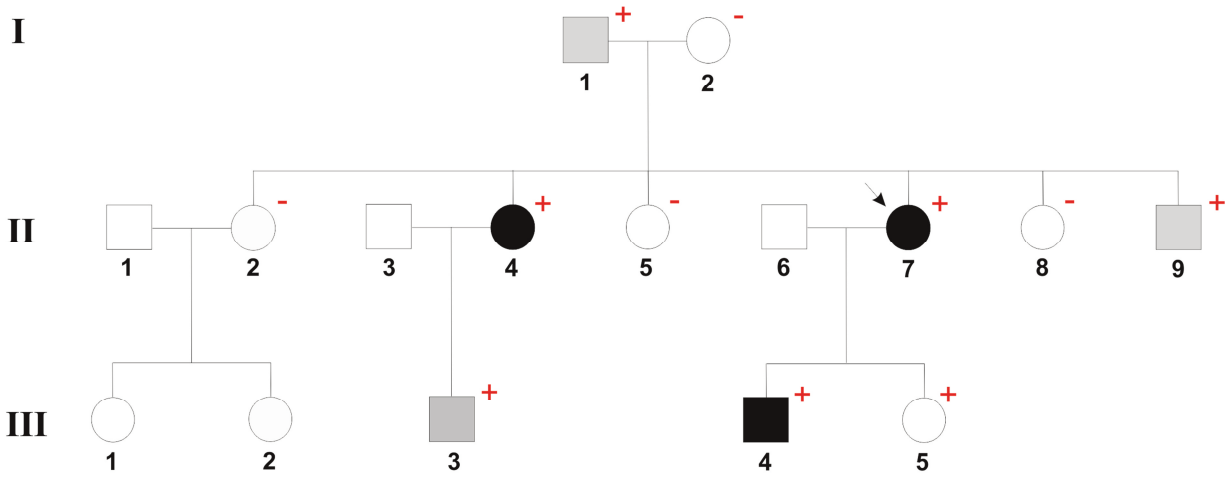


Figure 4.9. Pedigree of Family#3. Black and white symbols represent clinically affected and unaffected individuals, respectively; grey symbols denote subjects with minor signs of the disease. Proband is pointed by the arrow.

\+ *DSP* c.3889C>T p.Q1297*
\- negative for *DSP* c.3889C>T p.Q1297*

The proband (II-7) was diagnosed affected by ACM at the age of 30. She complained of palpitations. Twelve lead ECG showed intraventricular conduction delay and 24-Holter ECG documented the presence of numerous premature ventricular beats (PVBs) with left bundle branch block (LBBB) morphology, coming from the right ventricular outflow tract. The 2D-echocardiogram showed left and right ventricular dilatation with kinetic abnormalities in both ventricles. At the age of 37 she had a syncopal episode and a sustained ventricular tachycardia was recorded, thus an ICD was implanted.

All coding exons and 200 bp flanking intronic regions were captured and sequenced in two affected family members (II-4 and II-7, Figure 4.6). Overall, 89.5 and 91.3 million paired-end sequencing reads were generated for patients II-4 and II-7, respectively. Among the total reads, 89% (for patient II-4) and 86% (for patient II-7) were aligned to the UCSC human reference genome (hg19 release) and 64.5% of the total bases mapped to the targeted and enriched exons. Capture specificity (defined as percentage of uniquely mapped reads aligning to target region) was 73.1% and 72%. After filtering for alignment and sequence quality and

sequencing depth of the site , more than 50.000 SNPs and 5000 Insertion/Deletion (InDels) were identified in the whole exome of each patient. These high-confidence variants were compared against those present in several databases (dbSNPs141, 1000 Genomes Project, YanHuang project, and two databases belonging to BGI) and the variants with a MAF >1% were removed. Among the remaining variants, a total of 141 SNPs and 246 InDels common to both family members were found. Among the most candidate variants identified, a non-sense mutation in *DSP* gene was found (c.3889C>T, p.Q1297*) (Figure 4.10). This heterozygous nucleotide substitution causes the formation of a stop codon in exon 23 of *DSP*. The mutation is present in other 6 family members: two clinically affected (II-4 and III-4) who were diagnosed as having a biventricular form of the disease, two reporting minor signs of the disease as RV conduction delay at ECG (III-3) and PVBs (II-9), with normal 2D-echo findings in both cases. Moreover, one genotype-positive subject (III-5) had unremarkable ECG, echo and cardiac magnetic resonance findings. The father of the proband (I-1) was died at the age of 80 due to cardiac arrest with signs of heart failure, without cardiac evaluation.
This nucleotide substitution in exon 23 of *DSP* gene was not detected at dHPLC analysis due to its localization in a region with a high denaturing temperature.

Figure 4.10. IGV view: detection of *DSP* p.Q1297* heterozygous mutation.

Family #4

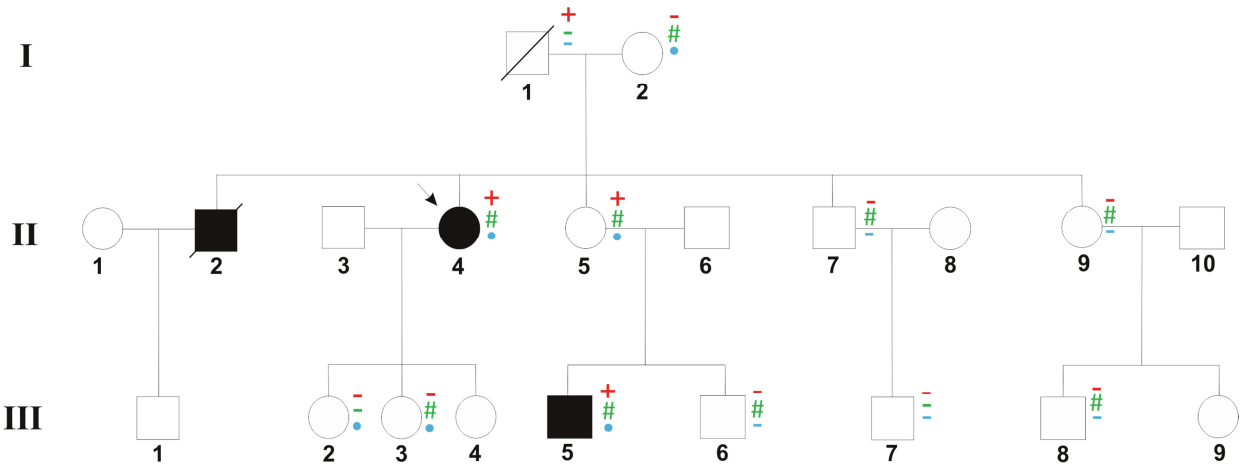


Figure 4.11. Pedigree of Family#4. Black and white symbols represent clinically affected and unaffected individuals, respectively. Proband is pointed by the arrow.

\+ *JUP* c.1942 G>A p.V648I
\- negative for *JUP* c.1942 G>A p.V648I
\# *CDH11* c.1814A>T p.N479I
\- negative for *CDH11* c.1814 A>T p.N479I
• *TTN* c.97717 C>T p.R32573C
\- negative for *TTN* c.97717 C>T p.R32573C

In Family #4, the proband reported minor alterations at echocardiography and inverted T waves in V1. The family had a case of sudden cardiac death (subject II-2, Figure 4.11) at the age of 35, diagnosed affected with ACM at autopsy. Previous genetic screening performed in the proband allowed to identify a variant in the gene encoding for *JUP* (p.V648I). This variant is present in ExAC database with a MAF= 0.7% and predicted tolerated by both prediction tools (PolyPhen2 and SIFT). The research of the genetic cause, reason of the expression of the disease in this family was carried out by whole exome sequencing in 2 individuals (II-4 and III-5).

The total clean reads produced were aligned to the human reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA). On average, 99.75% mapped successfully. The duplicate reads were removed, resulting in the average of 186,556,723 effective reads. Of total effective bases, 43.40% mapped on target regions (capture specificity). The mean sequencing depth on target regions was 111.72-fold. On average per sequencing individual, 47742 SNPs and 4919 InDels were identified.

All the variants present in one or more databases (dbSNPs141, 1000 Genomes Project, EVS and in-house Radboud University database) with a MAF >1% were filtered out. Moreover only coding variants that cause an amino acid change (missense), a formation of a stop codon (non-sense), an alteration of the splicing

or the reading frame were considered. Among the 228 variants shared by subjects II-4 and III-5, two were selected as candidates, for the predicted pathogenicity and the gene involved. A missense *TTN* variant (p.R32573C) and a single nucleotide variant in *CDH11* gene (p.N479I) encoding for a type II cadherin that mediate calcium-dependent cell-cell adhesion. The segregation of both variants is shown in Figure 4.11. Only the two affected subjects (II-4 and III-5) and the obligated carrier (II-5) present the combination of the three variants, whereas several healthy subjects carry the *CDH11* variant and two (III-2 and III-3) the *TTN* variant only.

Family #5

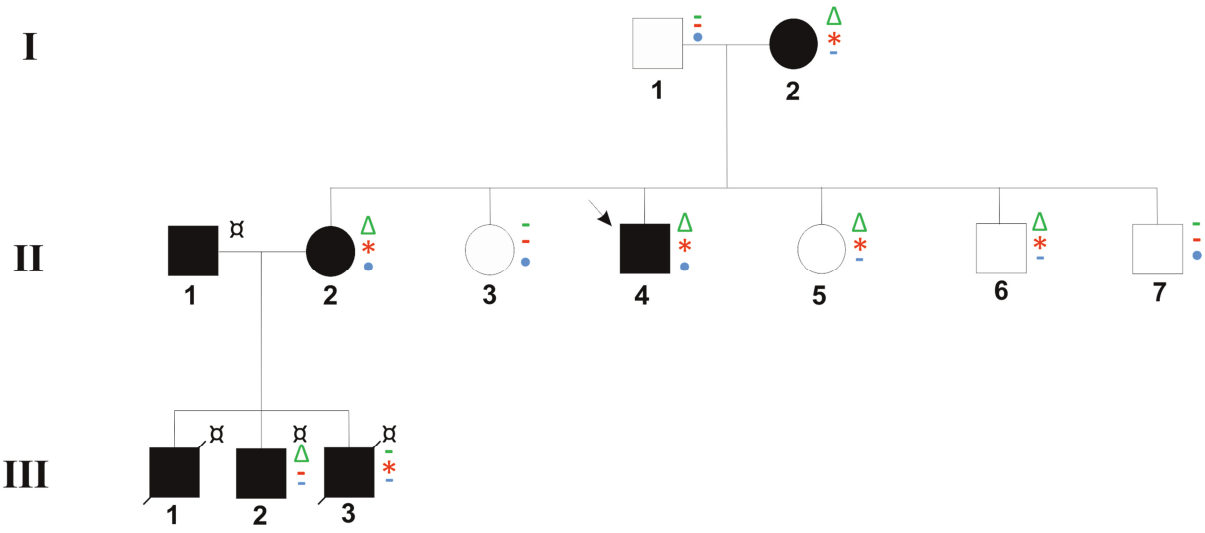


Figure 4.12. Pedigree of Family#5. Black and white symbols represent clinically affected and unaffected individuals, respectively. Proband is pointed by the arrow.

* *TTN* c.96592 C>A p.L32198M
- Negative for *TTN* c.96592 C>A p.L32198M
Δ *CMYA5* c.10789 A>T p.K3597*
- Negative for *CMYA5* c.10789 A>T p.K3597*
• *MYBPC3* c.2980 C>T p.L994F
- Negative for *MYBPC3* c.2980 C>T p.L994F
¤ *RYR2* c.1258 C>T p.R420W

In Family #5, two persons died suddenly at young age (III-1, III-3). Genetic analysis performed in the two subjects as well as in the affected brother (III-2), revealed the presence of a missense mutation in *RYR2* gene (p.R420W), underlying the characteristic phenotype commonly associated with *RYR2* mutations and expressed by the patients. The mutation is carried by the father (II-1), but also the mother (II-2) reported late potentials at ECG and a family history for ACM. The proband (II-4) was screened by dHPLC and direct sequencing for the 5 most important genes in ACM and resulted negative.

Whole exome sequencing of 3 DNA samples (subjects I-1, II-2 and II-4) was performed and averagely 18,813.86 Mb raw bases were produced by the Illumina sequencer. After removing low-quality reads averagely 188,079,352 clean reads were obtained. Total clean reads per sample were aligned to the human reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA). On average, 99.73% mapped successfully and from these the duplicate reads were removed by Picard tools. The capture specificity was 56.4%, 52.9% and 57.5% for sample I-1, II-2 and III-4, respectively. The mean sequencing depth on target regions was 140.30-fold. On average per sequencing individual, 99.50% of targeted bases were covered by at least 1X coverage and 97.70% of the targeted bases had at least 10X coverage. A total of about 47600 SNPs were identified per sample, 98% of which were present in dbSNP database. Moreover, about 5180 InDels were detected per sample, 270 of which, on average, alter the reading frame.

Variants present in either dbSNP141, 1000 Genomes Browser, Exome Variant Server, YanHuang project and databases belonging to BGI or to the in-house Radboud University database with a MAF >1% were excluded. Potentially pathogenic variants were searched among all the rare variants shared by both the affected subjects (II-2 and II-4) but not present in the healthy parent (I-1). Among the remaining 559 variants, 185 were located in a coding or splice site region and 131 were non-synonymous. Within the 13 known-disease genes, only a missense *TTN* variant (p.L32198M) has been identified. The variant is present in 4 affected subjects and in 2 family members with no clinical signs.

Filtering for the stopgain and stoploss variants, a SNV in *CMYA5* gene has been identified (c.10789 A>T, p.K3597*). It is reported in the SNP databases (rs185458523) with a MAF between 0.04% and 0.19%. This loss-of-function variant was found in all the affected family members, with exception of subject III-3, and in two healthy family members (II-5 and II-6).

Then, considering also the variants present in subject I-1, a rare missense variant has been found in *MYBPC3* gene (p.L994F) in the three exome data. In ExAC database this variant has a MAF=0.02% and is classified as a variant with “uncertain significance” in ClinVar database and associated with a HCM phenotype. Three healthy family members other than the two affected reported the same SNV (Figure 4.12).

4.4 Multiple approach

Family #6

The proband (IV-9, Figure 4.13) was diagnosed affected fulfilling ACM Revised Task Force Criteria. He reported RV dysfunction with ventricular arrhythmias of LBBB morphology. ECG showed late potentials other than inverted T waves in V1-V3. Clinical evaluation of 25 available family members was performed and 7 of them resulted affected. Mutation screening of the exons and intronic flanking regions of the 5 most common ACM mutated genes has been performed in the proband. Only an intronic variant located in exon 14 splice acceptor site of *PKP2* gene was identified (c.2578-3 T>C). Studying the segregation within the family, his father and his brother (subjects IV-10 and III-8) resulted positive for the same mutation (Figure 4.13).

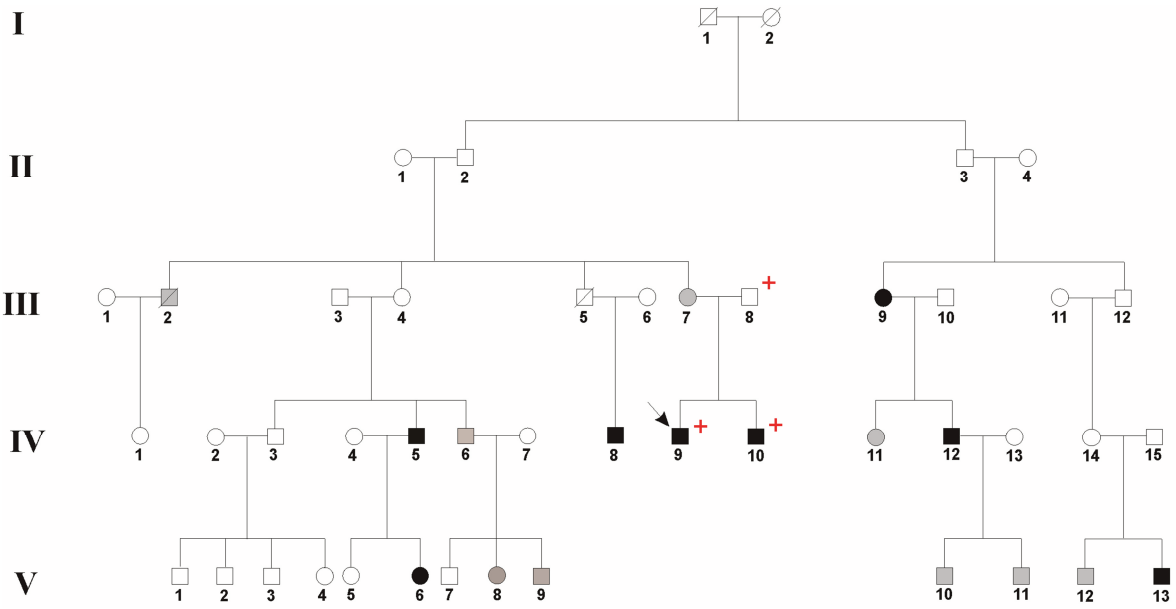


Figure 4.13. Pedigree of Family #6. Proband is pointed by the arrow. Black and white symbols represent clinically affected and unaffected individuals, respectively; grey symbols denote subjects not investigated (deceased individuals) or with minor signs of the disease.

\+ *PKP2* c.2578-3T>C

Considering that the *PKP2* mutation is carried by an external individual of the family (III-8) and is not present in other affected subjects, the analysis went on with genotyping and linkage analysis.

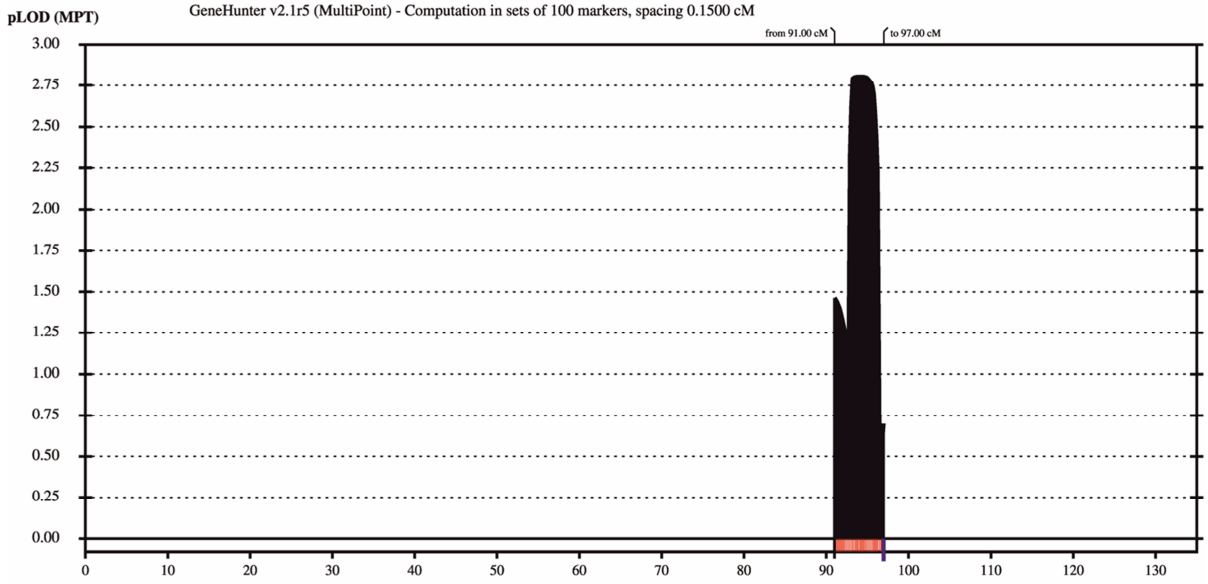
4.4.1 Linkage analysis

Individual genotyping was performed in 33 affected and unaffected family members using a SNP array containing approximately 370,000 markers.

Parametric linkage analysis was performed assuming autosomal dominant inheritance and an ‘‘affected only’’ design, which reduces the influence of unaffected persons carrying the disease gene on linkage results.

Two different linkage analyses have been performed, including or excluding the nucleus composed by individuals III-8, IV-9 and IV-10. In both cases, different marker spacing was considered, avoiding false LOD score values positives caused by number and set of selected SNPs.

Excluding the three subjects with the *PKP2* mutation, multipoint linkage analysis highlighted the presence of a candidate locus on chromosome 11q21 with a maximum parametric LOD score of 2.87 (Figure 4.14, Table 4.13).



Marker	chr.	position (cM)	pLODscore
rs1447352	11	92.72	1.322
rs1613	11	92.87	2.415
rs6483232	**11**	**93.03**	**2.697**
rs921983	**11**	**93.20**	**2.877**
rs4753485	**11**	**93.37**	**2.878**
rs644009	**11**	**93.55**	**2.879**
rs11020619	**11**	**93.70**	**2.879**
rs4073612	**11**	**93.86**	**2.879**
rs672392	**11**	**94.03**	**2.879**
rs499952	**11**	**94.18**	**2.879**
rs593690	**11**	**94.33**	**2.879**
rs4144617	**11**	**94.49**	**2.878**
rs7103994	**11**	**94.64**	**2.877**
rs547243	**11**	**94.83**	**2.874**
rs922020	**11**	**95.00**	**2.869**
rs541821	**11**	**95.16**	**2.859**
rs2155354	**11**	**95.31**	**2.849**
rs1255182	**11**	**95.47**	**2.839**
rs693830	**11**	**95.62**	**2.817**
rs7951485	**11**	**95.77**	**2.772**
rs569977	**11**	**95.92**	**2.666**
rs2186760	**11**	**96.09**	**2.509**
rs10501856	**11**	**96.24**	**2.286**
rs4594032	11	96.46	0.422

Figure 4.14 and Table 4.13.

Plot and pLOD values obtained for markers mapped on chromosome 11q21, in Family #6. Distance between markers: 0.15cM

Analysis of haplotype segregation on chromosome 11q21 showed a common region shared among all affected subjects (Figure 4.18). The critical region was delimited by crossovers in the affected individual IV-8 on the centromeric side (at 93.03 cM) and on the telomeric side (at 96.47 cM), refining the critical interval to a 3 cM region, between rs6483232 and rs4594032 markers (Figure 4.15).

