UNIVERSITÀ
DEGLI STUDI
DI PADOVA

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
CICLO XVII

Diagnostic Implications of Arrhythmogenic Cardiomyopathy Genetic Testing

Direttore della Scuola: Ch.mo Prof. Gaetano Thiene
Coordinatore d'indirizzo: Ch.mo Prof. Gaetano Thiene
Supervisore: Ch.mo Prof. Kalliopi Pilichou

Dottorando: Elisabetta Lazzarini

TABLE OF CONTENTS

ABSTRACT

Background & Aims: Arrhythmogenic Cardiomyopathy (AC) is a rare inherited heart muscle disease associated with mutations in genes encoding mainly components of the cardiac desmosome. We performed a comprehensive study of genetic variants in a cohort of AC subjects and the assessment of Next Generation Sequencing (NGS) strategies for molecular diagnosis of AC.

Methods: Ninety-nine unrelated index cases, of which 26 sudden cardiac death cases, underwent genetic screening for 5 desmosomal genes by denaturing high performance liquid chromatography and direct sequencing, whereas 46 probands were additionally screened for 3 extra desmosomal genes. A complementary analysis for copy number variants (CNVs) was performed by multiplex ligation-dependent probe amplification and quantitative real-time PCR in the entire cohort. A 4-step variant filtering strategy based on mutation type, frequency, evolutionary conservation and *in silico* analysis, was used. Whole Exome and Targeted NGS strategies were performed on Illumina platforms in order to test methods efficacy.

Results: Screening of 8 AC genes and subsequent 4-step variants filtering identified 37 different point desmosomal mutations in 42 AC probands (42%). The most frequently mutated genes resulted *PKP2* and *DSP,* with "radical" mutation type accounting for the 80% of the *PKP2* variants. No pathogenic mutations were identified in the extra desmosomal genes analyzed. CNVs analysis further revealed 3 different large genomic rearrangements in 5 probands (4%), increasing to 46 (46%) the number of positively genotyped patients. *PKP2* and *DSP* single mutation accounted respectively for 20% and 11% of the cohort, with *DSP* carriers showing a higher risk of sudden cardiac death. Eight multiple mutations carriers were observed (8%). NGS approaches identified 4 variants in extra desmosomal genes allowing a differential diagnosis in 4 patients.

Conclusions: A fine variant filtering avoids overrepresentation of putative pathogenic mutations and shows that radical and missense mutations should be equally interpreted with great caution in the setting of clinical diagnosis. NGS and CNVs analysis increased significantly the diagnostic yield in AC genetic testing. The genetics of AC is more complex than previously appreciated, with frequent requirement for more than one 'hit' for penetrant disease.

SOMMARIO

Introduzione & Scopo: La Cardiomiopatia Aritmogena (AC) è una malattia rara del muscolo cardiaco associata a mutazioni a carico di geni che codificano principalmente per componenti del desmosoma cardiaco. Abbiamo realizzato l`analisi genetica in una coorte di soggetti affetti da AC e lo sviluppo di strategie in Next Generation Sequencing (NGS) per la diagnosi molecolare di AC.

Metodi: Novantanove casi indice, di cui 26 soggetti di morte improvvisa, sono stati sottoposti a screening genetico per 5 geni desmosomiali mediante cromatografia denaturante in fase liquida ad alto rendimento e sequenziamento diretto, 46 probandi sono stati analizzati anche per 3 geni extra desmosomiali. Abbiamo inoltre ricercato nell`intera popolazione varianti del numero di copie (CNVs) mediante la tecnica MLPA (Multiplex Ligation-Dependent Probe Amplification) e PCR quantitativa. La strategia di filtraggio si è basata sul tipo di mutazione, frequenza, conservazione, e analisi *in silico*. Gli approcci NGS "Whole Exome" e "Targeted" sono stati eseguiti su piattaforme Illumina.

Risultati: L`analisi di 8 geni associati alla AC e il successivo filtraggio delle varianti ha individuato 37 diverse mutazioni puntiformi in 42 soggetti (42%). I geni più frequentemente mutati sono la *PKP2* e la *DSP*, le mutazioni "radicali" costituiscono l`80% delle varianti della *PKP2*. Non sono state riscontrate mutazioni patogene nei geni extra desmosomiali studiati. La ricerca di CNVs ha identificato 3 diversi riarrangiamenti cromosomici in 5 probandi (4%), portando a 46 (46%) i soggetti genotipo positivi. Si sono osservate mutazioni singole nei geni *PKP2* e *DSP* nel 20% e nell`11% dei soggetti, i portatori di mutazioni nella *DSP* presentano un rischio maggiore di morte improvvisa, 8 (8%) soggetti presentano mutazioni multiple. Le analisi mediante NGS hanno indentificato 4 varianti in geni extra desmosomiali permettendo la diagnosi differenziale in 4 pazienti.

Conclusioni: La diagnosi molecolare in ambito clinico rende necessaria un`attenta interpretazione del potenziale patogeno tanto delle mutazioni missenso quanto di quelle radicali soprattutto dopo il cospicuo aumento di varianti identificate con NGS e la ricerca di CNVs. La base genetica di AC è molto più complessa di quanto finora apprezzato, con una frequente presenza di più di una mutazione per una penetranza completa della malattia.

INTRODUCTION

1.1 Definition and Epidemiology

Arrhythmogenic Cardiomyopathy (AC) is an inherited disorder of the myocardium characterized by progressive adipose and fibrous tissue replacement of cardiomyocytes (Nava et al., 1988; Thiene et al., 1988). This process is associated with structural and functional changes involving predominantly the right ventricle (RV), structural manifestations mainly include ventricular dilatation, hypokinesia, and aneurysms of the ventricular wall, that in early disease are mostly located in the right ventricular inflow, outflow, and apical regions, called the "triangle of dysplasia" (Frank et al. 1978; Marcus et al. 1982; Fontaine et al., 1998) (Figure 1.1).

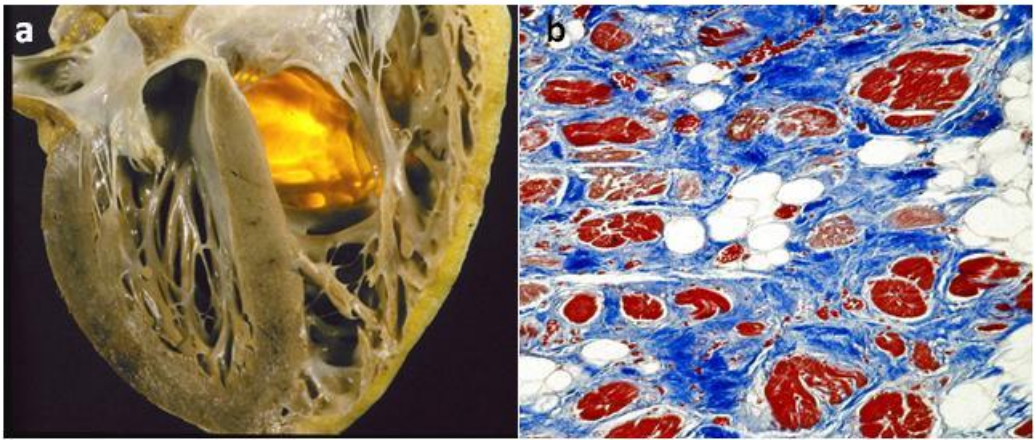

Figure 1.1: a) A case of AC in a 25-years old man who died suddenly. Four chamber cut of the heart specimen showing isolated fatty replacement of the right ventricular free wall and translucent infundibulum. B) Endomyocardial biopsy of the right ventricular free wall of a 39-years old patient who died suddenly, showing cadiomyocytes embedded in fibrous and fatty tissues (from Cardiovascular Pathology of Padua).

Clinical manifestations of AC mostly occur between the second and fourth decade of life, the mean age at diagnosis is estimated to be approximately 30 years, and males are more often affected than females, with an estimated gender ratio of 3:1 (Nava et al., 2000).

In the Veneto region, AC is the most common cause of sudden cardiac death (SCD) in individuals under the age of 35 and in young athletes (Thiene et al. 1988, Corrado et al., 1990).

In this geographic area, the prevalence of AC is estimated to range between 1:2000 and 1:5000 (Nava et al. 1988), however this value could be higher because

of the presence of many undiagnosed and misdiagnosed cases linked to the progressive nature of AC and its variable expressivity.

1.2 Clinical Manifestations

Progressive structural changes involving myocardium are preceded or accompanied by clinical manifestation including T-wave inversion in right precordial leads, epsilon waves, the presence of widening of the QRS complex observed in resting ECG, late potentials ventricular premature complexes which can be registered by signal averaged ECG and by Holter monitoring, respectively (McKenna et al., 1994). The most common clinical manifestations in AC are related to arrhythmias and conduction abnormalities, in particular ventricular tachycardia is a typical characteristic of AC patients that may lead to ventricular fibrillation and sudden death (SD) (Marcus et al., 1982; Thiene et al., 1988). However often SD can represent the first and only manifestation of AC or in some cases minor symptoms might occur making difficult its early identification.

Natural history of AC can be divided into four phases, based on clinical and pathological observations (Thiene et al., 1990; Corrado et al., 2000; Thiene et al., 2007; Basso et al., 2009).

1. A “concealed” phase typical of early AC, frequently characterized by the absence of clinical symptoms, although subtle structural changes and minor ventricular arrhythmia may be recognized. During this phase subjects are mainly asymptomatic and they may be at risk of SD, especially during physical exercise (Corrado et al., 1990).
2. The “overt electrical disorder” phase, characterized by the presence of functional and morphological abnormalities of the right ventricle (RV) and/or left ventricle (LV), usually detectable by cardiac imaging techniques. Patients manifest symptomatic ventricular arrhythmia, palpitation, syncope, and pre-syncope.
3. The third “right ventricular dysfunctional” phase is characterized by further extension of disease involving the RV myocardium that leads to impaired contractility and isolated right heart failure.

4. The most advanced phase is characterized by LV involvement leading to biventricular heart failure, which is difficult to distinguish from dilated cardiomyopathy (DCM). However, LV involvement may be detected also in earlier phases of the disease (Sen-Chowdhry et al., 2008).

1.3 Diagnostic criteria

Diagnosis of AC is based on classification of clinical findings into specific criteria, divided into major and minor according to specific association with the disease, as described in Table1.1. AC criteria were proposed by an International Task Force in 1994 (Mc Kenna et al., 1994) to facilitate AC diagnosis and were recently modified (Marcus et al., 2010), as the original criteria were highly specific but showed low sensitivity especially in evaluating asymptomatic subjects or family members affected by an early AC form (Marcus et al., 2010). The modified criteria focused on the detection of milder forms of the disease and aimed to facilitate the cascade screening of patients' relatives based on different clinical features comprising structural, histological and arrhythmic abnormalities along with a detailed familial history of disease and SD. However the clinical diagnosis of AC is still complicated, as a single diagnostic test is not enough to ascertain the disease. The definitive diagnosis requires the fulfilment of 2 major, 1 major plus 2 minor, or 4 minor criteria. A borderline diagnosis requires the fulfilment of 1 major plus 1 minor, or 3 minor criteria, and possible diagnosis fulfilment of 1 major or 2 minor criteria.

1.3.1 Clinical tests enabling AC diagnosis:

Baseline electrocardiography (12-lead ECG) findings associated with AC are abnormalities of depolarization, conduction and repolarization secondary to atrophy of ventricular walls that reflect the pathophysiology of the disease. ECG depolarization manifest as T-wave inversion in anterior precordial leads in the absence of RBBB (major criterion). Conduction abnormalities detected as prolongation of the QRS complex in right precordial leads in the absence of RBBB, (>110 ms in leads V1-V3) reflect slow conduction in the RV free wall (major criterion). The presence of epsilon wave, that represents a deflection between the end of the QRS complex and the beginning of the T wave (major

criterion) corresponds to a delayed electric potential that initiates in regions of healthy tissue surrounded by fibrofatty tissue indicative of intraventricular impulse conduction delay.

Signal-averaged electrocardiogram (SAECG) allows the registration of low-amplitude potentials within the end of the QRS complex (late potential) that are not wide enough to be evident on the 12-lead ECG.

Holter monitoring is used to register the electric activity of the cardiovascular system for a period of time of normally 24 h, with particular attention to the diurnal rhythm fluctuations.

Echocardiography represents the first-line imaging approach for evaluating patients with suspected AC or for screening of family members, allowing serial examination with the aim to assess the disease onset and progression.

Magnetic resonance imaging (MRI) is another non-invasive tissue characterization technique that allows the detection of morphological and structural abnormalities of the ventricular wall like microaneurysms and allows the calculation of RV and LV volumes and ejection fraction (Tandri et al., 2005; Sen-Chowdhry et al., 2006; Perazzolo Marra et al., 2014).

Endomyocardial biopsy is used to detect myocytes in diverse stages of cell death and of fibrofatty replacement (major criterion). However a negative biopsy is not enough to exclude AC because of the segmental nature of the disease, especially during early stages.

A comprehensive workup also involves the study of a full family history, with particular focus on cardiac symptoms, and the design of a detailed pedigree describing the occurrence of SD and unexplained heart failure. The value of the analysis of the family history and of AC genetic testing has been widely recognized and the detection of a pathogenic variants now contributes as a major diagnostic criterion for the diagnosis of AC (Thiene et al., 2007; Basso et al., 2009).

Original Task Force Criteria	Revised Task Force Criteria
I. Global or regional dysfunction and structural alterations*	
Major	
• 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	**By 2D echo:** • 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	
• Mild global RV dilatation and/or ejection fraction reduction with normal LV • Mild segmental dilatation of the RV • Regional RV hypokinesia	**By 2D echo:** • 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

Table 1.1 (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: Comparison of original and revised Task Force criteria for diagnosis of AC (from Marcus et al., 2010).

1.4 Treatment of AC patients

AC patients and athletes with probable or definite diagnosis of AC are strongly recommended to avoid physical activity and participation to competitive sports (Corrado et al., 2000).

The pharmacological treatment of patients showing mild and tolerated arrhythmias is palliative and consists of beta blockers and class I and class III antiarrhythmic drugs, targeted against arrhythmias and cardiac insufficiency. In particular sotalol has been reported as most efficient pharmacological antiarrhythmic treatment in AC (Basso et al., 2004) and amiodarone is recommended as add-on therapy if betablockers show to be unsuccessful in suppressing ventricular tachycardia (VT) (Sen-Chowdhry et al., 2004). The only

effective strategy against malignant ventricular tachyarrhythmias is the implantable cardioverter-defibrillator (ICD). ICD implantation is recommended as a first-line therapy after cardiac arrest or documented sustained VT, or its use as a primary prophylaxis in asymptomatic subjects showing a family history of SD or evidence of wide right ventricular (RV) dysfunction.

1.5 Molecular genetics of AC

To date, AC is considered an autosomal dominant trait showing reduced penetrance and expression variability (Nava et al., 1988; McKenna et al., 1994; Sen-Chowdhry et al., 2004; van Tintelen et al., 2006). However, compound and digenic heterozygosity are often described in patients with severe form of disease (Xu et al., 2011; Bauce et al. 2010; Rigato et al., 2013) and, rare autosomal recessive forms of the myocardial abnormalities are also observed, associated or not to cardiocutaneous syndromes (Naxos and Carvajal syndromes).
More than a decade elapsed between the recognition of familial AC (Nava et al., 1988), and the identification of the first disease-causing gene mutation (McKoy et al., 2000, Rampazzo et al., 2002). Linkage analysis and candidate gene approaches unmasked the genetic heterogeneity of AC with the identification of 13 genetic loci (Table 1.2).

The first molecular genetics studies involved the recessive syndromic form called Naxos disease. Naxos disease was first described in 1986 as a familial, autosomal recessive disease characterized by hair and skin abnormalities, woolly hair and palmoplantar keratoderma, and a form of cardiomyopathy showing AC-like features (Protonotarios et al., 1986). The typical woolly hair phenotype presented at birth, the erythema of the palms of the hands appeared within the first year of life, while the cardiac phenotype appeared only during adolescence or early adulthood. Nine affected individuals of four families from the Greek island of Naxos were originally described. Linkage analysis identified a locus at the long arm of the chromosome 17 (17q21.2) (Coonar et al., 1998). Subsequent sequencing of Plakoglobin (*JUP*) gene in affected subjects showed a 2 nucleotides

deletion leading to a frame shift effect which introduces a stop codon in the protein (McKoy et al., 2000).

Locus	Chromosome	Gene	Protein	Inheritance	Prevalence	Reference
AC1	14q24.3	TGFβ3	Transforming Growth factor β-3	AD	< 1%	Beffagna et al., 2005
AC2	1q.42-43	RYR2	Ryanodine receptor 2	AD	< 1%	Tiso et al., 2001
AC3	14q12-q22	Unknown				
AC4	2q32.1-q32.2	Unknown				
AC5	3p25.1	TMEM43	Transmembrane protein 43	AD	< 1%	Merner et al., 2008
AC6	10p12-p14	Unknown				
AC7	10q22.3	Unknown				
AC8	6p24	DSP	Desmoplakin	AD/AR	10-15%	Rampazzo et al., 2002
AC9	12p11.2	PKP2	Plakophilin-2	AD/AR	30-40%	Gerull et al., 2004
AC10	18q12.1	DSG2	Desmoglein-2	AD/AR	3-8%	Pilichou et al., 2006
AC11	18q12.1	DSC2	Desmocollin-2	AD/AR	1-5%	Syrris et al., 2006
AC12	17q21.2	JUP	Junction Plakoglobin	AD/AR	< 1%	Asimaki et al., 2007
AC13	10q21.3	CTNNA3	α-T-catenin	AD	NA	van Hengel et al., 2012
AC-like	2q35	DES	Desmin	AD	NA	Otten et al., 2010
AC-like	2q31.2	TTN	Titin	AD	NA	Taylor et al., 2011
AC-like	1q22	LMNA	Lamin A/C	AD	NA	Quarta et al., 2012
AC-like	6q22.1	PLN	Phospholamban	AD	NA	Van der Zwaag et al., 2012

Table 1.2: Genes associated with AC or overlap syndromes. AD: autosomal dominant, AR: autosomal recessive, NA: not available.

Soon after the identification of the *JUP* deletion involved in Naxos syndrome, another deletion this time in the Desmoplakin (*DSP*) gene, localized on chromosome 6p24.3, was described for another syndromic form, so called Carvajal syndrome (Carvajal-Huerta et al., 1998). Carvajal syndrome is an

autosomal recessive disorder first described in 1998, in eighteen subjects belonging to three families from Ecuador who showed epidermolytic palmoplantar keratoderma, woolly hair, and heart involvement. The homozygous variant linked to this disorder creates a premature stop codon which truncates early DSP protein without the entire C-domain (Norgett et al., 2000; Kaplan et al., 2004).

Subsequently, a variety of mutations were associated with dominant forms of AC in genes encoding desmosomal components, including: plakophilin-2 (*PKP2*), *DSP*, desmoglein-2 (*DSG2*), desmocollin-2 (*DSC2*), and *JUP* (Gerull et al., 2004; Rampazzo et al., 2002; Pilichou et al., 2006; Syrris et al., 2006; Asimaki et al., 2007). Thus, AC is now considered a disorder of the desmosome, as causative variants affecting different components of the cardiac desmosomes, have been reported in approximately 50% of AC subjects (Basso et al., 2012).

1.5.1 Desmosomes

Desmosome complexes are particularly abundant in tissues subjected to mechanical stress like the epithelium and the myocardium, where mediate mechanical anchorage of cardiomyocytes by connecting cytoskeleton to cell membranes of adjacent cells. In addition to cell adhesion, a function in cell-cell communication and tissue differentiation, and apoptosis, has been advanced (Chidgey et al., 2001; Merritt et al., 2002). These electron-dense symmetrical structures appear as dense membrane-associated plaques intercalated by 30 nm intercellular space which is divided by a central midline creating in this way the extracellular core domain, known as desmoglea, and the intracellular plaque. The intracellular plaque is commonly described as composed of two areas: the outer dense plaque, separated by a dense inner plaque (North et al., 1999). The desmosomal structure comprises transmembrane adhesive glycoproteins (components of the cadherin superfamily) and cytoplasmic proteins (components of the plakin and armadillo families). The outer dense plaque is where the cytoplasmic domains of the cadherins (DSG2, DSC2) attach to plakins (DSP) via armadillo proteins (PKP2 and JUP) (Figure 1.2). The inner dense plaque is where plakins attach to the intermediate filaments of the cell.

The connection between desmosomal components and intermediate filaments provides cellular adhesion and structural integrity.

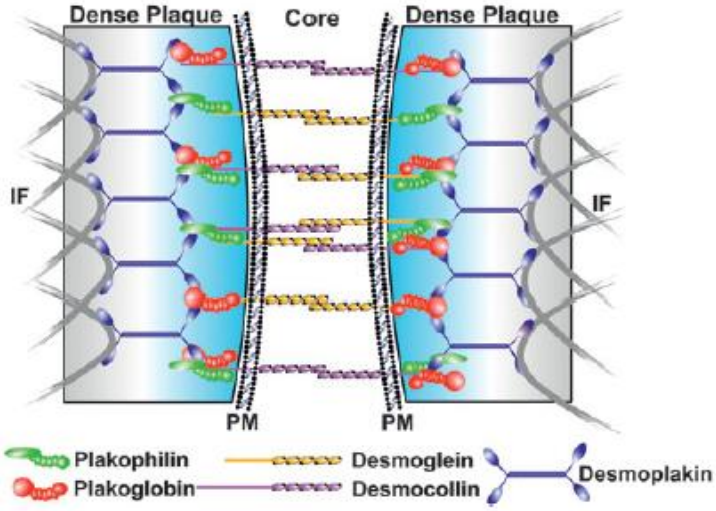

Figure 1.2: Schematic representation of desmosomal components (from Basso et al., 2009).

1.5.2 Desmosomal genes and proteins

Desmoplakin_DSP (6p24.3) belongs to the plakin protein family that includes a large number of proteins such as plectin, envoplakin and periplakin mediating the connection of different junctions with the cytoskeleton.

DSP is composed by an N-terminal domain which has a major role in targeting the protein to the membrane, a central coiled-coil rod domain involved in the protein dimerization and a C-terminal domain, that interacts directly to the intermediate filaments. It is expressed in all tissues containing desmosomes (Leung et al. 2002) and an alternative splicing of the DSP precursor mRNA produces two isoforms, differing in the length of the central α-helical domains: DSP I composed by 2871 amino acids and DSP II composed by 2271 amino acids.

The first pathogenic nucleotide variants described in DSP in families with autosomal dominant striate palmoplantar keratoderma without heart disease, were heterozygous nonsense (p.Q331X) and splice site (c.939+1G>A) variants (Armstrong et al., 1999). Soon after *DSP* was linked also to the autosomal dominant AC form, with the identification of a missense mutation in exon 7 (p.S299R) involving a high conserved amino acid and affecting a putative phosphorylation site (Rampazzo et al., 2002).

Since then, more than 100 *DSP* pathogenic variants have been detected in 5 to 16% of AC cases, many of them showing LV involvement (Pilichou et al. 2006; Fressart et al. 2010; Xu et al. 2010; Cox et al. 2011).

Plakophilin-2_PKP2 (12p11.21) the predominant protein isotype expressed in heart, belongs to the armadillo family of proteins and interacts directly to desmosomal cadherins and DSP (Chen et al. 2002). Two alternatively spliced mRNA transcripts gives origin to two protein isoforms: transcript 2b (881 amino acid long) and transcript 2a (837amino acid long) respectively (Gandjbakhch et al., 2011). PKP2 is comprise an amino-terminal head domain and nine armadillo repeat motifs and is necessary for heart morphogenesis and proper localization of DSP in mice (Grossmann et al., 2004).

In 2004, Gerull and colleagues described 26 *PKP2* different variants in 32 of 120 AC index patients, and suggested that the lack or aberrant inclusion of PKP2 into the desmosomes might affect cell-cell contacts (Gerull et al., 2004).

Up-to-date more than 120 pathogenic variants of *PKP2* have been associated with AC accounting for approximately 15 to 50% of reported cases (Gerull et al., 2004; Pilichou et al., 2006; Fressart et al., 2010; Xu et al. 2010; Cox et al., 2011). Most variants show a dominant inheritance with reduced penetrance, but recessive and compound heterozygous variants have also been identified in several patients (Xu et al., 2010). Moreover large deletions involving PKP2 have also been described (Cox et al., 2011; Li mura et al., 2013; Alcalde et al., 2014; see section 1.7.2).