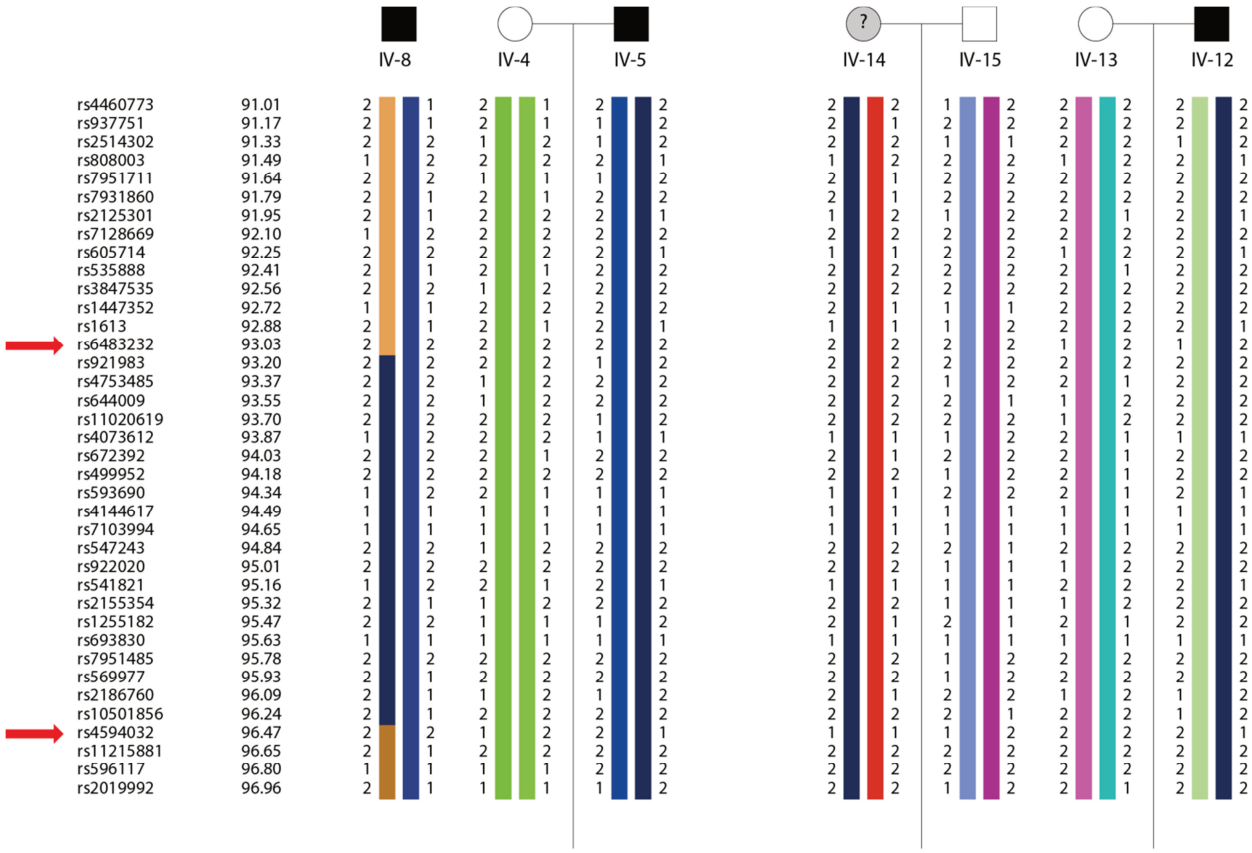
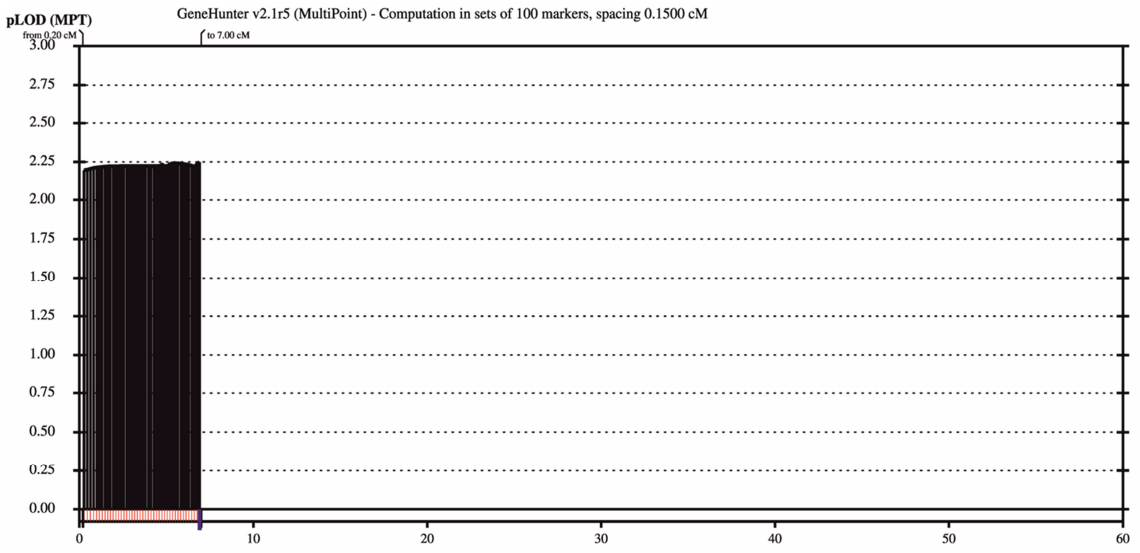


Figure 4.15. Partial haplotype segregation on chromosome 11q21 in subpedigree of Family #6. The at-risk haplotype, shared by all affected subjects, is shown in dark blue. Key recombination events occurred in subject IV-8 are pointed with the arrows.

The pLODscore in this region resulted negative if subjects III-8, IV-9 and IV-10 were included in the analysis, but from this second linkage analysis a new locus on chromosome 19p13.3 reported positive pLODscore values (Figure 4.16, Table 4.14).



Marker	chr.	position (cM)	pLODscore
rs2312724	**19**	**0.27**	**2.198**
rs740871	**19**	**0.46**	**2.199**
rs2283575	**19**	**0.63**	**2.205**
rs7250902	**19**	**0.82**	**2.210**
rs2240152	**19**	**1.00**	**2.213**
rs11667680	**19**	**1.15**	**2.215**
rs1558134	**19**	**1.31**	**2.217**
rs166451	**19**	**1.48**	**2.219**
rs2012125	**19**	**1.63**	**2.221**
rs757528	**19**	**1.79**	**2.221**
rs10500203	**19**	**1.94**	**2.222**
rs12104154	**19**	**2.10**	**2.222**
rs10407022	**19**	**2.25**	**2.222**
rs7254110	**19**	**2.41**	**2.222**
rs2907335	**19**	**2.56**	**2.222**
rs6510698	**19**	**2.72**	**2.222**
rs9304900	**19**	**2.88**	**2.222**
rs10412849	**19**	**3.03**	**2.224**
rs3786944	**19**	**3.18**	**2.224**
rs7247434	**19**	**3.33**	**2.225**
rs10126	**19**	**3.50**	**2.225**
rs2270083	**19**	**3.65**	**2.225**
rs2052059	**19**	**3.80**	**2.225**
rs7255123	**19**	**3.96**	**2.224**
rs350897	**19**	**4.12**	**2.222**
rs4807567	**19**	**4.28**	**2.222**
rs243372	**19**	**4.44**	**2.222**
rs10423119	**19**	**4.59**	**2.222**
rs7508325	**19**	**4.75**	**2.231**
rs2779173	**19**	**4.90**	**2.222**
rs858422	**19**	**5.05**	**2.222**
rs7260484	**19**	**5.21**	**2.234**
rs875552	**19**	**5.36**	**2.239**
rs11878567	**19**	**5.52**	**2.240**
rs4643459	**19**	**5.67**	**2.240**
rs3763045	**19**	**5.83**	**2.238**
rs4807830	**19**	**5.98**	**2.236**
rs10417646	**19**	**6.13**	**2.234**
rs8182608	**19**	**6.29**	**2.231**
rs173229	**19**	**6.46**	**2.226**
rs344580	**19**	**6.61**	**2.095**
rs7258236	**19**	**6.76**	**1.869**
rs466876	**19**	**6.92**	**1.522**
rs2967636	**19**	**7.07**	**0.463**
rs4804404	19	7.22	0.351

Figure 4.16 and Table 4.14.

Plot and pLOD values obtained for markers mapped on chromosome 19p13.3, in Family #6. Distance between markers: 0.15cM.

Analysis of haplotype segregation indicated the presence of the critical region on chromosome 19p13.3 in all affected subjects but one (V-13; Figure 4.15). The critical region was delimited by a centromeric cross-over in subject IV-9, thus defining a region of about 7 cM, between marker rs2312724 and marker rs2967636 (Figure 4.17).

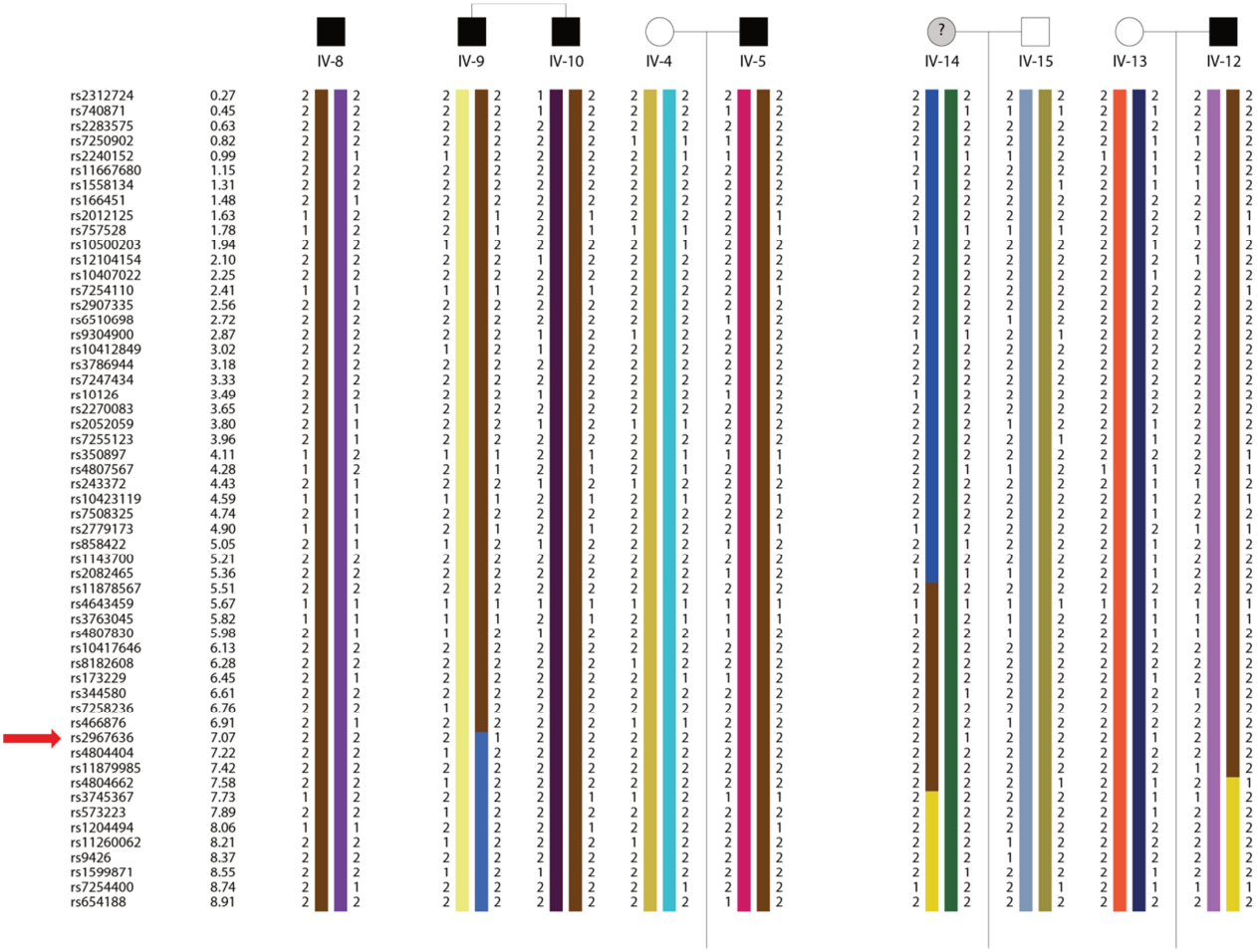


Figure 4.17. Partial haplotype segregation on chromosome 19p13.3 in subpedigree of Family #6. The at-risk haplotype is shown in brown. Key recombination event occurred in subject IV-9 is pointed with the arrows.

The segregations of both critical regions on chromosomes 11q21 and 19p13.3 are shown in Figure 4.18.

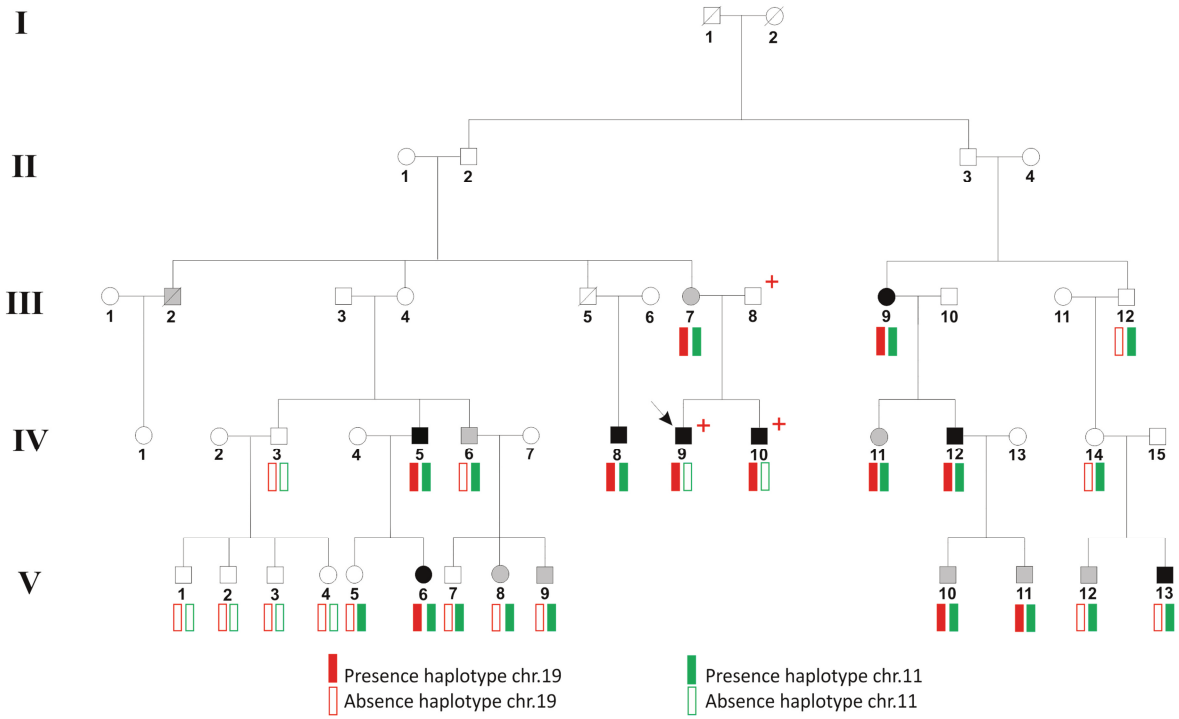


Figure 4.18. Pedigree of Family #6 and haplotypes segregation.

The two critical regions co-segregated in 5 out of 8 affected subjects but no one of the healthy family members. Some individuals reporting minor signs of the disease carried one or both critical haplotypes (III-7, IV-11, V-10, V-11, IV-6, V-8, V-9, V-12).

4.4.2 Whole exome sequencing (WES)

Considering the large number of genes mapped into the two critical regions, to identify the disease-genes in this family the whole exome of 4 affected subjects was sequenced (samples IV-5, IV-8, IV-9, IV-12) using the HiSeq2000 platform provided by Illumina. An average of 89 milion paired-end reads were generated, with a percent reads on target of about 70% and a mean coverage depth of 80X.
The detected variants were filtered for the alignment quality, sequence quality score, and sequencing depth of the site, to get high-confidence variants.
All the variants shared by the 4 patients were compared with those present in several SNP databases (dbSNP141, 1000 Genomes project, Exome Variant Server, HapMap eight exomes, YanHuang project) in order to discard common polymorphisms. The presence of any novel coding variant shared by the 4 patients was excluded. Among the 610 shared variants with a MAF≤ 0.01%, 395 mapped in a coding region or in a splice site, and 191 out of 395 were non-synonymous variants (missense, stop or frameshift). All of these filtered variants are present at least in one SNPs database, but no information about the frequency is available. None of these 191 SNVs is located into the critical regions or within a candidate

ACM gene. Considering all the exomes individually, the presence of novel variants predicted to be pathogenic by SIFT or PolyPhen2 and/or variants affected highly conserved residues was excluded.

To exclude the presence of a pathogenic mutation in all the ACM genes in the 4 affected subjects, the evaluation of the real sequence coverage depth on 11 known-disease genes (*PKP2*, *DSP*, *DSG2*, *DSC2*, *JUP*, *TMEM43*, *PLN*, *LMNA*, *DES*, *CTNNA3* and *TGFβ3*) was performed by using Integrative Genomic Viewer (IGV) for all the 4 samples separately. A total of 22 exons (Table 4.15) reported a low coverage (<20X) were re-sequenced by the dideoxy-chain termination method. All the four affected family members resulted negative for coding mutations in the 11 known-disease genes.

gene	**uncovered exons (<20X)**
PKP2	ex 1, ex 6
DSP	ex 1, ex 12
JUP	ex 5, ex 6, ex 11
DSC2	ex 1, ex 2
DSG2	ex 1, ex 5, ex 10
DES	ex 1, ex 8
TMEM43	ex 10, ex 12
CTNNA3	ex 10, ex 11
LMNA	ex 1, ex 9, ex 2, ex 10

Table 4.15. Exons of known-disease genes resulted not sufficiently covered by WES and re-sequenced with Sanger method. Primers used are listed in Table 1, Appendix D.

Only an intronic novel variant located near a splice donor site of *TMEM43* gene (c.766+8C>A) has been identified in subjects IV-5, IV-9 and IV-12. Analyzing the segregation within the family, this variant is absent in two affected individuals (IV-8 and V-13) and present in several family members without disease manifestations (IV-3, V-1, V-2, V-3, V-4, V-5).

4.4.3 Whole genome sequencing (WGS)

Whole genome sequencing (WGS) has been performed in two affected subjects (IV-10 and IV-12) and an healthy individual (IV-3) of Family#6 at Novogene Bioinformatics Technology Co.

According to the Illumina platform sequencing feature, for paired-end data the average percentage of Q30 above 80% and error rate below 0.1% is required. Sequencing data quality is summarized in Table 4.16 .

sample	Raw data(G)	Raw depth(x)	Effective(%)	Error(%)	Q30(%)
IV-3	172.41	59.57	99.73	0.05	85.37
IV-10	176.34	60.93	99.57	0.03	89.06
IV-12	173.05	59.79	99.75	0.04	86.61

Table 4.16. Overview of data production quality. Raw data= the original sequence data; Raw depth= mean coverage depth; Effective= the percentage of clean reads in all raw reads; Error rate= the average error rate of all bases; Q30= percentage of reads with average quality>30.

On average the sequencing coverage depth was 55.48X, 58.07X and 57.53X with 97.36%, 97.14% and 98.09% of bases with depth ≥20X for samples IV-3, IV-10 and IV-12, respectively.

Generally, the human whole genome has about 3.6M SNVs and 350K InDels (insertion and deletion, less than 50bp). Most (above 95%) SNVs with high frequency (MAF ≥5%) have records in dbSNP database. The main difference between WES and WGS is that no capture step is required in WGS and SNVs/InDels are detected in both the coding and non-coding part of the human genome (Table 4.17).

Sample	type	exonic	intronic	3'UTR	5'UTR	intergenic	Up stream	Down stream	splicing
IV-3	SNVs	20059	1106395	22287	4783	1842438	18977	19036	2346
	InDels	603	174569	4224	747	280099	3590	3503	368
IV-10	SNVs	20064	1111991	22187	4752	1837266	19237	19459	2351
	InDels	617	179349	4177	788	284743	3746	3602	370
IV-12	SNVs	20314	1123008	22690	4871	1861374	19335	19481	2364
	InDels	617	181361	4268	756	290046	3768	3618	379

Sample	type	ncRNA_exonic	ncRNA_intronic	ncRNA_3'UTR	ncRNA_5'UTR	ncRNA _splicing
IV-3	SNVs	8237	125703	531	89	228
	InDels	934	19978	91	18	37
IV-10	SNVs	7960	125997	499	88	240
	InDels	919	20303	94	18	35
IV-12	SNVs	7998	123693	564	78	230
	InDels	912	20278	98	13	41

Table 4.17. The number of SNVs and InDels in different genomic region. Upstream= the number of SNV/InDel in the 1kb upstream region of transcription start site; downstream= the number of SNV/InDel in the 1kb downstream region of transcription ending site; ncRNA= non-coding RNA.

The majority of SNVs detected for each sample were missense or synonymous variants, whereas the distribution of different InDels types was more equal, as shown in pie charts below (Figure 4.19).

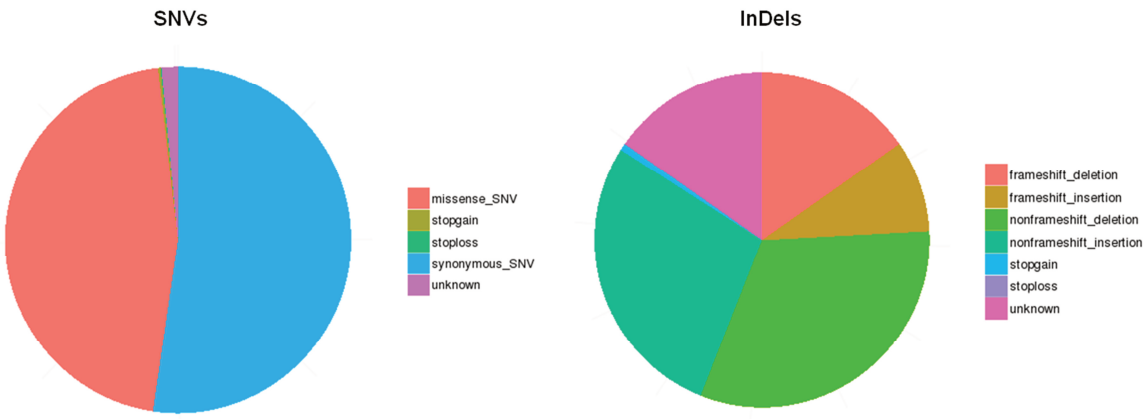


Figure 4.19. The number of different SNV types in coding region (left) and the number of different InDel types (right) for sample IV-3.

Moreover, structural variations (SVs) such as deletion, insertion, duplication, copy number variations, inversion and translocation of fragments between 1kb and 3Mb have been detected in each sample (Table 2, Appendix D).

The annotation step performed by ANNOVAR, allowed to extrapolate the whole exomes from the genome data and to filter for the frequency of each variant reported in public databases.

Considering the whole exome, among the variants shared only by the two affected subjects (IV-10 and IV-12) with a MAF ≤0.5%, 49 were located in conserved (phyloP>2.5) coding or splice site regions. The majority were classified as 'complex variants' and affected repetitive genomic regions. 3 out of 49 SNVs had no records in dbSNP database. None of these variants localized in a known or candidate ACM genes. (Table 3, Appendix D).

All the variants inside the critical region on chromosome 19p13.3 and shared by subjects IV-10 and IV-12 were then considered. 56 variants reporting a MAF ≤0.5% were identified, all but one (located in a splice acceptor site of *THEG* gene) mapped in an intronic or UTR region (Table 4, Appendix D). Among the genes included in this list of variants, *Protein Tyrosine Phosphatase Receptortype Sigma* (*PTPRS*) appeared the most candidate gene because of its possible role in intercellular adhesion and its myocardial expression. Three variants located in 3'UTR of this gene have been identified (rs116644020, rs200486129, rs139183127) other than a deeply intronic SNV. The phyloP score for all the 56 variants is below 2.5, indicating that the genomic regions where the variants mapped, are not conserved among species.

For both the analyses described above, the annotation was carried out using the in-house pipeline and databases of the Radboud University medical center (Nijmegen).

Taking into account that a positive pLODscore value is associated also to a 3 cM region on chr11q21, all the variants in this region identified in subject IV-12 were considered. Patient IV-10 was excluded because he did not carry the same haplotype at chromosome 11. All the variants shared by subjects IV-12 and IV-3 were excluded as well. Among the 378 unique variants called for sample IV-12 with a MAF ≤0.5%, 158 localized inside a gene (Table 5, Appendix D), but only one in an exonic region (*MTMR2*, rs146572467, MAF= 0.04%). The majority of these 158 variants mapped in introns and in repetitive genomic regions.

To summarize, WGS allowed to identify on average about 22500 SNPs and 984 InDels in exonic or splice site regions per sample. After filtering for sharing between samples IV-10 and IV-12, and exclusion with sample IV-3, a short list of putative variants was defined (Table 3, Appendix D). No potential causative mutations have been identified. Moreover all the variants inside both regions showing a positive LODscore value have been checked, even though no obvious candidate genes map in these regions. Most of the variants are complex and located in repetitive intronic or intragenic regions. Only one maps in the coding region of *MTMR2* gene, encoding for a member of the myotubularin family proteins and located on chromosome 11q21.

5. DISCUSSION

Arrhythmogenic cardiomyopathy (ACM) is an inherited myocardial disorder characterized by progressive cardiomyocyte death, followed by fatty or fibrofatty replacement (Thiene et al, 1988; Nava et al, 1988). It is mainly transmitted as an autosomal dominant trait, characterized by incomplete penetrance and variable expressivity. The most typical clinical presentation has two aspects: i. electrocardiographic abnormalities (ventricular tachycardia with LBBB, T-wave inversion in the V1 to V3); and ii. functional and structural abnormalities mostly of the right ventricle (RV) (Thiene et al., 1990). Even though left ventricle (LV) involvement was first described as a late consequence of the disease progression, today it is well-known that it could also appear as a first sign of the disease (Corrado et al, 1997). Beyond the classical form of ACM, a biventricular and a left dominant form are now considered as different patterns of the wide spectrum of ACM (Sen-Chowdhry et al., 2007). ACM is a devastating disease given the fact that the first symptom is often SCD, which makes early detection and a family screening test the cornerstone in the diagnostic evaluation. Familial occurrence is rather common; at least 15 independent loci and 13 disease genes have been associated with ACM thus far, with a large involvement of the desmosomal genes. Most ACM causing mutations have been identified in genes encoding for desmoplakin (*DSP*, Rampazzo et al., 2002), plakophilin-2 (*PKP2*, Gerull et al., 2004), desmoglein-2 (*DSG2*, Pilichou et al., 2006), desmocollin-2 (*DSC2*, Syrris et al., 2006) and plakoglobin (*JUP*, Asimaki et al., 2007). Recently, a novel ACM gene has been identified encoding for α-T-catenin, a protein highly expressed in cardiac *area composita*, a mixed type junctional structure, composed of both desmosomal and *adherens junction* proteins (van Hengel et al., 2013).
However, comprehensive mutation screening of all the 13 known ACM genes can detect causative mutations in ≈60% of probands (Groeneweg et al., 2015). Thus, in the remaining ≈40% of patients other genes are probably involved in the pathogenesis of the disease.

5.1 Targeted genes panels

The advent of next generation sequencing (NGS) technologies has revolutionized the field of medical genomics, enabling fast and cost-effective generation of genome-scale sequence data with exquisite resolution and accuracy. Rapid sequencing of the entire genome/exome of a patient removes the necessity to prioritize candidate genes for sequencing and therefore is largely applied for the identification of novel disease genes (Gilissen et al., 2012). For polygenic or multifactorial pathologies in which more genes are involved, such as inherited

cardiomyopathies, NGS approach is successfully used for genetic diagnosis and as a research tool to identify novel disease genes.
In the present study, three different NGS strategies have been considered to identify the genetic cause in both familial and sporadic ACM cases.

Two different Targeted gene panels have been used to screen a total of 59 index cases. The first panel, named 'Cardiomyopathies gene panel' comprises 56 genes frequently associated with different cardiomyopathies such as dilated and hypertrophic cardiomyopathies (DCM, HCM), long QT syndrome, Brugada syndrome, etc. Among the 19 probands analyzed, 10 were previously investigated for mutations in the 'three big ACM genes' (*PKP2*, *DSP* and *DSG2*) with Sanger sequencing (SS). These 10 subjects have been considered not only to complete the genotyping for all of the ACM genes included in the panel, but also to assess the specificity and sensibility of this approach. It allowed to identify 96% of the variants previously detected by SS: the 3 mutations, the rare variant and 22 out of 23 polymorphisms. The undetected polymorphism is located in a region showing low coverage (4X). Overall, 95% of the novel variants (38 out of 40) identified by NGS was confirmed by SS. The 2 unconfirmed variants were expected to be false positives since they correspond to unbalanced substitutions. Considering only the mutations (MAF ≤0.01%), in 15 out of 19 subjects at least one mutation has been identified. Taken together, these results in addition to the high representation of coverage (mean depth of 200X), suggest that it is a valuable tool for mutation screening in patients affected with inherited cardiomyopathies. Indeed, as reported by different research groups, customized targeted gene panels have been considered the gold standard method for the molecular diagnosis in hereditary cardiomyopathies because they are faster and cheaper than whole exome/genome approaches, they allow focusing on the most relevant genes for a particular disorder and minimize the problems associated with unsolicited findings (Sikkema-Raddatz et al., 2013).
By using the second panel, conventionally called 'ACM known and candidate genes panel', 44 samples were sequenced for a total of 69 genes, including the 13 known-disease genes and 56 potentially candidate genes. These genes have been selected on the basis of their cellular localization (*e.g.* mechanical junctions components), their role in pathways suggested to be involved in ACM (*e.g.* Wnt/β-catenin pathway, *TGFβ3* pathway, Hippo pathway), or their interacting network. This cohort of patients was previously screened by dHPLC and SS and mutations in the three most important ACM genes were excluded. In the past several studies have estimated dHPLC sensitivity and specificity around 98% and 100%, respectively (Ellis et al., 2000; Bagattin et al., 2004), and Sanger sequencing is still

considered the gold standard for mutation screening. Despite this, in 4 patients probably pathogenic mutations in one of these three genes have been identified (*DSP* c.3799C>T p.R1267*, *DSP* c.1267-2A>G, *PKP2* c.1063C>T p.R355*, *DSG2* c.1557C>A p.D519E).

In the last 10 years, from the introduction of the first high-throughput sequencing methods, the accuracy of these NGS techniques had a rapid increase, which corresponded to a constant decrease of their costs. A recent work reported an analytical sensitivity and specificity >99.5% for the NGS techniques (targeted gene panels and WES) comparable to Sanger sequencing, thus suggesting a possible shift from SS to massive parallel sequencing for routine genetic testing (Diekstra et al., 2014).

As observed in probands screened with the 'Cardiomyopathies gene panel', also in this cohort of ACM patients there are 6 subjects carrying one or more mutations in genes associated with other diseases (Table 4.9). Particularly, sample n° 4b carried a *LMNA* mutation (c.569G>A, p.Arg190Gln) previously reported in association with DCM (Perrot et al., 2009) and Emery-Dreifuss Muscular dystrophy (Cenni et al., 2005). The R190Q mutation is a semi-conservative amino acid substitution as these residues share similar properties, but differ in size, charge, or other properties which may impact the secondary structure. The R190 residue is conserved across species. Mutations involving the same residue (R190W) or the nearby residues (L183P, E186K, R189W, D192V, D192G, N195K) have been reported in association with DCM, further supporting the functional importance of this residue and region of the protein. It is not surprising that DCM and ACM patients result to carry different mutations in the same gene. Mutations in *DSP*, *PKP2* and *DSG2* genes, frequently involved in ACM, have also been identified in DCM patients (Elliott et al., 2010; Posch et al., 2008). In fact, even though DCM and ACM have been considered distinct clinicopathologic entities, both can be caused by mutations in desmosomal protein genes.

Another patient (n° 15b) showed a missense variant in *TNNC1* gene (c.435C>A, p. Asp145Glu), previously described in a patient with HCM and demonstrated to cause an increased calcium sensitivity of force recovery (Landstrom et al., 2008). Finally other mutations have been found in *SCN5A*, *MYL2* and *TPM1* genes commonly associated with Brugada syndrome (BrS), HCM and DCM phenotypes, respectively. In particular, a genetic overlap between BrS and ACM was recently confirmed by the group of Di Resta, which identified a significant enrichment of *DSG2* rare variants in a cohort of BrS patients (Di Resta et al., 2015). Overall, in the last 20 years, several examples clearly demonstrated that mutations in the same gene can be associated with different phenotypes. A classical example of this situation is the involvement of troponin and tropomyosin mutations both in

HCM and DCM. *In vitro* and *in vivo* functional studies demonstrated that mutations in these genes reduce Ca^{2+}-sensitivity, maximum activation, and Ca^{2+} affinity in DCM patients, whereas they have an opposite effect in HCM, in which calcium sensitivity is enhanced and predicted to increase contractility (Robinson et al., 2007; Lakdawala et al., 2010).

In addition, some phenotypes may be the result of variable expression of the same mutation and in these cases we should consider that such distinct phenotypes may be part of the same disease rather than being two or several separate diseases (Monserrat et al., 2015).

Several rare or low-frequency variants (0.01%≤ MAF ≤1%) in both ACM and candidate genes have been identified in our cohort of patients. The interpretation of these variants is still controversial, but interestingly some of these are present in more than one individual (Table 4.11), thus resulting in most cases in a MAF ten times higher than the MAF reported by common SNP databases. A possible pathogenic role or a modifier effect of these variants should be taken into account.

A particular case is represented by the giant gene encoding for titin and the huge amount of variants identified in this gene. The importance of titin in cardiomyopathies has only recently emerged because routine analysis of the whole *TTN* gene by traditional sequencing methods was very difficult due its enormous size and complexity (Chauveau et al., 2014). *TTN* is now considered the most common disease gene in dilated cardiomyopathy, accounting for up to 25% of patients (Herman et al., 2012), but it is also associated with an ACM form phenotype, characterized by greater risk for supraventricular arrhythmias and conduction disease (Brun et al., 2014).

Considering both the panels, a total of 81 'private' missense *TTN* variants (variants identified in only one sample of our cohort) have been detected. Among these, 62 could be considered mutations since they are absent in all SNP databases or they have a MAF ≤0.01%. Overall, an average of 1.2 *TTN* missense mutations per patient have been detected. The frequency of *TTN* rare missense variants in the general population is far above the expected frequency of disease causing mutations, suggesting that some of these could be either benign or could act as modifiers in genetically susceptible hosts (Brun et al., 2014; Golbus et al., 2012). Accordingly, the interpretation of these variants in our cohort of patients remains difficult as well.

The principal aim of the application of this targeted gene panel was to identify mutations in putative candidate genes for ACM in a selected cohort of patients. Our results demonstrated that at least one missense/non-sense mutation in a candidate gene was found in 40.6% of the index cases negative for mutations in known-disease genes. The most mutated genes resulted *FAT1* (4 mutations:

p.D919N, p.V2380I, p.A4541T, p.V2980L) and *DCHS1* (4 mutations: p.P557S, p.I2268T, p.G87S, p.I1069K) both encoding for members of the cadherin superfamily. However, the majority of these missense mutations were predicted 'tolerated' from prediction tools. Moreover many rare or low-frequency variants in these two genes were also found in our cohort of patients. Considering the RVIS (Residual Variation Intolerance Score) for these two genes, both are just below zero (-0.39 for *FAT1* and -1.38 for *DCHS1*), indicating that these genes have less common functional genetic variation relative to the genome wide expectation given the amount of apparently neutral variation. Taking into account all these predictions and the high number of variants in these two genes found among the 44 samples, it is difficult to assess a probable pathogenicity for these mutations.
Interesting missense and non-sense variants have been found in genes encoding for protein ZO-1, N-cadherin, p63, and talin1 genes.

5.2 Identification of novel mutations in *TP63* and *TLN1* genes

In Family #1, the proband (n°8b) resulted negative for mutations in the 5 most mutated genes in ACM, but showed two novel mutations in candidates genes: a missense variant in *TLN1* gene (c.3707A>G p.Q1236R) and a stop mutation in *TP63* gene (c.796C>T p.R266*). Even though *TP63* mutation was called by only one variant caller, both the variants were validated by Sanger sequencing. The lack of identification of this variant by VarScan variant caller is probably due to the unbalanced representation of the two alleles. This fact highlights the importance of the application of different tools in the variant calling process. c.3707A>G mutation is absent in all SNP databases and affects a gene that encodes for the cytoskeletal protein talin 1, a cytoplasmic protein concentrated in areas of cell-substratum and cell-cell contacts. The mutation is predicted to be possibly damaging by PolyPhen2 tool but tolerated by SIFT tool. The N-terminus of this protein contains elements for localization to cell-extracellular matrix junctions, whereas the C-terminus contains binding sites for proteins such as beta-1-integrin, actin, and vinculin (Burridge and Mangeat, 1984; Ben-Yosef and Francomano, 1999). The mutation causes an amino acid change at position 1236 (2541 total amino acids), which does not represent a functional protein domain. Given that the mutation is absent in the affected proband's mother, it is very likely that it has a tolerated effect.
On the contrary, *TP63* heterozygous mutation is carried also by the mother (II-2), who fulfilled ACM Task Force criteria, thus suggesting its possible pathogenic role. This gene encodes for different isoforms of the protein; two different N-termini and multiple C-termini are generated by alternative splicing, with a wide expression

among human tissues, such as in heart, kidney, placenta, prostate, skeletal muscle, testis and thymus (Yang et al., 1998). The mutation is located in an exon included in all the transcripts and it causes the formation of a stop codon in all of them.

TP63 (p63) belongs to the p53 family of transcription factors, together with p53 and p73 and acts as a sequence specific DNA binding transcriptional activator or repressor (Yang et al., 1998; Aylon and Oren, 2011). One of the key effector in the p63 developmental program is Perp, a critical component of the desmosome which plays an essential role in promoting a stable assembly of cell-cell adhesion complexes (Ihrie et al., 2005). In a case report, Valenzise et al. described a patient with ectodermal dysplasia associated with ACM in which a novel missense mutation of p63 has been identified (p.R298Q) (Valenzise et al., 2008). Nevertheless, a link between ACM and p63 mutation reported in this study remains debatable since mutations in the other desmosomal genes were not excluded. Conversely, mutations in p63 genes are widely associated with ectodermal syndromes such as ectrodactyly, ectodermal dysplasia, and cleft lip/palate syndrome 3, split-hand/foot malformation 4, ADULT syndrome and others (van Bokhoven and McKeon, 2002). More recently, TP63 has been demonstrated to be also a key endodermal cardiogenic factor, that modulates the specification and/or the proliferation of cardiac progenitors (Paris et al., 2012). In vitro studies of p63-null embryos revealed a severe dilated cardiomyopathy with a pronounced disorganization of the myofibrillar apparatus in cardiomyocytes (Paris et al., 2012). A large proportion of embryonic lethality was also observed in p63 KO mice, similarly to what was demonstrated for DSP-/- mice (Jonkman et al., 2005). All these findings are not sufficient to link p.R266* mutation to ACM phenotype in Family #1, even though the nature of the mutation, the segregation within the family and the recent involvement of p63 in heart development support its potential pathogenic role. Identification of additional mutation(s) in *TP63* gene by genetic screening of a large cohort of ACM patients, negative for mutations in the known-disease genes, could support the effective involvement of this gene in the genetic determination of the disease.

5.3 Identification of novel mutations in *TJP1* gene

In Family #2, the proband, who resulted negative for mutations in the 5 most important ACM genes, was then analyzed by using a Targeted gene panel. The analysis highlighted the presence of a novel missense mutation in *TJP1* gene (c.2006A>G, p.Tyr669Cys), which encodes for a tight junction protein (ZO-1). The segregation of the variant among the family members has been studied. The

mutation is carried by 5 individuals including the two sons of the dead proband's brother (III-1, III-2) and two other proband's brothers. Moreover, the mutation is carried by the mother (I-2), who showed minor signs of the disease, but reporting a family history for sudden cardiac death.
Another missense substitution (p.Arg265Trp) in *TJP1* gene has been identified in proband n° 37b, but in this case no family members were available.
Both mutations affect highly conserved amino acid residues and are predicted to be deleterious by PolyPhen2, SIFT, and Mutation Taster prediction tools. All these findings together with the evidence that protein ZO-1 colocalizes with N-cadherin at the intercalated discs of cardiomyocytes (Barker et al., 2002), suggest that *TJP1* gene may be a good candidate disease gene.
Protein ZO-1 is a member of the membrane-associated guanylate kinase (MAGUK) family of proteins that function in protein targeting, signal transduction, and determination of cell polarity (Anderson et al., 1996). It has been demonstrated the interaction between connexin43 (Cx43) and the second of three PDZ domain of ZO-1 in cardiac myocytes, and that ZO-1 regulates Cx43 plaque size (Toyofuku et al., 1998; Hunter et al., 2005). A key role in the localization and stabilization of adherens and gap junctions has also been ascribed to the cytoskeletal scaffold protein ZO-1 since it is a linker between the cadherin/catenin complex and gap junctions (Bruce et al., 2008; Palatinus et al., 2011). The two identified mutations are localized in different protein domains: Arg265 is located at the end of the second PDZ domain, whereas position 669 localizes inside the GUK domain. Due to its localization outside the PDZ2, mutation Arg265Trp may be apparently not relevant for the correct domain fold. Conversely, conservation analysis showed an accessory conserved extension beyond the canonical PDZ2 domain boundaries, suggesting a functional role for this PDZ2 extension. PDZ C-terminus flanking regions are known to frequently assume a regulative role in several PDZ containing proteins, such as improve domain stability and/or modulate the ligand binding specificity (Wang et al., 2010). The second mutation, p.Tyr669Cys, falls at the contact region between strands β2/β3, the two are members of a large five β-sheet, which forms the GUK hydrophobic core. For this domain a crystal structure is available. Molecular dynamics simulation revealed that the mutant protein is less stable than the wild-type. Moreover, GUK domain resulted to adopt a more opened conformation, closer to what observed for protein ZO-3 (Lye et al., 2010). Finally, the substitution of a tyrosine with a cysteine at position 669 causes a remarkable loss of hydrophobic interactions around the mutation area, promoting a clear structural rearrangement. All these predictions led to believe in a damaging effect of both mutations in the protein structure and function, thus causing a possible impairment and rearrangement of the gap

junctions at the intercalated discs. Laing and colleagues in 2007 demonstrated a significant reduction of ZO-1 at the IDs in failing hearts (patient with ischemic disease and dilated cardiomyopathy) coincident with reduced Cx43 staining in these hearts (Laing et al., 2007). The precise mechanism by which ZO-1 regulates gap junction formation is not completely clear, but different studies indicated that there is a link between the presence of adherence junctions and the formation of the gap junctions (Barker et al., 2002; Toyofuku et al., 1998). Subsequently, it can be speculated that loss of ZO-1 at IDs in heart failure may contribute to abnormal impulse propagation and arrhythmogenesis, thereby predisposing patients in heart failure to sudden cardiac death.

5.4 Identification of novel mutations in *CDH2* gene

Two novel nucleotide substitutions (c.1478A>G p.E493G, c.1472T>G p.V491G) were identified in the same proband (n°40b) in the gene encoding for N-cadherin (*CDH2*). The mutations are very close and localized in exon 10 of *CDH2* gene, which encodes for a portion of the extracellular protein domain, reported to determine the basic localization pattern together with E-cadherin in epithelial cells (Ozaki et al., 2010).

N-cadherin is a cell adhesion molecule, which mediates strong homophilic cell-cell adhesion via linkage to the actin cytoskeleton. Different research groups demonstrated the importance of N-cadherin in cardiac myocyte adhesion and myofibrillogenesis. Luo and colleagues reported that N-cadherin-deficient myocytes beat and myofibrils were well formed, but alignment of the myofibrils through regions of cell-cell contact was lost, resulting in their random orientation (Luo et al., 2003). Furthermore, a strong relation and interaction between N-cadherin and gap junction proteins, such as Cx43 was reported (Kostetskii et al., 2005; Li J et al., 2005 and 2008; Swope et al., 2012). Consequently, perturbation of the N-cadherin/catenin complex in heart disease may be an underlying cause, leading to the establishment of the arrythmogenic substrate by destabilizing gap junctions at the cell surface.

Since the proband resulted heterozygote for both mutations, two hypotheses must be considered. The mutations could be carried in *cis*, meaning that the patient has a wild-type allele and a double-mutated allele or in *tras.* In this latter case, the patient resulted to be a compound heterozygote, and this could better explain the manifestation of the disease in this subject and not in his parents. It is well-known that in some ACM cases, concomitant mutations in the same gene or in a second disease-gene are required for the overt clinical phenotype to develop or for modification of disease severity (Bauce et al., 2010). To discriminate between

these two hypotheses, clinical evaluation and genetic test in proband's parents will be necessary.

After the novel description of the intercalated disc (ID) organization, proposed by Franke and colleagues (Franke et al., 2006), ACM is today considered a 'disease of the '*area composita'* rather than a purely 'desmosomal disease', due to the recent involvement of αT-catenin mutations in the disease expression (van Hengel et al., 2013). In addition to the mechanical structures represented by *area composita* and desmosomes, IDs contain also gap junctions, fundamental structures in maintaining the electrical and chemical coupling between cells (Saffiz, 2005). Different evidences support the notion that gap junction formation requires the presence of neighbouring mechanical junctions, likely to provide stability for the docking and/or assembly of functional gap junction channels (Delmar and McKenna, 2010). The emerging picture is that ID is not a summation of independent structures, but rather an "organelle" where macromolecular complexes interact to maintain synchrony within cell populations (Delmar, 2004). In this scenario, novel mutations identified in protein ZO-1 and N-cadherin, two key proteins of the IDs functional unit, appear of particular interest.
Further essential studies for demonstrating the involvement of these candidate genes in the disease are necessary and will come from functional *in vitro* studies. To provide meaningful mechanistic information of novel ACM gene mutations, their effects on protein localization and protein-protein interaction will be analyzed using different cell lines (HL-1 and HEK293T cells).