Desmoglein-2_DSG2 (18q12.1) is a desmosomal cadherin belonging to the cadherin superfamily, which is involved in Calcium dependent cell-cell adhesion. In the desmosomes, there are four isoforms of desmogleins (DSG 1–4), showing tissue specific expression patterns (Green and Simpson, 2007). DSG2 is expressed in all tissues bearing desmosomes but is the only isoform expressed in cardiac tissue (Schäfer et al., 1994; Nuber et al., 1995). DSG2 has an intracellular anchor domain interacting with DSP, a transmembrane domain, four extracellular cadherin domains each one with a binding site to calcium that stabilizes structure and function of cadherins and a small signal and a preprotein domain.

DSG2 was linked for the first time to AC in 2006 with the detection of 9 pathogenic variants in a small Italian cohort (Pilichou et al., 2006). The majority of these variants co-segregating in 8 AC families were missense, located on highly

conserved amino acids spread in different domains of DSG2. In this gene was described for the first time associated with AC a patient with compound heterozygosity (Pilichou et al., 2006). Soon after another study confirmed *DSG2* association with AC by detecting 4 more variants in AC probands (Awad et al., 2006).

More than 50 *DSG2* variants have been detected in 2 to 20% of AC patients (Pilichou et al., 2006; Fressart et al., 2010; Xu et al., 2010; Cox et al. 2011).

Desmocollin-2_DSC-2 (18q12.1) is a glycoprotein, like DSG2, which belongs the cadherin superfamily mediating Calcium dependent cell-cell adhesion. Of the three desmocollin isoforms (DSC 1, 2,3), DSC2 is ubiquitously expressed in tissues bearing desmosomes and it is the only isoform present in cardiac tissue (Green and Simpson, 2007). DSC2 is composed of a signal domain, a preprotein domain followed by four highly conserved extracellular subdomains and an extracellular anchor domain at the N-terminus.

In 2006 Syrris and colleagues described the presence of heterozygous frameshift variants in *DSC2* in 4 affected AC probands not carrying variants in *DSP*, *JUP*, *PKP2* or *DSG2* (Syrris et al., 2006). In the same year, a second study involving 88 probands, identified in 1 AC patient a single heterozygous variant affecting an acceptor-splicing site in intron 5 (c.613-2A>G) of the *DSC2*, that resulted in skipping of the first 25 bp of exon 6 leading to a truncated protein (Heuser et al., 2006). Next 2 heterozygous missense variants in *DSC2* were found also in two Italian AC patients, resulting in the substitution of amino acids with different physicochemical properties and aberrant localization of the protein (Beffagna et al., 2007).

Less than 50 *DSC2* nucleotide variants have been reported in about 1 to 3% of AC cases (Syrris et al., 2006; Heuser et al., 2006, Fressart et al., 2010, Xu et al., 2010, Cox et al., 2011).

Plakoglobin_JUP (17q21.2), also called γ-catenin, is the major cytoplasmic protein of desmosomes which belongs to the armadillo proteins. It is present both in adherens junction and desmosomes, where it binds to the cytoplasmic domain of cadherins acting as a linker molecule between the inner and outer parts of the desmosomal plaque. JUP is formed by an N-terminal, a central containing highly conserved armadillo repeats and a C-terminal domain.

JUP was linked only recently to the dominant AC form followed by the description of a in frame insertion of a serine residue (p.S39_K40insS) in the N-terminus of the protein in a small German family (Asimaki et al., 2007).
Less than 20 *JUP* variants has been detected in approximately 1% of AC cases (Asimaki et al., 2007; Fressart et al., 2010; Xu et al., 2010; Cox et al., 2011).

1.5.3 Extra desmosomal genes and proteins

Although half of AC probands harbor a mutation in desmosomal encoding genes, approximately 1% of these patients carry mutations in non desmosomal genes (Alcalde et al., 2014).
Ryanodine Receptor_Ryr2 (1q43) is the only ryanodine receptor isoform present in the heart. Ryr2 is a homotetramer with a molecular weight of 565 kDa, that crosses the membrane of the sarcoplasmic reticulum and it is involved in the excitation-contraction coupling of the cardiomyocytes, as it induces the calcium release from the sarcoplasmic reticulum into the cytoplasm.
The cardiac ryanodine receptor has been linked to AC2 after the identification of six missense variants affecting highly conserved residues in families showing a dominant form of AC with effort-induced polymorphic tachycardias (Rampazzo et al. 1995; Tiso et al. 2001). The variants identified were supposed to increase the phosphorylation of the protein, with the consequent increase of the cytoplasmatic calcium.
Myocardial structural changes were once thought to result from altered intracellular calcium levels but the clinical phenotype is now recognized to be that of cathecholaminergic polymorphic ventricular tachycardia (CPVT) rather than AC, and is now considered a disease phenocopy (Laitinen et al., 2001; Priori et al., 2001).
Transforming Growth Factor-β3_TGFβ3 (14q24.3). In 2005, 2 nucleotide variants (c.-36G>A and c.1723C>T) were identified in the AC1 locus 14q23-q24, specifically in the 5' and 3' untranslated regions of TGFβ3 gene (Beffagna et al. 2005).
TGFβ3 is a multifunctional cytokine that belongs to the transforming growth factor superfamily, involved in the regulation of fibrosis, proliferation and cell adhesion. In vitro expression assays showed two-fold increase of the expression

level of mutated gene when compared to wild-type. Thus, mutations in TGFβ3 were thought to induce myocardial fibrosis, the hallmark AC feature, by stimulating the proliferation of mesenchymal cells and the production of extracellular-matrix components.

Although TGFβ3 is known to induce fibrosis, there is no evidence of TGFβ3 genetic variants causality in AC and no documented upregulation of TGFβ3 in vivo, therefore the effect of TGFβ3 variants in AC is still controversial.

Transmembrane Protein 43 _TMEM43 (3p25.1) was identified within the 2.36 Mb critical region of AC5 (3p25.1), through haplotype comparison of 15 clinically affected individuals with autosomal dominant AC from the island of Newfoundland in Canada (Merner et al. 2008).

TMEM43 encodes for a 400 amino acids protein containing a response element for PPARγ (an adipogenic transcription factor), which may explain the fibrofatty replacement of the myocardium, a characteristic pathological finding in AC. The gene product of *TMEM43* has been demonstrated to be the protein previously known as LUMA, which is a binding partner of emerin and the lamins, associated with Emery-Dreifuss muscular dystrophy (Liang et al., 2011).

Up-to-date the founder effect of one pathogenic variant in *TMEM43* (p.S358L) was observed in a fully penetrant and lethal form of AC and more recently, 3 additional variants were described in AC probands from UK without other causative variants in desmosomal encoding genes (Haywood et al., 2012).

Desmin_DES (2q35) is the main intermediate filament in heart muscle cells. The protein structure of DES is similar for all intermediate filaments composed by a central a-helical rod domain which is flanked by a globular head and a tail domain.

Mutations in *DES* have been described to underlie a heterogeneous spectrum of clinical-overlap syndromes (desminopathies) rather than typical AC (Basso et al., 2009).

The first *DES* mutation (p.R454W) demonstrated to affect the localization of DSP and PKP2 at the intercalated disk was identified in a family showing an early onset of conduction system disorder without mutations in other desmosomal genes. This evidence suggested a link between desmosomal and desmin-associated cardiomyopathies (Otten et al., 2010). Up to date 10 pathogenic

missense mutation have been associated to myopathy disorders with AC phenotype.

Titin_TTN (2q31.2), is a giant myofilament protein composed by 38.138 amino acids, essential to maintain structural integrity of the sarcomere. *TTN* was considered a good candidate gene because of the proximity to the locus 2q32.1-q32.3 (AC4) and its connection with the transitional junctions at the intercalated disk.

Eight unique missense *TTN* variants were identified in 7 families with AC-like phenotype and without desmosomal genes mutations, including a prominent p.T2896I variant which showed complete cosegregation in one large family (Taylor et al., 2011).

Lamin A/C_ LMNA (1q22) are intermediate filament-type proteins, the main building blocks of the nuclear lamina, that creates a meshwork underlying the inner nuclear membrane, and the nuclear interior.

In 2012 it was reported that mutations in *LMNA* gene mimic AC phenotype. Genetic screening of 186 UK patients with borderline or definite diagnosis of AC identified 4 nucleotide variants in 4 (4%) analyzed index cases in the absence of other desmosomal gene mutations. ECG findings of these patients showed severe structural and conduction abnormalities whereas 2 of them exhibited fibrofatty replacement at the endomyocardial biopsy (Quarta et al., 2012).

Phospholamban_PLN (6q22.1) is a protein involved in calcium homeostasis in the sarcoplasmic reticulum of the cardiac muscle cell. Dephosphorylated PLN interacts with the SERCA pump inhibiting the Calcium pump activity.

The founder p.R14del mutation in the PLN was identified in 15% of Dutch patients with DCM and in 12% with AC negative for other desmosomal gene mutations (van der Zwaag et al., 2012). AC mutation carriers showed typical clinical features with RV and LV involvement and consegregation of the mutation in a large AC family.

α-T-catenin_CTNNA3 (10q21.3) is a cytoplasmic molecule necessary for dynamic maintenance of tissue morphogenesis, which is involved in cell-cell adhesion of contractile cardiomyocytes by binding PKPs and by contributing to the formation of the area composit,.

CTNNA3, located on 10q21.3, was only recently linked to AC13 by analyzing 76 AC negative Italian probands for mutations desmosomal-related genes. Up to date only 2 variants has been identified, a missense mutation (p.V94D) and an in-frame 3-bp deletion (p.L765del) , both located in important domains of CTNNA3 affecting the interaction between the mutant protein and β-catenin (van Hengel et al., 2012).

1.6 Mecchanisms in Arrhythmogenic Cardiomyopathy pathogenesis

Several theories have been advocated in the pathogenesis of AC: dysontogenic, degenerative, inflammatory, transdifferentiation, as causative or secondary factors for the trigger and progression of the disease.

The dysontogenic theory considered AC as a milder form of Uhl's anomaly, which is a congenital heart defect charactherized by hypoplasia of the RV myocardium at birth (Dokuparti et al., 2005); at the contrary, in AC has been demonstrated that myocyte loss occurs progressively starting from childhood (Daliento et al.,1995).

The degenerative theory which was postulated in 1996, considered AC as a consequence of myocyte death, either by necrosis or apoptosis, due to inherited ultra-structural defects (Mallat et al., 1996; Valente et al., 1998). Experimental data have shown that the key initiating phenomenon in the cascade of events that lead to fibrofatty replacement of the normal myocardium is myocyte necrosis (Pilichou et al., 2009; Rizzo et al., 2012).

In the inflammatory theory the disease is considered the result of preceding myocarditis, since myocardial inflammation is a common feature in hearts with AC (Bowles et al., 2002; Thiene et al., 1991). More recent studies considered that viral myocarditis overlay on an already affected heart accelerating the disease progression, rather than being involved as a primary factor in the etiology of the disease (Calabrese et al., 2006).

The transdifferentiation theory suggests that cardiac myocytes undergo a metamorphosis and switch to the fate of adipocytes (D'Amati et al., 2000). Even though this theory seems questionable due to the limited dedifferentiation capabilities of adult cardiomyocytes, recent studies supported the idea that adipocytes in AC derive from progenitor cells of the second heart field, which

give rise to the bulbus cordis and the pulmonary infundibulum (Lombardi et al., 2009). According to this hypothesis, the progenitor cells of the second heart field switch into adipocytes because of suppressed Wnt/β -catenin signaling as a result of the JUP translocation to the nucleus (Garcia-Gras et al., 2006). This pathway is known to regulate adipogenesis, fibrogenesis and apoptosis, however contradictory results are coming from experimental studies regarding the activated of this pathway (Li J et al., 2011; Li D et al., 2011; Lombardi et al., 2011).
Most recently, desmosome disruption in AC was linked to activation of Hippo/Yes-associated protein (YAP) pathway. The activation or inhibition of this pathway regulates cardiomyocyte proliferation and thus heart size (Chen SN et al., 2014), suggesting that molecular changes at the intercalated discs in AC patients modulate the cros talk between Wnt/β-catenin and Hippo/YAP pathway.

However the role of desmosomal gene mutations in the pathophysiology of myocardial injury remains elusive (Basso et al., 2011). Pathogenic variants located in the major components of cardiac desmosomes (DSP, DSG2, DSC2, PKP2, JUP) may influence desmosome composition and function, organization of junction assemblies at the intercalated disc level, or alter in the Wnt/β-catenin signalling pathway:
- The pathogenic effects of desmosomal genes variants may result in the incorporation of abnormal protein affecting the correct assembly and function of desmosomes (dominant negative effect), or in a insufficient incorporation of a normal protein in the desmosomal structure (haploinsufficiency), or in the loss of essential protein-protein interactions. Pathogenic variants in the desmosomal components may cause the disruption of these structures resulting in the reduction of the mechanical contacts between cells. The impairment of the desmosomal organization would lead to loss of myocyte adhesion, and a subsequent cell death, that could be enhanced by physical activity (Delmar and Mc Kenna, 2010). As the cardiac myocytes have a limited regenerative capacity cell death may activate a repair mechanism implying replacement by fibrous and adipose tissue.
Another possible effect of abnormal desmosomal components may result in the structural re-organisation of the cardiac junctions known as area composita, which comprise both desmosome and gap junction components (Franke et al., 2006).

The remodeling of gap junction assembly (Patel et al., 2014; Asimaki et al., 2014; Kleber et al., 2014) may lead to impair localization of the gap junctional protein connexin 43, resulting in heterogeneous conduction of the electric impulse and an increased propensity for arrhythmias.

- Desmosome components involved in cell-cell adhesion, such as JUP, are also mediators in intercellular signaling. Desmosome dysfunction leads to the translocalization of JUP from the desmosome to the nucleus, where it suppresses the canonical Wnt/β-catenin signaling pathway (Garcia-Gras et al., 2006). This would cause increased expression of transcriptional regulators of adipogenesis, that have been suggested to mediate the differentiation of second heart field cardiac stem cells into adipocytes (Garcia-Gras et al., 2006; Lombardi et al. 2009).

1.7 Genetic Screening in AC

1.7.1 Conventional mutation analysis techniques

Since the 70s direct sequencing developed by Sanger (Sanger et al., 1977) has been the method of choice for DNA sequencing. Sanger sequencing is based on the amplification of specific DNA regions. Despite the technique underwent several modifications and improvements including its partial automation, it remains relatively time-consuming and expensive to keep pace with the growing need of sequencing larger DNA portions.

In genetically heterogeneous diseases, the cost and effort of DNA sequencing is often considerable and numerous DNA pre-analytic techniques have been developed for the detection of point mutations and small deletions. Pre-analytic techniques, such as denaturing gradient gel electrophoresis (DGGE), single-strand conformation polymorphism (SSCP), chemical cleavage method (CCM), denaturing high performance liquid chromatography (DHPLC), all followed by direct sequencing of aberrant conformers or elution profiles allowed to minimize sequencing to a subset of abnormal PCR products.

DGGE is a technique developed by Fischer and Lerman (Fischer and Lerman, 1983) that can identify homoduplex molecules that differ by single bp substitutions. In DGGE, double-stranded (ds) DNA is electrophoresed through a

gradient of increasing concentration of a denaturing agent (urea or formamide) or of increasing temperature. In this electrophoresis gradient, DNA domains dissociate according to their melting temperature, resulting in decreased electrophoretic mobility. Mutant DNAs denature at a different isoelectric point and thus are distinguished from normals. DGGE however appears not suitable for the detection of somatic mutations where few mutants allele compete with an excess of normal alleles (Nollau et al., 1997).

SSCP analysis is based on the differential electrophoretic mobility of single-strand DNA molecules that differ by a single base (Sidransky et al., 1997). Thermally denatured DNA is electrophoresed, and those single stranded DNA fragments that take up an altered conformation show up as aberrantly migrating bands on the electrophoretic gel. It is relatively simple, quick and cheap. However, for optimal results, fragment size should be in the range of 150 to 200 bp (Sheffield et al., 1993) and SSCP analysis becomes increasingly inefficient at detecting mutations with the increasing size of the PCR product.

In CCM PCR heteroduplexes are incubated with two mismatch-specific reagents: Hydroxylamine modifies unpaired cytosine and potassium permanganate modifies unpaired thymine. The samples are then incubated with piperidine, which cleaves the DNA backbone at the site of the modified mismatched base. Cleavage products are separated by electrophoresis, revealing the identity and location of the mutation. Compared with other mutation-detection techniques (such as SSCP, DGGE) that detect mutations in short DNA fragments and require highly specific melting temperatures, CCM has a higher diagnostic sensitivity, and can analyze amplicons < or = 2 kb in length (Tabone et al., 2006), although it is technically challenging.

DHPLC technique is based on the detection of heteroduplexes containing a mutation or polymorphism by reduced column retention of heteroduplexes compared to the respective homoduplexes under partially denaturing conditions (Oefner and Underhill, 1995). DHPLC, under optimized conditions, is cost-effective, highly accurate, rapid, and efficient for mutation detection. A disadvantage of DHPLC is the requirement and maintenance of a specialized and expensive equipment and optimization of each reaction required achieving the

highest sensitivity for mutation detection, moreover it can also be difficult to identify homozygotes unless the sample is spiked with a known control.

For many years the use of techniques such as SSCP or DHPLC analysis together with Sanger sequencing was an absolute necessity, due to the lack of affordable and high-throughput methods, until the field of genetic screening was changed in 2008 by the launch of the Next Generation Sequencing (NGS) technology.

1.7.2 Copy number variants

Copy number variants (CNVs) including deletions, duplications, and other complex genomic rearrangements that lead to a change in the number of copies of a specific chromosomal region, represent a major source of genetic variability in human DNA (Beckmann et al., 2007). CNVs may be benign polymorphic variants (Copy Number Polymorphism) or pathogenic variants underlying a wide range of disorders (Mendelian or complex diseases) by various molecular mechanisms such as gene dosage, gene disruption, gene fusion, position effects (Zhang et al., 2009). The extent to which copy number variants contribute to human disease is still unknown.

The techniques commonly used for the detection of point mutations, such as direct sequencing and DHPLC, usually fail to find CNVs. Different molecular assays have been developed to detect CNVs such as, Southern blot, quantitative PCR assays (qPCR), Fluorescence *in-situ* hybridization (FISH), array comparative genomic hybridization (CGH-array), single nucleotide polymorphism genotyping platforms (microarray analysis), Ligation-dependent Probe Amplification (MLPA).

Traditionally, Southern blots have been used to determine gene copy number (Southern et al., 1975). Southern blotting is a powerful method for typing structural rearrangements. The principle of the technique relies upon fragmentation of DNA with a restriction endonuclease enzyme, followed by gel electrophoresis and transfer to a nylon membrane (blotting) (Mellars et al., 2011). However, this techniques requires a significant quantity of gDNA, it is laborious and time consuming.

Quantitative PCR (qPCR) is a high throughput technique for determining gene copy number, it is based on the measurement of PCR amplicon accumulation in real time. Amplicon accumulation is measured by fluorescent based chemistry, which primarily consists of either DNA intercalating dyes such as SYBR green. In qPCR the target DNA is amplified in a quantitative reaction for around 90 min, without requiring digesting, blotting or overnight steps.

FISH is a technique typically used to identify chromosomal abnormalities from metaphase or interphase spreads using fluorescent probes. The strength of FISH lies in the direct visualization of DNA copy number at the single cell level (Cantsilieris et al., 2013), however because of its low resolution (about 5-10 Mbp for FISH) short CNVs are still difficult to detect (Duan et al., 2013).

CGH-array is a technique based on dual hybridization of test and reference DNA to either short or long oligonucleotides immobilized on a glass slide (Carter et al., 2007). The signal ratio between test and reference sample is normalized and used to infer copy number. Although this technique allows the determination of breakpoints, the resolution of CGH-array, around 1-25 kb with 1 million probes for CGH-array (Yoon et al., 2009), still makes difficult the detection of short CNVs (Duan et al., 2013).

SNP microarrays are also hybridization based and have the advantage of analyzing both single nucleotide differences and in some cases non-polymorphic copy number probes that are not restricted by sequence properties of SNPs (McCarroll et al., 2008). The main strength of microarray platforms is the ability to screen CNVs on a genome-wide level at relatively low cost in large data sets, however there are several limitations: the ability of microarray platforms to resolve breakpoints and detect small rearrangements is generally poor, and is dependent on probe location and spacing, moreover structural rearrangements that do not affect copy number (such as inversions and translocations) are not detected (Cantsilieris et al., 2013).

MLPA is a versatile semi-quantitative method used to determine the relative copy number of several nucleic acid sequences in a single multiplex PCR-based reaction (Schouten et al., 2002), from relatively low amounts of gDNA. The technique relies upon hybridization and ligation of two adjacent oligonucleotides to a specific gDNA sequence. The strengths of MLPA lie in the number of loci

that can be analyzed in a single reaction, the specificity of the ligation step, the reliability and accuracy of CNVs measurement, and the relative low cost for conducting large scale association studies. Moreover MLPA, compared to FISH and Southern blot analysis, targets short (50-70 nt) fragments allowing also the identification of single gene/exon aberrations and it is a low cost and straightforward technique compared to CGH-array (Figure 1.3). However it does not allow the exact determination of the breakpoints.

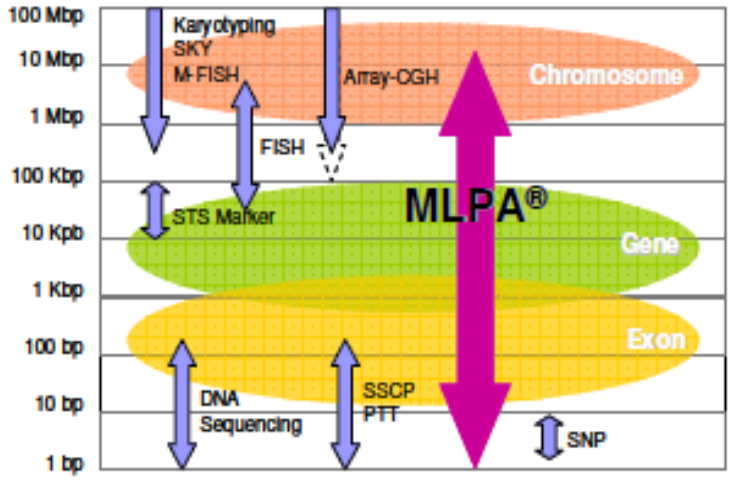

Figure 1.3: MLPA in comparison to other techniques. MLPA can detect a wide range of genomic alterations, from single point mutations to large chromosomal deletions/duplications (from MRC Holland).

Recently CNVs have been associated to different cardiomyopathies (Koopmann et al., 2006; Bisgaard et al. 2006; Gupta et al., 2010) including AC (Cox et al., 2011, Roberts et al., 2013; Li Mura et al., 2013). In the first study, MLPA revealed 3 *PKP2* exon deletions in 3 of the 149 AC cases: rispectivly the deletion of exon 8, the deletion of a region including exon 1 and 4 and the deletion of the entire coding region of *PKP2* gene (Cox et al., 2011). The frequency of CNVs in this study was about 2%, comparable to *DSG2* or *DSC2* mutations in the Dutch cohort. Later, Roberts and colleagues described large deletions involving *PKP2* in 2 AC cases identified both by microarray and MLPA analysis. The *PKP2* deletion including exons 2-14 was identified in one index case and his son, while the other AC patient carried a large de novo 7.9 Mb deletion on chromosome 12p12-1p11, which comprise the entire *PKP2* gene (Roberts et al., 2013). More recently, another 122.5 kb deletion in the chromosomal region 12p11.21 encompassing

PKP2 was identified in an Italian AC patient and his relatives by SNP-array and qPCR analysis (Li Mura et al., 2013). Hence, CNVs screening is now strongly recommended when conventional sequencing does not identify pathogenic variants in AC-related genes (Alcalde et al., 2014).