5.5 Whole exome sequencing

In some familial cases, in which traditional mutation screening failed to identify mutations in the 5 most mutated genes for ACM (*DSP*, *PKP2*, *DSG2*, *DSC2* and *JUP*), the entire coding part of the genome was sequenced for at least two members of each family by using the whole exome sequencing (WES) approach. In Family #3, NGS technique allowed to identify a stop mutation in *DSP* gene (c.3889C>T p.Q1297*), which was not identified with the previous mutation screening. The variant is located in a very wide *DSP* exon (exon 23, 2294 bp) and the analysis by dHPLC was carried out fragmenting the exon in different amplicons. The failure of c.3889C>T mutation detection is probably due to the high denaturing temperature of the region where the variant is located. The mutation co-segregated with the disease phenotype in this family, even if the penetrance is incomplete.

WES analysis in Family #4 allowed to identify a missense variant in *CDH11* (NM_001797, c.1436 A>T, p.N479I) and another one in *TTN* gene (c.97717 C>T p.R32573C), shared by both affected subjects (II-4 and III-5). *CDH11* variant is present in ExAC database with a MAF=0.005% and located in a conserved site (phyloP=3.24). This gene encodes a type II classical cadherin from the cadherin superfamily, integral membrane proteins that mediate calcium-dependent cell-cell adhesion. Other than in osteogenic, chondrogenic and adipogenic differentiation, *CDH11* (OB-cadherin) together with *CDH2* plays an important role in wound healing when fibroblasts turn into myofibroblasts to increase wound contraction and promote wound closure (Alimperti et al., 2015). *CDH11*, but not *CDH2* is also necessary for mesenchymal stem cell differentiation into smooth muscle cell and mediates cell-cell contacts in both valvular fibroblasts and myofibroblasts (Alimperti et al., 2015, Wang et al., 2014). Despite the fact that *CDH11* may be a good candidate gene in ACM, and the mutation p.N479I is predicted damaging by different tools, it is identified also in 6 healthy subjects of Family #4 discouraging its potential pathogenic effect. *TTN* missense variant is also very rare in the general population (dbSNP MAF= 0.02%, ExAC MAF=0.02%) and was found in other 3 family members with no clinical signs. Considering both the variants together with the rare variant in *JUP* gene (c.1942 G>A p.V648I) previously identified, only the two affected subjects (II-4 and III-5) and the obligate carrier (II-5) present the combination of the three. Thus, we can speculate that none of the variants is pathogenic by itself, but the combination of the three could be involved in the genetic determination of the disease in this family. While in the past genetic screening through traditional methods forced researches to select a subset of most candidate genes, now with NGS technology, this process is no more necessary. Consequently, more and more frequently, multiple mutations in known and candidate disease genes are detected in each patient, and the difficulty to assess the pathogenicity to one of them increases. On the other hand, it is not surprising that many cardiomyopathy variants initially published as causative, turn out to be present at low frequency in large exome sequencing data sets and result false positive mutations (Norton et al., 2012).

In Family #5, after the premature death of two young members (III-1, III-3) a genetic test was performed and revealed the presence of a *RYR2* mutation (p.R420W) in subjects II-1, III-1, III-2 and III-3. In the genetic-negative affected mother (II-2), as well as in her affected brother (II-4) and in the healthy father (I-1), WES was performed. Also in this family a novel missense *TTN* variant (c.96592 C>A, p.L32198M) co-segregated with the disease phenotype. Moreover, among the most putative variants identified, a stop-mutation in *CMYA5* gene (NM_1536104, rs185458523, c.10789 A>T, p.K3597*) was selected. This variant

is reported in all considered SNP databases with a MAF that span from 0.04% to 0.19%. This gene encodes for myospryn, a rather novel protein, found to interact with dysbindin, implicated in the molecular pathology of Duchenne muscular dystrophy (DMD) (Benson et al., 2004). Western and northern blot analysis detected myospryn both in skeletal and heart muscle. In the heart, it has been demonstrated that it is localized to intercalated discs and colocalized with desmin in a perinuclear punctate distribution (Kouloumenta et al., 2007). More precisely, the N-terminal domain of mouse desmin interacted with the C-terminal of human myospryn. The p. K3597* mutation detected in this family, causes a formation of a stop codon in exon 4 of the protein composed by 13 total exons, thus losing the C-terminal domain. The C-terminal domain of myospryn was found to interact with the C-terminal domains of titin in human fetal and adult skeletal muscle other than with the muscle-specific protease calpain-3 (CAPN3) (Sarparanta et al., 2010). Despite the fact that c.10789 A>T myospryn variant has a frequency of about 0.1% in the general population, all these findings lead to speculate a possible reduction of the interactions between these proteins at the intercalated discs. In addition, the novel *TTN* mutation found in these patients could contribute to this probable rearrangement at the cellular junctions. To better understand the role of these mutations, a clinical re-evaluation of subjects II-5 and II-6 will be necessary and *in vitro* functional studies would be helpful in the characterization of their possible functional effects.

Due to the reduced costs, the less amount of data obtained and the easier interpretation of the results, targeted NGS approaches remain the most common approaches for genetic diagnosis. Despite the large application of customized gene panels, WES allows to discover novel variants in unsuspected disease genes. It is in fact estimated that only 44.51% of pathogenic mutations are found in the exonic region of candidate genes compared to the theoretical 85% in the entire exome (Biswas et al., 2014).

5.6 Multiple approach

In the present study, a large family (Family #6) in which ACM segregated among 5 generations has been considered and studied by using different genetic approaches.

Traditional mutation screening of the 5 genes most frequently mutated in the proband (IV-9) revealed the presence of a novel mutation near exon 14 splice acceptor site of *PKP2* (c.2578-3T>C). The same mutation is present also in the affected brother (IV-10), but carried by the father (III-8), who is an external family's member. Considering that this mutation has not been previously described and is

located in a splice site of *PKP2* gene, it could be considered a probable pathogenic mutation.

Due to the high number of affected and unaffected subjects in this family, a multipoint linkage analysis has been carried out. Supposing that c.2578-3T>C *PKP2* mutation may have a role in the phenotypic expression in subjects IV-9 and IV-10, the first linkage analysis was performed excluding this branch. The analysis highlighted the presence of a candidate locus on chromosome 11q21 with a maximum parametric LOD score of 2.87. Analysis of haplotype segregation showed a common region shared by all the affected subjects and delimited by crossovers in the affected individual IV-8, refining the critical interval to a 3 cM region. Extending the linkage analysis to the whole family, a new locus on chromosome 19p13.3 reported positive pLODscore values (2.22). Both the analyses have been carried out assuming an "affected-only'' design, which classifies all individuals with no or minor clinical manifestations as "unknown'', thus avoiding decrease of LOD score values due to low penetrance. Last analysis allowed to define a critical region of 7cM shared by all the affected subjects exception made for individual V-5. In both cases the pLOD score value is positive but lower than 3.0, the threshold of statistical significance for accepting linkage.

According to the data obtained by the two linkage analyses performed in the family, two different hypotheses could be taken into account. The first analysis pointed out the identification of a novel locus on chromosome 11q21 shared by all the affected subjects, exception made for the two brothers sharing the same *PKP2* mutation, that could be involved in their phenotypic expression. The second analysis pointed out the identification of a novel locus on chromosome 19p13.3 shared by all the affected subjects, but one. In this last subject, the clinical expression could be due to a mutation in a different locus or could be due to a different disease mimicking ACM.

On the other hand, the two critical regions co-segregated in 5 out of 8 affected subjects but no one of the healthy family members (Figure 4.18). There were no evidences for a specific correlation between the presence of the critical haplotypes and the expression of the phenotype. Moreover, no candidate genes mapped inside these critical regions.

As reported by different studies, traditional linkage analysis was combined to NGS approaches in order to increase the likelihood to discover the genetic cause of the disease expression in this family (Bailey-Wilson and Wilson, 2011; Ott et al., 2015). The whole exomes of 4 affected subjects were sequenced (samples IV-5, IV-8, IV-9, IV-12), but no novel variants shared by all the 4 individuals have been found, neither inside both critical regions, nor in the entire exome. By now, it is well-known that one of the limits of WES approach is the lack coverage in a small

percentage (usually <5%) of the exonic regions, thus resulting in a loss of information (Gilissen et al., 2012; Rehm et al., 2013). Therefore, to guarantee the absence of coding mutations in one of the known-disease genes, the coverage in these genes was checked for the 4 samples and all the exons with a coverage <20X have been re-sequenced.

Only an intronic novel variant located near a splice donor site of *TMEM43* gene (c.766+8C>A) has been identified in subjects IV-5, IV-9 and IV-12. It is possible to exclude its potential pathogenic role due to the genotype-negative affected individuals IV-8 and V-13, other than for the presence of the variant in several family members without disease manifestations.

With the rapid improvements in technology and decreasing costs, whole genome sequencing (WGS) has become a feasible and effective approach for mutation detection. The absence of the capture/enrichment step in the process allows to better analyze all the coding regions other than to identify possible non-coding variants. Moreover, WGS allows more reliable detection of structural variants and copy-number variations (Gilissen et al., 2014; Spielmann and Klopocki, 2013). Recently, it has also been demonstrated that WGS performed better than WES, providing a more even coverage, no strand bias, and a higher proportion of coding regions completely covered (Lelieveld et al., 2015).

For all these reasons, WGS has been performed in two affected members (IV-10 and IV-12) and an healthy individual (IV-3) of Family#6. Firstly, exome data was extrapolated from the genomes and, among the variants shared by the two affected individuals, all those present also in subject IV-3 were excluded. Only variants reporting a MAF ≤0.5% and located in a conserved (phyloP >2.5) coding or splice site regions have been considered as the most probable “causative” mutations. Different filters have also been applied to increase the probability to identify the causative mutation. None of these remaining variants was located in a putative candidate gene and the majority are classified as ‘complex variants’, represented by repetitive elements. In these subjects WGS didn’t detect any novel variant not previously identified with WES. Secondly, due to the presence of two linkage regions on chromosome 11 and chromosome 19, all the coding and non-coding variants inside these regions were checked. In the 7 cM region on chromosome 19p13.3, 56 variants were selected for the frequency (MAF ≤0.5%), but all of them were located in unconserved genomic regions (phyloP <2.5). All the variants mapped in an intronic or UTR region, exception made for a single nucleotide substitution in a splice donor site of *THEG* gene, a gene specifically expressed in the nucleus of haploid male germ cells and then excluded to be associated with ACM. Among the mutated genes *Protein Tyrosine Phosphatase Receptortype Sigma* (PTPRS) appeared the most candidate gene because of its

possible role in intercellular adhesion and its myocardial expression. Three variants located in 3'UTR of this gene have been identified (rs116644020, rs200486129, rs139183127) other than an intronic SNV. Looking at the alignments, two of the UTR variants are close and located in a poly-A stretch, whereas the SNV is located between exon 1 and exon 2, in a site not predicted to be spliced. Actually, it is very difficult to assess a correlation between one of these variants and the expression of the disease.

Finally, positive LODscore values were reported also in a 3cM region on chromosome 11q21, shared by all affected family members with exclusion of subjects IV-9 and IV-10. 378 variants with a MAF≤0.5% have been identified in subject IV-12, with the large representation of intergenic and deeply intronic variants. No putative pathogenic mutations have been identified also in this genomic critical region.

Although WGS is considered the most comprehensive genetic test so far, it obviously has some limitations. The lack of the identification of the causative mutation in this family may be due to low sequencing coverage, bioinformatics variant calling issue or the misinterpretation of the variants. The huge amount of non-coding variants identified in both critical regions as well as into the entire genome remains of difficult interpretation since the knowledge of all protein-coding exons in the genome is still incomplete and the disease gene could be yet unknown or not annotated. Therefore, even though this data does not allow, at present, to perform a genotype-phenotype correlation, it should be considered 'permanent data', that may be re-analysed in the future.

6. CONCLUSIONS

Arrhythmogenic cardiomyopathy is an inherited cardiomyopathy commonly transmitted as an autosomal dominant trait, characterized by incomplete penetrance and variable expressivity.

Due to its significant clinical and genetic heterogeneity, the diagnosis of ACM is challenging and requires a multifaceted approach to patient and families evaluation. Whilst, on the one hand, NGS technologies allows to parallel sequence a large amount of genes, on the other hand, several thousand of genetic variants are identified per patient. It appears evident that the crucial point is no more the identification of a possible causative mutation but the interpretation and validation of the huge amount of variants of uncertain clinical significance (VUS) that we found by using these approaches. As a patient affected with ACM or reporting minor signs of the disease requires life-long follow-up, NGS results need to be re-evaluated in the future considering new insights and published data.

This is especially true for wider NGS strategies, such as whole exome/genome sequencing. Therefore, customized targeted gene panels seem to be the gold standard method for the molecular diagnosis in hereditary cardiomyopathies. While this heterogeneous group of diseases is composed by clinical distinct entities, the genetic overlap between them appears increasingly more evident.

In the present study, targeted gene panel approach pointed out two novel putative candidate genes: *TJP1* and *CDH2*. Both genes encode for proteins localized at the cardiac intercalated disc, which was recently described as a macromolecular complex that acts to maintain synchrony and adhesion between cells. It is within this “organelle” that the desmosome resides and where most ACM-relevant mutations have been found. According to the literature, these findings suggest to consider the overall functional unit of the intercalated disc and its interacting proteins to understand the pathogenetic mechanism of ACM.

7. REFERENCES

Adessi C, Matton G, Ayala G, et al. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res. 2000; 28(20):E87.

Ai D, Fu X, Wang J, et al. Canonical Wnt signaling functions in second heart field to promote right ventricular growth. Proc Natl Acad Sci USA. 2007; 104(22):9319-24.

Alcalai R, Metzger S, Rosenheck S, Meiner V, Chajek-Shaul T. A recessive mutation in desmoplakin causes arrhythmogenic right ventricular dysplasia, skin disorder, and woolly hair. J Am Coll Cardiol. 2003; 42(2):319-27.

Alimperti S and Andreadis ST. *CDH2* and *CDH11* act as regulators of stem cell fate decisions. Stem Cell Research. 2015; 14(3):270-82.

Anderson, J. M. Cell signalling: MAGUK magic. Curr. Biol. 1996; 6(4):382-4.

Angst BD, Nilles LA, Green KJ. Desmoplakin II expression is not restricted to stratified epithelia. J Cell Sci. 1990; 97(2):247–57.

Asano Y, Takashima S, Asakura M, et al. *Lamr1* functional retroposon causes right ventricular dysplasia in mice. Nature genetics. 2004; 36(2):123-130.

Asimaki A, Syrris P, Wichter T, Matthias P, Saffitz JE, McKenna WJ. A novel dominant mutation in plakoglobin causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet. 2007; 81(5):964-73.

Asimaki A, Tandri H, Huang H, et al. A new diagnostic test for arrhythmogenic right ventricular cardiomyopathy. N Engl J Med. 2009; 360(11):1075-84.

Asimaki A, Kléber AG, MacRae CA, Saffitz JE. Arrhythmogenic Cardiomyopathy - New Insights into Disease Mechanisms and Drug Discovery. Prog Pediatr Cardiol. 2014; 37(1-2):3-7.

Aylon Y and Oren M. New plays in the p53 theater. Curr Opin Genet Dev. 2011; 21(1):86-92.

Awad MM, Dalal D, Tichnell C, et al. Recessive arrhythmogenic right ventricular dysplasia due to novel cryptic splice mutation in *PKP2*. Human Mutat. 2006; 27(11):1157.

Azaouagh A, Churzidse S, Konorza T, Erbel R. Arrhythmogenic right ventricular cardiomyopathy/dysplasia: A review and update. Clin Res Cardiol. 2011; 100(5):383-94.

Bagattin A, Veronese C, Bauce B, et al. Denaturing HPLC-based approach for detecting *RYR2* mutations involved in malignant arrhythmias. Clin Chem. 2004; 50(7):1148-55.

Bailey-Wilson JE and Wilson AF. Linkage Analysis in the Next-Generation Sequencing Era. Hum Hered. 2011; 72(4):228-36.

Barker RJ, Price RL, Gourdie RG. Increased association of ZO-1 with connexin43 during remodeling of cardiac gap junctions. Circ Res. 2002; 90(3):317–24.

Basso C, Thiene G, Corrado D, Angelini A, Nava A, Valente M. Arrhythmogenic right ventricular cardiomyopathy. Dysplasia, dystrophy, or myocarditis? Circulation. 1996; 94(5):983-91.

Basso C, Fox PR, Meurs KM, et al. Arrhythmogenic right ventricular cardiomyopathy causing sudden cardiac death in boxer dogs: a new animal model of human disease. Circulation. 2004; 109(9):1180-5.

Basso C and Thiene G. Adipositas cordis, fatty infiltration of the right ventricle, and arrhythmogenic right ventricular cardiomyopathy. Just a matter of fat? Cardiovasc Pathol. 2005; 14(1):37-41.

Basso C, Czarnowska E, Della Barbera M, et al. Ultrastructural evidence of intercalated disc remodelling in arrhythmogenic right ventricular cardiomyopathy: an electron microscopy investigation on endomyocardial biopsies. Eur Heart J. 2006; 27(15):1847-54.

Basso C and Thiene G. Autopsy and endomyocardial biopsy findings. In: Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia. Marcus FI, Nava A, Thiene G, editors. Milan, Italy Springer Verlag; 2007: pp. 29-44.

Basso C, Corrado D, Marcus FI, Nava A, Thiene G. Arrhythmogenic right ventricular cardiomyopathy. Lancet. 2009; 373(9671):1289-300.

Bauce B, Frigo G, Marcus FI, et al. Comparison of clinical features of arrhythmogenic right ventricular cardiomyopathy in men versus women. Am J Cardiol. 2008; 102(9):1252-7.

Bauce B, Nava A, Beffagna G, et al. Multiple mutations in desmosomal proteins encoding genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia. Heart Rhythm. 2010; 7(1):22-29.

Beffagna G, Occhi G, Nava A, et al. Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. Cardiovasc Res. 2005; 65(2):366-73.

Beffagna G, De Bortoli M, Nava A, et al. Missense mutations in desmocollin-2 N-terminus, associated with arrhythmogenic right ventricular cardiomyopathy, affect intracellular localization of desmocollin-2 in vitro. BMC Med Genet. 2007; 8:65.

Bennet PM, Maggs AM, Baines AJ, Pinder JC. The Transitional Junction: A New Functional Subcellular Domain at the Intercalated Disc. Molecular Biology of the Cell. 2006; 17(4):2091-100.

Benson MA, Tinsley CL, Blake DJ. Myospryn is a novel binding partner for dysbindin in muscle. J Biol Chem. 2004; 279(11):10450-8.

Ben-Yosef T and Francomano CA. Characterization of the human talin (TLN) gene: genomic structure, chromosomal localization, and expression pattern. Genomics. 1999; 62(2):316-9.

Ben-Ze'ev A and Geiger B. Differential molecular interactions of beta-catenin and plakoglobin in adhesion, signaling and cancer. Curr Opin Cell Biol. 1998; 10(5):629-39.

Bhonsale A, Groeneweg JA, James CA, et al. Impact of genotype on clinical course in arrhythmogenic right ventricular dysplasia/cardiomyopathy-associated mutation carriers. Eur Heart J. 2015; 36(14):847-55.

Bierkamp C, Mclaughlin KJ, Schwarz H, Huber O, Kemler R. Embryonic heart and skin defects in mice lacking plakoglobin. Dev Biol. 1996; 180(2):780-5.

Biswas A, Rao VR, Seth S, Maulik SK. Next generation sequencing in cardiomyopathy: towards personalized genomics and medicine. Mol Biol Rep. 2014; 41(8):4881-8.

Bonne G, Di Barletta MR, Varnous S, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. Nat Genet. 1999; 21(3):285-8.

Borrmann CM, Grund C, Kuhn C, Hofmann I, Pieperhoff S, Franke WW. The area composita of adhering junctions connecting heart muscle cells of vertebrates. II. Colocalizations of desmosomal and fascia adhaerens molecules in the intercalated disk. Eur J Cell Biol. 2006; 85(6):469-85.

Bowles NE, Ni J, Marcus F, Towbin JA. The detection of cardiotropic viruses in the myocardium of patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy. J Am Coll Cardiol. 2002; 39(5):892-95.

Bruce AF, Rothery S, Dupont E, Severs NJ. Gap junction remodelling in human heart failure is associated with increased interaction of connexin43 with ZO-1. Cardiovasc Res. 2008; 77(4):757-65.

Brun F, Barnes C V, Sinagra G, et al. Titin and desmosomal genes in the natural history of arrhythmogenic right ventricular cardiomyopathy. J Med Genet. 2014; 51(10):669-76.

Burridge K and Mangeat P. An interaction between vinculin and talin. Nature. 1984; 308(5961):744-6.

Calabrese F, Basso C, Carturan E, Valente M, Thiene G. Arrhythmogenic right ventricular cardiomyopathy/dysplasia: is there a role for viruses? Cardiovasc Pathol. 2006; 15(1):11-17.

Calore M, Lorenzon A, De Bortoli M, Poloni G, Rampazzo A. Arrhythmogenic cardiomyopathy: a disease of intercalated discs. Cell Tissue Res. 2015; 360(3):491-500.

Caspi O, Huber I, Gepstein A, et al. Modeling of Arrhythmogenic Right Ventricular Cardiomyopathy With Human Induced Pluripotent Stem Cells. Circ Cardiovasc Genet. 2013; 6(6):557-68.

Cenni V, Sabatelli P, Mattioli E, et al. Lamin A N-terminal phosphorylation is associated with myoblast activation: impairment in Emery-Dreifuss muscular dystrophy. J Med Genet. 2005; 42(3):214-20.

Cerrone M, Noorman M, Lin X, et al. Sodium current deficit and arrhythmogenesis in a murine model of plakophilin-2 haploinsufficiency. Cardiovasc Res. 2012; 95(4):460-8.

Chauveau C, Rowell J, Ferreiro A. A rising titan: *TTN* review and mutation update. Hum Mutat. 2014; 35(9):1046-59.

Chen SN, Gurha P, Lombardi R, Ruggiero A, Willerson JT, Marian a. J. The hippo pathway is activated and is a causal mechanism for adipogenesis in arrhythmogenic cardiomyopathy. Circ Res. 2014; 114(3):454-68.

Chen X, Bonne S, Hatzfeld M, van Roy F, Green KJ. Protein binding and functional characterization of plakophilin 2. Evidence for its diverse roles in desmosomes and beta-catenin signaling. J Biol Chem. 2002; 277(12):10512-22.

Christensen AH, Benn M, Tybjaerg-Hansen A, Haunso S, Svendsen JH. Missense variants in plakophilin-2 in arrhythmogenic right ventricular cardiomyopathy patients-disease-causing or innocent bystanders? Cardiology. 2010; 115(2):148-54.

Christensen AH, Andersen CB, Tybjaerg-Hansen A, Haunso S, Svendsen JH. Mutation analysis and evaluation of the cardiac localization of *TMEM43* in arrhythmogenic right ventricular cardiomyopathy. Clin Genet. 2011; 80(3):256-64.

Corrado D, Thiene G, Nava A, Rossi L, Pennelli N. Sudden death in young competitive athletes: Clinicopathologic correlations in 22 cases. Am J Med. 1990; 89(5):588-96.

Corrado D, Basso C, Thiene G, et al. Spectrum of clinicopathologic manifestations arrhythmogenic right ventricular cardiomyopathy / dysplasia: a multicenter study. J Am Coll Cardiol. 1997; 30(6):1512-20.