1.7.3 Next-generation sequencing (NGS)

Since the conclusion of the Human Genome Project in 2000, based on Sanger technology, which required about 13 years (Lander et al. 2001; Venter et al. 2001) and 70 million dollars (Metzker et al., 2010), deep sequencing technologies were developed to sequence in short range of time whole genomes and exomes. NGS, also known as 'deep sequencing' and 'high-throughput sequencing', employ new sequencing apparatuses able to produce millions of small DNA sequences (reads) in a single run.

In the last years NGS ability to increase throughput and to decrease sequencing costs shifted the interest from the research of variants in specific DNA regions to the identification of variants from a genome-wide sequencing data. This has led to advances in diagnostics and scientific research with the detection of variants linked to Mendelian and complex diseases (Sobreira et al., 2010; Rabbnai et al., 2012; Yang et al., 2013; Krawitz et al., 2014; Nijman et al., 2014). NGS has been successfully applied for instance to the identification of genetic defects of a variety of disorders such as cancer (Mwenifumbo et al., 2013, Kougioumtzi et al.,2014), neurological diseases (Della Mina et al., 2014; Zhao et al., 2013), intellectual disability (Iqbal et al., 2012; Tsurusaki et al., 2011), mitochondrial dysfunctions (Dames et al., 2013; Tang et al., 2013), cardiovascular diseases (Mook et al., 2013; Dorn et al., 2014).

1.7.3.1 Applications of NGS

NGS has the potential to focus on the analysis of entire genomes (Whole Genome Sequencing-WGS), on the sole coding part of the genome (Whole Exome Sequencing, WES) or on specific target genes (Targeted Resequencing). NGS technologies may also be employed for the detection of DNA methylation sites (Methylation Sequencing, Methyl-seq), for DNA-protein interactions studies (Chromatin Immunoprecipitation Sequencing, ChIp-seq), for transcription factor

profiling and gene expression quantification studies (RNA-seq), ribosomal Sequencing (Ribo-seq), small RNA profiling including microRNAs and promoter-associated RNAs (sRNA-Seq), thus making possible epigenetic and transcriptome analysis.

Whole Genome Sequencing (WGS)

With the availability of NGS technologies, WGS experienced a remarkable growth in the last few years. WGS was first applied to bacterial genomes using 454 technology (Smith et al., 2007) and since then, a large number of different genomes have been sequenced. As of January 2014, for instance, the number of genome sequencing projects in the Genome Online Database reached 37.540 (Pagani et al., 2012). In addition to the sequencing of the genomes of different organisms, such as animals (Yang et al., 2014; Zhou et al., 2014; Schroyen et al., 2014), plants (Cao et al., 2011; Matsumoto et al., 2008; Baev et al., 2014; Pacurar et al., 2014), bacteria (Fournier et al., 2014; Pylro et al., 2014; Mathers et al., 2015; Wang et al., 2015) and viruses (Ericsen et al., 2014; Lwande et al., 2014; Marcacci et al., 2014), projects to characterize the DNA sequence of individuals have gathered pace, and WGS of humans is becoming commonplace (Gonzaga-Jauregui et al., 2012). Several groups are studying human genomic variation by sequencing or genotyping large number of individuals, including multi-institute consortia projects such as the 1000 Genomes Project (1000 Genomes Project Consortium, 2010), the 10K project (Genome 10K Community of Scientists, 2009), the Personal Genome Project (Ball et al,. 2012), the HapMap project (International HapMap Consortium, 2003).

Whole Exome Sequencing (WES)

Whole exome sequencing (WES) study design can provide a new strategy to explore the aetiological roles of rare variants in complex diseases. At first WES has been successfully applied to the research setting for the elucidation of candidate genes and causal variants in individuals and families with different genetic disorders. In 2009, Ng and coll. successfully identified by WES one missense mutation in the Myosin Heavy Chain 3 gene (*MYH3)* among 4 unrelated affected individuals with the rare autosomal dominant Freeman-Sheldon syndrome (Ng et al., 2009). Since this initial study, a growing number of reports have described the successful application of WES in identifying causative variants

in disorders with recessive and dominant inheritance patterns as well as in disorders caused by *de novo* variants (Choi et al., 2009; Ng et al., 2010; Vissers et al., 2010; Vilariño-Güell et al., 2010; Girard et al., 2011; Choquet et al., 2015; Schottmann et al., 2015; Menezes et al., 2015). These achievements have led to increasing interest in the application of WES in the clinical practice. Among the advantages of WES, the selective sequencing of the coding sequence of the human genome is less expensive in comparison to WGS since the exome represents only about 1% of the genome. In the clinical setting, WES can enable a new approach to diagnose patients with an unclear clinical diagnosis and allow the study of the entire exome than a relatively small subset of genes traditionally associated with the patient's condition. Moreover, genotyping or resequencing the entire exome in patients who are negative from classical screening of disease-related genes subset enables the research in genes and nucleotide variants not yet associated to the disease, together with the possibility of storing NGS data for potential future studies.

However, while attractive, the introduction of WES for clinical laboratories remains challenging, especially because of the laborious and accurate process of bioinformatics analysis and data interpretation that are required to identify candidate genes and causal variants.

Targeted Resequencing (TR)

The success of NGS in research resulted in its adoption in clinical care. Although WGS and WES are the most comprehensive strategies for genetic analysis, in the clinical setting, are still unaffordable. In comparison, TR can reach higher coverage of exonic regions of interest while reducing the sequencing cost and time. These rapid, accurate, and relatively low cost method allows a high-throughput, genotype-based approach to molecular diagnosis, infact TR may rapidly screen at once large panels of genes associated with a particular phenotype or may provide differential diagnosis in diseases that present with atypical manifestations, or for which not all genetic variants are known.

The transfer of NGS from research to diagnostic laboratories already allowed the development, validation and implementation of TR Gene Panel assays for the NGS-based genetic testing and diagnosis of patients with different disorders

(Jones et al., 2013; Blue et al., 2014; Verma et al., 2014; Glod et al., 2014; Morgan et al., 2014; Ankala et al., 2014; Sikkema-Raddatz et al., 2013).

Chromatin Immunoprecipitation Sequencing (ChIp-seq)

NGS technologies offer the potential to substantially accelerate also epigenomic research, including analysis of post-translational modifications of histones, of the interaction between transcription factors and their direct targets, of nucleosome positioning on a genome-wide scale and of the characterization of DNA methylation patterns (Bormann et al., 2010; Fouse et al., 2010; Bhaijee et al., 2011). Using ChIP-Seq technology, post-translational modifications of histones and the location of transcription factors can be studied at the whole-genome level (Neff and Armstrong, 2009). Originally ChIP-enriched sequences were analyzed using array technology. However limited array content required the selection of the sequences to analyze, generally genes promoter and CG-rich regions were selected. ChIP-seq technique gains two main advantages from the NGS approach, firstly, it is not limited by the microarray content, so that all sequences bound are identified, and secondly, it does not depend on the efficiency of probe hybridization. Many ChIP-seq studies have been performed, such as the ENCODE (Consortium, 2004) and FANTOM5 (Andersson et al., 2004) projects, that revealled genome wide profiles and binding sites for a range of DNA binding proteins.

Transcription factor profiling

The development of NGS technologies has provided a unique opportunity to study transcriptional regulatory networks. Transcription factors are essential elements that regulate the transcription and the spatio-temporal expression of genes, thereby ensuring the accurate development and functioning of an organism. Earlier, the quantification of the production of nascent RNAs was achieved by bromo-uridinating nuclear run-on RNA molecules followed by sequencing (Core et al., 2008) or by immunoprecipitation of RNA polymerase followed by sequencing of the bound RNA fragments (Churchman and Weissman, 2001).

The use of NGS for transcription factor profiling have greatly advanced the understanding of gene regulation and revealed additional mechanisms in the regulation of gene expression (Sharma et al., 2013; Driessen et al., 2013), and large international consortia (Adams et al., 2012; ENCODE, 2012) and numerous

laboratories are generating and releasing such genome-wide datasets into the public domain.

Methylation Sequencing (Methyl-seq)

NGS has also enabled the mapping of epigenetic marks such as DNA methylation and histone modification patterns in a genome-wide manner. Methylation of cytosine residues in DNA is the most studied epigenetic marker and is known to silence parts of the genome by inducing chromatin condensation (Newell-Price et al., 2000).

Precise genome wide mapping of methylation patterns has been made possible by various NGS techniques, including methylated DNA immunoprecipitation (Taiwo et al., 2012), MethylC-seq (Lister et al., 2009) and reduced representation bisulfite sequencing (Meissner et al, 2005). With respect to the technologies used to isolate the methylated sequences, i.e. methylated DNA immunoprecipitationsequencing and methylated CpG island recovery assay, NGS-analysis clearly reveals all sequences enriched. Many NGS methylation studies have been presented among which those of the ENCODE consortium (Consortium, 2004).

RNA-sequencing (RNA-seq)

Beyond DNA sequencing and epigenetic analyses, NGS has also enabled the global mapping of the transcriptome by RNA-sequencing (RNA-seq). High-throughput methods have facilitated the detection and quantification of transcripts and discovery of novel isoforms. Significant interest also lies in uncovering the role of various regulatory factors in controlling the expression of genes, such as transcription factors and non-coding RNAs.

Microarray technologies provided the first practical technique for measuring genome-wide transcript levels. However, microarrays were only applicable to studying known genes and had significant problems with cross-hybridization and high noise levels (Wang et al., 2009). Much more accurate measurement of mRNA levels became possible with the introduction of RNA-seq, which was developed in both yeast (Nagalakshmi et al., 2008) and mammalian cells (Cloonan et al., 2008). This method uses the high-throughput sequencing of cDNA fragments generated from a RNA library. It allows precise quantification of transcripts and exons, moreover it is a great tool for identifying novel genes and RNAs.

The ability to detect and accurately quantify transcript levels using NGS technologies has significant impacts for research and clinic (McManus et al., 2010; Li et al., 2014; Shin et al., 2014; Zhao et al., 2014; Liu et al., 2014).

Ribosomal Sequencing (Ribo-seq)

The impact of NGS technologies on Ribosomal RNA characterization has been noteworthy. Ribosomal Sequencing (Ribo-Seq) was first developed by Ingolia in 2009 in order to measure the quantities of ribosome-bound fragments by freezing ribosomes and using the translation inhibitor cycloheximide 9 (Ingolia et al., 2009), the mRNA is then digested and the resulting fragments are sequenced to reveal mRNA regions occupied by ribosomes. The quantification of ribosome-bound regions reflects the translation efficiency.

The successive use of NGS technologies for metatranscriptomic profiling has been so successful that at present scientists have access to the full set of coding and non-coding RNA in a community of organisms (Bomar *et al.*, 2011; Shi *et al.*, 2009; Stewart *et al.*, 2011). Sequencing of the small sub-unit rRNA genes have also been helpful in the determination of bacterial diversity and structure highlighting the microbial diversity of various ecosystems (Taib et al., 2013; Toma et al., 2014; Chen et al.,2014).

Small RNA profiling (sRNA-Seq)

NGS has also emerged as a direct profiling method for small RNA (sRNA-Seq), that in most eukaryotic organisms function as guides in association with specific proteins for regulation at the post-transcriptional or transcriptional level. Compared to platforms such as microarray or PCR-based assays, sRNA-Seq allows semi-open-ended analysis of both known and unknown sRNAs. Most initial sRNA-discovery studies used pyrosequencing (Margulies et al., 2005; Rajagopalan et al., 2006; Ruby et al., 2006). Subsequently, different NGS platforms with higher throughput and improved methods such as short-read sequencing-by-synthesis of amplified DNA colonies, that facilitate sequencing depth with respect to previous methods (Bentley et al., 2006), were applied for sRNA-Seq and resulted in the discovery of an ever-growing number of sRNA (Seila et al., 2008; Affymetrix ENCODE Transcriptome Project et al., 2009; Taft et al., 2009; Klimczak et al., 2014).

1.7.3.2 NGS Platforms

The main NGS platforms available in the market at present are HiSeq and MiSeq sequencers product by Illumina/Solexa, San Diego, CA, USA, Ion Torrent Proton, Personal Genome Machine- PGM and Solid3 by Life Technologies, Carlsbad, CA, USA, 454 Genome Sequencer-GS FLX and GS Junior System by Roche Applied Science, Mannheim, Germany. Each platform employs different sequencing chemistries and amplification methods, and this causes different strength and weakness points. The choice of a platform depends upon the expectations of research question, accordingly one may be concerned about the performance such as read lengths, data output, accuracy and cost.

Illumina platforms

Illumina platforms are based on bridge amplification after library preparation and the use of reversible dye terminator for sequencing.

The Genome Analyzer- GA was first introducded by Solexa in 2006 and further developed by Illumna, initially reads were very short (3+5-bases) in comparison with Sanger sequencing. Since then many technical improvements have been made and presently Illumina platforms can produce reads ranging from 100 to 250 bases, but with high throughput. Read-lengths are however limited by multiple factors that cause signal decay and dephasing, such as incomplete cleavage of fluorescent labels.

In early 2010 Illumina launched the HiSeq2000, that mantains the same sequencing strategy of the GA and can produce up to 200 Gb of 100-base paired-end reads in the course of a 8 days run.

In 2011 Illumina introduced the MiSeq, that implements the sequencing by synthesis technology and is similar to the HiSeq system. It is a benchtop sequencer, suitable for single-day experiments, that combines the processes of cluster generation, sequencing by synthesis, and data analysis into a single machine generating up to 5.1 Gb of 150-base paired-end reads per day (http://www.illumina.com/).

The main error type of Illumina technology is substitution, due to the simultaneously addition of all the four nucleotides in the reaction mix. However, because of the lower cost per sequence than other platforms, interest in amplicon

sequencing on Illumina platforms is growing (Guilhamon et al., 2014; Manson-Bahr et al., 2014; Igartua et al., 2015; Schirmer et al., 2015).

Roche platforms

Roche platforms use pyrosequencing and detect chemiluminescent signals of pyrophosphate release on nucleotide incorporation. They can produce read lengths close to Sanger (400-1000 bp), however, the price per base on the Roche 454 platform is much higher than other NGS platforms.

The first commercially successful NGS sequencer was the pyrosequencing platform developed by 454 Life Sciences in 2005, with a read length of 100–150 bp and output reads exceededing 200,000 reads and 20 MB in a single run. The 454 GS FLX platform launched in 2008 can generate 40 Mb of 400-base paired-end reads in a 10-hours run.

The benchtop sequencer 454 GS Junior released by Roche in early 2010 produces 35 Mb of sequence data from 400 to 600-base paired-end reads in a 10-hours run (http://454.com/products/gs-junior-system/).

The main disadvantage of Roche sequencers is due to the difficulty of the pyrosequencing technique to solve long homopolymeric regions, so the principal errors are base insertion or deletion in these regions during sequencing, rather than substitutions as in the Illumina platforms. On the other hand, thanks to the addition of a single dNTP each time, substitution errors are rarely encountered in these reads.

Life Technologies platforms

Life Technologies uses a similar sequencing method that detects signals of H^+ ions release due to binding of nucleotides by semiconductor detector (Biswas et al., 2014).

Ion Torrent Proton is a high-capacity sequencer that employs a semiconductor-based technology interfaced with sequencing-by-synthesis (Rothberg et al., 2011). It permits 200-bp reads in 2 hours and can generate up to 10 Gb per run (http://www.lifetechnologies.com/).

Its lower-capacity counterpart, the Ion PGM, released by Ion Torrent at the end of 2010, is a bench-top sequencer that implements semiconductor sequencing technology without requiring fluorescence and camera scanning, resulting in higher speed, lower cost, and smaller instrument size. Currently, it enables up to

400-bp reads in 2-4 hours and can generate up sequence data form 30Mb to 1 G per run (http://www.lifetechnologies.com/).

The SOLiD system, available from 2007, uses a chemistry based on the polony sequencing technique published in the same year as the 454 method (Shendure et al., 2005), with the difference that in the SOLiD method base detection of the DNA fragments is achieved by sequencing by ligation instead of through polymerase reaction. SOLiD can achieve very high-throughput, up to 320 Gb of sequence data in 50-base paired-end reads per an 10-day run (http://www3.appliedbiosystems.com/). The 2-base enconding system allows to determine and to correct almost all sequencing errors. However short read length (50 bases) and resequencing only in applications is still its major shortcoming (Liu et al., 2012).

1.7.3.3 Limitations of NGS

Despite the numerous advantages of the NGS technologies, thare are still technical limitations, expected to decrease in the next few years.

Limitation in the read lengths remains among the major technical drawbacks. Although recent sequencers achieve up to 1000 bases, the average length ranges between 100 and 400 bases, and short read-lengths are known to involve more complicated analyses. Reads are assembled using overlapped ends to build longer stretches of DNA, which are attached to each other until the assembly of an entire chromosome. Short reads include shorter overlapping ends, which makes the accurate determination of the previous and subsequent reads difficult. Therefore, the analysis and assembly efforts required increases by several folds for short read-lengths in comparison to longer lengths.

Another challenge concerns reads coverage. Reads produced by NGS do not completely cover the entire target region, either exome or a gene set: some sequences such as GC-rich sequences can be difficult to capture and some sequences cannot be targeted at all because of insufficient or inadequate probes. Moreover uneven capture efficiency across exons influences the reads depths of specific exons. Although the off-target hybridization value is often important (20%) not all sequences can be aligned with the reference sequence to allow SNP calling.

Furthermore, NGS creates lots of background noise, consisting in mainly sequencing errors, which can be platform-specific since technically linked to the sequencing chemistrie implemented in a particular sequencing platform (Coonrod et al., 2012). The most common error type in NGS is substitution, which complicates the distinction of a true single nucleotide variant. The indel (insertion/deletion) error is another common error type especially in the Ion Torrent and SOLiD platforms. Errors can also arise from mistakes during the library preparation and sequencing or during the alignment. For instance read misalignments often arise in presence of repetitive regions and pseudogenes.

Another limitation of NGS is the difficult detection of structural variants, such as copy-number variants, inversions and translocations. However recent studies show the sucelssufull identification of CNV from NGS data (Zhao et al., 2013; Wang et al., 2014; Duan et al., 2014; Horpaopan et al., 2015).

Another issue concerns the reproducibility of data, regarding data replication and verification. The possibility of obtaining different mutation patterns from two NGS experiments carried out by different groups without any overlap between the two sets of mutations (Agrawal et al. 2011; Stransky et al. 2011) suggests data reproducibility as a major limit of NGS (Faita et al., 2012; Nekrutenko et al., 2012).

Furthermore, one of the more evident challenges of NGS is the huge amount of data that require proper analysis and storing. Storage of computational data is becoming an expensive problem, hampered by increasing data volumes and frequent updates of analysis methods and tools (Lampa et al., 2013). Long term archiving of large amounts of data is becoming an important task, as scientists are reluctant to discard raw data since improved algorithms may help extract further information from them in the near future.

Management, filtering and prioritization of variants and clinical interpretation of genetic findings require expertise. NGS data management and interpretation require a large extent of bioinformaticians and biologists who have learnt basic bioinformatics.

1.7.3.4 NGS and Cardiomyopathies

NGS represents a valuable approach for the molecular analysis of genetically heterogeneous disorders, such as inherited cardiomyopathies, and SD-associated disorders. These diseases are linked to multiple causative genes, often showing rare and private variants segregating within small families, and whose direct sequencing would be laborious, time consuming and expensive.

In the last years linkage analysis performed in large families and genome-wide association studies performed in diverse populations have been successful in detecting causal loci, genes and variants of inherited cardiomyopathies. However, the wide locus and allelic heterogeneity typical of cardiomyopathies requires to fully sequencing the whole coding region of several genes, making NGS the most appropriate technique for this purpose.

Although cardiomyopathies usually present different clinical features, they exhibit some level of genetic and, especially in their end-stage, phenotypic overlap. For instance the detection of causative variants in desmosomal genes sometimes also occur in DCM patients in addition to ACM patients (Posch et al., 2008; Elliott et al., 2010; Zhang et al. 2011; Garcia-Pavia et al., 2011; Garcia-Pavia et al., 2013), so that variants in the same gene or gene set correspond to a wider phenotypic spectrum. Thus, NGS allows a differential diagnosis approach for correctly identifying patients with unclear clinical characteristics.

With different NGS approaches genomic regions of interest can be targeted to enable clinical approach for the diagnostic purposes. Cardiomyopathy associated genes are sequenced by performing customized targeted sequencing on NGS platforms and forming a panel and presented as diagnostic tool for the cardiomyopathy in a clinical setup (Biswas et al., 2014).

Different NGS approaches, such as targeted resequencing or WES have been successful in finding causative mutations in different cardiomyopathies.

A study involving 223 unrelated probands with Hypertrophic cardiomyopathy, using targeted resequencing of 41 cardiovascular genes, reported 152 distinct candidate variants in sarcomeric or associated genes (89 novel) in 143 patients (64%), and an additional 94 candidate variants (73 novel) in desmosomal, and ion-channel genes in 96 patients (43%) (Lopes et al., 2013).

Another study described the use of NGS in order to search for de novo genetic variants in 2 unrelated infants presenting with recurrent cardiac arrest and prolonged QTc interval. The 2 parent-child trios were investigated by WES and follow-up candidate gene screening was performed on an independent cohort of 82 subjects with long-QT (LQT) syndrome without mutations in LQT genes. 3 heterozygous de novo mutations were detected in calmodulin1 (*CALM1*) or calmodulin 2 (*CALM2*) in the 2 probands and in 2 additional subjects with recurrent cardiac arrest. Additional biochemical studies were performed to determine the functional consequences of these mutations (Crotti et al., 2013).
A recent study identified a novel disease-causing variant in α-actinin 2 by targeted resequencing 48 disease genes for HCM, which was shown to segregate with the cardiomyopathic trait in an Italian family with HCM. (Girolami et al., 2014), emphasizing the potential of NGS approach in diagnostic screening.

1.7.3.5 NGS and AC

The potential of NGS in the identification of causal mutations also in AC or overlap syndromes has been so far highlighted by different reports.
Hedberg and coll., using WES, identified a heterozygous mutation p.P419S in *DES* gene in a Swedish family affected by autosomal dominant myofibrillar myopathy with Arrhythmogenic Cardiomyopathy. The mutation was then detected by Sanger sequencing in 17 additional affected family members (Hedberg et al., 2012).
Another study describing the NGS application to extend the genetic analysis in an AC proband who was negative for mutations in five desmosomal genes revealed a heterozygous missense mutation p.T351A in *LDB3*. This result was confirmed by subsequent Sanger DNA sequencing, which also detected the mutation in other six relatives with a definite or borderline diagnosis of AC. This study highlights the usefulness of NGS to point to new causative genes in AC (Lopez-Ayala et al., 2014).
Moreover, the efficiency of NGS in detecting mutations in AC is emphasized in a very recent study. Gréen and coll. designed and validated a NGS test panel for parallel sequencing 10 genes in AC patients. All the samples were successfully sequenced and mutations found by Sanger sequencing were also found using

NGS, which showed a sensitivity varying from 99.3% to 100%, and a specificity varying from 99.9% to 100%, depending on the bioinformatics pipeline (Gréen et al., 2015).

1.7.3.6 NGS and Molecular Autopsy

The identification of the molecular basis of SD during molecular autopsy has the potential to highlight the causes of death and especially permit cascade screening, genetic conselling and clinical management of at-risk family members. NGS application at post mortem will enable a comprehensive analysis of all known SD-associated genes or even the entire exome or genome from a relative small quantity of DNA, and in a cost-effective way.

NGS-based molecular autopsies have already been successful in detecting genetic causes of SD.

Campunzano and coll. describe the genetic analysis of 104 SD-related genes using a NGS custom panel for a case of sudden infant death syndrome (SIDS) after negative autopsy. NGS identified seven variants in 6 different genes in the index case, two of them were previously described as pathogenic. Familial segregation showed that the index case's mother carried 5 of the 6 genetic variants, this last inherited from his father; the sister carried 3 of the 6 variants identified in the index case. This study underlines the utility of NGS to identify potentially genetic variants and the crucial role of familial genotyping to clarify the pathogenic role of unknown variants and to identify other genetic carriers at risk of SCD (Campuzano et al., 2014).