Corrado D, Leoni L, Link MS, et al. Implantable cardioverter-defibrillator therapy for prevention of sudden death in patients with arrhythmogenic right ventricular cardiomyopathy/dysplasia. Circulation. 2003; 108(25):3084-91.

Corrado D, Basso C, Leoni L, et al. Three-dimensional electroanatomic voltage mapping increases accuracy of diagnosing arrhythmogenic right ventricular cardiomyopathy/dysplasia. Circulation. 2005; 111(23):3042-50.

Cox MG, van der Smagt JJ, Noorman M, et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy diagnostic task force criteria: impact of new task force criteria. Circ Arrhythm Electrophysiol. 2010; 3(2):126-33.

Dalal D, Nasir K, Bomma C et al. Arrhythmogenic right ventricular dysplasia: A United States experience. Circulation. 2005; 112(25):3823-32.

Dalal D, Molin LH, Piccini J, et al. Clinical features of arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in plakophilin-2. Circulation. 2006; 113(13):1641-9.

Daliento L, Turrini P, Nava A, et al. Arrhythmogenic right ventricular cardiomyopathy in young versus adult patients: similarities and differences. J Am Coll Cardiol. 1995; 25(3):655-64.

d'Amati G, Bagattin A, Bauce B, et al. Juvenile sudden death in a family with polymorphic ventricular arrhythmias caused by a novel *RYR2* gene mutation: Evidence of specific morphological substrates. Hum Pathol. 2005; 36(7):761-7.

Delmar M, Coombs W, Sorgen P, Duffy HS, Taffet SM. Structural bases for the chemical regulation of Connexin43 channels. Cardiovasc Res. 2004; 62(2):268-75.

Delmar M. The intercalated disk as a single functional unit. Heart Rhythm. 2004;1(1):12-13.

Delmar M and Sorgen PL. Molecular organization and regulation of the cardiac gap junction channel Connexin43. 2009. In: Zipes DP, Jalife J (eds) Cardiac electrophysiology: from cell to bedside. Saunders Elsevier, Philadelphia, pp 85–92.

Delmar M and McKenna WJ. The cardiac desmosome and arrhythmogenic cardiomyopathies: From gene to disease. Circ Res. 2010; 107(6):700-14.

den Haan AD, Tan BY, Zikusoka MN, et al. Comprehensive desmosome mutation analysis in north americans with arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Cardiovasc Genet. 2009; 2(5):428-35.

Diekstra A, Bosgoed E, Rikken A, et al. Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. Clin Chem. 2015; 61(1):154-62.

Di Resta C, Pietrelli A, Sala S, et al. High-throughput genetic characterization of a cohort of Brugada syndrome patients. Hum Mol Genet. 2015; 24(20):5828-35.

Dusek RL, Godsel LM, Green KJ. Discriminating roles of desmosomal cadherins: Beyond desmosomal adhesion. J Derm Science. 2007; 45(1):7-21.

Elliott P, Andersson B, Arbustini E, et al. Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. Eur Heart J. 2008; 29(2):270-6.

Elliott P, O'Mahony C, Syrris P, et al. Prevalence of desmosomal protein gene mutations in patients with dilated cardiomyopathy. Circ Cardiovasc Genet. 2010; 3(4):314-22.

Ellis LA, Taylor CF, Taylor GR. A comparison of fluorescent SSCP and denaturing HPLC for high throughput mutation scanning. Hum Mutat. 2000; 15(6):556-64.

Eshkind L, Tian Q, Schmidt A, Franke WW, Windoffer R, Leube RE. Loss of desmoglein 2 suggests essential functions for early embryonic development and proliferation of embryonal stem cells. Eur J Cell Biol. 2002; 81(11):592-8.

Fatkin D, MacRae C, Sasaki T, et al. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction- system disease. N Engl J Med. 1999; 341(23):1715-24.

Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006; 34(3):e22.

Fontaine G, Frank R, Guiraudon G, et al. Signification des troubles de conduction intraventriculaire observes dans la dysplasia ventriculaire droite arhythmogene. Arc Mal Coeur. 1984; 77:872-9.

Forbes MS and Sperelakis N. Intercalated discs of mammalian heart: a review of structure and function. Tissue Cell. 1985; 17(5):605-48.

Fox PR, Maron BJ, Basso C, Liu S, Thiene G. Spontaneously occurring arrhythmogenic right ventricular cardiomyopathy in the domestic cat. A new animal model similar to the human disease. Circulation. 2000; 102(15):1863-70.

Franke WW, Borrmann CM, Grund C, Pieperhoff S. The area composita of adhering junctions connecting heart muscle cells of vertebrates. I. Molecular definition in intercalated disks of cardiomyocytes by immunoelectron microscopy of desmosomal proteins. Eur J Cell Biol. 2006; 85(2):69-82.

Franke WW, Dörflinger Y, Kuhn C, et al. Protein LUMA is a cytoplasmic plaque constituent of various epithelial adherens junctions and composite junctions of myocardial intercalated disks: A unifying finding for cell biology and cardiology. Cell Tissue Res. 2014; 357(1):159-72.

Fressart V, Duthoit G, Donal E, et al. Desmosomal gene analysis in arrhythmogenic right ventricular dysplasia/cardiomyopathy: spectrum of mutations and clinical impact in practice. Europace. 2010; 12(6):861-8.

Fürst DO, Osborn M, Nave R, Weber K. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. J Cell Biol. 1988; 106(5):1563-72.

Gallicano GI, Kouklis P, Bauer C, et al. Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. J Cell Biol. 1998; 143(2):2009-22.

Gallicano GI, Bauer C, Fuchs E. Rescuing desmoplakin function in extra-embryonic ectoderm reveals the importance of this protein in embryonic heart, neuroepithelium, skin, and vasculature. Development. 2001; 128(6):929-41.

Garcia-gras E, Lombardi R, Giocondo MJ, et al. Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhytmogenic right ventricular cardiomyopathy. J Clin Invest. 2006; 116(7):2012-21.

Garrod D, Chidgey M. Desmosome structure, composition and function. Biochim Biophys Acta. 2008; 1778(3):572-87.

Gerull B, Gramlich M, Atherton J, et al. Mutations of *TTN*, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. Nat Genet. 2002; 30(2):201-4.

Gerull B, Heuser A, Wichter T, et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. Nat Genet. 2004; 36(11):1162-4.

Gerull B, Atherton J, Geupel A, et al. Identification of a novel frameshift mutation in the giant muscle filament titin in a large Australian family with dilated cardiomyopathy. J Mol Med. 2006; 84(6):478-83.

Gilissen C, Hoischen A, Brunner HG, Veltman J A. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012; 20(5):490-7.

Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014; 511(7509):344-7.

Golbus JR, Puckelwartz MJ, Fahrenbach JP, et al. Population-Based Variation in Cardiomyopathy Genes. Circ Cardiovasc Genet. 2012; 5(4):391-9.

Gomes J, Finlay M, Ahmed AK, et al. Electrophysiological abnormalities precede overt structural changes in arrhythmogenic right ventricular cardiomyopathy due to mutations in desmoplakin-A combined murine and human study. Eur Heart J. 2012; 33(15):1942-53.

Goossens S, Janssens B, Bonne S, et al. A unique and specific interaction between alphaT-catenin and plakophilin-2 in the area composita, the mixedtype junctional structure of cardiac intercalated discs. J Cell Sci. 2007; 120(12):2126-36.

Groeneweg JA, Bhonsale A, James CA, et al. Clinical Presentation, Long-Term Follow-Up, and Outcomes of 1001 Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy Patients and Family Members. Circ Cardiovasc Genet. 2015; 8(3):437-46.

Grossmann KS, Grund C, Huelsken J, et al. Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. J Cell Biol. 2004; 167(1):149-60.

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet. 2005; 37(5):549-54.

Haghighi K, Kolokathis F, Gramolini AO, et al. A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. Proc Natl Acad Sci U S A. 2006; 103(5):1388-93.

Hamid MS, Norman M, Quraishi A, et al. Prospective evaluation of relatives for familial arrhythmogenic right ventricular cardiomyopathy/dysplasia reveals a need to broaden diagnostic criteria. J Am Coll Cardiol. 2002; 40(8):1445-50.

Hatzfeld M. Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion? Review. Biocm.Biophys Acta. 2006; 1773(1):69-77.

Herman DS, Lam L, Taylor MR, et al. Truncations of titin causing dilated cardiomyopathy. N Engl J Med. 2012; 366(7):619-628.

Heuser A, Plovie ER, Ellinor PT, et al. Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet. 2006; 79(6):1081-8.

Hulot JS, Jouven X, Empana JP, Frank R, Fontaine G. Natural history and risk stratification of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation. 2004; 110(14):1879-84.

Hunter AW, Barker RJ, Zhu C, Gourdie RG. Zonula occludens-1 alters connexin43 gap junction size and organization by influencing channel accretion. Mol Biol Cell. 2005; 16(12): 5686–98.

Ihrie RA, Marques MR, Nguyen BT, et al. Perp is a p63-regulated gene essential for epithelial integrity. Cell. 2005; 120(6):843-856.

Itoh-Satoh M, Hayashi T, Nishi H, et al. Titin mutations as the molecular basis for dilated cardiomyopathy. Biochem Biophys Res Commun. 2002; 291(2):385-393.

Jonkman MF, Pasmooij AM, Pasmans SG, et al. Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa. Am J Hum Genet. 2005; 77(4):653-60.

Kannankeril PJ, Mitchell BM, Goonasekera SA, et al. Mice with the R176Q cardiac ryanodine receptor mutation catecholamine induced ventricular tachycardia and cardiomyopathy. Proc Natl Acad Sci U S A. 2006; 103(32):12179-84.

Kant S, Krusche C, Gaertner A, Milting H, Leube RE. Loss of plakoglobin immunoreactivity in intercalated discs in arrhythmogenic right ventricular cardiomyopathy: Protein mislocalization versus epitope masking. Cardiovasc Res. 2016; 109(2):260-71.

Kaplan SR, Gard JJ, Carvajal-Huerta L, Ruiz-Cabezas JC, Thiene G, Saffitz JE. Structural and molecular pathology of the heart in Carvajal syndrome. Cardiovasc Pathol. 2004a; 13(1):26-32.

Kaplan SR, Gard JJ, Protonotarios N, et al. Remodeling of myocyte gap junctions in arrhythmogenic right ventricular cardiomyopathy due to a deletion in plakoglobin (Naxos disease). Heart Rhythm. 2004b; 1(1):3-11.

Kapoun AM, Liang F, O'Young G, et al. B-type natriuretic peptide exerts broad functional opposition to transforming growth factor-beta in primary human cardiac fibroblasts: fibrosis, myofibroblast conversion, proliferation, and inflammation. Circ Res. 2004; 94(4):453-61.

Kim C, Wong J, Wen J, et al. Studying arrhythmogenic right ventricular dysplasia with patient-specific iPSCs. Nature. 2013; 494(7435):105-10.

Kirchhof P, Fabritz L, Zwiener M, et al. Age- and training-dependent development of arrhythmogenic right ventricular cardiomyopathy in heterozygous plakoglobin-deficient mice. Circulation. 2006; 114(17):1799-806.

Klauke B, Kossmann S, Gaertner A, et al. De novo desmin-mutation N116S is associated with arrhythmogenic right ventricular cardiomyopathy. Hum Mol Genet. 2010; 19(23):4595-4607.

Klymkowsky MW, Williams BO, Barish GD, Varmus HE, Vourgourakis YE. Membrane-anchored plakoglobins have multiple mechanisms of action in Wnt signaling. Mol Biol Cell. 1999; 10(10):3151-69.

Kostetskii I, Li J, Xiong Y, et al. Induced deletion of the N-cadherin gene in the heart leads to dissolution of the intercalated disc structure. Circ Res. 2005; 96(3):346-54.

Kouloumenta A, Mavroidis M, Capetanaki Y. Proper perinuclear localization of the TRIM-like protein myospryn requires its binding partner desmin. J Biol Chem. 2007; 282(48):35211-21.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996; 58(6):1347-63.

Krusche CA, Holthofer B, Hofe V, et al. Desmoglein 2 mutant mice develop cardiac fibrosis and dilation. Basic Res Cardiol. 2011; 106(4):617-33.

Lacroix D, Lions C, Klug D, Prat A. Arrhythmogenic right ventricular dysplasia: catheter ablation, MRI, and heart trans- plantation. J Cardiovasc Electrophysiol. 2005; 16(2):235-6.

Lahtinen AM, Lehtonen A, Kaartinen M, et al. Plakophilin-2 missense mutations in arrhythmogenic right ventricular cardiomyopathy. Int J Cardiol. 2008; 126(1):92-100.

Laing JG, Saffitz JE, Steinberg TH, Yamada KA. Diminished zonula occludens-1 expression in the failing human heart. Cardiovascular Pathology. 2007; 16(3):159-64.

Lakdawala NK, Dellefave L, Redwood CS, et al. Familial dilated cardiomyopathy caused by an alpha-tropomyosin mutation: the distinctive natural history of sarcomeric dilated cardiomyopathy. J Am Coll Cardiol. 2010; 55(4):320-9.

Lammerding J, Schulze PC, Takahashi T, et al. Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction. J Clin Invest. 2004; 113(3):370-8.

Landstrom AP, Parvatiyar MS, Pinto JR, et al. Molecular and functional characterization of novel hypertrophic cardiomyopathy susceptibility mutations in TNNC1-encoded troponin C. J Mol Cell Cardiol. 2008; 45(2):281-8.

Lelieveld SH, Spielmann M, Mundlos S, Veltman J a., Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Hum Mutat. 2015; 36(8):815-22.

LeWinter MM and Granzier H. Cardiac Titin - A Multifunctional Giant. Circulation. 2010; 121(19):2137-45.

Li D, Liu Y, Maruyama M, et al. Restrictive loss of plakoglobin in cardiomyocytes leads to arrhythmogenic cardiomyopathy. Hum Mol Genet. 2011; 20(23):4582-96.

Li J, Patel VV, Kostetskii I, et al. Cardiac-Specific Loss of N-Cadherin Leads to Alteration in Connexins With Conduction Slowing and Arrhythmogenesis. Circ Res. 2005; 97(5):474-81.

Li J, Levin MD, Xiong Y, Petrenko N, Patel V V, Radice GL. N-cadherin haploinsufficiency affects cardiac gap junctions and arrhythmic susceptibility. J Mol Cell Cardiol. 2008; 44(3):597-606.

Li J, Swope D, Raess N, Cheng L, Muller EJ, Radice GL. Cardiac tissue-restricted deletion of plakoglobin results in progressive cardiomyopathy and activation of {beta}-catenin signaling. Mol Cell Biol. 2011; 31(6):1134-44.

Liang WC, Mitsuhashi H, Keduka E, et al. *TMEM43* Mutations in Emery-Dreifuss Muscular Dystrophy-Related Myopathy. Am Neurol Ass. 2011; 69(6):1005-13.

Lindner TH and Hoffmann K. EasyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analysis. Bioinformatics. 2005; 21(3):405-7.

Lobo F, Silver MD, Butany J et al. Left ventricular involvement in right ventricular dysplasia/cardiomyopathy. Can J Cardiol. 1999; 15(11):1239-47.

Lombardi R and Marian AJ. Molecular genetics and pathogenesis of arrhythmogenic right ventricular cardiomyopathy: A disease of cardiac stem cells. Pediatr Cardiol. 2011; 32(3):360-5.

Lorenzon A, Pilichou K, Rigato I, et al. Homozygous Desmocollin-2 Mutations and Arrhythmogenic Cardiomyopathy. Am J Cardiol. 2015; 116(8):1245-51.

Luo Y, Radice GL. Cadherin-mediated adhesion is essential for myofibril continuity across the plasma membrane but not for assembly of the contractile apparatus. J Cell Sci,2003; 116(8):1471-9.

Lye MF, Fanning AS, Su Y, Anderson JM, Lavie A. Insights into Regulated Ligand Binding Sites from the Structure of ZO-1 Src Homology 3-Guanylate Kinase Module. J Biol Chem. 2010; 285(18):13907-17.

Ma D, Wei H, Lu J, et al. Generation of patient-specific induced pluripotent stem cell-derived cardiomyocytes as a cellular model of arrhythmogenic right ventricular cardiomyopathy. Eur Heart J. 2013; 34(15):1122-33.

Marcus FI, Fontaine GH, Guiraudon G, et al. Right ventricular dysplasia: a report of 24 adult cases. Circulation. 1982; 65(2):384-98.

Marcus FI, McKenna WJ, Sherrill D, et al. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia Proposed Modification of the Task Force Criteria. Eur Heart J. 2010; 31(7):806-14.

Maron BJ, Towbin JA, Thiene G, et al.; American Heart Association; Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translation Biology Interdisiplinary Working Groups, Council on Epidemiology and Prevention. Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional

Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. Circulation. 2006; 113(14):1807-16.

Martherus R, Jain R, Takagi K, et al. Accelerated cardiac remodeling in desmoplakin transgenic mice in response to endurance exercise is associated with perturbed Wnt/β-catenin signaling. Am J Physiol Heart Circ Physiol. 2016; 310(2):H174-87.

Martin ED, Moriarty MA, Byrnes L, Grealy M. Plakoglobin has both structural and signalling roles in zebrafish development. Dev Biol. 2009; 327(1):83-96.

McKenna WJ, Thiene G, Nava A, et al. Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society. Br Heart J. 1994; 71(3):215-8.

McKoy G, Protonotarios N, Crosby A, et al. Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). Lancet. 2000; 355(9221):2119-24.

McLendon PM and Robbins J. Desmin-related cardiomyopathy: an unfolding story. Am J Physiol Heart Circ Physiol. 2011; 301(4):H1220-8.

Merner ND, Hodgkinson KA, Haywood AF, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the *TMEM43* gene. Am J Hum Genet. 2008; 82(4):809-21.

Mertens C, Hofmann I, Wang Z, et al. Nuclear particles containing RNA polymerase III complexes associated with the junctional plaque protein plakophilin 2. Proc Natl Acad Sci USA. 2001; 98(14):7795-800.

Meurs KM, Ederer MM, Stern JA. Desmosomal gene evaluation in Boxers with arrhythmogenic right ventricular cardiomyopathy. Am J Vet Res. 2007; 68(12):1338-41.

Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988; 16(3):1215.

Missiaen L, Robberecht W, van den Bosch L, et al. Abnormal intracellular Ca2+ homeostasis and disease. Cell Calcium. Review.2000; 28(1):1-21.

Monserrat L, Ortiz-genga M, Lesende I, et al. Genetics of cardiomyopathies: novel perspectives with next generation sequencing. Curr Pharm Des. 2015; 21(4):418-30.

Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955; 7(3):277-318.

Nasir K, Bomma C, Tandri H, et al. Electrocardiographic features of arrhythmogenic right ventricular dysplasia/cardiomyopathy according to disease severity: a need to broaden diagnostic criteria. Circulation. 2004; 110(12):1527-34.

Nava A, Thiene G, Canciani B, et al. Familial occurrence of right ventricular dysplasia: a study involving nine families. J Am Coll Cardiol. 1988; 12(5): 1222-8.

Norgett EE, Hatsell SJ, Carvajal-Huerta L, et al. Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. Hum Mol Genet. 2000; 9(18):2761-6.

Norton N, Robertson P, Rieder M, et al.; National Heart, Lung and Blood Institute GO Exome Sequencing Project. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circ Cardiovasc Genet. 2012; 5(2):167-74.

Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015; 16(5):275-284.

Otten E, Asimaki A, Maass A, et al. Desmin mutations as a cause of right ventricular heart failure affect the intercalated disks. Heart Rhythm. 2010; 7(8):1058-64.

Ozaki C, Obata S, Yamanaka H, Tominaga S, Suzuki ST. The extracellular domains of E- and N-cadherin determine the scattered punctate localization in epithelial cells and the cytoplasmic domains modulate the localization. J Biochem. 2010; 147(3):415-25.

Palatinus JA, O'Quinn MP, Barker RJ, Harris BS, Jourdan J, Gourdie RG. ZO-1 determines adherens and gap junction localization at intercalated disks. Am J Physiol Heart Circ Physiol. 2011; 300(2):H583-94.

Palka H and Green K. Roles of plakoglobin end domains in desmosome assembly. J Cell Sci. 1997; 110(19):2359-71.

Paris M, Rouleau M, Pucéat M, Aberdam D. Regulation of skin aging and heart development by TAp63. Cell Death Differ. 2012;19(2):186-93.

Perazzolo Marra M, Rizzo S, Bauce B, De Lazzari M, et al. Arrhythmogenic right ventricular cardiomyopathy. Contribution of cardiac magnetic resonance imaging to the diagnosis. Herz. 2015; 40(4):600-6.

Perrot A, Hussein S, Ruppert V, et al. Identification of mutational hot spots in *LMNA* encoding lamin A/C in patients with familial dilated cardiomyopathy. Basic Res Cardiol. 2009; 104(1):90-9.

Pieperhoff S and Franke WW. The area composita of adhering junctions connecting heart muscle cells of vertebrates. VI. Different precursor structures in non-mammalian species. Eur J Cell Biol. 2008; 87(7):413-30.

Pilichou K, Nava A, Basso C, et al. Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy. Circulation. 2006; 113(9):1171-79.

Pilichou K, Remme CA, Basso C, et al. Myocyte necrosis underlies progressive myocardial dystrophy in mouse *DSG2*-related arrhythmogenic right ventricular cardiomyopathy. J Exp Med. 2009; 206(8):1787-802.

Pinamonti B, Di Lenarda A, Sinagra G, Silvestri F, Bussani R, Camerini F. Long-term evolution of right ventricular dysplasia-cardiomyopathy. The Heart Muscle Disease Study Group. Am Heart J. 1995; 129(2):412-5.

Pinamonti B, Miani D, Sinagra G et al. Familial right ventricular dysplasia with biventricular involvment and inflammatory infiltration. Heart Muscle Disease Study Group. Heart. 1996; 76(1):66-9.

Pinamonti B, Pagnan L, Bussani R et al. Right ventricular dysplasia with biventricular involvement. Circulation. 1998; 98(18):1943-5.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010; 20(1):110-21.

Posch MG, Posch MJ, Geier C, et al. A missense variant in desmoglein-2 predisposes to dilated cardiomyopathy. Mol Genet Metab. 2008; 95(1-2):74-80.

Priori SG, Napolitano C, Tiso N, et al. Mutations in the cardiac ryanodine receptor gene (h*RYR2*) underlie catecholaminergic polymorphic ventricular tachycardia. DNA Seq. 2001; 103(2):196-200.

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81(3):559-75.

Quarta G, Syrris P, Ashworth M, et al. Mutations in the Lamin A/C gene mimic arrhythmogenic right ventricular cardiomyopathy. Eur Heart J. 2012; 33(9):1128-36.

Rampazzo A, Nava A, Malacrida S, et al. Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet. 2002; 71(5):1200-06.

Rampazzo a., Calore M, van Hengel J, van Roy F. Intercalated Discs and Arrhythmogenic Cardiomyopathy. Circ Cardiovasc Genet. 2014; 7(6):930-40.

Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet. 2013; 14(4):295-300.

Robinson P, Griffiths PJ, Watkins H, Redwood CS. Dilated and hypertrophic cardiomyopathy mutations in troponin and alpha-tropomyosin have opposing effects on the calcium affinity of cardiac thin filaments. Circ Res. 2007; 101(12):1266-73.

Romero J, Mejia-Lopez E, Manrique C, Lucariello R. Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC/D): A Systematic Literature Review. Clin Med Insights Cardiol. 2013; 7:97 114.

Ross SE, Hemati N, Longo KA, et al. Inhibition of adipogenesis by Wnt signaling. Science. 2000; 289(5481):950-3.

Ruiz P, Brinkmann V, Ledermann B, et al. Targeted mutation of plakoglobin in mice reveals essential functions of desmosomes in the embryonic heart. J Cell Biol. 1996; 135(1):215-25.

Saffitz JE. Dependence of electrical coupling on mechanical coupling in cardiac myocytes. Insights gained from cardiomyopathies caused by defects in cell-cell communication. Ann NY Acad Sci. 2005; 1047:336-44.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74(12):5463-7.

Sarparanta J, Blandin G, Charton K, et al. Interactions with M-band titin and calpain 3 link myospryn (*CMYA5*) to tibial and limb-girdle muscular dystrophies. J Biol Chem. 2010; 285(39):30304-315.

Sato PY, Musa H, Coombs W, et al. Loss of plakophilin-2 expression leads to decreased sodium current and slower conduction velocity in cultured cardiac myocytes. Circ Res. 2009; 105(6):523-26.

Sato PY, Coombs W, Lin X, et al. Interactions between ankyrin-G, plakophilin-2, and connexin43 at the cardiac intercalated disc. Circ Res. 2011; 109(2):193-201.

Sen-Chowdhry S, Syrris P, Ward D, Asimaki A, Sevdalis E, McKenna WJ. Clinical and Genetic Characterization of Families With Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy Provides Novel Insights Into Patterns of Disease Expression. Circulation. 2007; 115(13):1710-20.

Shendure J and Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008; 26(10):1135-45.

Sikkema-Raddatz B, Johansson LF, de Boer EN, et al. Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics. Hum Mutat. 2013; 34(7):1035-42.

Simpson MA, Mansour S, Ahnood D, et al. Homozygous mutation of desmocollin-2 in arrhythmogenic right ventricular cardiomyopathy with mild palmoplantar keratoderma and woolly hair. Cardiology. 2009; 113(1):28-34.

Spielmann M, Klopocki E. CNVs of noncoding cis-regulatory elements in human disease. Curr Opin Genet Dev. 2013; 23(3):249-56.

Stokes DL and Wagenknecht T. Calcium transport across the sarcoplasmic reticulum: structure and function of Ca2+-ATPase and the ryanodine receptor. Eur J Biochem. Review. 2000; 267(17):5274-9.

Swope D, Cheng L, Gao E, Li J, Radice GL. Loss of Cadherin-Binding Proteins β-Catenin and Plakoglobin in the Heart Leads to Gap Junction Remodeling and Arrhythmogenesis. Mol Cell Biol. 2012; 32(6):1056-67.

Syrris P, Ward D, Evans A, et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. Am J Hum Genet. 2006; 79(5): 978-84.

Tandri H, Saranathan M, Rodriguez ER, et al. Noninvasive detection of myocardial fibrosis in arrhythmogenic right ventricular cardiomyopathy using delayed-enhancement magnetic resonance imaging. J Am Coll Cardiol. 2005; 45(1):98-103.

Tandri H, Macedo R, Calkins H, et al. Role of magnetic resonance imaging in arrhythmogenic right ventricular dysplasia: insights from the North American arrhythmogenic right ventricular dysplasia (ARVD/C) study. Am Heart J. 2008; 155(1):147-53.

Taylor M, Graw S, Sinagra G, et al. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. Circulation. 2011; 124(8):876-85.

Thiele H and Nurnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics. 2005; 21(8):1730-2.

Thiene G, Nava A, Corrado D, et al. Right ventricular cardiomyopathy and sudden death in young people. N Engl J Med. 1988; 318:129-33.

Thiene G, Nava A, Angelini A, Daliento L, Scognamiglio R, Corrado D. Anatomoclinical aspects of arrhythmogenic right ventricular cardiomyopathy. In Advances in cardiomyopathies. Edited by Baroldi G, Camerini F, Goodwin JF.Milano: Springer Verlag; 1990:397-408.
Thiene G, Corrado D, Nava A, et al. Right ventricular cardiomyopathy: is there evidence of an inflammatory aetiology? Eur Heart J. 1991; 12 Suppl D: 22-5.

Thiene G, Basso C, Danieli GA, Rampazzo A, Corrado D, Nava A. Arrhythmogenic right ventricular cardiomyopathy: a still underrecognised clinical entity. Trends Cardiovascular Med. 1997; 7(3):84-90.

Thiene G, Corrado D, Basso C. Arrhythmogenic right ventricular cardiomyopathy/dysplasia. Orphanet J Rare Dis. 2007; 2(1):45.

Thiene G, Nava A, Angelini A et al. Anatomo-clinical aspects of arrhythmogenic right ventricular cardiomyopathy. In: Baroldi G, Camerini F, Goodwin JF (eds) Advances in cardiomyopathies. Springer Verlag, pp 397-408.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14(2):178-92.

Tiso N, Stephan DA, Nava A, et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). Hum Mol Genet. 2001; 10(3):189-94.

Toyofuku T, Yabuki M, Otsu K, Kuzuya T, Hori M, Tada M. Direct association of the gap junction protein connexin-43 with ZO-1 in cardiac myocytes. J Biol Chem. 1998; 273(21):12725-31.

Turrini P, Corrado D, Basso C, Nava A, Bauce B, Thiene G. Dispersion of ventricular depolarization-repolarization: a non-invasive marker for risk stratification in arrhythmogenic right ventricular cardiomyopathy. Circulation. 2001; 103(25):3075-80.

Uhl HS. A previously undescribed congenital malformation of the heart: almost total absence of the myocardium of the right ventricle. Bull John Hopkins Hosp. 1952; 91(3):197-209.

van Tintelen JP, Entius MM, Bhuiyan ZA, et al. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation. 2006; 113(13):1650-8.

Valente M, Calabrese F, Thiene G, et al. In vivo evidence of apoptosis in arrhythmogenic right ventricular cardiomyopathy. Am J Pathol. 1998; 152(2):479-84.

Valenzise M, Arrigo T, De Luca F, et al. R298Q mutation of p63 gene in autosomal dominant ectodermal dysplasia associated with arrhythmogenic right ventricular cardiomyopathy. Eur J Med Genet. 2008; 51(5):497-500.
van Bokhoven H, McKeon F. Mutations in the p53 homolog p63: allele-specific developmental syndromes in humans. Trends Mol Med. 2002; 8(3):133-9.

van der Zwaag PA, Jongbloed JD, van den Berg MP, et al. A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. Hum Mutat. 2009; 30(9):1278-83.

van der Zwaag PA, van Rijsingen IAW, Asimaki A, et al. Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. Eur J Heart Fail. 2012; 14(11):1199-1207.

van Hengel J, Calore M, Bauce B, et al. Mutations in the area composita protein alphat-catenin are associated with arrhythmogenic right ventricular cardiomyopathy. Eur Heart J. 2013; 34(3):201-10.

van Spaendonck-Zwarts KY, van der Kooi AJ, van den Berg MP, et al. Recurrent and founder mutations in the Netherlands: the cardiac phenotype of *DES* founder mutations p.S13F and p.N342D. Neth Heart J. 2012; 20(5):219-28.

van Tintelen JP, Entius MM, Bhuiyan ZA, et al. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation. 2006; 113(13):1650-8.

van Tintelen JP, Van Gelder IC, Asimaki A, et al. Severe cardiac phenotype with right ventricular predominance in a large cohort of patients with a single missense mutation in the *DES* gene. Heart Rhythm. 2009; 6(11):1574-83.

Wang CK, Pan L, Chen J, Zhang M. Extensions of PDZ domains as important structural and functional elements. Protein Cell. 2010; 1(8):737-51.

Wang H, Leinwand LA, Anseth KS. Roles of transforming growth factor-β1 and OB-cadherin in porcine cardiac valve myofibroblast differentiation. FASEB J. 2014; 28(10):4551-62.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38(16):e164.

Whittock NV, Ashton GH, Dopping-Hepenstal PJ, et al. Striate palmoplantar keratoderma resulting from desmoplakin haploinsufficiency. J Invest Dermatol. 1999; 113(6):940-6.

Wichter T, Paul M, Wollmann C, et al. Implantable cardioverter/defibrillator therapy in arrhythmogenic right ventricular cardiomyopathy. Single-center experience of long-term folow-up and complications in 60 patients. Circulation. 2004; 109(12):1503-8.

Xu T, Yang Z, Vatta M, et al. Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy. J Am Coll Cardiol. 2010; 55(6):587-97.

Yang A, Kaghad M, Wang Y, et al. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. Mol Cell. 1998; 2(3):305-16.

Yang Z, Bowles NE, Scherer SE, et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Res. 2006; 99(6):646-55.

Yoerger DM, Marcus F, Sherrill D, et al. Echocardiographic findings in patients meeting task force criteria for arrhythmogenic right ventricular dysplasia: new insights from the multidisciplinary study of right ventricular dysplasia. Am Coll Cardiol. 2005; 15;45(6):860-5.

Yoshida M, Romberger DJ, Illig MG, et al. Transforming growth factor-beta stimulates the expression of desmosomal proteins in bronchial epithelial cells. Am J Respir Cell Mol Biol. 1992; 6(4):439-45.

Zhurinsky J, Shtutman M and Avri Ben-Ze'ev. Plakoglobin and β-catenin: protein interactions, regulation and biological roles. Journal of Cell Science. 2000; 113(18):3127-39.

Zou J, Cao K, Yang B, et al. Dynamic substrate mapping and ablation of ventricular tachycardias in right ventricular dysplasia. J Interv Card Electrophysiol. 2004; 11(1):37-45.

APPENDIX A

Table 1. Genes included in the 'Cardiomyopathies gene panel'

Gene	NM Refseq	Related disease
PLN	NM_002667	DCM; HCM; ACM
DES	NM_001927	DCM; ACM
LMNA	NM_001282625	DCM; ACM; LVNC
MYBPC3	NM_000256	DCM; HCM; LVNC
MYH7	NM_000257	DCM; HCM; LVNC
TNNT2	NM_001276345; NM_001001432; NM_001276347	DCM; HCM; LVNC
TNNI3	NM_000363	DCM; HCM; LVNC
TPM1	NM_001018004; NM_001018006; NM_001018007; NM_001018020; NM_000366; NM_001018005;NM_001018008;NM_001018020	DCM; HCM; LVNC
TNNC1	NM_003280	DCM; HCM
MYH6	NM_002471	DCM; HCM
VCL	NM_014000	DCM; HCM
TAZ	NM_000116	DCM; LVNC
LDB3	NM_001171610; NM_001080116; NM_007078	DCM; HCM; LVNC
SCN5A	NM_198056; NM_001099404	DCM; BrS; LQTS; LVNC
PSEN1	NM_000021	DCM
PSEN2	NM_000447	DCM
SGCD	NM_000337	DCM
ACTC1	NM_005159	DCM; HCM; LVNC
ABCC9	NM_020297; NM_005691	DCM; BrS
DMD	NM_000109	DCM; LVNC
ANKRD1	NM_014391	DCM
NEXN	NM_144573	DCM
CSRP3	NM_003476	DCM; HCM
TCAP	NM_003673	DCM; HCM
ACTN2	NM_001103; NM_001278344	DCM; HCM
PRKAG2	NM_016203	HCM
MYOZ2	NM_016599	HCM
MYL2	NM_000432	HCM
MYL3	NM_000258	HCM
CACNA2D1	NM_000722	SQTS; BrS
LAMP2	NM_013995; NM_002294; NM_001122606	HCM; DCM
CASQ2	NM_001232	CPVT
PKP2	NM_004572	ACM; DCM; BrS
DSP	NM_004415	ACM; DCM; LVNC
DSG2	NM_001943	ACM; DCM
DSC2	NM_024422	ACM; DCM
JUP	NM_002230	ACM; DCM
TGFB3	NM_003239	ACM
CTNNA3	NM_013266	ACM
TMEM43	NM_024334	ACM
TTN	NM_001267550	ACM ; DCM; HCM
RYR2	NM_001035	CPVT; DCM; ACM
KCNQ1	NM_000218; NM_181798	LQTS, SQTS
KCNH2	NM_000238; NM_001204798	LQTS, SQTS
CAV3	NM_033337	LQTS, DCM
ANK2	NM_001127493; NM_001148	LQTS, CPVT
KCNE1	NM_000219; NM_001270405; NM_001127668; NM_001127670	LQTS
KCNE2	NM_172201	LQTS
KCNJ2	NM_000891	LQTS, SQTS, CPVT
CACNA1C	NM_001167624; NM_001167625; NM_199460; NM_001129829	BrS, LQTS, SQTS,
GPD1L	NM_015141	BrS
CACNB2	NM_201597; NM_001167945; NM_000724; NM_201590; NM_201590; NM_201570	BrS, SQTS
SCN1B	NM_001037; NM_199037	BrS
KCNE3	NM_005472	BrS
SCN3B	NM_018400	BrS
KCND3	NM_004980	BrS

Table 2. Genes included into 'ACM known and candidate genes panel'

Gene	NM Refseq
PLN	NM_002667
DES	NM_001927
LMNA	NM_001282625
TAZ	NM_000116
SCN5A	NM_198056; NM_001099404
ACTC1	NM_005159
ACTN2	NM_001103; NM_001278344
MYL2	NM_000432
MYL3	NM_000258
CASQ2	NM_001232
PKP2	NM_004572
DSP	NM_004415
DSG2	NM_001943
DSC2	NM_024422
JUP	NM_002230
TGFB3	NM_003239
CTNNA3	NM_013266
TMEM43	NM_024334
TTN	NM_001267550
RYR2	NM_001035
KCNE1	NM_000219; NM_001270405; NM_001127668; NM_001127670
CDH2	NM_001792
CTNNA1	NM_001903
CTNNB1	NM_001904
MYZAP	NM_001018100
CTNND1	NM_001206885; NM_001206886
TJP1	NM_003257
ARVCF	NM_001670
TJP2	NM_001170414; NM_001170630; NM_001170416; NM_001170415.1
TP63	NM_003722; NM_001114982
PERP	NM_022121
GJA1	NM_000165
GJD4	NM_153368
GJC1	NM_005497; NM_001080383
JAM3	NM_032801
XIRP1	NM_194293
CASR	NM_000388; NM_001178065
CDH11	NM_001797
LAMR1/RPSA	NM_002295
PTPLA	NM_014241
PNN	NM_002687
PKP4	NM_003628
KCNJ11	NM_000525, NM_001166290
PRKCA	NM_002737
NF2	NM_181825.2, NM_016418.5
YAP1	NM_001195045.1, NM_001282101.1
TEAD1	NM_021961.5
FAT1	NM_005245
DCHS1	NM_003737
ATP2A2	NM_170665, NM_001681
TLN1	NM_006289
STRN	NM_003162
DLG1	NM_001290983; NM_001204387.1 ; NM_001204386.1
MST1	NM_020998
MST2/STK3	NM_001256312.1; NM_006281.3
EB1/MAPRE1	NM_012325
AXIN1	NM_003502.3
AXIN2	NM_004655
GSK3B	NM_002093.3
KAZN	NM_201628.2, NM_001018000.3, NM_001018001.2, NM_001017999.2
HSPA2	NM_021979

ANK3	NM_001149.3, NM_020987.3, NM_001204403.1, NM_001204404.1
TNNC1	NM_003280
GPD1L	NM_015141
TNNI3	NM_000363
ABCC9	NM_020297; NM_005691
VCL	NM_014000
TPM1	NM_001018004; NM_001018006; NM_001018007; NM_001018020; NM_000366; NM_001018005; NM_001018008
PSEN2	NM_000447

APPENDIX B

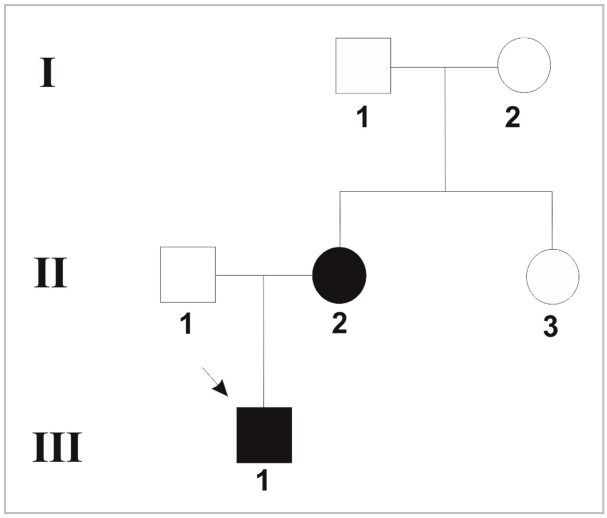


Figure 1. Pedigree of Family #1. Proband is indicated by the arrow. Both subject III-1 and II-2 present a clear ACM form reporting infundibular arrhythmias and late potentials (III-1) and fibrous substitution at RMN (II-2). DNA sample from subject II-3 is available, but not clinical data.

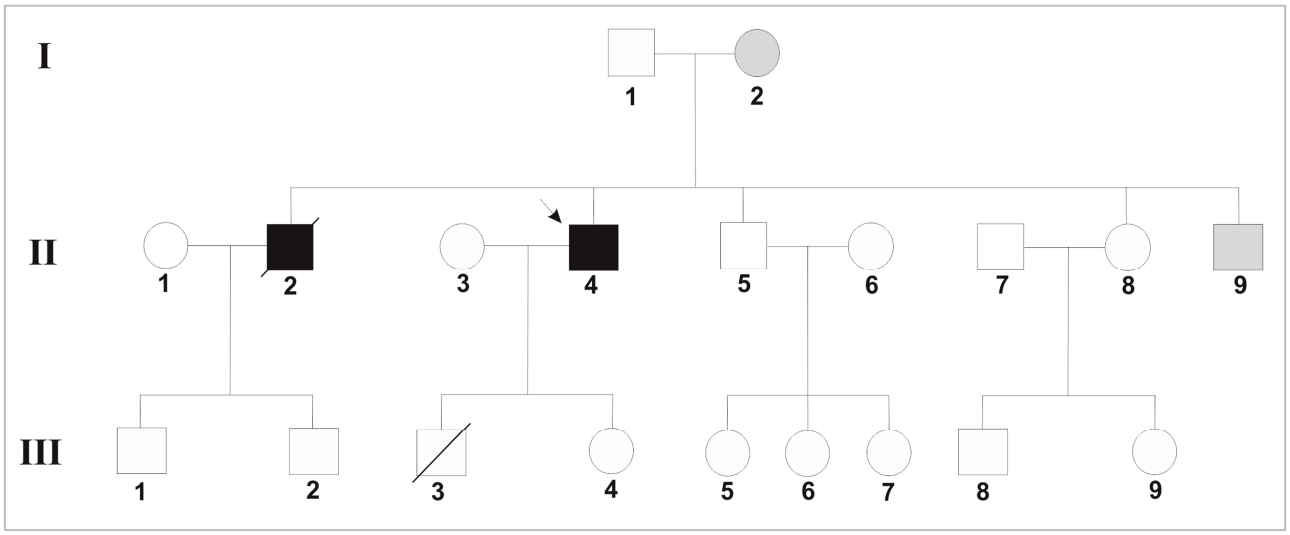


Figure 2. Pedigree of Family #2. Proband is indicated by the arrow and present a moderate ACM form with initial biventricular involvement. Black and white symbols represent clinically affected and unaffected individuals, respectively. Deceased individual II-2 presented negative T waves at ECG and experienced 3 syncope before death. Grey symbols denote subjects with minor signs of the disease. Subject III-3 died at young age but cardiac involvement was excluded.

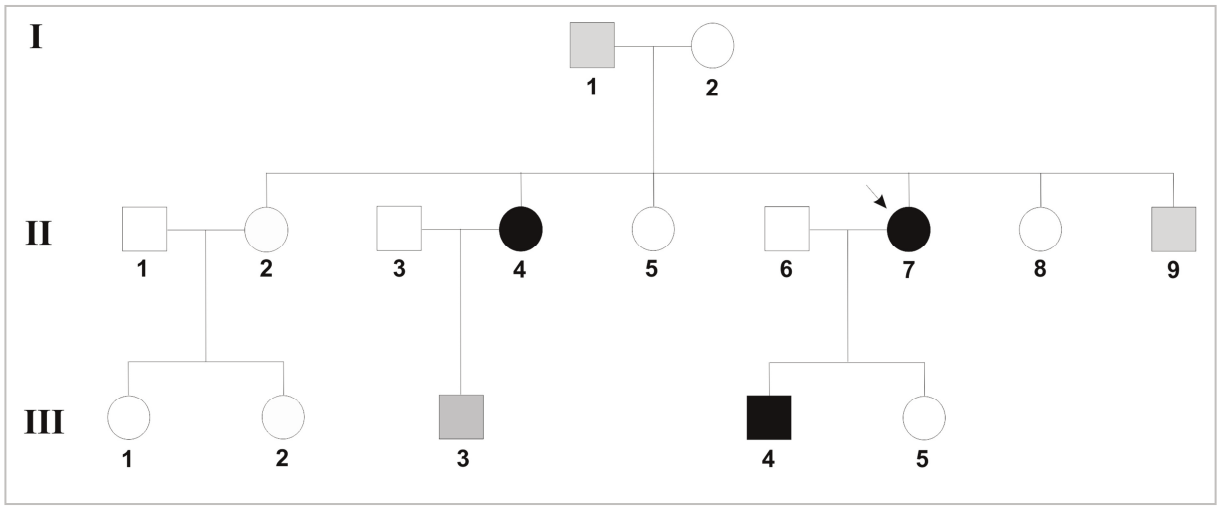


Figure 3. Pedigree of Family #3. Proband is indicated by the arrow and present a severe ACM form with sustained ventricular tachycardia and biventricular dilatation with kinetic abnormalities in both ventricles. Black symbols represent clinically affected individuals (II-4 and III-4) as having biventricular form of the disease. Grey symbols denote subjects with minor signs of the disease: RV conduction delay at ECG (III-3) and premature ventricular beats (II-9); or not investigated subjects(I-2).