Another study performed by the same group described the application of NGS to identify the genetic cause of SD in a juvenile cohort with no-conclusive cause of death after comprehensive autopsy. Twenty-nine cases were analyzed, low quality DNA cases were analyzed for 7 main SD- associated genes using Sanger technology, while good quality cases were analyzed for 55 SD-associated genes using NGS technology. They identified 35 pathogenic and/or potentially pathogenic genetic variants in 12 cases (41.37%): 10 genetic variants in genes encoding cardiac ion channels were identified in 8 cases (27.58%), and 25 genetic variants in genes encoding structural cardiac proteins in 9 cases (31.03%).(Campuzano et al., 2014).

Loporcaro and coll. describe the use of WES as an efficient and cost-effective approach to incorporate molecular studies into the conventional postmortem examination. WES and gene-specific surveillance of 90 known major cardiac channelopathy/cardiomyopathy genes were performed on a sudden death victim after inconclusive autopsy. WES analysis revealed the p.R249Q mutation in the myosin heavy chain 7 gene (*MYH7*) previously associated with familial hypertrophic cardiomyopathy (HCM) and SD. Six additional family members were found to be at risk for HCM and were recommended to clinical follow-up, however no clinical signs were described. (Loporcaro et al., 2014).

Another study describes the application of NGS to estimate the frequency of pathogenic variants in the genes most frequently associated with SD. Fifteen SD cases and 29 patients with channelopathies were analyzed for 34 SD associated genes. Likely pathogenic variants were identified in three out of 15 (20 %) forensic SD cases and in 12 out of 29 (41 %) patients with channelopathies, highlighting the potential of NGS to increase the diagnostic rate significantly in the clinical setting (Hertz et al., 2014).

A recent study by Bagnall and coll. describes the use of WES to identify a causal mutation in a family with cardiac disease showing phenotype heterogeneity. This approach detected the p.A119T mutation in the alpha-actinin-2 (*ACTN2*) gene that segregated with disease, highlighting the value of WES in the cardiac genetic testing in families with mixed clinical presentations (Bagnall et al., 2014).

Another recent study reports post-mortem WES and gene-specific analysis of 117 SD susceptibility genes for 14 SD victims. Overall, 8 ultra-rare variants (7 missense, 1 in-frame insertion) in 6 genes absent in 3 publically available exome databases were identified in 7 of 14 cases (50%). Of the 7 missense alterations, 2 (p.T171M in Voltage-Dependent L-Type Ca^{2+} Channel Alpha 1 Subunit-*CACNA1C*, p.I22160T in *TTN*) were predicted damaging by 3 independent *in silico* tools. This study emphasize the utility of WES to detect rare genetic variants underlying SD, and the complexity of accurate interpretation of each variant (Narula et al., 2014).

1.7.3.7 Interpretation of the clinical significance of variants in NGS

The detection of novel genetic variants is rapidly increasing with the introduction of high throughput technologies into the clinical diagnostics, making challenging the interpretation of variants' clinical significance.

Although the lack of a comprehensive and collectively-accepted interpretation protocol, variants classification is aided by available interpretation guidelines (Kazazian et al., 2009), recently improved and updated (Richards et al., 2008). American College of Medical Genetics (ACMG) recommends a correct annotation and a standardized terminology to report nucleotide variants (den Dunnen et al., 2000) and to assess their pathogenicity carefully evaluating a string of parameters such as evolutionary conservation, population studies, functional in vitro studies, clinical presentation, family history and co-segregation.

The availability of reliable databases collecting genetic variants and information on genotype-phenotype associations help to discern between benign variants and disease-causing variants, by reporting for instance whether a specific variant had been already described in another patients. A number of established public databases focusing on inherited diseases report information on genotype-phenotype associations are available, such as the Human Gene Mutation Database (HGMD) (Stenson et al., 2013), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) (Landrum et al 2014). Several web-based, gene-centered, locus-specific databases have also been developed, such as the Leiden Open Variation Database (LOVD) (Fokkema et al., 2005), that provides a flexible, freely available tool for gene-centered collection and display of DNA variants. In particular, comprehensive information about all known AC-associated variants are collected and shared in the ARVD/C database since 2009 (van der Zwaag et al., 2009). Further, the 1000 Genomes Project and Exome Variant Server (EVS) report million of single nucleotide variants (SNVs) obtained from deep sequencing of large population cohorts (1000 Genomes Project Consortium, 2010; Exome Variant Server, 2011), and also report information about the variant frequencies in the population.

1.7.4 Assessing pathogenicity of desmosomal genetic variants

The evaluation of the DNA variants has to be extremely accurate in AC cases because disease-associated genes have high background genetic variation rate. A

recent study demonstrated that the overall yield of mutations among AC cases was 58% versus 16% in controls (Kapplinger et al., 2011) by proving that previous AC-associated pathogenic genetic variants are present in the general population. Specifically, missense mutations were hosted by 16% of controls versus a similar 21% of AC cases. This poses several challenges to interpreting the AC genetic test, as 1 in 6 healthy individuals would meet current criteria for a so-called positive AC genetic test result, even with proper qualification of these rare "mutations" with the clinically ambiguous designation as a "variant of uncertain significance." Another study provided confirmatory data by reporting that 18% (38 genetic variants) of previously reported AC pathogenic genetic variants were identified in the Exome Sequencing Project (ESP) population (1 nonsense and 37 missense) (Andreasen et al., 2013). It is noteworthy that convincing family co-segregation was reported in only 1 of these 38 pathogenic genetic variants and 3 variants had also functional characterization showing significant differences between mutant and wild-type transfected cells. Thus far, 190 pathogenic missense genetic variants have been associated with AC, all reported in Human Gene Mutation Database (HGMD) and distributed as follows: distributed as follows: 56 in *DSP* ,49 in *DSG2* ,48 in *PKP2*, and 24 in *DSC2.* From all reported missense pathogenic genetic variants associated with AC, only 62 % predicted to be deleterious by *in silico* platforms, using PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org/ index.php). Moreover, 57 % are predicted to be deleterious using Condel (CONsensus DELeteriousness score of non-synonymous single nucleotide variants) (http://bg.upf.edu/ condel/analysis), which integrates five different platforms (Polyphen-2, SIFT, MAPP, LogR Pfam E -value, and Mutation Assessor).

On the other hand, "radical" mutations (insertions or deletions, splice junction mutations, and nonsense mutations) in the Kapplinger study, were significantly more prevalent in AC cases compared with controls (50 % vs 0.5 % respectively, $p=9.8x10^{-44}$), indicating that the presence of this type of genetic variant has a high likelihood of being associated with AC pathogenicity (Kapplinger et al., 2011). Notably, "radical" mutations constituted the majority (75 of 102, 73.5%) of genetic alterations identified in mutation-positive AC cases.

Thus far, 155 “radical” mutations have been associated with AC, all reported in HGMD, and distributed as: 75 in *PKP2*, 52 in *DSP*, 16 in *DSG2*, and 12 in *DSC2*. Despite of being reported in the HGMD database, not all “radical” genetic variants denote a pathogenic role in clinical practice, especially if no family segregation has been reported. A recent study in Japanese population with 35 AC cases showed that carriers with *PKP2* premature stop codon developed the disease at a significantly younger age than other mutations carriers (Ohno et al., 2013), while another more recent study performed in European population suggests later onset in carriers with *PKP2* premature stop codon (Alcalde et al., 2014).

Hence, the most particularly weakness to misclassification are mutations identified in large cohorts of individual unrelated probands (without data on familial segregation). For this purpose guidelines in genetics (Kazazian et al., 2009; Richards et al., 2008), highly recommend to perform accurate *in silico* analysis, *in vitro* assays, and, as key point, co-segregation studies in families to assess the pathogenicity on AC.

1.7.5 Assessing pathogenicity of extra desmosomal genetic variants

In the last 4 years, extra desmosomal genes have been also associated with AC however all together have a lower incidence of <5% of AC patients (Alcalde et al., 2014). Extra desmosomal genes associated with AC are a heterogeneous group of genes with different functions and the underlying mechanism in AC needs still to be clarified. It is interesting that most of extra desmosomal genes encode cytoskeleton-associated proteins such as, *DES*, *LMNA*, *CTNNA3*, and *TTN*, previously described in other cardiomyopathies, i.e. DCM (Garcia-Pavia et al., 2013). Therefore, recent studies support the concept of “Arrhythmogenic Cardiomyopathy” as an entity encompassing DCM an AC, given by the evidence that extra desmosomal genes are commonly associated with DCM and AC, whereas desmosomal genes had not yet been associated with any other cardiomyopathy (Saffitz et al., 2011). Contradictory studies have been published since then regarding the incidence of desmosomal mutations in other cardiomyopathies (Posch et al., 2008; Elliott et al., 2010; Zhang et al. 2011; Garcia-Pavia et al., 2011; Garcia-Pavia et al., 2013; Lopes et al., 2013). Consequently, assessing pathogenicity in extra desmosomal genes without

cascade family screening, clinical data and/or is in vitro assays is problematic, since overlapping symptoms are common among inherited cardiomyopathies.

1.7.6 Clinical implications of genetics

Genetic screening is gaining ground in the identification of patients and family members at increased risk of AC. Identification of a misclassified genetic variant in cardiomyopathy patients might lead to erroneous risk stratification, misdiagnosis of family members and this could have potentially devastating clinical consequences. It is therefore important that variants being reported as causative of cardiomyopathies are truly disease causing in order to avoid overrepresentation of AC-associated genetic variants.

Genetic information in AC is helping to provide a potential cause of the pathology, but clinical findings still remains the main basis for diagnosis and treatment of patients. Thus, the identification of a potential pathogenic genetic variant cannot override clinical judgment regarding AC diagnosis. In addition, lack of pathogenic genetic variant in the setting of convincing clinical evidence neither should call the diagnosis into question nor rule out the disease (Kapplinger et al., 2011). All AC diagnosed patients or suspicious of suffering AC should be genetically analyzed, and for all genes associated with AC due to reported digenic/compound cases (Nakajima et al., 2012; Barahona-Dussault et al., 2010). Identification of the pathogenic genetic variant may help to clarify the cause of the disease, and posterior familial genetic analysis may identify genetic carriers that could remain asymptomatic but at risk of SD. The lack of 'gold standard' tool in clinical diagnosis makes genetic testing a complementary tool in diagnosis, even though lacks clear pathogenic interpretation. Determining the pathogenicity of a genetic variant is the main genetic challenge in current clinical practice. Establishing solid potential therapeutic and prognostic implications of gene variants associated with AC is not available due to lack of clinical-genetic correlations. Despite being accepted that "radical" variants should be considered more dangerous than missense variants due to truncation of proteins associated with AC (Kapplinger et al., 2011), genetic results should be interpreted with extreme caution and by multidisciplinary teams including at least cardiologists and geneticists.

AIM OF THE STUDY

AC is a clinically and genetically heterogeneous disorder associated with arrhythmias and sudden death, in particular in young adults and athletes. Mutations in the desmosomal genes *DSP*, *PKP2*, *DSG2*, *DSC2*, *JUP* have been associated with the disease in approximately 50% of total cases. Since SCD may occur as the first symptom of the disease, an early genetic diagnosis is highly relevant for the identification of affected subjects and especially of genetically affected relatives before the manifestation of the clinical phenotype.

The overall aim of the present investigation was the development of a fast, cost-effective and comprehensive screening strategy for the molecular diagnosis of AC.

Specific objectives of the present study include:

1. Assessment of the spectrum and prevalence of desmosomal mutations in a large cohort of unrelated index cases,
2. Assessment of the presence and prevalence of large genomic structural rearrangements (CNVs) in a large cohort of unrelated index cases,
3. Analysis of potential disease causing mutations pathogenicity,
4. Evaluation of mutation type role in the development of AC,
5. Comprehensive genotype-phenotype study in order to stratify the risk of developing the disease in asymptomatic mutation carriers,
6. Investigation of Next Generation Sequencing (NGS) potential and limitations,
7. Application of NGS for molecular diagnosis in AC patients, extending the spectrum to genes not related with the disease.

2. METHODS

2.1 Cohort and Clinical Examination

The study involved a total of 99 subjects with a clinical or post mortem diagnosis of AC referred to the Referential Clinical Genetic Centre of Arrhythmic Cardiomyopathies in Padua.

Twenty-six probands were sudden death victims enrolled from the Registry of Cardio-Cerebro-Vascular Pathology, Veneto Region, Venice, Italy. For the rest 73 probands, Arrhythmogenic cardiomyopathy clinical diagnosis was based on established major and minor criteria revised by the European Society of Cardiology and the Federation of Cardiology Task Force (Marcus et al., 2010). Clinical evaluation included physical examination, 12-lead ECG, signal-averaged ECG, cardiac magnetic resonance imaging, 24-hour ambulatory ECG monitoring. Exercise tests and electrophysiological studies were performed when considered necessary. AC diagnosis was considered definite when 2 major, or 1 major and 2 minor criteria, or 4 minor criteria from different categories, were fulfilled; borderline when 1 major and 1 minor or 3 minor criteria from different categories were fulfilled; possible: when 1 major or 2 minor criteria from different categories were fulfilled. When possible, the first-degree and second-degree relatives from families where a disease-causing variant was detected were called for prospective evaluation and genetic screening. Written informed consent was obtained from all study participants before blood sampling and genetic screening.

Variants frequency, type, and localization were compared to a group of 500 unrelated healthy ethnically matched volunteers (1000 alleles).

2.2 DNA Extraction

Genomic DNA (gDNA) isolation was obtained from different matrices including whole blood, frozen myocardial tissue, formalin-fixed and paraffin-embedded (FFPE) tissue through specific onboard protocols of MagNA Pure Compact System (Roche Applied Science, Mannheim, Germany), an automated benchtop device that uses the magnetic bead technology for nucleic acids isolation. The MagNA Pure Compact performs all nucleic acids isolation steps preventing cross-

contamination between samples. It automatically pipets into the sample a pre-aliquoted volume of lysis buffer containing chaotropic salts and proteinase K, nucleic acids are bound to the surface of the magnetic glass beads and after several washing steps the purified DNA is eluted in DNasi-free H_20.

2.2.1 DNA extraction from blood

gDNA extraction was carried out from 200 µl of whole blood by using Nucleic acid isolation kit I (Roche Diagnostics GmbH, Mannheim, Germany) on the MagNA Pure Compact System by setting the ``DNA_Blood_100_400`` protocol on the machine software. The typical DNA yield from 200 µl of human whole blood using the ``DNA_Blood_100_400`` protocol and an elution volume of 100 µl is 6 µg.

2.2.2 DNA extraction from frozen and FFPE tissue

gDNA extraction was performed from approximately 5 mg of snap frozen tissue or from 2 FFPE section with 8µm thickness by using Nucleic acid isolation kit I (Roche Applied Science, Mannheim, Germany) with the ‘’DNA-culture Cells’’ protocol. Paraffin removal from FFPE tissue, commonly achieved by consecutive washes of xylene and ethanol, revealed to be unnecessary when using this protocol. An initial pre-treatment step was performed to enhance cell lysis. Briefly 200 µl of MagNA Pure Compact DNA Lysis buffer (Roche Applied Science, Mannheim, Germany) was added in the tissue sample and a 2 min disruption on a Tissue Lyzer System (Qiagen Venlo, Limburg, The Netherlands) was carried out at 25 Hz. The homogenized sample was then incubated with 20 µl of Proteinase K (20 mg/mL; Roche Diagnostics GmbH, Mannheim, Germany) for 10 min at 56 °C and transferred on the MagNA Pure Compact workstation. The typical DNA yield using the “DNA-culture Cells” protocol from 5 mg of snap frozen tissue and an elution volume of 50 µl is 14 µg, and the typical DNA yield from 2 FFPE section (8µm thickness) is 4 µg.

2.2.3 DNA extraction from saliva

The 500 µl saliva sample was mixed with 200 µl of MagNA Pure Compact DNA Lysis buffer and, after vigorous mixing by vortex, incubated at 56 °C for 10 min. The lysate was then transferred on the MagNA Pure Compact workstation and

extracted with the Nucleic acid isolation kit I (Roche Diagnostics GmbH, Mannheim, Germany) and the "DNA_blood_external_Lysis" protocol. The typical DNA yield from 500 µl of saliva is normally 6 µg, comparable with the yield from whole blood extraction.

2.3 RNA Extraction

RNA extraction was obtained from whole blood or frozen myocardial tissue, when available, by using the MagNA Pure Compact System. RNA was eluted in 50 µL of RNasi free H_20 and stored at -80 °C.

2.3.1 RNA extraction from blood

RNA was isolated from 200 µl of whole blood by using RNA isolation kit I (Roche Diagnostics GmbH, Mannheim, Germany) and 20 µl of DNasi (Roche Diagnostics GmbH, Mannheim, Germany) on the MagNA Pure Compact System according to the "RNA_Blood'' protocol. The mean RNA yield from 200 µl of human whole blood is 1µg.

2.3.2 RNA extraction from frozen tissue

RNA was isolated from approximately 5 mg of snap frozen tissue on the MagNA Pure Compact System employing RNA isolation kit I and 20 µl of DNasi (Roche Diagnostics GmbH, Mannheim, Germany) following the "RNA_Cells" protocol. The pre-treatment step was performed as previously described after adding the samples with 200 µl of MagNA Pure Compact RNA Lysis buffer (Roche Diagnostics GmbH, Mannheim, Germany). The typical RNA yield using the "RNA_Cells" protocol from 5 mg of snap frozen tissue and an elution volume of 50 µl is 15 µg.

2.4 Nucleic acids Quantification

Nucleid acid isolation is followed by a quantity and quality (purity, integrity) check prior before downstream preparation steps. DNA quantification allows standardization of the PCR input material and is crucial in NGS DNA library preparation since influence run quality and efficiency.

2.4.1 Spectrophotometric method

Nucleic acids quantification is performed by spectrophotometers by measuring absorption of ultraviolet light at the wavelength of 260 nanometers and applying the Lambert-Beer law that correlates absorbance, molar extinction co-efficient and nucleic acids concentration. Absorbance measurements were carried on a Nanovue spectofotometer (GE Healthcare Life Sciences, UK) by directly pipetting onto the pedestals 2 µl of the DNA or RNA sample dissolved in DNasi or RNasi-free H_20, after an initial blank measurement. Nanovue automatically calculates the nucleic acids concentration by applying specific extinction coefficients (50 for dsDNA, 40 for RNA).

The ratio $A_{260/280}$ and $A_{260/230}$ are used as indicators of the sample purity. The $A_{260/280}$ ratio of the nucleic acid sample is generally used as indicator of protein contamination. Infact the 280 nm is the absorbance wavelength of aromatic amino acid side chains and phenol groups. Pure DNA should present a ratio 260/280 between 1.8 and 2, pure RNA should present a ratio 260/280 between 2 and 2.2. The $A_{260/230}$ ratio is generally used as indication of organic contaminants. The 230 nm is the absorbance wavelength of many organic compounds (i.e. phenol, TRIzol, and chaotropic salts present in the most common lysis buffers. In pure samples the ratio $A_{260/230}$ should be between 2 and 2.2.

2.4.2 Qubit Fluorometer

The quantity and quality of DNA was further assessed by using Qubit 2.0 Fluorometer (Invitrogen, Life technologies, Carlsbad, CA, USA) in combination with dsDNA BR Assay kit (Invitrogen, Life technologies, Carlsbad, CA, USA) for DNA measurements and RNA Assay kit (Invitrogen, Life technologies, Carlsbad, CA, USA) for RNA measurements according to the manufacturer's instructions. Briefly, 1 µl of the sample is mixed with 200 µl of working solution containing specific fluorescent dyes, composed by 199 µl of dsDNA BR buffer or RNA buffer and 1 µl of dsDNA BR reagent or RNA reagent respectively. After an incubation of 2 min the measurement is performed on the Qubit 2.0 by setting the DNA or RNA protocol. The Qubit 2.0 is a benchtop fluorometer that uses fluorescent dyes that specifically bind to either DNA or RNA therefore it is able to selective quantify either DNA or RNA. Moreover, as the DNA dye exclusively

binds to double strand DNA (dsDNA), it further allows the selective quantification of dsDNA that will be exclusively used for downstream applications, minimizing the effect of contaminant RNA and ssDNA.

2.4.3 Agilent Bioanalyzer

The Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) is a microfluidics-based platform that can perform quantification and quality control of nucleic acids by an electrophoretic separation of samples on micro channels containing fluorescent dyes. One µl of DNA or RNA samples is sufficient for analysis on 2100 Bioanalyzer with the DNA 1000 chip (Agilent, Santa Clara, Carlsbad, USA) and the RNA Nano Chip (Agilent, Santa Clara, Carlsbad, USA). All the assays were performed according to the manufacturer's guide. Briefly, DNA and RNA chips were prepared by adding 9 µl of gel matrix to the assigned well under pressure, and 9 µl of gel-dye mix to the next 2 wells. After adding 5 µl of marker to each well, 1 µl of DNA or RNA ladder and 1 µl of each sample were added to the separate wells. Chips were then vortexed for 1 min at 2400 rpm on a IKA vortex mixer and were run on the Agilent 2100 Bioanalyzer. The run data were analyzed by Agilent 2100 expert software version B.02.08.SI648(SR2).

The DNA 1000 Kit allows the separation with high resolution, sizing and quantification of dsDNA fragments ranging from 25 to 1000 bp. The Agilent RNA Nano kit evaluates the ratio between the 18S and the 28S ribosomal subunits, and the presence of degraded small RNA fragments, in order to calculate a RNA integrity value (RIN) that expresses an estimation of the integrity of the RNA sample. RIN values range from 10 (intact) to 1 (totally degraded) (Figure 2.1)

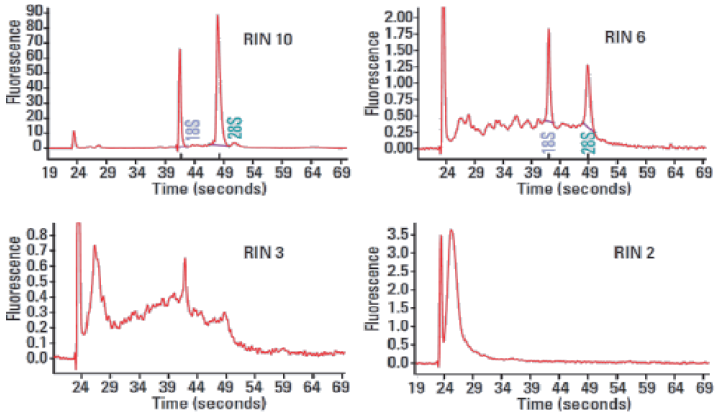

Figure 2.1: Samples range from intact (RIN10) to degraded (RIN2) (from www.chem.agilent.com)

2.4.4 Agarose gel 1%

The integrity of isolated gDNA is checked by on a 1% agarose gel. The presence of a strong single band indicate high quality and integer dsDNA, on the contrary a smear indicates the presence of degraded DNA.

2.5 Genetic Screening

The entire cohort of 99 AC probands was screened for pathogenic variants in five desmosomal encoding genes (*DSG2, DSC2, DSP, PKP2, JUP*), whereas only 46 AC subjects were subsequently screened for causative variants in 3 extra-desmosomal genes (*CTNNA3, DES, PLN*).

Each exon and exon-intron boundaries of the 8 genes was amplified by polymerase chain reaction (PCR) and analyzed by denaturing high-performance liquid chromatography (DHPLC) and Sanger sequencing.

2.5.1 Polymerase chain reaction (PCR)

PCR is a method developed by Mullis in the 80s (Mullis et al., 1990) that allows the exponential amplification of specific target DNA regions (Figure 2.2). A PCR reaction requires: template DNA, deoxynucleotide triphosphates (dNTPs), oligonucleotide primers flanking the target DNA sequence, DNA polymerase enzyme, reaction buffer and magnesium. PCR is composed by different consecutive reactions:

1 Template denaturation by heating DNA at 95 °C,

2 Annealing temperature (Ta) usually ranges from 55 to 65 °C; oligonucleotide primers may align on the target region of the single strand DNA.

3 Elongation at 72 °C, the DNA polymerase synthesizes new strands of DNA starting from the 3' of annealed primers. The newly synthesized DNA of the first cycle will be the template of the next and so on, reaching a million-fold increase of the DNA quantity at the end of the reaction.