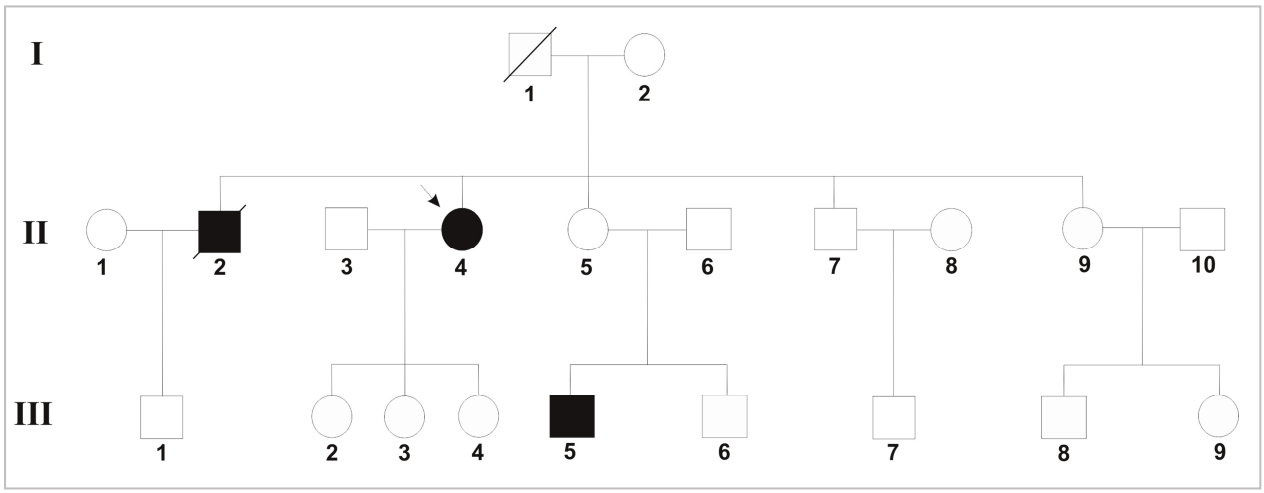


Figure 4. Pedigree of Family #4. Proband is indicated by the arrow and reported minor alterations at echocardiography and inverted T waves. Subject II-2, died suddenly aged 35 and was diagnosed affected with ACM at autopsy.

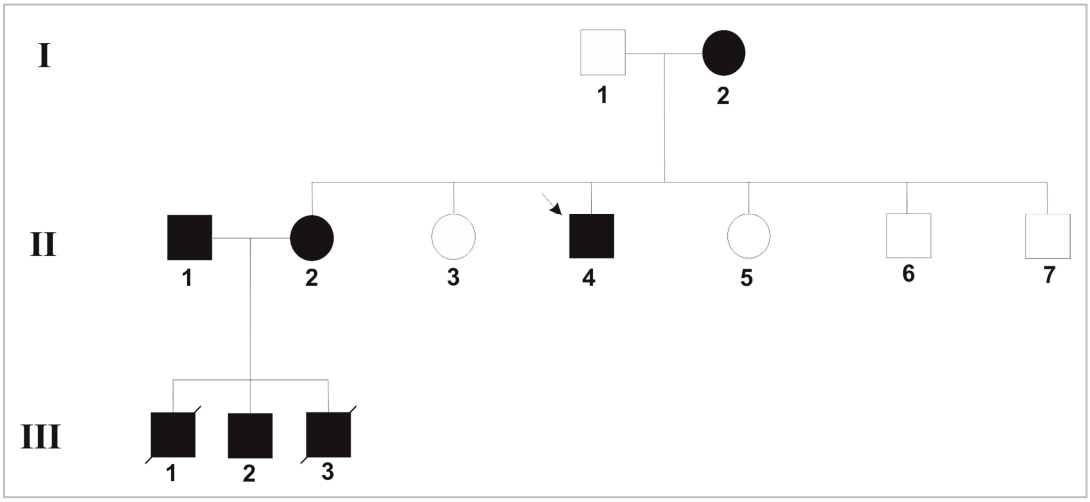


Figure 5. Pedigree of Family #5. Proband is indicated by the arrow and present a minor form of the disease. Black and white symbols represent clinically affected and unaffected individuals, respectively. Two sudden cardiac death at young age occurred (III-1, III-3).

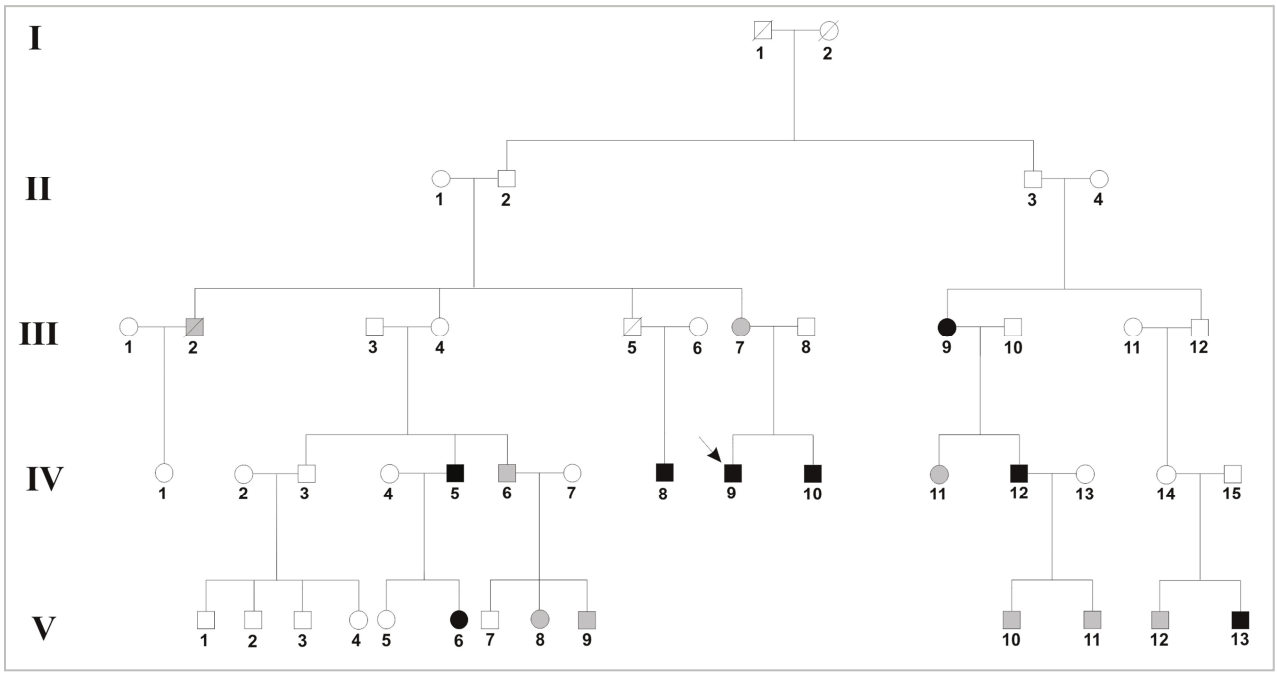


Figure 6. Pedigree of Family #6. Proband is indicated by the arrow. Black and white symbols represent clinically affected and unaffected individuals, respectively. Proband reported RV dysfunction with ventricular arrhythmias of LBBB morphology, late potentials and inverted T waves in V1-V3. Proband's brother present a severe form of the disease as well. Grey symbols denote subjects with minor signs of the disease such as ECG alterations (IV-6, V-8) or hypokinesia (V-11, V-12); or not investigated subjects (III-2).

APPENDIX C

Table 1. 'Private' *TTN* variants identified in 19 probands with the 'Cardiomyopathies gene panel'

#proband	variant	frequences	predictions
1a	rs199901448 c.37735G>A p.Ala12579Thr		Polyphen: benign SIFT: tolerated
	c.21844G>A p.Glu7282Lys		Polyphen: possibly damaging SIFT: deleterious
2a	rs200166942 c.100400T>G p.Val33467Gly		Polyphen: p.damaging SIFT: deleterious
	rs186624523 c.11788G>A p.Glu3930Lys		Polyphen: benign SIFT: deleterious
4a	c.37583C>T p.Ala12528Val		Polyphen: benign SIFT: tolerated
5a	rs376874956 c.33856G>A p.Glu11286Lys		Polyphen: p.damaging SIFT: deleterious
7a	c.82021C>T p.Arg27341Trp		Polyphen: p.damaging SIFT: deleterious
8a	c.84411G>T p.Lys28137Asn		Polyphen: benign SIFT: deleterious
	rs71423569 c.56483A>C p.Asn18828Thr		Polyphen: possibly damaging SIFT: deleterious
9a	rs56173891 c.102751A>G p.Met34251Val		Polyphen: benign SIFT: deleterious
	rs2857279 c.96409G>A p.Val32137Ile		Polyphen: possibly damaging SIFT: deleterious
	c.73261G>A p.Asp24421Asn		Polyphen: possibly damaging SIFT: deleterious
	rs372975579 c.68210C>T p.Ala22737Val		Polyphen: p.damaging SIFT: deleterious
	rs77351975 c.60232G>A p.Val20078Met		Polyphen: p.damaging SIFT: deleterious
	rs72677243 c.48953T>C p.Ile16318Thr		Polyphen: p.damaging SIFT: deleterious
10a	rs150223722 c.38929C>T p.Pro12977Ser		
12a	c.99833C>T p.Thr33278Ile		Polyphen: p.damaging SIFT: deleterious
	c.63139G>A p.Gly21047Arg		Polyphen: p.damaging SIFT: deleterious
13a	c.100460C>T p.Pro33487Leu		Polyphen: p.damaging SIFT: deleterious
	rs182428755 c.34474C>A p.Pro11492Thr	dbSNP:0.3%	
15a	rs72650011 c.30274C>T p.His10092Tyr	EVS:0.3%	Polyphen: benign SIFT: deleterious
	rs72647894 c.9359G>A p.Arg3120Gln	EVS:0.3%	Polyphen: p.damaging SIFT: deleterious
16a	rs72646885 c.69676A>G p.Ser23226Gly	dbSNP:0.2% EVS:0.7%	Polyphen: p.damaging SIFT: deleterious
	rs72677237 c.47723G>A p.Arg15908His	dbSNP:0.2% EVS:0.7%	Polyphen: p.damaging SIFT: deleterious
17a	c.104564C>A p.Ser34855Tyr		Polyphen: p.damaging SIFT: deleterious
	rs267599025 c.95806G>A p.Asp31936Asn		Polyphen: p.damaging SIFT: deleterious
	c.66311A>T p.Lys22104Met		Polyphen: p.damaging SIFT: deleterious
	rs139636676 c.4291C>T p.Arg1431Trp		Polyphen: p.damaging SIFT: deleterious

#proband	variant	frequences	predictions
18a	rs72650035 c.33059A>G p.Tyr11020Cys		Polyphen: benign SIFT: deleterious
19a	c.68083G>A p.Ala22695Thr		Polyphen: possibly damaging SIFT: deleterious
	rs199954570 c.48751G>A p.Asp16251Asn		Polyphen: p.damaging SIFT: deleterious
	rs72650031 c.32624C>T p.Pro10875Leu		

Table 2. Rare and low-frequency variants identified in our cohort of patients by using the 'ACM known and candidate genes panel'.

#proband	gene	variant	frequences	predictions	class
2b	*SCN5A*	rs45489199 c.6016C>G p.Pro2006Ala	EVS:0.1% ExAC:0.1%	Polyphen: benign SIFT: tolerated	IIb
	FAT1	rs116784674 c.7700G>A p.R2567H	dbSNP:0.5% EVS:1% ExAC:0.9%		
3b	*ANK3*	rs61845768 c.10055A>G p.Glu3352Gly	dbSNP:0.2% EVS:0.4% ExAC:0.4%		
5b	*STK3*	rs187757501 c.203G>A p.Ser68Asn	dbSNP:0.2% EVS:0.4% ExAC:0.3%		VIII/IV
	CTNNA3	rs190073606 c.779A>G p.Gln260Arg	dbSNP:0.1% ExAC:0.1%		
8b	*ABCC9*	rs61688134 c.2200G>A p.V734I	dbSNP:0.6% EVS:0.9% ExAC:0.7%		IIb/IV
7b	*ANK3*	rs148904927 c.7225T>C p.Ser2409Pro	dbSNP:0.5% EVS:0.5% ExAC:0.5%		
9b	*ANK3*	rs147527383 c.10688A>G p.Glu3563Gly	dbSNP:0.2% EVS:0.2% ExAC:0.2%		
	ANK3	rs140183285 c.9997A>T p.Thr3333Ser	dbSNP:0.2% EVS:0.2% ExAC:0.2%		
	SCN5A	rs41311123 p.Gly1406Gly c.4218G>A	EVS:1% ExAC:0.8%		IX
10b	*SCN5A*	rs45489199 p.Pro2006Ala c.6016C>G	EVS:0.1% ExAC:0.1%	Polyphen: benign SIFT: tolerated	IIb
	TJP1	rs2229517 c.1412A>G p.Asn471Ser	dbSNP:0.7% EVS:1% ExAC:0.9%		
13b	*STK3*	rs187757501 p.Ser68Asn c.203G>A	dbSNP:0.2% EVS:0.4% ExAC:0.3%		VIII/IV

#proband	gene	variant	frequences	predictions	class
16b	*AXIN2*	rs115931022 c.1235A>G p.N412S	dbSNP:0.3% EVS:0.5% ExAC:1%		
	ANK3	rs61845768 c.10055A>G p.E3352G	dbSNP:0.2% EVS:0.4% ExAC:0.4%		
	ABCC9	rs61688134 c.2200G>A p.V734I	dbSNP:0.6% EVS:0.9% ExAC:0.7%		IIb/IV
18b	*KAZN*	rs146471978 c.1406C>T p.Thr469Met	EVS:0.2% ExAC:0.2%		
19b	*FAT1*	rs56007012 c.3373A>C p.I1125L	EVS:0.1% ExAC:0.1%		
20b	*PNN*	rs145307846 c.1576G>A p.Glu526Lys	EVS:0.4% ExAC:0.4%		
21b	*SCN5A*	rs45489199 c.6016C>G p.Pro2006Ala	EVS:0.1% ExAC:0.1%	Polyphen: benign SIFT: tolerated	IIb
	TJP1	rs2229517 c.1412A>G p.Asn471Ser	dbSNP:0.7% EVS:1% ExAC:0.9%		
	ANK3	rs61845768 c.10055A>G p.Glu3352Gly	dbSNP:0.2% EVS:0.4% ExAC:0.4%		
22b	*XIRP1*	rs145699338 c.1156G>A p.Glu386Lys	dbSNP:0.1% EVS:0.5% ExAC:0.4%		
23b	*ANK3*	rs61845768 c.10055A>G p.Glu3352Gly	dbSNP:0.2% EVS:0.4% ExAC:0.4%		
	CTNND1	rs145191455 c.155A>G p.Asn52Ser	EVS:0.3% ExAC:0.3%		
	ABCC9	rs61688134 c.2200G>A p.Val734Ile	dbSNP:0.6% EVS:0.9% ExAC:0.7%		IIb/IV
24b	*ARVCF*	rs113625788 c.448G>A p.D150N	dbSNP:0.1% EVS:0.3% ExAC:0.3%		
26b(2a)	*ANK3*	rs141939315 c.5558G>A p.R1853Q	dbSNP:0.3% EVS:0.5% ExAC:0.5%		
	FAT1	rs200633985 c.3503C>T p.S1168L	EVS:0.4% ExAC:0.4%		
	TJP2	rs28556975 c.2810T>C p.L937P	dbSNP:1% EVS:1% ExAC:0.3%		
28b	*CTNNA3*	rs140913916 c.1453A>T p.Thr485Ser	EVS:0.1% ExAC:0.1%		
	XIRP1	rs61736135 c.1741C>A p.Pro581Thr	dbSNP:0.6% EVS:0.9% ExAC:0.7%		
29	*FAT1*	rs373140687 c.10498T>A p.Ser3500Thr	ExAc:0.013%		
	TMEM43	rs35028636 c.82C>T p.Arg28Trp	dbSNP:0.5% EVS:0.7% ExAC:0.3%		III
	ARVCF	rs199498113 c.1822C>T p.Arg608Cys	EVS:0.1% ExAC:0.1%		

#proband	gene	variant	frequences	predictions	class
30b	*SCN5A*	rs36210423 c.1715C>A p.A572D	EVS:0.2% ExAC:0.4%		I/VIII
31b	*PKP2*	rs143004808 c.76G>A p.D26N	dbSNP:0.3% EVS:0.5% ExAC:0.6%		
	ATP2A2	rs140234740 c.3064A>G p.I1022V	EVS:0.1% ExAC:0.1%		
	KCNE1	rs1805128 c.253G>A p.D85N	EVS:0.9% ExAC:0.9%		
	XIRP1	rs61736135 c.1741C>A p.P581T	dbSNP:0.6% EVS:0.9% ExAC:0.7%		
34b	*PKP4*	rs148019751 c.329A>G p.Gln110Arg	ExAC:0.1%		
37b	*DSC2*	rs200056085 c. 28647999_28648000ins2 p.A897KfsX4	dbSNP:0.6% EVS:1% ExAC:0.8%		
	DCHS1	rs142972252 c.1546G>A p.A516T	ExAC:0.1%		
	AXIN1	rs367788267 c.1868C>T p.S623L	ExAC:0.04%	PolyPhen2: benign SIFT: tolerated	
39b	*DSC2*	rs148185335 c.1787C>T p.Ala596Val	EVS:0.1% ExAC:0.1%		IX
40b	*XIRP1*	rs146284335 c.5352G>A p.M1784I	dbSNP:0.5% EVS:1% ExAC:0.7%		
41b	*PKP2*	rs143004808 c.76G>A p.D26N	dbSNP:0.3% EVS:0.5% ExAC:1.4%		I
42b(16a)	*TMEM43*	rs145619906 c.424G>A p.Glu142Lys	dbSNP:0.1% EVS:0.1%	Polyphen: p.damaging SIFT:deleterious	III
	CTNND1	rs76817459 c.257A>G p.Asn86Ser	EVS:0.1% ExAC:0.1%		
	DCHS1	rs60443131 c.8672G>A p.Arg2891Gln	ExAC:0.02%	Polyphen: benign SIFT:deleterious	
	XIRP1	rs146284335 c.5352G>A p.Met1784Ile	dbSNP:0.5% EVS:1% ExAC:0.7%		
	FAT1	rs142805532 c.3749A>G p.Tyr1250Cys	dbSNP:0.3% EVS:0.2% ExAC:0.2%		

Table 3. 'Private' *TTN* variants identified in 44 probands with the 'ACM known and candidate genes panel'.

#proband	variant	frequences	predictions
1b	c.27133A>G p.Thr9045Ala		SIFT: deleterious
3b	rs375198596 c.87367A>C p.Ser29123Arg		Polyphen: p.damaging SIFT: deleterious
	rs56391938 c.77216C>Gp.Ala25739Gly	EVS:0.1% ExAC:0.1%	SIFT: deleterious
	c.33259C>G p.Pro11087Ala		SIFT: deleterious
4b	c.102841A>C p.Ile34281Leu		Polyphen: p.damaging SIFT: deleterious
	rs376988498 c.74063G>A p.Arg24688His		Polyphen: p.damaging SIFT: deleterious
7b	c.75053T>C p.Val25018Ala		Polyphen: p.damaging SIFT: deleterious
8b	rs374093805 c.10457C>T p.Ala3486Val		Polyphen: p.damaging SIFT: deleterious
12b	rs201738153 c.32557C>T p.Pro10853Ser		SIFT: deleterious
14b	rs72648925 c.14870C>G p.Thr4957Ser	EVS:0.3% ExAC:0.3%	SIFT: deleterious
	rs397517756 c.92782G>C p.Asp30928His		Polyphen: p.damaging SIFT: deleterious
	rs72647868 c.3133G>A p.Val1045Met		SIFT: deleterious
	c.105787_105788delGCinsTT p.Ala35263Phe		
16b	c.37735G>A p.Ala12579Thr		
17b	c.87922T>C p.Tyr29308His		Polyphen: p.damaging SIFT: deleterious
21b	rs200118743 p.Gly6933Ala c.20798G>C	EVS:0.1% ExAC:0.1%	SIFT: deleterious
23b	rs200697681 c.81892G>A p.Asp27298Asn		Polyphen: p.damaging SIFT: deleterious
	rs201614524 p.Arg21635Cys c.64903C>T		Polyphen: p.damaging SIFT: deleterious
	rs72648940 c.16546G>T p.Asp5516Tyr		Polyphen: p.damaging SIFT: deleterious
24b	c.107657A>G p.Lys35886Arg		SIFT: deleterious
	c.65351A>G p.Asp21784Gly		Polyphen: p.damaging SIFT: deleterious
	c.7613G>A p.Gly2538Asp		Polyphen: p.damaging SIFT: deleterious
25b	rs55945684 c.103688T>C p.Val34563Ala	EVS:0.1% ExAC:0.1%	SIFT: deleterious
	rs372787601 c.74305A>G p.Asn24769Asp		SIFT: deleterious
26b	rs200166942 c.100400T>G p.Val33467Gly		Polyphen: p.damaging SIFT: deleterious
	rs186624523 c.11788G>A p.Glu3930Lys	dbSNP:0.2%	SIFT: deleterious
27b	rs72646808 c.53096G>A p.Arg17699His	EVS:0.2% ExAC:0.2%	Polyphen: p.damaging SIFT: deleterious
28b	c.75127G>C p.Val25043Leu		Polyphen: p.damaging SIFT: deleterious

#proband	variant	frequences	predictions
29b	c.3241G>T p.Ala1081Ser		Polyphen: p.damaging SIFT: deleterious
	c.4076A>G p.Glu1359Gly		Polyphen: p.damaging SIFT: deleterious
	c.94079A>G p.Asn31360Ser		Polyphen: p.damaging SIFT: deleterious
	rs181189778 c.43690T>A p.Ser14564Thr	dbSNP:0.1%	Polyphen: p.damaging SIFT: deleterious
31b	c.68642G>A p.Arg22881Gln		SIFT: deleterious
32b	rs72648923 c.14698G>A p.Ala4900Thr	EVS:0.2% ExAC:0.3%	Polyphen: p.damaging SIFT: deleterious
33b	rs199761901 c.98500G>A p.Glu32834Lys		Polyphen: p.damaging SIFT: deleterious
	c.60197C>T p.Pro20066Leu		Polyphen: p.damaging SIFT: deleterious
35b	rs368607833 c.36391C>T p.Arg12131Cys		SIFT: deleterious
36b	rs72648273 c.98294C>G p.Ala32765Gly	EVS:0.3% ExAC:0.3%	Polyphen: p.damaging SIFT: deleterious
	rs199895260 c.61922G>A p.Arg20641Gln	EVS:0.3% ExAC:0.3%	Polyphen: p.damaging SIFT: deleterious
39b	rs201043950 c.69821G>A p.Gly23274Asp	EVS:0.2% ExAC:0.1%	Polyphen: p.damaging SIFT: deleterious
	c.4102A>C p.Asn1368His		Polyphen: p.damaging SIFT: deleterious
40b	rs149001703 c.98390A>G p.Asn32797Ser	dbSNP:0.6% EVS:0.6% ExAC:0.4%	SIFT: deleterious
	c.72254C>A p.Thr24085Asn		Polyphen: p.damaging SIFT: deleterious
41b	rs55866005 c.104560G>C p.Val34854Leu	dbSNP:0.4% EVS:0.6% ExAC:0.5%	SIFT: deleterious
	rs55886356 c.101766G>C p.Gln33922His	dbSNP:0.6% EVS:0.8% ExAC:0.7%	Polyphen: p.damaging SIFT: deleterious
	c.21166G>T p.Val7056Leu		
44b	c.104564C>A p.Ser34855Tyr		Polyphen: p.damaging SIFT: deleterious
	rs267599025 c.95806G>A p.Asp31936Asn		Polyphen: p.damaging SIFT: deleterious
	rs139636676 c.4291C>T p.Arg1431Trp		Polyphen: p.damaging SIFT: deleterious

APPENDIX D

Table 1. Primers used to resequence the uncovered exons of the 11 know-disease genes in the 4 affected individuals of Family#6.

gene	exon	primer forward	primer reverse
PKP2	1	ACTCGAGCGGGGCGGGGCTCG	ACTCCCAGCACGCGGGGTGAG
	6	TTCAGGGGAGGTGATGTTTG	GGATTACAGGCGCAGACC
DSP	1	GGTAGCGAGCAGCGACCTC	AAAACCTTTCCACCTTCGGG
	12	TCAGCTTCATTTGAGGGGAAA	GGCAAGGCATCGTGTGTCTA
JUP	5	ATATTTCCAGAGGGGCGTCA	CCAAGGCTGTGCAGGATAGA
	6	GAACTATTCGTGCTTTAGCTG	AGATTTAGAGCCCCCAGTCC
	11	ATTGCTACAGGGTGGAGGTG	AGGAGACCCCCAAAAGTGTT
DSC2	1	TCAGACCTCGCTCTGTAATTGA	TATCCCCGTTCCCCTAGTTT
	2	TTGTCCTGATTTCTTTCTCCAA	CTCCATCCACATCCCTTCC
DSG2	1	CCAGGGAGGAGCCGAGTG	GATTTTCCGAAGCCCCAGGT
	5	TCTTGATCGAGAAGAAACACCA	AAGCAATGGCATGTAAAGTCC
	10	AGAGGTTTCCAATTCATGCAG	CTTGACCTCGTGATCCACCT
DES	1a	GGGGGCTGATGTCAGGAG	GTAGTTGGCGAAGCGGTCAT
	1b	TCTAAGGGCTCCTCCAGCTC	AGGCCTAGCCTCCTGTGC
	8	CTTGGTCAGGCTGAGTGTGC	CACAGCTGGGTGGAAGACAG
TMEM43	10	TGGAGCTTTGAATGGAGCA	CTGGCTATGACCCCTGGTAG
	12	GATCTGCTGAGCTGGTGAGG	AAAATTGACCGGGGCTCCTG
CTNNA3	10	TTTTGTTTTTCCAGTCTCACAA	AGGAAGCAGAAGTTGCAGTC
	11	TTCCTATTGTCGATGATGAAT	ACCATGCCTGTCCCAGTATT
LMNA	1	CCGAGCAGTCTCTGTCCTTC	GCCCTCTCACTCCCTTCCT
	2	CAAACCAACCTAATGCAAGGA	AGGACAGGTGAATGGCTCTG
	9	ATGGAGATGATCCCTTGCTG	ATTGGCAAACTTGGGTTGG
	10	CCATCACCACAGAGGACAGA	CCAGCGAGTAAAGTTCCAAAA

Table 2. Number of structural variants (SVs) identified by WGS.

sample	varType	total	cds	splicing	utr5	utr3	intron	Up stream	Down stream	Nc RNA	intergenic
IV-12	Inversion	199	67	0	0	0	35	2	0	13	82
	Transl.	476	10	0	1	13	133	2	2	28	287
	Deletion	3372	76	2	2	9	1109	23	26	108	2017
	Insertion	373	4	0	3	3	139	4	5	13	202
IV-10	Inversion	185	50	0	2	0	38	2	0	11	82
	Transl.	459	4	0	1	13	128	4	3	30	276
	Deletion	2960	66	2	1	9	974	14	25	117	1752
	Insertion	587	8	0	2	2	222	2	7	22	322
IV-3	Deletion	3135	70	3	2	10	1006	19	29	111	1885
	Inversion	185	41	0	0	1	43	1	0	12	87
	Transl.	424	8	0	1	9	136	1	0	23	246
	Insertion	538	2	2	3	1	190	3	9	19	309

Table 3. Variants identified in the exomes extracted from the genomes and shared by subjects IV-10 and IV-12 with a MAF ≤0.5% and phyloP>2.5.

Chr.	Start position	Variant type	Variant/SNP ID	SNP Freq	Gene name	Gene component	phyloP
16	71956511	cx	rs112755068	-1	*IST1*	EXON	5,23
5	74021846	cx	rs76339998	-1	*GFM2*	EXON	4,774
4	88231392	cx	rs72613567	-1	*HSD17B13*	SD_SITE	3,087
12	103352171	cx	rs71438488	-1	*ASCL1*	EXON	4,413
16	18908104	cx	rs35487751	-1	*SMG1*	SD_SITE	3,651
7	286468	cx	rs139364156	-1	*FAM20C*	EXON	5,264
2	3749151	cx	rs201406139	-1	*ALLC*	EXON	3,267
1	240255568	cx	rs71168893	-1	*FMN2*	EXON	3,895
6	170871046	cx	rs113440919	-1	*TBP*	EXON	4,325
1	248059779	cx	rs80255919	-1	*OR2W3*	EXON	4,525
11	55861277	cx	rs138372442	-1	*OR8I2*	EXON	4,436
12	974308	cx	rs72649808	-1	*WNK1*	EXON	5,681
5	60628153	cx	rs374788320	-1	*ZSWIM6*	EXON	3,331
7	15725797	cx	rs138679259	-1	*MEOX2*	EXON	5,18
8	96047806	cx	rs34960210	-1	*NDUFAF6*	SD_SITE	4,421
11	48285981	cx	rs76457745	-1	*OR4X1*	EXON	3,063
14	105996049	cx	rs5811180	-1	*TMEM121*	EXON	3,885
1	185108542	sb	A>T	-1	***TRMT1L***	EXON	5,025
2	20867122	cx	rs71839776	-1	*GDF7*	EXON	2,974
17	67190536	cx	rs147311261	-1	*ABCA10*	EXON	3,595
11	56344843	cx	rs148438199	-1	*OR5M10*	EXON	5,183
1	154842199	cx	rs56352724	-1	*KCNN3*	EXON	5,12
19	1621907	cx		-1	***TCF3***	EXON	2,971
3	42251577	cx	rs10546421	-1	*TRAK1*	EXON	3,996
1	170135646	sb	C>T rs183687510	0,1377	*METTL11B*	EXON	2,698
21	34860749	cx	rs71804651	-1	*DNAJC28*	EXON	5,229
12	7045891	cx	rs142596770	-1	*ATN1*	EXON	4,281
8	145617534	cx	rs148509143	-1	*ADCK5*	EXON	3,897
1	44125962	cx	rs74416524	-1	*KDM4A*	SA_SITE	4,637
19	41173892	cx	rs72024235	-1	*NUMBL*	EXON	2,605
2	85360828	cx		-1	***TCF7L1***	EXON	3,609
12	51723598	cx	rs75442020	-1	*CELA1*	EXON	5,839
14	77493761	cx	rs78762909	-1	*IRF2BPL*	EXON	3,703
17	67145190	cx	rs77569683	-1	*ABCA10*	EXON	4,529
6	16327915	cx	rs199744696	-1	*ATXN1*	EXON	4,398
7	149518532	cx	rs11353848	-1	*SSPO*	EXON	4,74
7	149503916	cx	rs66470151	-1	*SSPO*	EXON	3,928
3	63898360	cx	rs34631248	-1	*ATXN7*	EXON	4,152
3	49395673	cx	rs2836258862	-1	*GPX1*	EXON	3,027
11	73020375	cx	rs10682858	-1	*ARHGEF17*	EXON	4,183

Chr.	Start position	Variant type	Variant/SNP ID	SNP Freq	Gene name	Gene component	phyloP
22	37465385	cx	rs368080841	-1	*TMPRSS6*	SA_SITE	5,306
22	38482352	cx	rs113792005	-1	*BAIAP2L2*	EXON	4,074
1	109792735	cx	rs144034706	-1	*CELSR2*	EXON	2,566
3	8671319	cx	rs6147701	-1	*SSUH2*	EXON	2,794
12	53184619	cx	rs140653778	-1	*KRT3*	SD_SITE	3,356
11	124750447	cx	rs71859853	-1	*ROBO3*	EXON	3,539
11	130298117	cx	rs55808430	-1	*ADAMTS8*	EXON	2,903
12	89866054	cx	rs191378952	-1	*POC1B*	SA_SITE	5,119
1	226924875	cx	rs200752967	-1	*ITPKB*	EXON	3,815

For complex variants only one dbSNP record (rs#) is listed in the table. In bold type are the 3 variants without a correspondence in dbSNP database. Sb=substitution; cx=complex.

Table 4. Variants identified inside the critical region of chr19p13.3 shared by subject IV-10 and IV-12 with a MAF ≤0.5%.

Chr.	Start position	Ref.	Variant	Var. type	SNP ID	SNP Freq.	Gene name	Gene component	phyloP
19	1414297	G	A	sb		-1	*DAZAP1*	INTRON	0,187
19	5051840	C	A	sb	rs2620815	-1	*KDM4B*	INTRON	-3,772
19	632004	A	T	sb		-1	*POLRMT*	INTRON	0,08
19	4996450	A	G	sb		-1	*KDM4B*	INTRON	0,366
19	5206273	T	A	sb	**rs116644020**	-1	***PTPRS***	**UTR**	-0,669
19	2942172	C	T	sb	rs77664718	-1	*ZNF77*	INTRON	0,178
19	3635708	T	C	sb		-1	*PIP5K1C*	INTRON	0,121
19	615247	C	T	sb	rs62132645	-1	*HCN2*	INTRON	-0,408
19	1075035	A	C	sb	rs3214170	-1	*HMHA1*	INTRON	-0,782
19	5206276	A	G	sb	**rs200486129**	-1	***PTPRS***	**UTR**	-0,08
19	5469849	G	C	sb		-1			0,543
19	4372207	G	C	sb		-1	*SH3GL1*	INTRON	0,646
19	1295883	C	A	sb	rs201782617	0,04591	*EFNA2*	INTRON	0,219
19	3152391	C	A	sb		-1	*GNA15*	INTRON	-0,06
19	1930489	A	T	sb	rs76335243	-1			-1,284
19	5319616	C	A	sb		-1	***PTPRS***	**INTRON**	-0,363
19	2938682	C	T	sb		-1	*ZNF77*	INTRON	0,122
19	4729069	A	G	sb		-1			0,159
19	1295921	G	T	sb		-1	*EFNA2*	INTRON	0,546
19	3928004	C	T	sb		-1	*ATCAY*	UTR	-0,632
19	1593292	G	T	sb		-1			0,17
19	2960913	G	A	sb		-1			-100
19	5469404	C	A	sb	rs185545586	0,1837			0,317
19	2411841	G	T	sb	rs12609648	-1	*TMPRSS9*	INTRON	-3,734
19	2411841	G	C	sb	rs12609648	-1	*TMPRSS9*	INTRON	-3,734
19	5698541	G	T	sb	rs117261656	-1	*LONP1*	INTRON	0,642

Chr.	Start position	Ref.	Variant	Var. type	SNP ID	SNP Freq.	Gene name	Gene component	phyloP
19	604988	G	T	sb	rs113541645	-1	*HCN2*	INTRON	0,222
19	2580141	C	A	sb	rs4335872	-1	*GNG7*	INTRON	0,444
19	5786073	C	T	sb	rs190862396	0,04591	*DUS3L*	INTRON	-0,238
19	6329458	G	A	sb		-1	*ACER1*	INTRON	-0,409
19	4705100	C	G	sb		-1	*DPP9*	INTRON	-0,111
19	4836610	A	C	sb	rs80320354	0,04591			-0,433
19	6241686	T	G	sb		-1	*MLLT1*	INTRON	1,155
19	5658862	A	G	sb		-1	*SAFB*	INTRON	0,1
19	1085608	T	C	sb	rs139132266	0,04591	*HMHA1*	INTRON	-1,089
19	1216565	G	A	sb		-1	*STK11*	INTRON	0,21
19	4835884	G	A	sb		-1			-1,233
19	4745422	C	T	sb		-1			0,189
19	6172483	G	C	sb		-1	*ACSBG2*	INTRON	0,135
19	1980376	G	A	sb	rs377669285	-1	*CSNK1G2*	UTR	0,512
19	4548013	A	G	sb	rs200320836	0,3673	*SEMA6B*	INTRON	0,097
19	375655	G	A	sb		-1	*THEG*	SD_SITE	-2,7
19	1792210	C	T	sb	rs117609099	-1	*ATP8B3*	INTRON	-2,71
19	387281	T	C	sb		-1			-100
19	5206274	G	T	sb	**rs139183127**	-1	***PTPRS***	**UTR**	-0,721
19	2210328	C	T	sb		-1	*DOT1L*	INTRON	-0,803
19	946630	G	A	sb		-1	*ARID3A*	INTRON	0,2
19	3081764	G	T	sb		-1			0,317
19	3930761	G	A	sb		-1			-0,611
19	3131446	C	A	sb	rs10164384	-1			-1,659
19	1930487	A	T	sb	rs60152900	-1			0,162
19	4374743	G	A	sb		-1	*SH3GL1*	INTRON	-0,164
19	4521456	G	A	sb	rs189471903	0,3673			-0,805
19	689146	T	C	sb	rs12978248	-1	*PRSS57*	INTRON	-1,48
19	493369	G	C	sb		-1			0,164
19	5759325	T	A	sb	rs190201024	0,04591	*CATSPERD*	INTRON	0,068

Sb= substitution. In bold type are the variants identified in PTPRS gene.

Table 5. Variants identified inside the critical region on chr11q21 in subject IV-12 with a MAF ≤0.5%

Chr.	Start position	Variant type	Variant/SNP ID	Gene name	Gene component
11	93112917	cx		*CCDC67*	intronic
11	93118413	cx		*CCDC67*	intronic
11	93122711	cx	rs140335781	*CCDC67*	intronic
11	93125844	cx		*CCDC67*	intronic
11	93126086	cx		*CCDC67*	intronic
11	93133579	cx		*CCDC67*	intronic
11	93139645	cx		*CCDC67*	intronic
11	93151510	sb	A>C rs374560834	*CCDC67*	intronic
11	93164261	cx		*CCDC67*	intronic
11	93170909	cx		*CCDC67*	UTR3
11	93211879	cx		*SMCO4*	UTR3
11	93211879	cx		*SMCO4*	UTR3
11	93266015	cx	rs71978895	*SMCO4*	intronic
11	93267379	cx		*SMCO4*	intronic
11	93267535	cx		*SMCO4*	intronic
11	93267537	cx		*SMCO4*	intronic
11	93272455	sb	G>C rs72966968	*SMCO4*	intronic
11	93275070	sb	A>G rs72966970	*SMCO4*	intronic
11	93401618	cx		*CEP295*	intronic
11	93402580	cx		*CEP295*	intronic
11	93410461	sb	G>A	*CEP295*	intronic
11	93418866	cx		*CEP295*	intronic
11	93446022	cx	rs56969423	*CEP295*	intronic
11	93446022	cx		*CEP295*	intronic
11	93447068	sb	G>A	*CEP295*	intronic
11	93454831	cx	rs71064775	*SCARNA9*	ncRNA_exonic
11	93482001	cx		*C11orf54*	intronic
11	93483802	sb	G>A	*C11orf54*	intronic
11	93486498	cx		*C11orf54*	intronic
11	93486526	cx		*C11orf54*	intronic
11	93493679	cx		*C11orf54*	intronic
11	93557080	cx		*VSTM5*	intronic
11	93570177	cx		*VSTM5*	intronic
11	93576624	cx	rs59069636	*VSTM5*	intronic
11	93764497	cx		*HEPHL1*	intronic
11	93782784	cx		*HEPHL1*	intronic
11	93782784	cx		*HEPHL1*	intronic
11	93789822	cx		*HEPHL1*	intronic
11	93789822	cx		*HEPHL1*	intronic
11	93789823	cx		*HEPHL1*	intronic

Chr.	Start position	Variant type	Variant/SNP ID	Gene name	Gene component
11	93793258	cx		*HEPHL1*	intronic
11	93817502	cx		*HEPHL1*	intronic
11	93826266	cx		*HEPHL1*	intronic
11	93826266	cx		*HEPHL1*	intronic
11	93827137	cx		*HEPHL1*	intronic
11	93831127	cx		*HEPHL1*	intronic
11	93862252	cx		*PANX1*	UTR5
11	93875290	sb	G>A	*PANX1*	intronic
11	94112137	cx		*GPR83*	UTR3
11	94114724	cx	rs56180709	*GPR83*	intronic
11	94114724	cx		*GPR83*	intronic
11	94152631	cx		*MRE11A*	UTR3
11	94184604	sb	C>T	*MRE11A*	intronic
11	94191053	cx	rs59279005	*MRE11A*	intronic
11	94218676	cx	rs71459730	*MRE11A*	intronic
11	94218791	sb	G>T	*MRE11A*	intronic
11	94246110	cx		*C11orf97*	intronic
11	94247416	cx	rs373026301	*C11orf97*	intronic
11	94248857	sb	C>T	*C11orf97*	intronic
11	94249794	cx		*C11orf97*	intronic
11	94250020	sb	G>T rs676521	*C11orf97*	intronic
11	94251022	cx		*C11orf97*	intronic
11	94252608	cx	rs61528058	*C11orf97*	intronic
11	94264603	cx		*C11orf97*	intronic
11	94264603	cx		*C11orf97*	intronic
11	94279080	sb	T>A	*FUT4*	UTR3
11	94305383	sb	A>C	*PIWIL4*	intronic
11	94307722	cx		*PIWIL4*	intronic
11	94313801	cx		*PIWIL4*	intronic
11	94316198	cx		*PIWIL4*	intronic
11	94316210	cx	rs58072951	*PIWIL4*	intronic
11	94316342	cx		*PIWIL4*	intronic
11	94334520	cx	rs372273598	*PIWIL4*	intronic
11	94345106	cx	rs71459738	*PIWIL4*	intronic
11	94345346	cx	rs144175515	*PIWIL4*	intronic
11	94350519	cx		*PIWIL4*	intronic
11	94527703	sb	T>C	*AMOTL1*	intronic
11	94606941	cx		*AMOTL1*	UTR3
11	94699391	cx	rs58215154	*CWC15*	intronic
11	94699391	cx		*CWC15*	intronic
11	94708334	sb	A>G rs6483383	*KDM4D*	intronic
11	94708335	sb	G>A rs6483384	*KDM4D*	intronic

Chr.	Start position	Variant type	Variant/SNP ID	Gene name	Gene component
11	94717787	sb	A>G	*KDM4D*	intronic
11	94824141	cx		*ENDOD1*	intronic
11	94831936	sb	G>A rs75098296	*ENDOD1*	intronic
11	94836260	cx		*ENDOD1*	intronic
11	94836264	sb	G>T rs674973	*ENDOD1*	intronic
11	94841658	cx		*ENDOD1*	intronic
11	94841658	cx		*ENDOD1*	intronic
11	94855753	cx		*ENDOD1*	intronic
11	94856752	cx		*ENDOD1*	intronic
11	94889358	cx		*LOC101929295*	ncRNA_intronic
11	94889358	cx		*LOC101929295*	ncRNA_intronic
11	94906293	sb	A>C	*SESN3*	UTR3
11	94915046	sb	T>C rs191030205	*SESN3*	intronic
11	94919358	cx		*SESN3*	intronic
11	94919358	cx		*SESN3*	intronic
11	94939003	sb	G>T rs2199057	*SESN3*	intronic
11	94939529	cx	rs3032056	*SESN3*	intronic
11	94959199	sb	A>T rs7944086	*SESN3*	intronic
11	95527512	sb	C>A	*CEP57*	intronic
11	95547233	cx		*CEP57*	intronic
11	95547233	cx	rs369609536	*CEP57*	intronic
11	95548682	cx		*CEP57*	intronic
11	95553074	cx		*CEP57*	intronic
11	95578167	**sb**	**C>T rs146572467**	***MTMR2***	**exonic**
11	95583439	cx		*MTMR2*	intronic
11	95586200	cx		*MTMR2*	intronic
11	95632800	cx		*MTMR2*	intronic
11	95638491	cx		*MTMR2*	intronic
11	95656529	sb	A>C	*MTMR2*	intronic
11	95737611	cx		*MAML2*	intronic
11	95741204	sb	T>A rs4998067	*MAML2*	intronic
11	95748747	cx		*MAML2*	intronic
11	95748747	cx		*MAML2*	intronic
11	95752846	cx	rs373815878	*MAML2*	intronic
11	95767572	cx		*MAML2*	intronic
11	95767596	cx		*MAML2*	intronic
11	95801602	cx		*MAML2*	intronic
11	95801602	cx		*MAML2*	intronic
11	95809365	sb	G>A rs76820499	*MAML2*	intronic
11	95809594	cx		*MAML2*	intronic
11	95809594	cx		*MAML2*	intronic
11	95810936	sb	G>A	*MAML2*	intronic

Chr.	Start position	Variant type	Variant/SNP ID	Gene name	Gene component
11	95814678	cx		*MAML2*	intronic
11	95846348	cx		*MAML2*	intronic
11	95855661	cx		*MAML2*	intronic
11	95855661	cx		*MAML2*	intronic
11	95881823	cx	rs71040133	*MAML2*	intronic
11	95881823	cx		*MAML2*	intronic
11	95887116	sb	T>A rs11021451	*MAML2*	intronic
11	95899318	cx		*MAML2*	intronic
11	95899318	cx	rs372154381	*MAML2*	intronic
11	95899341	cx		*MAML2*	intronic
11	95902823	cx	rs35274204	*MAML2*	intronic
11	95911080	cx		*MAML2*	intronic
11	95917603	cx		*MAML2*	intronic
11	95917773	sb	G>T	*MAML2*	intronic
11	95920956	cx		*MAML2*	intronic
11	95945272	cx		*MAML2*	intronic
11	95945272	cx		*MAML2*	intronic
11	95958996	cx	rs36112117	*MAML2*	intronic
11	95965635	cx		*MAML2*	intronic
11	95985338	cx		*MAML2*	intronic
11	96016768	sb	G>A rs76683564	*MAML2*	intronic
11	96024191	cx		*MAML2*	intronic
11	96048385	cx		*MAML2*	intronic
11	96059522	cx	rs35138919	*MAML2*	intronic
11	96092755	cx		*CCDC82*	intronic
11	96110521	cx		*CCDC82*	intronic
11	96123734	cx		*JRKL*	UTR5
11	96125435	cx		*JRKL*	UTR3
11	96207166	cx		*JRKL-AS1*	ncRNA_intronic
11	96222166	cx		*JRKL-AS1*	ncRNA_intronic
11	96226382	cx		*JRKL-AS1*	ncRNA_intronic
11	96230212	cx		*JRKL-AS1*	ncRNA_intronic
11	96231183	cx		*JRKL-AS1*	ncRNA_intronic
11	96231580	cx		*JRKL-AS1*	ncRNA_intronic

For complex variants only one dbSNP record (rs#) is listed in the table. In bold type is the only one exonic variant inside the critical region. Sb=substitution; cx=complex.