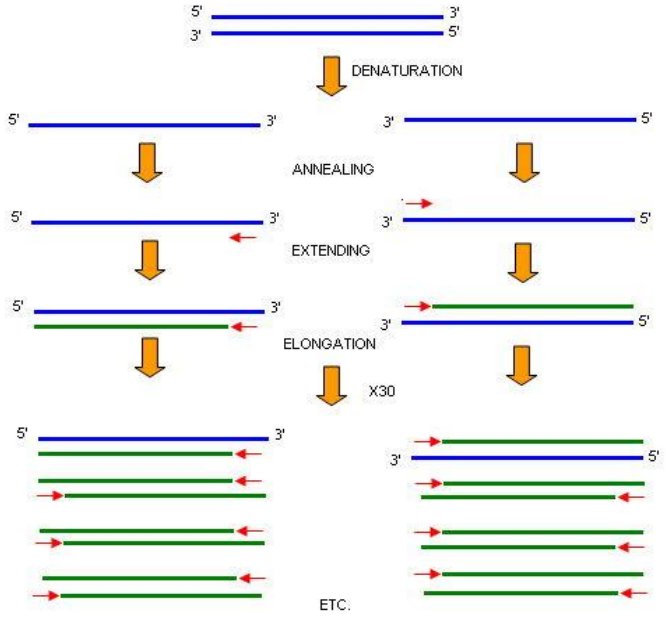

Figure 2.2: Schematic representation of the PCR reaction

2.5.1.1 Standard PCR

Amplification reactions were carried out in a final volume of 12.5 µl containing 50 ng of template (final concentration 4ng/µl), 1X of PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of dNTPs, 0.8 µM of forward and reverse primer, and 0.32 U of Taq DNA polymerase (Life Technologies, Carlsbad, CA, USA). For some amplicons, amplification was enhanced by the use of 1µl of Dimethyl sulfoxide (DMSO) in the reaction mix. DMSO destabilizes the double helix structure by hydrogen binding to the major and minor grooves of DNA and reduces secondary structure formation in the DNA template, thus facilitating primers annealing.

Amplification reactions were performed on Mastercycler Pro (Eppendorf, Hamburg, Germany) at specific Ta for every PCR product, optimized based on the GC content of the sequence and melting temperatures (Tm) of primers. Briefly, each sample was denatured at 95 °C for 10 min to allow the activation of the hot start DNA polymerase, and exposed to 40 amplification cycles of denaturation for 30 sec at 95 °C, annealing for 30 sec at a range of Ta comprised between 55 °C and 65 °C and extension at 72 °C for 30 sec; followed by a final extension step of 7 min at 72 °C to enhance the amplicon elongation.

The specificity and the amplification yield of some PCR products was enhanced by the use of touch-down PCR (TD-PCR).

TD-PCR protocols set annealing temperatures above the anticipated melting temperature of a primer/template pair and then decrease the annealing temperature of subsequent cycles stepwise. The initial annealing temperatures of the first few cycles is higher than the expected Tm of primers, and then for the subsequent cycles the temperature progressively decreases to lower Ta to allow the correct hybridization of primer to the template. Thermal cycling conditions used for TD-PCR are as follows: an initial incubation at 95 °C for 10 min; followed by 10 cycles at 95 °C for 30 sec, annealing temperature stepdowns every cycle of 1 °C (from 70 °C to 60 °C); extension at 72 °C for 30 sec. The successive 30 cycles are performed at a Ta of 60 °C with denaturation and extension steps as above; followed by the final extension at 72 °C for 7 min.

PCR products (ranging from 150-700 bp) were analyzed on 2% agarose gel.

2.5.1.2 GC rich PCR

GC rich amplicons were amplified by using AmpliTaq Gold 360 Master Mix kit (Life Technologies, Carlsbad, CA, USA) containing all premixed PCR components according to the manual. Amplification reactions were performed in a final volume of 12.5 µl by adding 7 µl of mix, 2.5 µl of GC solution provided, 2 µl of forward and reverse primer to 50 ng of template DNA.

2.5.1.3 Primers design

The oligonucletide primers used for both amplification for subsequent Sanger sequencing were designed by PRIMER 3 software (http://primer3.ut.ee/) (Rozen and Skaletsky, 2000), on the following reference sequences retrieved from the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/):

DSG2 gene has 15 exons and spans on ~51 kb of chromosome 18q12.1, NM_001943 is the unique DSG2 transcript variant.

DSC2 gene contains 17 exons and spans on ~36 kb of chromosome 18q12.1, NM_024422 represents the transcript variant DSC2a.

DSP gene contains 24 exons and spanson ~45 kb of chromosome 6p24.3, NM_004415 represents the longest transcript variant type 1, DSP.

PKP2 genecontains 14 exons and spans on ~106 kb of chromosome 12p11.21, NM_004572 represents the predominant and longest isoform in the heart.

JUP gene contains 14 exons and spans on ~32 kb of chromosome 17q21.2, NM_002230 represents transcript variant 1.

CTNNA3 gene contains 18 exons and spans on ~ 10kb of chromosome 10q21.3, NM_013266 represents the transcript variant 1.

PLN gene has 2 exons and spans on ~ 12kb of chromosome 6q22.31, NM_002667 is the unique PLN transcript variant.

DES gene contains 9 exons and spans on ~ 8kb of chromosome 2q35, NM_001927 is the unique DES transcript variant.

Primers pairs (length 18-22 bases) were designed to maintain a Tm comprised between 58 and 62 °C, a maximum Tm difference of 1 °C, self complementarity, that is the tendency of a primer to bind to itself, and pair complementarity, that is the tendency of the left primer to bind to the right primer comprised between 3 and 0 and when possible a GC content below 60%.

The designed primers were analyzed using different bioinformatics tools in order to check their reaction specificity and accuracy. They were aligned by BLAST (http://blast.be-md.ncbi.nlm.nih.gov) and BLAT (http://genome-euro.ucsc.edu/cgi-bin/hgBlat) on the genome reference sequence (GRCh37/hg19) in order to verify their specific annealing and avoid the presence of SNPs in predicted PCR primer binding sites. To confirm the amplicon length and Tm, they were also checked by *in silico* PCR tool on UCSC website (http://genome-euro.ucsc.edu/index.html), and then produced by Invitrogen (Life Technologies, Carlsbad, CA, USA).

2.5.2 Agarose gel

For standard 2% agarose gel electrophoresis, 2 g of agarose were added to 100 ml of 1X Tris-Acetate-EDTA (TAE) buffer. The solution was heated in a microwave to dissolve agarose, gel was added with 5 µl of Nancy-520 DNA Gel Stain (Sigma-Aldrich, Saint Louis, MO, USA), cast in a sealed tray and a proper comb was inserted. Aliquots of 5 µl of PCR product and 3 µl bromophenol blue loading dye were mixed and loaded into each gel well. In addition, 2 µl of DNA Marker VIII (Roche Applied Science, Manheim, Germany) were loaded in order to determine the fragment sizes. Electrophoretic run was performed at 100 V in 1X

TAE buffer. Visualization was achieved on a LAS mini 4000 (Fujifilm, Tokyo, Japan).

2.5.3 Denaturing High Performance Liquid Chromatography (DHPLC)

DHPLC is a pre analytical technology that can detect the presence of single nucleotide variants and small insertions and deletions (Oefner and Underhill et al., 1995). It separates heteroduplex molecules from homoduplex molecules by ion-pair reverse-phase liquid chromatography on a column containing nonporous alkylated polystyrene-divinylbenzene particles. Heteroduplexes are made by double stranded amplification products composed by two almost complementary strands, containing a mismatch at the variant level, derived from two heterozygous alleles. The four configurations possible for heterozygous DNA are shown in figure 2.3. The analysis is performed under partial denaturation temperature, the slow DNA renaturation allows the formation of homoduplexes and heteroduplexes. The samples are run into the buffer flow that contains triethylammonium acetate (TEAA) and acetonitrile. The positively charged portion of TEAA interacts with the negatively charged phosphate group of DNA, that binds to the hydrophobic chromatographic column. With an increasing linear acetonitrile gradient injected into the column, the DNA binding capacity to the TEAA ions decreases and the DNA fragments are released. Heteroduplexes, because of the mismatch, have a lower column binding affinity and thus a reduced retention time with respect to the homoduplexes. Eluted samples pass through the UV detector, that registers absorbance at 260 nm over time. In absence of DNA variants all the homoduplex molecules will show the same retention time and will elute as a single peak, on the contrary if a DNA variant is present, the chromatogram will show two or four peaks.

DHPLC was used to investigate the presence of variants in the 8 target genes, the analysis was performed using a Transgenomic WAVE System (Transegenomic, USA) with a DNASep HT Cartridge column.

After PCR reactions, amplified fragments were denatured at 96 °C for 5 min, and then slowly cooled at room temperature to allow the formation of the heteroduplexes. Exons longer than 500 bp were splitted in more amplicons to

allow DHPLC analysis. The optimal temperature for fragments analysis was calculated using Navigator Software (Transgenomic, USA). Runs were carried out at flow rate of 0.9 ml/min with the mutation detection application.
Samples with abnormal DHPLC profiles were purified and sequenced.

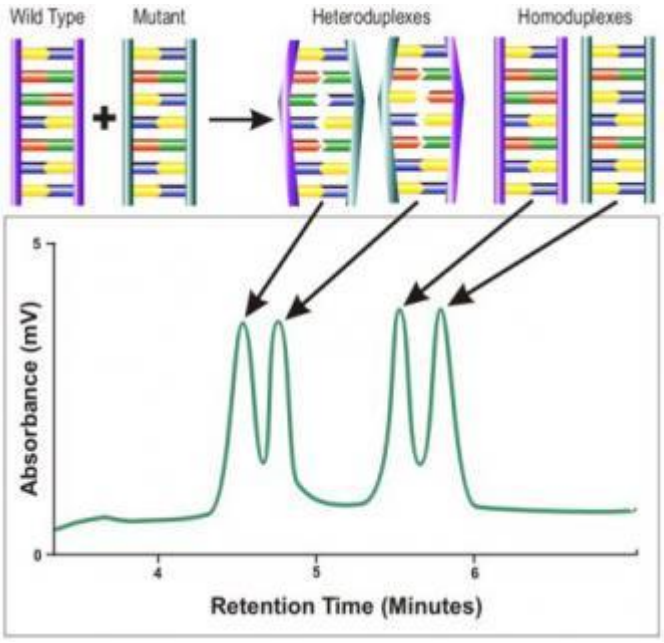

Figure 2.3: Differences between homoduplex and heteroduplex (modified from http://www.transgenomic.com)

2.5.4 Purification of PCR products

Before sequencing, PCR products were purified by a treatment with Exonuclease 1 (EXO) and Shrimp Alkaline Phosphatase (SAP) enzymes, to eliminate residues of primers and dNTPs from the previous PCR reaction, as they could interfere with the subsequent sequence reaction.
The purification reaction was carried out with "Illustra ExoProStar 1-Step" kit (GE Healthcare Life Sciences, UK) by mixing 5 µl of PCR product and 2 µl of reaction mix, containing SAP enzyme that dephosphorylates dNTPs and EXO that hydrolyzes residual primers and aspecific oligonucleotides that could have been amplified during the PCR. The reaction conditions were: an incubation at 37 °C for 30 min followed by 80 °C for 15 min to inactivate the enzymes.

2.5.5 Direct sequencing

The DNA sequencing method developed by Sanger in the 1970s (Sanger et al., 1977) is based on the DNA chain-termination by chemically altered bases called di-deoxynucleotidetriphosphates (ddNTPs) carrying four different fluorophores. The method requires the presence of normal deoxynucleoside triphosphates (dNTPs), and dideoxynucleotide triphosphates (ddNTPs) lacking a 3'-hydroxy group that, when incorporated into a newly synthesized DNA fragment, terminate the DNA strand elongation at specific bases (A, C, T, or G). This process produces DNA fragments with different sizes that can be separated by capillary electrophoresis and detected with laser-induced fluorescence (Figure 2.4).

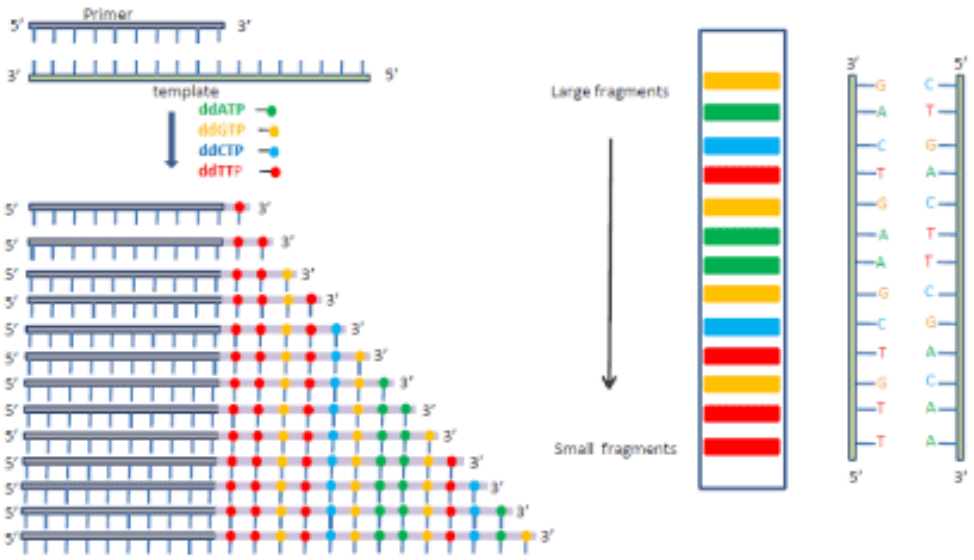

Figure 2.4: Schematic representation of the Sanger sequencing technology

Sequencing reactions were performed with BigDye terminator v3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA, USA) following the manufacturer'instructions. 1 µl of each purified PCR products was mixed with 2 µl of sequencing buffer at a final concentration 1X, 1 µl of BigDye Terminators solution (containing dNTPs, fluorophore-conjugated ddNTPs, and polymerase), 10 pm of forward or reverse primer in a total volume of 10 µl. Sequencing reaction was performed as follows: 96 °C for 1 min, 30 cycles of 96 °C for 10 sec, 55 °C for 5 sec, 60 °C for 2 min 30 sec. Fluorescent-labeled sequences were then cleaned-up to eliminate unincorporated big dye terminators, dNTPs, and salts

using by ethanol/EDTA precipitation. Briefly, 50 µl of EtOH 96% and 2 µl of EDTA 125 mM were added to each sample, they were incubated at room temperature for 20 min and centrifuged at maximum speed for 15 min at 20 °C. samples were washed with 150 µl of EtOH 96% and resuspended in 10 µl of H_20. 5 µl of sequence reaction were mixed with 15 µl of Hi-Di Formamide (Life Technologies, Carlsbad, CA, USA), denatured for 5 min at 95 °C and submitted to sequencing analysis on an ABIPRISM 310 Genetic Analyser (Life Technologies, Carlsbad, CA, USA).

2.5.5.1 Sequences Analysis

CHROMAS software (Technelysium) was used to view Sanger sequencing electropherograms, while comparison with genomic sequences (GRCh37, hg19) obtained from UCSC Genome Browser (http://genome.ucsc.edu/) was done by SeqMan II (DNASTAR, Madison, WI, USA). AlaMut software (Interactive Biosoftware, Rouen, France) was used to evaluate nucleotide variants by integrating information from different sources (*in silico* prediction algoritms, conservation, literature infos etc).

Human Genome Variation Society (http://www.hgvs.org/) recommendations (den Dunnen and Antonarakis, 2000) was used for nucleotide variant nomenclature.

Whenever a putative pathogenic variant was detected, the presence of the variant was confirmed by performing a new DNA extraction, PCR and sequence reaction.

For every novel nucleotide variant detected, 500 ethnically matched unrelated control subjects were screened for the presence of the variant in the healthy population. Screening of control subjects was performed by DHPLC with Rapid protocol and Sanger sequencing whenever an abnormal DHPLC pattern was observed.

Sequence nucleotide variants found in dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) were considered known polymorphisms and their frequency in the general population was determined by the data available from the Exome Variant Server (NHLBI GO Exome Sequencing Project, http://evs.gs.washington.edu/EVS/) and in the 2188 control chromosomes available from the 1000 Genomes Project. Intronic variants located less than 50 bp from the exon boundaries were in silico analyzed at least by three different

algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer), to predict a possible effect on the splicing sites of the transcript.

2.5.6 Filtering and prioritization of variants identified by Sanger sequencing

In order to identify potential causal variants, nucleotide variants were prioritized by 3 different filters, based on

a. the frequency of variants in the population,
b. the conservation of the amino acid residue among species,
c. and the pathogenetic predition of the variants obtained from *in silico* analysis.

2.5.6.1 Variant Filtering based on Frequency

The first filter applied to exclude common variants in the general population is based on the frequency of the nucleotide change in our group of healthy controls, the current default global population according to the data available in the freeware variant databases developed as results of the 1000 Genomes Project and the NHLBI Exome Sequencing Project (ESP), and literature information.

The 1000 Genomes dataset (phase 1, released in the May 2011, http://www.1000genomes.org/) provide genotype data from 1094 worldwide individuals. The Exome Variant Server (http://evs.gs.washington.edu/EVS/) contains exome sequence data on a total of 6503 individuals: 2203 African-Americans individuals and 4300 European-Americans.

These databases provide the minor allele frequency (MAF) of each variant in the respective populations. Referring to those frequency data we classified variants as: pathogenic when they are detected with a MAF < 0.02%; rare variants when MAF is comprised between 0.02% and 0.1%; and polymorphisms when they are present with a frequency > 0.1%.

We arbitrary set up frequency cut-off frequency to 0.02% to classify a variant as 'mutation' given that the prevalence of AC in the Veneto Region is estimated as 1:5000 (0.0002, 0.02%). Based on the frequency of AC in the general population, the estimated number of individuals in the ESP data (European-Americans section) that can be expected to be affected by AC is 1.

Coding and splicing variants that did not fulfill this criterion, were excluded to be potential pathogenic and were considered to be rare variants if present in

population and in control alleles with MAF comprised between 0.02% and 0.1% , or as polymorphisms when MAF was > 0.1%.

2.5.6.2 Variant Filtering based on the conservation of nucleotides and amino acids among species

The second filter is based on the phylogenetic conservation of the substituted amino acid. The relationship between the level of evolutionary conservation of an amino acid and its functional importance is well know, so variants occurred at the level of conserved nucleotide and/or amino acids are predicted to be less tolerated. The analysis was performed using AlaMut that integrates both nucleotide and amino acid evolutionary conservation information, with PhastCons and phyloP scores and ClustalW alignments. The PhastCons and PhyloP scores are conservation scores derived from the multiple alignment of different vertebrate species, ClustalW is a tool that performs multiple sequence alignment of DNA or proteins. Amino acids conserved among mammals and almost all lower animals were classified as 'Highly conserved`, amino acids conserved among mammals and a few lower animal were classified 'moderately conserved', amino acids conserved among almost all mammals and no lower animals were classified as 'weakly conserved', and amino acids not conserved among mammals or other lowere species were classified as 'not conserved'.

Nonsense variants, frameshift variants and variants located on the splice site (±1, ±2) which respectively introduce in premature stop codon, distrupt the transcript's reading frame, interfere with the correct splicing processes were considered potentially pathogenic.

2.5.6.3 *In silico* analysis

The third filter is a combination of statistical algorithms which aim to evaluate a string of variant-specific features, as proposed by van Spaendonck-Zwarts and coll. (van Spaendonck-Zwarts et al., 2013).

In silico analysis of missense variants is taking into account the outcome of different prediction tools:

- Polyphen-2 (http://genetics.bwh.harvard.edu/pph/). Polyphen is a tool that predicts the possible impact of an amino acid substitution on the structure and function of a protein (Adzhubei et al., 2013). Three outcomes are possible:

probably damaging (the variant is expected with high confidence to affect protein structure), possibly damaging (the variant supposed to affect protein structure), benign (the variant is not expected to have any effect on protein structure).

- Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/www/SIFT_enst_submit.html). SIFT is a tool that predicts the possible impact of an amino acid substitution on the protein function. (Ng and Henikoff 2001). It classifies amino acids substitutions as tolerated or deleterious.
- Grantham Score. The Grantham score expresses the difference in the physicochemical properties of the amino acids (side chain atomic composition, polarity, molecular volume) (Grantham et al., 1974). Grantham score is comprised between 0 and 215. A high score indicates a great difference in chemical properties between the two amino acids and thus a greater impact on protein structure.
- Align-Grantham Variation with Grantham Deviation (A-GVGD) (http://agvgd.iarc.fr/). It combines the biophysical characteristics of amino acids and protein multiple sequence alignments to predict where missense substitutions fall in a spectrum from enriched deleterious to enriched neutral. There are 7 possible outcomes: C0, C15, C25, C35, C45, C55, C65. A higher score indicates likely deleterious substitutions.
- BLOSUM (BLOcks SUbstitution Matrix) 62. It is an identity scoring matrix based on local alignments of protein sequences. (Henikoff and Henikoff, 1992). Scores for each position are derived from observations of the frequencies of substitutions in blocks of local alignments in related proteins. In the BLOSUM62 matrix, the alignment from which scores were derived was created using sequences sharing no more than 62% identity. The Blosum62 substitution matrix contains scores for all possible exchanges of one amino acid with another, where the lowest possible score (-4) indicates a low probability of substitution, and the highest score (11) indicates a high probability of substitution in the alignments. A high score indicates substitutions affecting conserved amino acids and thus likely deleterious.

- Conservation among species was evaluated using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). It was used to align amino acid sequences among different species in order to evaluate the conservation of the changed amino acids.
- Conservation between isoforms was evaluated using ClustalW. It was used to align amino acid sequences of different isoforms in order to evaluate the conservation of the changed amino acids.
- Frequency in control population is based on the minor allele frequency of the variant given by the Exome variant server and in a number of controls.
- In silico evaluation of splicing is based on at least three different Splice Site prediction softwares (SpliceSiteFinder, MaxEntScan and GeneSplicer) using AlaMut.
- Family information was considered when available
- Evidences from functional analysis were considered when available from the literature reporting in vitro studies or animal models.

When a feature was not available, it was not calculated and considered for the final classification.

In silico analysis of nonsense and frame-shift variants do not consider the first 6 steps and are based on general features of the variant, predictions about functional effects, splice predictions, and frequency in a control population. Family information about co-segregation, functional analysis and phenotypic features were considered when available.

Variants are classified as follows based on a specific score:

- (putative) pathogenic: with proven co-segregation in the family .
- % score ≥ 70%: Variant of unknown clinical significance (VUS) likely to be pathogenic (VUS3)
- 45% ≤ % score < 70%: VUS of uncertain significance (VUS2)
- 25% ≤ % score < 45%: VUS unlikely to be pathogenic (VUS1)
- % score < 25%: Not pathogenic (VUS0)

2.5.6.4 Final assessment of the variants' pathogenicity

The final pathogenicity of variants was assessed further integrating information from a detailed literature review about each variant and family cosegregation analysis.

A variant was considered disease-causing mutation when it fulfilled at least one of the following criteria:

- Previously associated with AC
- Not detected in a large healthy ethnically matched control population
- Cosegregated in the family with the disease phenotype.

2.6 Multiplex Ligation-dependent Probe Amplification (MLPA)

Multiplex ligation-dependent probe amplification (MLPA) is a method based on the multiplex amplification of specific probes developed for the relative quantification of nucleic acid sequences (Schouten et al., 2002) that allows the detection of copy number variants of specific target regions.

MLPA analysis were performed in the Cardiovascular Genetics Laboratory (University of Padua), after a preliminary study on 8 samples carried out in the Cardiogenetics Laboratory (UMCG, University of Groningen).

To investigate for large deletions/duplications within desmosomal genes, MLPA was performed by using "SALSA MLPA P168 ARVC-PKP2 probemix kit" (MRC-Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions. The probemix contains oligonucleotide probes for every exon of *PKP2* (NM_004572.3) and its promoter, 6 probes for exons 1, 5, 7, 9, 21, 24 of *DSP* (NM_004415.2), 3 probes for exons 2, 9, 12 of *JUP* (NM_002230.2), 3 probes for exons 1, 7, 17 of *DSC2* (NM_004949.3), 3 probes for exons 1, 6, 15 of *DSG2* (.NM_001943.3). It also contains 3 probes for exons 1, 6, 7 of *TGFβ3* (NM_003239.2) and 2 probes for exons 3 and 97 of *RYR2* (NM_001035.2). It also contains nine control fragments and eight reference probes located on different chromosomes.

The technique is based on the simultaneous amplification of probes after hybridization on targets and can be divided in five distinct steps (Figure 2.5):

1) DNA denaturation
2) DNA hybridization with the oligonucleotide probes. MLPA probe pairs are composed by two oligonucleotides, that are contiguous to the region of interest and hybridize to immediately adjacent target sequences.
3) Ligation reaction. Only when both probes are hybridized can be ligated and trigger PCR reaction
4) PCR reaction. Probes contain at their 5' and 3' ends specific sequences complementary to universal PCR primers, so a universal pair of primers is used for the subsequent amplification step.
5) Analysis of PCR products by capillary electrophoresis.

One extremity of the 2 probes of every oligonucleotide pairs is conjugated with a fluorescent marker, and the other one contains a "stuffer" sequence different in length between the different probes, thus allowing the simultaneous analysis of several targets, as sequences differing for even one nucleotide can be recognized. The amount of ligation products is directly proportional to the input DNA due to the specific amplification of probes correctly hybridized and ligated.

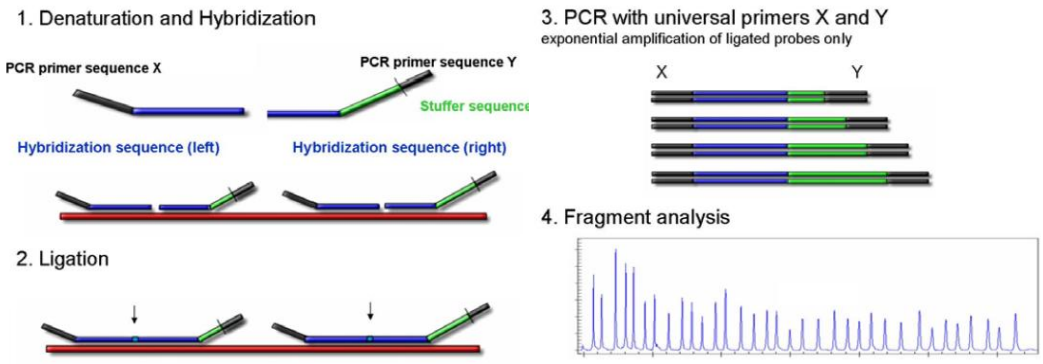

Figure 2.5: Outline of the MLPA recation (modified from Schoutern et al., 2002). After hybridisation to the target sequence in the DNA sample, the oligonucleotide probes are enzymatically ligated. Ligation products are amplified using PCR primers for sequences X and Y. Amplification products from each probe have a unique length and are separated by electrophoresis. Relative amounts of probe amplification products, in comparison to a reference DNA sample, correspond to the relative copy number of the target sequences.

Briefly, the entire MLPA procedure comprise the following reaction steps:

1. DNA denaturation.

   5 µl DNA sample were denatured for 5 minutes at 98 °C
2. Hybridization of probes to the DNA sample.

3ul of Hybridization mix (1.5 µl SALSA probemix and 1.5 µl MLPA) was added to the samples, followed by an incubation at 95 °C for 1 minute and hybridisation at 60 °C for 16 hours

3. Ligation.

   32 µl of ligase master mix (3 µl Ligase Buffer A + 3 µl Ligase Buffer B + 25 µl water + 1 µl Ligase-65 were added to the sample, followed by incubation at 54 °C for 15 minutes and 5 minutes at 98 °C to inactivate the ligase enzyme

4. PCR amplification of ligated probes.

   10 µl of polymerase master mix (7.5 µl water + 2 µl PCR Primer mix + 0.5 µl SALSA Polymerase) was added to the samples, followed by PCR reaction: 35 cycles of 30 s at 95 °C, 30 s at 60 °C, and 60 s at 72 °C, final 20 min at 72 °C)

5. Capillary electrophoresis of PCR products.

   1 µl of MLPA reaction was mixed with formamide and loaded on an ABI310 Sequencer (Applied Biosystems, Foster City, CA), using ABI POP-4 polymer, and GS400HD size standard.

In each run, 1 healthy control sample was included every 7 samples, and negative controls (water) to exclude contamination were included.

2.6.1 Run Quality control and data analysis

The MLPA reaction products were run on ABI310, and GeneMapper software (Applied Biosystems) was used in order to check that the run was successful, data were analyzed by the Coffalyser Software (MRC Holland, Amsterdam, The Netherlands).

This software first performs an evaluation of raw data by checking the DNA amount, denaturation, ligation efficiency and includes a baseline correction and peak identification.

The probemix, in addition to the target and reference probes, contains nine control probes, that probes prooduce, after the reaction, the DNA Denaturation Fragments (D-fragments, 88 nt and 96 nt), the Quantity Fragments (Q-fragments, at 64-70-76-82 nt), the 92 nt benchmark probe, and the X and Y fragments (100 and 105 nt) (Table 2.1).

The D-fragments and Q-fragments were visualized and analysed for every sample. The D-fragments (88nt and 96 nt) peak area was less than 40% lower than the 92nt fragment, indicating the execution of a correct denaturation step. The Q-fragments peak areas was verified to be lower (< 33%) than the 92 nt control fragment, indicating that the DNA amount in the reaction was enough and the ligation reaction was successful. Whenever a technical problem occurred, the entire MLPA reaction procedure was repeated.

The D-fragments hybridize to GC rich genomic sequences that are difficult to denature, they therefore work as positive controls for the denaturation step. If the 88nt and 96 nt fragments are lower (< than 40 %) than the 92 nt control fragment, a denaturation problem occurred. The Q-fragments are fragments that do not need hybridization to DNA or ligation to be amplified during the PCR step. If the Q-fragment signals are higher (> 33%) than the 92 nt control fragment, the DNA amount in the reaction was not sufficient or the ligation step failed.

Name	Length (nt)	Interpretation
92 nt benchmark probe	92	Normal probe; forms a benchmark to compare other quality control fragments to.
Q-fragments	64, 70, 76, 82	**High** when DNA amount is too low or ligation failed. All four Q-fragment signals ≥ 33% of the 92 nt control fragment → DNA quantity insufficient.
D-fragments	88, 96	**Low** in case of poor sample DNA denaturation. Signal ≤ than 40 % of the 92 nt control fragment → DNA denaturation problems.
X & Y fragments	100, 105	Control for sample swapping.[1]

[1] Rare cases are known of males lacking this Y-specific sequence and females carrying this Y-sequence on an X-chromosome.

Table 2.1 : Internal references probes (modified from MRC-Holland).

The presence of large deletions in the target genes can be observed from raw data, and are verified by the subsequent data analysis performed by Coffalyzer software (Figure 2.6), which performs 2 normalization steps:

- an intra-sample normalization, where the probe peaks are compared to the ones of the reference probes, within each sample,
- an inter-sample normalization, where the probe peaks in the target sample are compared to the ones of the control samples.

This process allows the calculation of final probe ratios, called Dosage quotient (DQ). A DQ is close to 1 indicates a wild type sample, when it is close to 0.5 it indicates an heterozygous deletion (1 allele) and when it is close to 0 it indicates an homozygous deletion, as indicated by the MRC-Holland (Table 2.2).

Table 1 – Relation Dosage Quotients and Copy Number (based on normal status of 2 copies)

Dosage Quotient Distribution	Copy Number Status
DQ = 0	**0 copies** (homozygous deletion)
0.40 < DQ < 0.65	**2⇨1 copy** (heterozygous deletion)
0.80 < DQ < 1.20	**NORMAL** (identical to reference samples)
1.30 < DQ < 1.65	**2⇨3 copies** (heterozygous duplication)
1.75 < DQ < 2.15	**2⇨4 copies** (or 1⇨2 copies)
All other values	**Ambiguous result**

Table 2.2: Relation Dosage Quotients and Copy Number (based on normal status of 2 copies) Dosage Quotient Distribution Copy Number Status (modified from MRC-Holland).

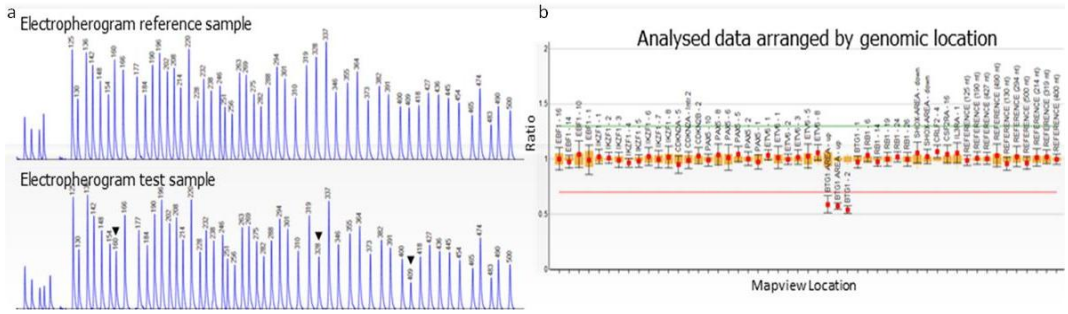

Figure 2.6: a) Comparison between electropherograms of a test sample (bottom) and a reference sample (top) showing the decrease of three probes in the test sample (arrows). b) Probe ratios of the same test sample calculated by Coffalyser.Net software after analysis of the two samples: arranging probes by chromosomal location shows a reduced copy number for the three adjacent probes in the test sample. (modified from MRC-Holland)

2.7 Quantitative real-Time PCR (qPCR)

Deletions detected by MLPA were validated by relative quantitative PCR (qPCR) performed on gDNA on an Light Cycler 480 II (Roche Applied Science, Mannheim, Germany) by SYBR Green-based quantification according to the manufacturer's protocol (Figure 2.7). The reactions were prepared in a final volume of 20 µl with 1x Master SYBR Green I (Roche Applied Science, Maanheim, Germany), 20 pmol of forward and reverse primers and 100 ng of gDNA (5 ng/µl). The amplification was carried out under the following conditions: an initial preincubation of 95 °C for 10 min followed by 45 cycles of denaturation at 95 °C for 30 sec, annealing at 60 °C for 20 sec and extension at 72 °C for 10 sec, and a final extension of 72 °C for 10 min. qPCR experiments runs were carried out in triplicate and included a negative control. The housekeeping gene GAPDH was used as reference. Primers for exons 2, 9, 15 of *DSG2*; 2, 6, 15 of *DSC2*; 2, 7, 13 of *PKP2* were developed by the Primer3, primers were designed to have a Tm around 60 °C and produce amplicons approximately 200 bp long.

After the amplification, melting curve analysis and a run on agarose gel were performed to verify the absence of aspecific products. The copy number differences between the samples were calculated by the Crossing threshold (Ct) method. Ct values were automatically calculated by the LightCycler software, using the second derivative maximum. The efficiency of each PCR reaction was automatically determined by the formula: $E = 10^{-1/slope}$. For relative quantification, the ratio of the target gene or exon with respect to the housekeeping gene of the patient's sample is then divided by the ratio of a control sample.

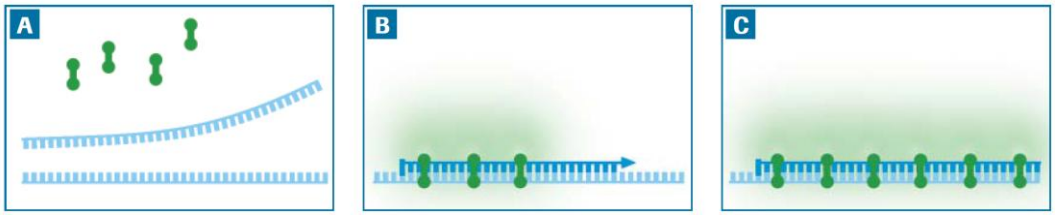

Figure 2.7: PCR in the presence of SYBR Green I. a) SYBR Green I dye only fluoresces when it is bound to dsDNA and excited by blue light. SYBR Green I does not bind to single-stranded DNA, so fluorescence is minimal during denaturation. b) As dsDNA forms and is synthesized, SYBR Green I binds the dsDNA and the fl uorescent signal from the bound SYBR Green I (green light) increases. c) At the end of elongation, all DNA is double-stranded, the maximum amount of SYBR Green I is bound and the fluorescent signal is at its maximum for that PCR cycle. Therefore, the fluorescent signals from SYBR Green I are measured at the end of eachelongation phase (modified from http://lifescience.roche.com/).

2.8 Transcript Analysis

2.8.1 Reverse transcription

Total RNA was isolated from from peripheral lymphocytes of the patient and reverse transcribed.

200 ng of total RNA was mixed with 1 µl of 50 µM random hexamers (Life Technologies, Carlsbad, CA, USA), 1 µl of 10 mM dNTP mix hexamers (Life Technologies, Carlsbad, CA, USA) and H_2O to a total volume of 10 µl and incubated on a thermal cycler at 65 °C for 5 min. 2 µl of 5X first strand buffer (Life Technologies, Carlsbad, CA, USA), 1 µl of 40U/µL RNaseOUT (Life Technologies, Carlsbad, CA, USA) and 2 µl of H_2O were added to the mix, mixed by pipetting and heated at 42 °C for 2 minutes. 1 µl of SuperScript III Reverse Transcriptase (200 U/µL) (Life Technologies, Carlsbad, CA, USA) was added and mixed by pipetting to start the retrotranscription. The reaction mix was incubated at 42 °C for 50 min, followed by inactivation at 72 °C for 15 min.

2.8.2 PCR

Targeted regions of the cDNA were amplified with the standard PCR protocol, with specific primers flanking the deleted region, located on exon 3 and exon 5 of PKP2, developed using Primer3 software (F 5'-3'AGCCCGTCACTCAGAACAG, R 5`-3`ATGCCACGAAGCTGGTTA). PCR products were run on 2% agarose gel and bands were excised from the gel.

2.8.3 DNA extraction from agarose gel

Extraction of DNA fragments from agarose gel allows to select the correct band and to remove the excess of primers and dNTPs coming from the previous PCR. Extraction was performed by using QIAquick Gel Extraction Kit (Qiagen Venlo, Limburg, The Netherlands) based on the absorption of nucleic acids to silica-gel particles in presence of salts, according to the manufacturer`s protocol.

Briefly, the DNA band was excised from the agarose gel, dissolved in 3 volumes of QC buffer and incubated at 50C C for 10 min, before adding 1 volume of isopropanol. Dissolved gel extracts were transferred on a QIAquick spin column that was centrifuged at 17,000 x g for 1 min, discarding the flowthrough. The column was washed with 600 µl of PE buffer and centrifuged for 1 min at 17,000g. it was placed on new collection tube and re-centrifuged for 3minutes at 17,000g to remove residual ethanol. Purified DNA was eluted in 20 µl of H_2O.

5 µl of purified DNA were used to set up a sequencing reaction as previously described. Sequence run was carried out on an ABI310 Sequencer (Life Technologies, Carlsbad, CA, USA) and analyzed using Chromas (Technelysium), Seqman II (DNASTAR, Madison, WI, USA), and AlaMut (Interactive Biosoftware, Rouen, France) softwares .

2.9 Next Generation Sequencing (NGS)

2.9.1 The NGS Workflow

Different NGS platforms are commercially available with different sequencing approaches, however the workflow can be summarized as follows.

1) DNA fragmentation and library preparation: gDNA is randomly fragmented and universal adaptors are ligated to the produced fragments in order to enrich targeted regions of the genome by hybridization.

2) Sequencing: parallel polymerase chain reaction and sequencing of the target DNA fragments produce clonally clustered amplicons. Sequencing methods vary and consist of several NGS technologies, including reversible terminator reactions, pyrosequencing, sequencing by ligation and real-time sequencing. These sequencing methods, applied on the enriched target samples, originate millions of reads. These reads are expected to cover the portions of the reference genome that was targeted, and the number of reads covering every target nucleotide is called coverage. Multiple algorithms have been developed in order to align the obtained sequences with the reference sequence.

3) Alignment: reads are aligned to a reference genome (resequencing) or assembled (de novo sequencing) (Flicek et al., 2009). The standard formats for storing aligned reads from NGS experiments are the SAM/BAM file formats. The sequence alignment/map (SAM) file format is used to store sequence data and alignment information for short read sequence, a binary encoding of a SAM format file in the BAM file. IGV can be used to view SAM and BAM files, which are used to store read alignments in a smaller file size (Li et al., 2009), or Variant Call Format (VCF) files. IGV allows to view through large sets of reads and to observe the overall coverage throughout the target region, to zoom into the sequence till the nucleotide level. The Burrows-Wheeler Aligner (BWA) for instance is the most used next-generation sequence alignment program that uses the Burrows-Wheeler transform to align paired-end short reads to the genome. The resulting aligned sequence is then analyzed in search of nucleotides that vary from the human reference genome and are identified as SNPs, on a process called SNP Calling. Commercial software packages, such as CLC bio (CLC Genomics Workbench), offer a complete solution from importing a FASTQ file to creating

alignments and calling variants. SNP discovery tools include programs such as the Genome Analysis Toolkit Unified Genotyper.

4) SNP calling: highlights nucleotides different from the reference sequence.

A NGS projects typically identifies 20000 to 30000 variants per exome, reflecting the great amount of genetic variation of the human genome. This makes the identification of disease causing variants among the thousands of variants very challenging. This process is further complicated by the necessity of distinguish true variants from sequencing errors or false-positive variants. Many filtering methods have been developed to prioritize variant and facilitate the research of the causal ones. All prioritization methods consider the possible effect of the nucleotide variant on the protein, and that the variant is expected to be particularly rare in the population. There are several freely available variant databases providing information on variants such as their frequency in different populations. In particular, raw sequence data from the 1000 Genomes Project and the Exome Sequencing Project, available from several sources, provide useful information for identifying common polymorphisms from NGS projects.

2.9.2 The Illumina workflow

Illumina sequencers are based on sequencing-by-synthesis using a fluorescently labeled method. The general workflow includes library preparation, single molecule amplification by bridge PCR and reversible terminator sequencing-by-synthesis.

The process starts with DNA library preparation, where DNA is randomly fragmented, and Illumina-specific adaptors are ligated at both ends of the fragments. The fragments are attached on a oligo-derivatized surface of a flow cell, that contained oligos that bound with the adaptors. Amplification of the clusters is called bridge amplification: DNA polymerase is used to produce clusters of approximately one million copies of the original fragment. During sequencing the four labeled nucleotides are simultaneously added to the flow cell channels with the DNA polymerase, to be incorporated into the cluster fragments. The four nucleotides are labeled with base-specific fluorescents, the label contains a 3’-OH group, that inhibits the fluorescent, the polymerase ligates the fluorescent

labeled nucleotides in the clusters and the 3'-OH group detaches and the fluorescence is detected. Illumina sequencer produces reads around 100 bases.

2.9.3 Next Generation Sequencing - Whole Exome Sequencing (WES)

Whole Exome Sequencing (WES) of 12 out of the 99 genotyped subject was performed in collaboration with BMR Genomics (Padua, Italy) in 2 consecutive sequencing runs, each including 6 samples.

WES is composed by three main steps: library preparation, sequencing, bioinformatic data analysis of results including SNP Calling.

2.9.3.1 Library preparation

Libraries were constructed according to the Illumina recommendations, and enriched for the coding regions of 20846 genes using the TruSeq DNA Sample Preparation kit (Illumina, San Diego, CA, USA) according to the "low sample protocol", used to process a maximum of 48 samples at one time.

The protocol for the library preparation is composed by different steps (Figure 2.8).

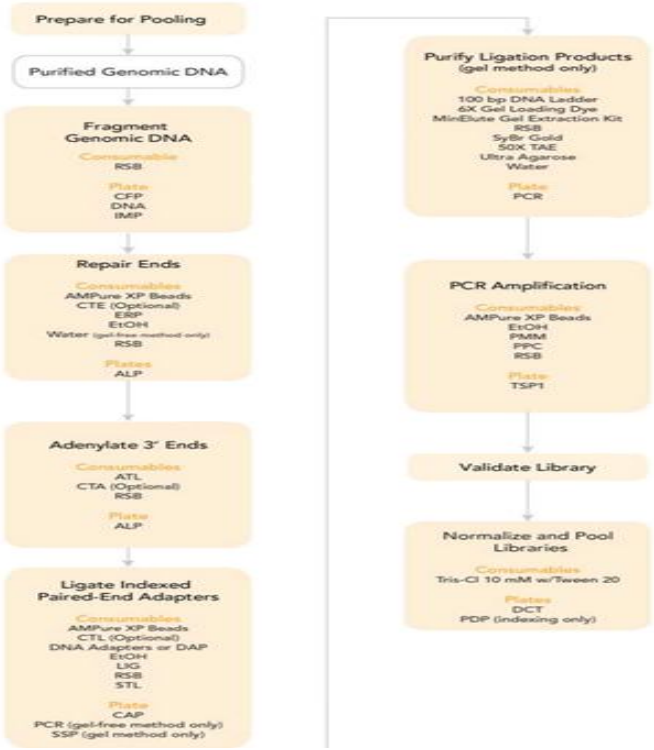

Figure 2.8: Overview of the WES sample preparation workflow (modified from www.illumina.com/).

- **DNA Fragmentation**

According to TruSeq Exome Enrichment recommendation the final library should contain DNA fragments of 200-300 bp average insert size: 1 µg (52.5 µl at 20 ng/µl) of gDNA was sheared on a Covaris S220 (Covaris, Woburn, MA, USA) using the following settings: Duty cycle 10%, Intensity 5.0, Bursts per second 200, Duration 120 sec, Mode Frequency sweeping, Power 23W, Temperature 6 °C, according to the manual. After a brief centrifugation at 600g for 5 sec, 50 µl of fragmented gDNA were transferred to every well of a PCR plate.

- **End Repair**

Covaris shearing produces dsDNA fragments with 3'or 5' overhangs, therefore this process is necessary to change the overhangs resulting from fragmentation into blunt ends.

40 µl of End Repair Mix (Illumina) was added to each well of the plate containing the fragmented DNA, mixing thoroughly. The 3' -5' exonuclease activity of the End Repair Mix removes the 3'overhangs and the polymerase activity fills in the 5' overhangs. The plate was incubated on the thermal cycler for 30 min at 30 °C.

The sample clean up was performed according to the following procedure: 160 µl of AMPure XP Beads were added to each well of the plate containing the End Repair Mix and the plate was incubated at room temperature for 15 min. The plate was placed on the magnetic stand at room temperature for 15 min, and the supernatant was discarded from each well. After two washing steps with 80% EtOH, the plate was dried at room temperature for 15 minutes, and the dried pellet was resuspended in 17.5 µl of Resuspension Buffer. 15 µl of the clear supernatant was transfer from each well of the plate to the corresponding well of the new PCR plate.

- **3' Ends Adenylation**

This process adds a single adenine nucleotide to the 3' ends of the blunt fragments, in order to avoid their ligation to one another during the adapter ligation reaction.

Briefly 2.5 µl of A-Tailing Control and 12.5 µl of A-Tailing Mix were added to the bottom of each well of the plate, the plate was incubated on the thermal cycler for 30 min at 37 °C.

- **Adapters Ligation**

During this step multiple indexing adapters are getting ligated to the ends of the DNA fragments, preparing them for hybridization on a flow cell. On the 3' end of the adapter a single timine nucleotide is located, to provide a complementary overhang for ligating the adapter to the fragment. 2.5 µl of Ligase Control, 2.5 µl of DNA Ligase Mix and 2.5 µl of each thawed DNA Adapter Indexes were added to each well of the plate, the volume was adjusted to 37.5 µl and the plate was incubated on the thermal cycler for 10 min at 30 °C. Then 5 µl of Stop Ligase Mix were added to each well to inactivate the ligation, and the volume was adjusted to 42.5 µl.

The clean up was performed as previously described, the dried pellet was resuspended with 22.5 µl Resuspension Buffer, and 20 µl of the clear supernatant were transferred to a new PCR plate.

- **Ligation Products Purification and size selection**

This step is performed to purify the products of the ligation reaction and to remove unligated adapters, adapters that may have ligated to one another, and in order to select a size-range of sequencing library for the subsequent cluster generation. For exome enrichment, a 200–300 bp insert size target was chosen. This process was carried out by an elettrophoretic run on a 2% agarose with SyBr Gold gel in 1X TAE Buffer. 7 µl of 4X Loading Buffer were added to each well of the plate, 17 µl of Resuspension Buffer and 7 µl of 4X Loading Buffer were added to 3 µl of DNA ladder. The ladder solution was loaded in one lane of the gel, and the samples on the other lanes of the gel. The gel was run at 120 V constant voltage for 120 min. After the gel visualization, the bands from the gel spanning from 300–400 bp were excised. DNA extraction was carried out following the instructions of the MinElute Gel Extraction Kit. Briefly, the gel slices were incubated in the QG solution at room temperature until they have completely dissolved and purification was performed on one MinElute spin column, eluting in 25 µl of QIAGEN EB. 20 µl of each sample from the MinElute collection tube were transferred to a new PCR plate.

- **DNA Fragments Enrichment**

In this step a PCR is performed with specific primers that anneals to the ends of the adapters, therefore DNA fragments that have adapters on both ends are selectively enriched and the DNA library is amplified.

5 µl of thawed PCR Primers and 25 µl of thawed PCR Master Mix were added to each well of the plate. Amplification was carried out in the thermal cycler as follows: 98 °C for 30 sec, 10 cycles of 98 °C for 10 sec, 60 °C for 30 sec, 72 °C for 30 sec and 72 °C for 5 min. The Clean Up of the PCR products was performed as previously described. The pellet was resuspended in 32.5 µl of Resuspension Buffer and 30 µl of the supernatant were transferred on a new PCR plate.

- **Library Validation**

Quality control analysis of the sample library includes quantification of the DNA library templates and quality control.

Quantification

An accurate quantification of DNA library is essential to create optimal cluster densities across every lane of the flow cell and obtain high quality sequencing data.

Successful enrichment (at least 10 µg) was verified by qPCR according to the Illumina Sequencing Library qPCR Quantification Guide. Briefly, a 100-fold dilution of the 2mM qPCR control template was prepared. 100 µl of 0.1% Tween 20 were added to 100 µl of the diluted template and a titration curve of six 2x serial dilutions was prepared (20 pM, 16 pM, 8 pM, 6 pM, 4 pM, 2 pM, 1 pM). Then 998 µl of 0.1% Tween 20 were added to 2 µl of the library template to make a 500-fold dilution (~4 pM), three independent dilutions of the library template were prepared to make triplicate measurements. 18 µl of the SYBR reaction mix (10 µl of 2XKAPA SYBR FAST Master Mix Universal, 0.2 µl of 10 µM qPCR Primer 1.1 and 2.1, 7.6 µl of Nuclease-free Water) were added to each well of the plate, 2 µl of the control template dilutions, the unknown library dilutions, or water were added to each well. The plate was placed on the qPCR machine with the following conditions: 95 °C for 10 min (Hot start), and 40 cycles at 95 °C for 10 sec, 60 °C for 30 sec. finally, qPCR results are analyzed and the initial

concentration of the unknown library templates is automatically calculated based on the standard curve.

Quality Control

The size of the enriched fragments was checked by running 1 µl of 1:100 diluted library aliquot of the enriched library on a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) using a DNA 1000 chip as previously described.

- **Libraries Pooling**

Multiplexed DNA libraries are normalized to 10 nM and then pooled in equal volumes for the cluster generation. The concentration of sample library were normalized to 10 nM, then 10 µl of each normalized sample library were transferred to be pooled on a new PCR plate.

2.9.3.2 Sequencing

Cluster generation was achieved according to the Illumina Cluster Generation User Guide.

In this procedure the template is attached to the surface of an oligonucleotide-coated flow cell and amplified to produce a cluster bound to the surface of the flow cell. Cluster generation workflow includes cluster amplification, linearization, blocking, primer hybridization. during cluster amplification the sample is hybridized on the flow cell and amplified, then the amplified sample is prepared for sequencing: one of the two adapters are cleaved off from the surface of the flow cell, the 3' OH ends of the linearized dsDNA clusters is blocked, dsDNA is denatured and sequencing primers can hybridize.

Briefly the template DNA is denatured with 0.1 N NaOH to a DNA concentration of 20 pM, and added with Hybridization Buffer. The DNA library was sequenced on a flow cell with paired end 2x100 bp run a on an HiSeq2000 (Illumina, San Diego, CA, USA).

Data Analysis

Image acquisition, and image and signal processing were performed during the run. A first data analysis was performed using the Illumina software package for Illumina Systems. Consecutive analysis was performed by using three different pipelines.

- A first pipeline mapped the read sets obtained against the human reference genome NCBI37/hg19 using Burrows-Wheeler Aligner tool (BWA), a freeware software package that efficiently aligns short sequencing reads against a large reference sequence, and required a previous indexing of the reference. PCR duplicates were removed and all alignment BAM files (Binary Alignment/Map) have been indexed and filtered. BAM files can be uploaded on Integrative Genomics Viewer (IGV) software, visualization tool for interactive exploration of large, integrated genomic datasets (Robinson et al., 2011).

Genome Analysis ToolKit (GATK) was used for base quality score recalibration and realignment. Variant calling was performed with GATK Unified Genotyper by using default parameters, according to the GATK recommendations, and produced Variant Caller Format files (VCF; Danecek et al., 2011) from BAM files. Annotation was carried out by uploading VCF files on the web version of the variant annotation program ANNOVAR (Wang et al., 2010). ANNOVAR provides for every variant allele frequencies available from dbSNP, 1000 genomes, and the Exome Sequencing Project. It also provides predictions of their effects on the amino-acid sequence, and pathogenicity prediction scores from SIFT, Polyphen, MutationTaster, MutationAssessor, and attaches annotations of conserved regions.

The number of reads on target, that is the number of reads mapping on the exome, was calculated from the total sequenced reads and the reads resulting after filtering and mapping. The average coverage was calculated as: the total bases aligned on the exome/ total bases of the exome.

- A second pipeline was the Illumina suite. HiSeq reads were mapped with ELANDv2 to the hg19 reference sequence, a mapping tool included in the Illumina software package that performs multiseed and gapped alignment of paired reads, using default parameters. Variant calling was performed with Consensus Assessment of Sequence And VAriation (CASAVA) with default parameters. CASAVA is a Linux application principally developed to call alleles, SNPs,indels. It also automatically generates a range of statistics (mean depth and percentage target coverage). CASAVA outputs were imported into “Genome Studio” for visualization.

-The third analysis was carried out by using CLC Genomics Workbench, an analysis package for analyzing, comparing, and visualizing next generation sequencing data. CLC carried out aligment, filtering, and variant calling by applying default parameters.

The sensitivity of WES with respect to Sanger sequencing was calculated considering the number of variants correctly identified by WES with respect to the total number of variants of the samples. The specificity was calculated considering the false positives calls of WES among the 719885 true negative bases.

2.9.4 Next Generation Sequencing - Targeted Resequencing (TR)

Targeted resequencing (TR) for 55 genes associated with different cardiomyopathies was carried out on 2 genotyped subjects at the University of Groningen.

2.9.4.1 Library preparation

Library preparation and enrichment were carried out by using the SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library, according to the manufacturer's instructions (Figure 2.9).

- **DNA Fragmentation**

3 µg of high-quality gDNA were diluted with Tris-EDTA Buffer to a total volume of 130 µl and sheared on a Covaris S220 (Covaris, Woburn, MA, USA), using the following settings: Duty cycle 10%, Intensity 5.0, Bursts per second 200, Duration 80 sec, Mode Frequency sweeping, Treatment Time 360 sec, Temperature 4 °C, according to the manual. The target peak for base pair size is between150 and 200 bp.

The sample was purified by using the QIAquick protocol (Qiagen, Hilden, Germany), according to the following procedure. Briefly 5 volumes of Buffer PB were added to 1 volume of sample, the sample was transferred on a QIAquick spin column to bind DNA on the silica membrane, and centrifuged for 30 sec, discarding the flow-through. DNA was washed by adding 750 µl of Buffer PE to the column and centrifuged for 30sec. DNA was eluted in 20 µl of nuclease-free H_2O.

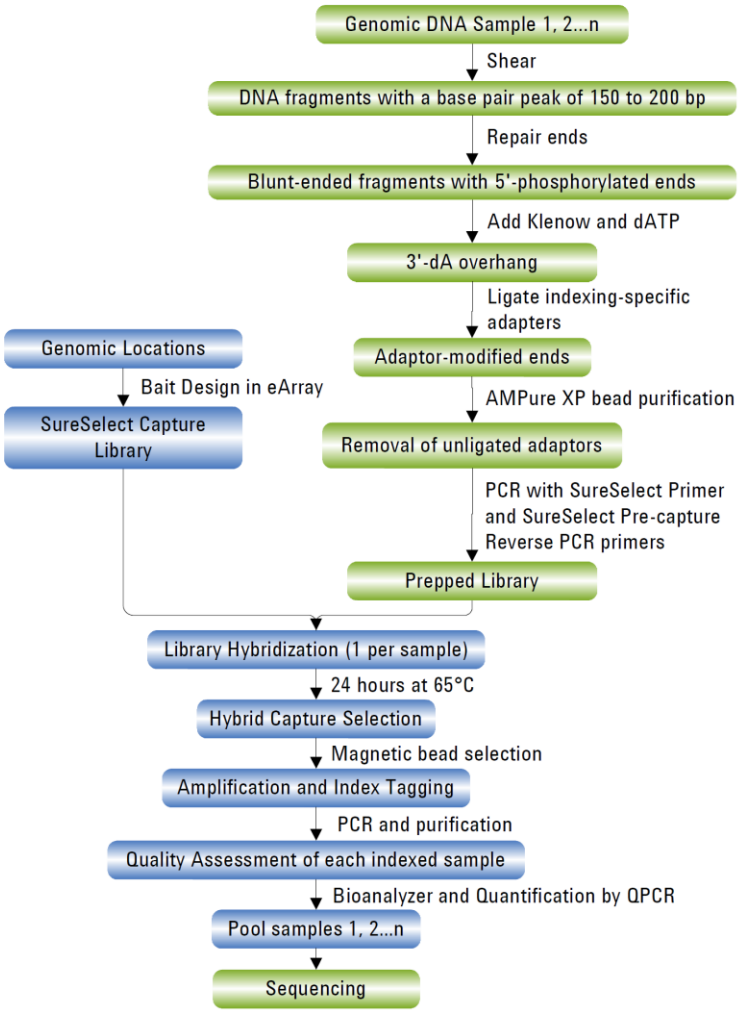

Figure 2.9: Overview of the TR sample preparation workflow (modified from http://www.genomics.agilent.com/).

- **Quality control**

The quality of the library was assesse with the 2100 Bioanalyzer using a DNA 1000 chip and reagent kit, according to the Agilent DNA 1000 Kit Guide. The electropherogram should show a distribution with a peak height comprised between 150 and 200 bases.

- **End Repair**

End repair was performed by combining 48 µl of each DNA sample with 100 µl of End Repair Mix (35.2 µl of Nuclease-free water, 10 µl of 10X End Repair Buffer, 1.6 µl of dNTP Mix, 1 µl of T4 DNA Polymerase, 2 µl of Klenow DNA Polymerase, 2.2 µl of T4 Polynucleotide Kinase). Every tube was incubated in a

thermal cycler for 30 min at 20 °C. Repaired DNA was purified using the QIAquick protocol , according to the previously described protocol.

- **3' Ends Adenylation**

Fragment dA tailing was performed by adding 20 µl of reaction mix (11 µl of Nuclease-free water, 5 µl of 10× Klenow Polymerase Buffer, 1 µL of dATP, 3 µl of Exo(-) Klenow) to 30 µl of each end repaired blunt DNA sample. Reactions were incubated on a thermal cycler for 30 min at 37 °C, DNA was purified using the QIAquick protocol.

- **Adapters Ligation**

Ligation of the adaptors was performed by adding 37 µl of the reaction mix (15.5 µl of Nuclease-free Water, 10 µL of 5X T4 DNA Ligase Buffer, 10 µl of SureSelect Adaptor Oligo Mix, 1.5 µl of T4 DNA Ligase) to 13 µl of each DNA sample. A 10:1 molar ratio of adaptor to genomic DNA insert is used, based on the initial DNA quantity of 3 µg. Samples were incubated for 15 min at 20 °C on a thermal cycler and purified by using Agencourt AMPure XP beads.

- **Size Selection**

The size selection was performed on a LabChip XT DNA Assay (750 chip, Caliper Life Sciences, Hopkinton, MA, USA) according to the manual. The LabChip XT XT DNA assays are based on traditional gel electrophoresis using a microfluidic network on a chip. Briefly, DNA samples are mixed with a sample buffer and loaded into each wells of the chip. DNA fragments are electrophoretically separated and stained with an intercalating dye. Sizing and concentration for the sample is determined using a ladder and an internal marker. After every step, the fragments were purified following to the QIAquick protocol.

- **Library Amplification**

Briefly, 50 µl of the reaction mix (15 µl of Indexing Adaptor-ligated library, 21 µl of Nuclease-free water, 1.25 µl of SureSelect Primer, 1.25 µL of SureSelect ILM Indexing Pre Capture PCR Reverse Primer, 10 µl of 5X Herculase II Rxn Buffer, 0.5 µl of 100 mM dNTP Mix, 1 µl of Herculase II Fusion DNA Polymerase) were added to 15 µl of each DNA sample. A third of the adaptor-ligated fragments are used for amplification. Samples were incubated in a thermal cycler with the

following program: 98 °C for 2 min, 10 cycles of 98 °C for 30 sec, 65 °C for 30 sec, 72 °C for 1 min, and 72 °C for 10 min. DNA was purified by using the QIAquick protocol.

- **Quality control**

Quantity, quality and size distribution of the PCR products were assessed with 2100 Bioanalyzer with the Agilent DNA 1000 Kit. The electropherogram should show a distribution with a peak size approximately 250 to 275 bp. Figure 3 Analysis of amplified prepped library DNA using a DNA 1000 assay. The electropherogram shows a single peak in the size range of 250 to 275 bp.

SureSelect hybridisation buffer was prepared according to the manufacturer instructions and heated to 65 °C for 5 minutes to prevent precipitate formation. 5 µl of SureSelect capture library and 2 µl of 25% RNase block were mixed and placed on ice. 500 ng of DNA prepped library in a volume of 3.4 µl were added to the second row in the PCR plate, combined with 5.6 µl SureSelect block mix, heated on a thermal cycler at 95 °C for 5 min and held at 65 °C. Maintaining the tubes at 65 °C, 40 µl of hybridisation buffer were added to the first row in a PCR plate and temperature was maintained at 65 °C for additional 5 minutes. 7 µl of the capture library mix were added to the third row in the PCR plate, holding the tubes at 65 °C for additional 2 min. 13 µl of hybridisation mix were transferred from the first raw of the plate to the SureSelect capture library mix contained in the third raw, and the entire contents of each prepped library mix in the second raw was transferred to the hybridization solution in the third raw of the plate. The hybridization mixture was incubated for 24 hours at 65 °C.

- **Streptavidin bead capture**

Magnetic beads prepared by adding 200 of µl SureSelect binding buffer and vortexing tubes for 5 sec. Tubes were placed on a magnetic rack and beads allowed to settle before removing the supernatant, beads were washed 3 times and resuspended in 200 µl of SureSelect binding buffer. After the 24 hour incubation at 65 °C, the hybridisation mixture was added directly into the streptavidin bead solution. Tubes were incubated for 30 min at room temperature. Beads were separated by placing tubes on a magnetic rack and the supernatant was removed. Beads were then washed with 500 µl of SureSelect wash buffer 1, incubating at

room temperature for 15 min. After spin centrifuging, tubes were placed on a magnetic rack and the supernatant was removed. The second washing stage was performed by resuspending the beads in 500 µl of SureSelect wash buffer 2 and incubating tubes at 65 °C for 15 min. Beads were separated on a magnetic rack and the supernatant was removed. This step was repeated twice. 50 µl of SureSelect elution buffer was then added to the beads, that were vortex mixed and incubated at room temperature for 5 min.

- **Addition of Index Tags by Post-Hybridization Amplification**

The amplification reaction mix (22.5 µl of Nuclease-free H_2O, 10 µl of 5X Herculase II Rxn Buffer, 0.5 µl of 100 mM dNTP Mix, 1 µl of Herculase II Fusion DNA Polymerase, 1 µl of SureSelect ILM Indexing Post Capture Forward PCR Primer, 10 µl of PCR Primer Index 1 through Index 16) was prepared according to the manufacturer's instructions and kept on ice. 35 µl of the reaction mix and 1 µl of the appropriate index PCR Primer (Index 1 through Index 16) were added to each tube, containing 14 µl of each DNA. The tubes were incubated in a thermal cycler with the following parameters: 98 °C for 2 min, 10 cycles of 98 °C for 30 sec, 57 °C for 30 sec, 72 °C 1 min, and 72 °C for 10 min. Reactions were purified using using the QIAquick protocol, with a final elution volume of 15 µl.

- **Libraries Pooling**

Multiplexed DNA libraries were normalized to 10 nM and combined such that each index-tagged sample is present in equimolar amounts in the pool.

2.9.4.2 Sequencing

3µl of the 10 nM multiplexed sample pool was diluted with 16 µl of Buffer EB (10mM Tris-Cl, ph 8.5), added with 1 µl of 2 N NaOH solution and incubated for 5 minutes at room temperature. A 2 mM PhiX sample was also denatured with 1 µl of 2 N NaOH. To 8 µl of denatured library, 992 µl of Hybridization Buffer were added to give a library concentration of 12 pM. The library and the PhiX were further diluted with Hybridization Buffer to give a concentration of 4 pM and and 1% of PhiX was added into the sample. The MiSeq reagent cartridge was prepared and 600 µl of the library were loaded into the cartridge. The flow cell,

the hybridization manifold, and the tube strip holder were installed onto the Cluster Station. A 150 bp paired-end reads was performed on a MiSeq sequencer, according to the manufacturer's instructions.

Data Analysis

MiSeq Reporter generates intermediate analysis files in a FASTQ format, files have been demultiplexed (separated by sample), so that every file contains sequences of one sample. FASTQ files have been exported and were analysed using the NextGENe software (Softgenetics, PA, USA) with default parameters, which derived a consensus sequence for each sample. NextGENe software performed quality control of sequencing data, the removal of PCR duplicates and reads filtering for wrong orientation, wrong insert size, one mate or both mates unmapped, and for multiple hits on the reference genome, and output reads were aligned to the human reference genome NCBI37/hg19. The NextGENe software output VCF files that were uploaded on Cartagenia Bench Lab (Cartagenia, Leuven, Belgium), a commercial software set up in house that uses an automated filtering procedure in order to prioritize variants. Cartagenia BenchLab applies various filters developed by the Cardiogenetics laboratory to the large numbers of variants from NGS platforms in order to select only candidates that meet specific filtration criteria and to identify clinically relevant variants. The filters take into consideration the presence and the frequency of the variants in several public databases (1000 genomes, EVS, dbSNP), and an in house variant database. The output of the filtering pipeline is normally composed of a dozen of variants.

The sensitivity and specificity of TR with respect to Sanger sequencing were already assessed (Sikkema-Raddatz et al., 2013).

2.10 Statistical Analysis

Kruskal Wallis test was used to assess the significance of differences between subgroups. A 2-tailed probability value <0.05 was regarded as statistically significant.

3. RESULTS

3.1 Cohort

Ninty-nine unrelated index cases with a clinical diagnosis of AC, 77 males, 22 females; ratio 4:1; average age at diagnosis 41±17 years (range 12-78), referred for molecular genetic screening to the Cardiovascular Genetics Laboratory (University of Padua) between years 2012 and 2014.

Index cases N=99			
Autopsy diagnosis N=26			
Gender: male, n (%)			26 (100%)
Age			28±11 (12-60)
	LD	N=13	22±6 (17-40)
	RD	N=0	
	BIV	N=13	29±12 (15-60)
Clinical diagnosis N=73			
Gender: male, n (%)			51 (70%)
Age			41±17 (12-78)
Cardiac transplantation			14 (14%)
2010 AC criteria	Definite		63 (87%)
	Borderline		3 (4%)
	Possible		7 (9%)
		I Global and/or Regional Dysfunction and Structural Alterations	60 (36M, 24m)
		II Fibrofatty replacement on endomyocardial biopsy, n (%)	6
		III Repolarization abnormalities, n (%)	44 (36M,8m)
		IV Depolarization/Conduction Abnormalities	32 (7M,25m)
		V Arrhythmias	55 (22M, 32m)
		VI Family History	14 (9M, 5m)

Table 3.1: Demographic (sex, age) and clinical characteristics of 99 patients. LD- left dominant pattern of AC, RD- right dominant pattern of AC, BIV- AC with biventricular involvement.

Seventy-three of the 99 subjects were enrolled at the Referential Clinical Genetic Centre of Arrhythmic Cardiomyopathies (Cardiology Unit), of which 14

underwent cardiac transplantation and the rest 26 of the 99 index cases were enrolled at the Veneto Registry of Cardio-Cerebro-Vascular Pathology (Cardiovascular Pathology Unit).

Clinical diagnosis of AC was defined according the 2010 Task Force major and minor criteria (Marcus et al., 2010), as previously described. Diagnostic criteria classified the 73 clinically affected index cases as follows: 63 definite, 3 borderline, 7 possible (Table 3.1). Twenty-six SCD probands, all males, with an abnormal disease pattern were grouped based on morphological features as AC with left dominant form (n=13) and AC with biventricular involvement (n=13).

3.2 Genetic screening of 5 desmosomal and 3 extra desmosomal genes

All 99 subjects underwent genetic screening for 5 desmosomal encoding genes, whereas 3 extra desmosomal genes were investigated in 46 of the 99 subjects. DNA was extracted from blood in 76 samples, from frozen myocardium in 9 samples, and from FFPE myocardium in 14 samples. Screening was performed by DHPLC, followed by Sanger sequencing of DNA samples showing abnormal elution pattern and GC rich amplicons.

DHPLC analysis detected overall 193 anomalous elution profiles and subsequent direct sequencing confirmed the presence of nucleotide variants within the 8 genes as follows: 69 (36%) were located in *DSP*, 37 in *DSG2* (19%), 35 in *PKP2* (18%), 19 in *DSC2* (10%), 13 in *JUP* (7%), 12 in *DES* (6%), 7 in *CTNNA3* (4%) and 1 only in *PLN*.

Seventy-two of the 193 (37%) variants were intronic, 6 (3%) resided in splicing sites and 3 (2%) were located in the untranslated regions (UTR) of the gene. The rest 112 (58%) variants were located within the coding regions, of which 60 (54%) were missense, 37 (33%) synonymous, 9 (8%) frame shift, 6 (5%) nonsense. A summary of nucleotide variants detected in the 8 genes is reported in table 3.2 (in appendix), and in figure 3.1.

Variants in *DSP*

Genetic screening of all 24 *DSP* gene exons identified 69 (36%) variants in 99 index subjects of which 23 were intronic, 2 were splicing site variants, 2 were

located in the UTR, 18 were missense, 19 synonymous, 3 frameshift and 2 nonsense.

Variants in *PKP2*

The 14 exons of *PKP2* gene were examined for variants in all 99 index subjects. A total of 35 (18%) variants were identified, 10 of them were intronic, 2 were located on splicing sites, 14 of them were missense, 4 frameshift, 1 synonymous and 3 nonsense.

Variants in *DSG2*

The 15 exons of *DSG2* gene were examined for variants in all 99 index subjects. A total of 37 (19%) variants were identified in *DSG2*, 18 of them were intronic, 2 were located on splicing sites, while 16 were located within the coding region. Twelve of them were missense, 4 synonymous, 1 nonsense variant that was found in homozygosity.

Variants in *DSC2*

The 14 exons of *DSC2* gene were examined for variants in all 99 index subjects. A total of 19 (10%) variants were identified, 6 of them were intronic, and the remaining 13 were located in the coding region. 8 variants were missense, 1 frameshift, 4 synonymous, 0 nonsense. Eighteen of the 19 nucleotide variants were found in heterozygosis.

Variants in *JUP*

The 14 exons of *JUP* gene were examined for variants in all 99 index subjects. A total of 13 (7%) variants were identified, 5 of them were intronic, 1 was located in the UTR, and the remaining 7 were exonic: 6 variants were missense and 1 was synonymous.

Variants in *DES*

The 7 exons of *DES* were examined for variants in all 46 index subjects. 12 (7%) variants were identified, 4 intronic variant and 8 missense exonic variants .

Variants in *CTNNA3*

The 14 exons of *CTNNA3* gene were examined for variants in 46 index subjects. 7 (3%) variants were identified, 5 of them were intronic, and the 2 exonic variants were missense.

Variants in *PLN*

Ninty-nine index AC cases underwent genetic screening for the unique coding exon of *PLN,* but only an intronic variant was identified.

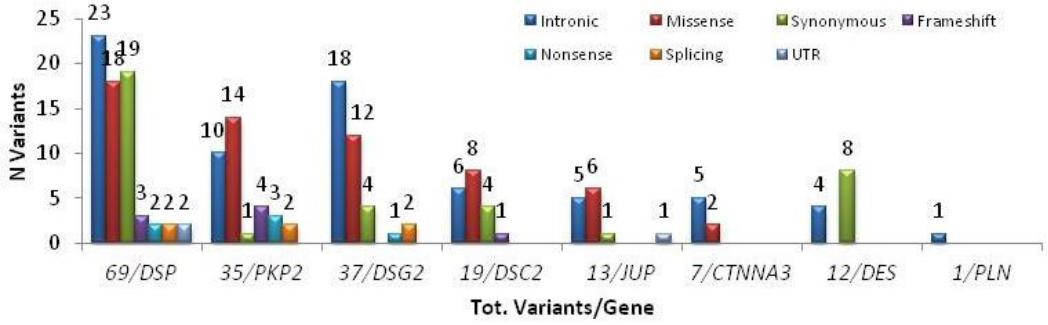

Figure 3.1: The 193 variants detected in 8 genes, sorted per gene and type (intronic, missense, synonymous, frameshift, nonsense, splicing, UTR).

3.2.1 Variants filtering

A first filtering step based on the position and function of nucleotide variants, excluded 72 of the 193 unique variants due intronic localization without predicted functional effects on the normal splicing of the transcript, 3 variants were excluded due to UTRs localization and 37 synonymous variants were excluded as not changing the functional properties of the protein. The rest 81 variants are listed below (Tabe 3.3, Figure 3.2).

Gene	cDNA change	Predicted protein change	Ex	dbSNP ID	Type	Index case/ tot	Reference
DSP	c.88 G>A	p.V30M	1	rs121912998	M	2/99	Yang et al., 2006
DSP	c.423-1 G>A	r.spl	4		Sp	1/99	Bauce et al., 2005
DSP	c.448 C>T	p.R150X	4		N	1/99	Pilichou et al., 2014
DSP	c.897 C>G	p.S299R	7	rs121912992	M	1/99	Rampazzo et al., 2002
DSP	c.913 A>T	p.I305F	7	rs17604693	M	7/99	Rampazzo et al., 2002
DSP	c.939+1 G>A		7		Sp	1/99	Sen-Chowdhry et al., 2008
DSP	c.2203 G>A	p.G735S	16		M	1/99	Novel

DSP	c.2956 C>T	p.Q986X	21		N	1/99	Campuzano et al., 2013
DSP	c.3297_3298 insTTGT	p.C1100LfsX11	23		Fs	1/99	Novel
DSP	c.3774 C>A	p.D1258E	23		M	1/99	Rampazzo et al., 2008
DSP	c.3862 A>C	p.K1288Q	23	rs138907450	M	1/99	Bao et al., 2013
DSP	c.4372 C>G	p.R1458G	23	rs28763965	M	1/99	Cox et al., 2011
DSP	c.4535 A>G	p.Y1512C	23	rs2076299	M	2/99	Yu et al., 2008
DSP	c.4961 T>C	p.L1654P	23		M	1/99	Rampazzo et al., 2008
DSP	c.4973 C>T	p.S1658F	23	rs202084959	M	1/99	Rigato et al., 2013
DSP	c.5134 A>C	p.N1712H	23		M	1/99	Novel
DSP	c.5178 C>A	p.N1726K	23	rs147415451	M	1/99	Klauke et al., 2010
DSP	c.5213 G>A	p.R1738Q	23	rs6929069	M	13/99	Den Haan et al., 2009
DSP	c.5218 G>A	p.E1740K	23	rs142885240	M	2/99	Cox et al., 2011
DSP	c.5498 A>T	p.E1833V	24	rs78652302	M	2/99	Gehmlich et al.,2010
DSP	c.5511 dup	p.R1838SfsX19	24		Fs	1/99	Novel
DSP	c.6701 A>G	p.N2234S	24		M	63/99	Novel
DSP	c.7039 A>G	p.I2347V	24		M	1/99	Rigato et al., 2013
DSP	c.7461_7464 del	p.D2489MfsX17	24		F	1/99	Novel
DSP	c.7622 G>A	p.R2541K	24	rs142078450	M	1/99	Rampazzo et al., 2008
PKP2	c.76 G>A	p.D26N	1	rs143004808	M	5/99	van Tintelen et al., 2006
PKP2	c.83 del	p.S29AfsX10	1		Fs	1/99	Novel
PKP2	c.109 A>T	p.K37X	1		N	1/99	Novel
PKP2	c.147_150 del	p.T50SfsX61	1		Fs	3/99	Novel
PKP2	c.175 C>T	p.Q59X	1		N	2/99	Rigato et al., 2013
PKP2	c.184 C>A	p.Q62K	1	rs199601548	M	1/99	Van Tintelen et al., 2006
PKP2	c.209 G>T	p.S70I	1	rs75909145	M	4/99	Koopman et al., 2007
PKP2	c.535 C>A	p.H179N	3		M	1/99	Novel
PKP2	c.1012 A>G	p.T338A	3	rs139851304	M	1/99	Bhuiyan et al., 2013
PKP2	c.1097 T>C	p.L366P	4	rs1046116	M	21/99	Koopman et al., 2007
PKP2	c.1576 A>G	p.T526A	7		M	1/99	Ostrowska Dahlgren et al., 2012
PKP2	c.1583 C>T	p.T528M	7		M	1/99	Novel
PKP2	c.1592 T>G	p.I531S	7	rs148240502	M	2/99	Lahtinen et al., 2008
PKP2	c.1643 del	p.G548VfsX15	7		Fs	2/99	Gerull et al., 2004
PKP2	c.1820 G>A	p.C607Y	9		M	1/99	Novel
PKP2	c.1841 T>C	p.L614P	9		M	1/99	Ma et al., 2013
PKP2	c.2009 del	p.N670TfsX14	10		Fs	1/99	Basso et al., 2006
PKP2	c.2111 G>A	p.G704E	10		M	1/99	Novel
PKP2	c.2119 C>T	p.Q707X	10	rs397517017	N	1/99	Basso et al., 2006
PKP2	c.2145+3_+6 dup		10		I	1/99	Novel
PKP2	c.2299+1 G>T		11		Sp	1/99	Novel
PKP2	c.2365 A>G	p.I789V	12		M	1/99	Novel
PKP2	c.2489+1 G>A		12	rs111517471	Sp	1/99	Gerull et al., 2004
PKP2	c.2552 C>T	p.T851M	13	rs146118033	M	1/99	

DSG2	c.44 T>A	p.L15Q	1		M	1/99	Bhuiyan et al., 2009
DSG2	c.166 G>A	p.V56M	3	rs121913013	M	2/99	Syrris et al. 2007
DSG2	c.245 G>T	p.G82V	4		M	1/99	Novel
DSG2	c.378+2 T>G		4		Sp	1/99	Cox et al., 2011
DSG2	c.722 C>T	p.A241V	7		M	1/99	Novel
DSG2	c.797 A>G	p.N266S	7	rs121913011	M	1/99	Pilichou et al., 2006
DSG2	c.875 G>A	p.R292H	8	rs185821167	M	1/99	Jiménez-Jáimez et al., 2014
DSG2	c.877 A>G	p.I293V	8	rs2230234	M	21/99	Posch et al., 2008
DSG2	c.1543 G>A	p.V515I	11	rs2230235	M	1/99	
DSG2	c.1652-1 G>T		12		Sp	1/99	Novel
DSG2	c.1672 C>T	p.Q558X	12		N	1/99	Pilichou et al., 2006
DSG2	c.2137 G>A	p.E713K	14	rs79241126	M	8/99	Basso et al., 2006
DSG2	c.2318 G>A	p.R773K	14	rs2278792	M	18/99	Yu et al., 2008
DSG2	c.2368 C>T	p.H790Y	15	rs114544564	M	1/99	Den Haan et al., 2009
DSG2	c.2759 T>G	p.V920G	15	rs142841727	M	1/99	Syrris et al., 2007
DSC2	c.32 A>G	p.N11S	1	rs868333	M	1/99	Den Haan et al., 2009
DSC2	c.536 A>G	p.D179G	5		M	2/99	De Bortoli et al., 2010
DSC2	c.1787 C>T	p.A596V	12	rs148185335	M	2/99	den Haan et al., 2009
DSC2	c.2194 T>G	p.L732V	14	rs151024019	M	1/99	Bhuiyan et al., 2009
DSC2	c.2326 A>G	p.I776V	15	rs1893963	M	9/99	Yu et al., 2008
DSC2	c.2393 G>A	p.R798Q	15	rs61731921	M	5/99	Posch et al., 2008
DSC2	c.2596 G>C	p.A866P	16		M	1/99	Novel
DSC2	c.2603 C>T	p.S868F	16	rs141873745	M	1/99	Fressart et al., 2010
DSC2	c.2686_2687 dup	p.A897KfsX4	16	rs200056085	Fs	1/99	Syrris et al., 2006
JUP	c.425 G>A	p.R142H	2	rs41283425	M	4/99	Koopmann et al., 2007
JUP	c.1372 G>A	p.A458T		rs139559495	M	1/99	
JUP	c.1960 G>A	p.E654K	12		M	1/99	Novel
JUP	c.2069 A>G	p.N690S	13	rs147628503	M	1/99	
JUP	c.2089 A>T	p. M697L	14	rs1126821	M	50/99	Koopmann et al., 2007
JUP	c.2124 T>A	p.D708E	14		M	1/99	Novel
CTNNA3	c.478 T>A	p.S160T	5	rs61749223	M	1/99	
CTNNA3	c.1787 G>A	p.S596N	13	rs4548513	M	24/99	

Table 3.3: The 81 unique variants resulting from the I-filtering step, will be subjected to 3 more filtering steps, Ex: exon, I: Intronic, S: synonymous, M: missense, UTR: untranslated region, N: nonsense, Fs: frameshift, Sp: splicing.

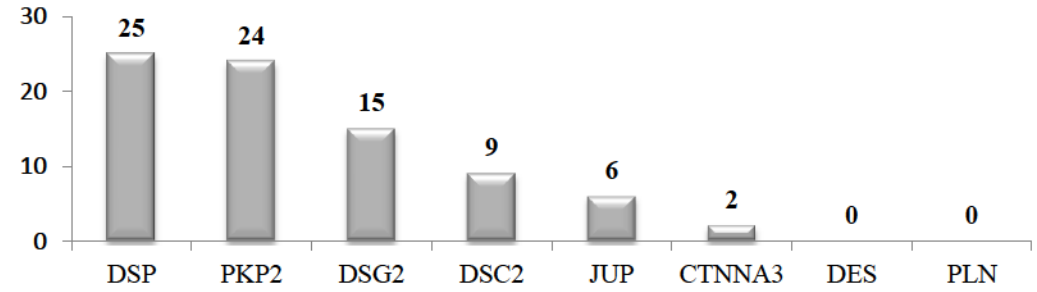

Figure 3.2: Distribution among genes of the 81 variants putative as pathogenic after the I filtering step.

3.2.2 Allelic frequency variant filtering

A second step filtering based on allelic frequency over the disease specific cut-off (0.02%), excluded 29 (36%) of the 81 variants. The excluded variants, listed on table 3.4 and figure 3.3, were searched in the 1000 genomes and EVS databases or in the internal control dataset, and were classified as not pathogenic due to high frequency in the general population.

Of these worth mentioning that the duplication c.2686_2687dup in *DSC2* gene leading to a frame-shift (p.A897KfsX4) was also excluded due to high MAF (0.006/12). Indeed this variant which was recently identified in 1.5% of the Italian healthy control subjects, was demonstrated to affect only DSC2 isoform a, while isoform b, more expressed in the heart, was not altered (De Bortoli et al., 2010).
One variant however (p.V30M in *DSP*), excluded due to high MAF (0.00220), was re-admitted as pathogenic given the current literature information (Yang et al., 2006). This p.V30M variant, located in the N-terminal of DSP, was demonstrated *in vitro* to disrupt the binding ability of DSP; this last, showed a prevalent cytoplasmic localization instead of the normal cell membrane distribution, suggesting p.V30M is affecting the interaction and localization of DSP. According to these evidences and its labeling as 'Pathogenic' in the ClinVar database (last update Aug 9, 2013), the p.V30M variant was considered potentially pathogenic in this study.

Gene	cDNA change	Predicted protein change	Ex	dbSNP ID	MAF	Type	Index case/ tot	Reference
DSP	c.88 G>A	p.V30M	1	rs121912998	0.0022 /11	M	2/99	Yang et al., 2006
DSP	c.913 A>T	p.I305F	7	rs17604693	0.0184 /92	M	7/99	Rampazzo et al., 2002
DSP	c.3862 A>C	p.K1288Q	23	rs138907450	0.0004 /2	M	1/99	Bao et al., 2013
DSP	c.4372 C>G	p.R1458G	23	rs28763965	0.0018 /9	M	1/99	Cox et al., 2011
DSP	c.4535 A>G	p.Y1512C	23	rs2076299	0.2033 /1018	M	2/99	Yu et al., 2008
DSP	c.5213 G>A	p.R1738Q	23	rs6929069	0.222 /483	M	13/99	Den Haan et al., 2009
DSP	c.5498 A>T	p.E1833V	24	rs78652302	0.008 /18	M	2/99	Gehmlich et al.,2010
PKP2	c.76 G>A	p.D26N	1	rs143004808	0.003 /6	M	5/99	van Tintelen et al., 2006
PKP2	c.209 G>T	p.S70I	1	rs75909145	0.009 /20	M	4/99	Koopman et al., 2007
PKP2	c.1012 A>G	p.T338A	3	rs139851304	0.001 /2	M	1/99	Bhuiyan et al., 2013
PKP2	c.1097 T>C	p.L366P	4	rs1046116	0.143 /312	M	21/99	Koopman et al., 2007
PKP2	c.1592 T>G	p.I531S	7	rs148240502	0.003 /6	M	2/99	Lahtinen et al., 2008
DSG2	c.875 G>A	p.R292H	8	rs185821167	0.001 /1	M	1/99	Jiménez-Jáimez et al., 2014
DSG2	c.877 A>G	p.I293V	8	rs2230234	0.039 /84	M	21/99	Posch et al., 2008
DSG2	c.1543 G>A	p.V515I	11	rs2230235	0.0098 /49	M	1/99	
DSG2	c.2137 G>A	p.E713K	14	rs79241126	0.038 /82	M	8/99	Basso et al., 2006
DSG2	c.2318 G>A	p.R773K	14	rs2278792	0.272 /594	M	18/99	Yu et al., 2008
DSG2	c.2368 C>T	p.H790Y	15	rs114544564	0.002 /4	M	1/99	Den Haan et al., 2009
DSG2	c.2759 T>G	p.V920G	15	rs142841727	0.0032 /16	M	1/99	Syrris et al., 2007
DSC2	c.32 A>G	p.N11S	1	rs868333	0.077 /169	M	1/99	Den Haan et al., 2009
DSC2	c.1787 C>T	p.A596V	12	rs148185335	0.001 /3	M	2/99	den Haan et al., 2009
DSC2	c.2194 T>G	p.L732V	14	rs151024019	0.001 /1	M	1/99	Bhuiyan et al., 2009
DSC2	c.2326 A>G	p.I776V	15	rs1893963	0.19 /415	M	9/99	Yu et al., 2008
DSC2	c.2393 G>A	p.R798Q	15	rs61731921	0.029 /64	M	5/99	Posch et al., 2008
DSC2	c.2686_2687 dup	p.A897KfsX4	16	rs200056085	0.006 /12	Fs	1/99	Syrris et al., 2006
JUP	c.425 G>A	p.R142H	2	rs41283425	0.028 /61	M	4/99	Koopmann et al., 2007
JUP	c.2089 A>T	p. M697L	14	rs1126821	0.429 /936	M	50/99	Koopmann et al., 2007
CTNNA3	c.478 T>A	p.S160T	5	rs61749223	0.011 /25	M	1/99	
CTNNA3	c.1787 G>A	p.S596N	13	rs4548513	0.473 /1032	M	24/99	

Table 3.4: 29 variants excluded due to high MAF (>0.02%) after the II- filtering step, Ex: exon, M: missense, Fs: frameshift.

The remaining 53 unique variants after the II filtering step, are distributed exclusively in the 5 major desmosomal genes. The spectrum of heterozygous variants distribution is now changed showing a higher prevalence of *PKP2* (19, 36%), followed by *DSP* (19, 36%), *DSG2* (8, 15%), *DSC2* (3, 6%), *JUP* (4, 7%) (Figure 3.3).

After MAF-filtering, *DSP* variants were reduced to 19 instead of 25 previously observed (see 3.2.1); of these, 12 were missense, 2 nonsense, 3 frameshift, 2

splicing site variants. Regarding *PKP2* variants, 19 of 24 passed this filtering step: 9 missense, 3 nonsense, 4 deletions leading to a frameshift and 3 variants predicted to affect splicing sites. *DSG2* gene filtered variants were 8 instead of 15, of which 5 missense, 1 nonsense, and 2 splicing site variants. Finally only 3 *DSC2* and 4 *JUP* missense variants passed to the next filtering step and none in the extra desmosomal genes *CTNNA3*, *DES* and *PLN*.

The 53 nucleotide variants obtained from MAF-filtering were observed in 54 AC index cases (54%). Of note, assuming a less stringent MAF as 0.02%-0.1% 20 more variants were observed in 20 additional AC patients (73%).

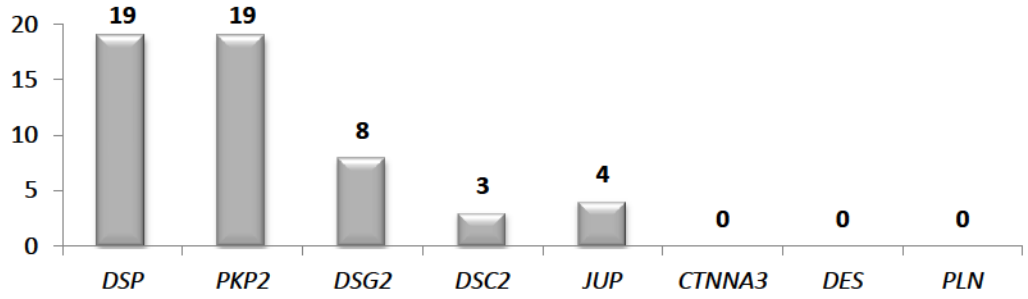

Figure 3.3: Distribution among genes of the 53 remaining variants found in our AC cohort after the II-filtering step.

3.2.3 Evolutionary conservation based filtering

Amino acid conservation analysis excluded 9 more missense variants, as shown in table 3.5 and figure 3.4, and subsequently reduced the amount of putative pathogenic variants into 44.

Among the 44 different putative pathogenic variants obtained, 18 occurred in *DSP*, 16 in *PKP2*, 6 in *DSG2*, 3 in *DSC2*, 1 in *JUP*. Twenty-four (54%) of them changed a single nucleotide resulting in a codon that codes for a different amino acid (missense mutations), 7 (16%) of them were small insertions or deletions that cause a shift of the reading frame (frameshift), 6 (14%) were nonsense mutations creating a premature stop codon (nonsense), 7 (16%) occurred in putative donor or acceptor splicing site located within the exon-intron boundaries (±10bp).

Thus one or more of the 44 variants resulting from the evolutionary conservation-filtering step were present in 49 index AC cases (49%) whereas 9 variants were excluded in 8 AC probands.

Gene	cDNA change	Predicted protein change	Ex	Amino acid conservation	Type	Index case/ tot	Reference
DSP	c.7622 G>A	p.R2541K	24	Weakly conserved	M	1/99	Rampazzo et al., 2008
PKP2	c.535 C>A	p.H179N	3	Not Conserved	M	1/99	Novel
PKP2	c.2365 A>G	p.I789V	12	Weakly conserved	M	1/99	Novel
PKP2	c.2552 C>T	p.T851M	13	Weakly conserved	M	1/99	
DSG2	c.44 T>A	p.L15Q	1	Not Conserved	M	1/99	Bhuiyan et al., 2009
DSG2	c.245 G>T	p.G82V	4	Weakly conserved	M	1/99	Novel
JUP	c.1960 G>A	p.E654K	12	Weakly conserved	M	1/99	Novel
JUP	c.2069 A>G	p.N690S	13	Weakly conserved	M	1/99	
JUP	c.2124 T>A	p.D708E	14	Weakly conserved	M	1/99	Novel

Table 3.5: Variants excluded after the III- filtering step. the 9 variants with MAF >0.02% affecting not conserved or weakly conserved amino acids among species, , Ex: exon, M: missense.

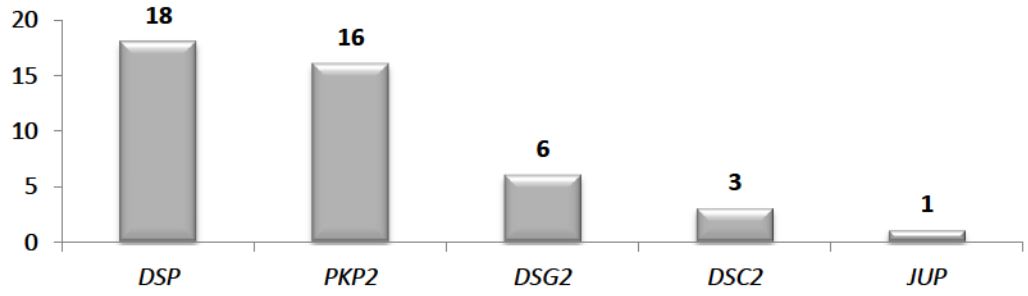

Figure 3.4: Distribution among genes of the 44 resting variants found in our AC cohort after the III- filtering step.

3.2.4 *In silico*-based filtering of variants

The resting 44 variants resulting from the previous steps were further classified according to specific features using an *in silico* algoritm proposed by van Spaendonck-Zwarts (see 2.5.6, van Spaendonck-Zwarts et al., 2013), into 4 different classes VUS0 VUS1 VUS2 VUS3. Variants pathogenic classification in table 3.6 and figure 3.5, is based on additional evaluation of the protein conformational and functional alteration.

Gene	VUS0	VUS1	VUS2	VUS3	Pathogenic
DSP		c.5218G>A, p.E1740K c.5134A>C, p.N1712H c.7039A>G, p.I2347V	c.2203G>A, p.G735S c.3774C>A, p.D1258E c.4961T>C, p.L1654P c.4973C>T, p.S1658F c.5178C>A, p.N1726K	c.423-1G>A, r.spl c.448C>T, p.R150X c.939+1G>A, r.spl c.2956C>T, p.Q986X c.3297_3298insTTGT, p.C1100LfsX11 c.5511dup, p.R1838SfsX19 c.6701A>G, p.N2234S c.7461_7464del, p.D2489MfsX17	c.88G>A, p.V30M c.897C>G, p.S299R
PKP2	c.1576A>G, p.T526A c.184C>A, p.Q62K	c.2145+3_2145+6dup, r.spl?		c.83del, p.S29AfsX10 c.109A>T, p.K37X c.147_150del, p.T50SfsX61 c.175C>T, p.Q59X c.1583C>T, p.T528M c.1643del, p.G548VfsX15 c.1820G>A, p.C607Y c.1841T>C, p.L614P c.2009del, p.N670TfsX14 c.2111G>A, p.G740E c.2119C>T, p.Q707X c.2299+1G>T, r.spl c.2489+1G>A, r.spl	
DSG2		c.166G>A, p.V56M	c.722C>T, p.A241V	c.378+2T>G, r.spl c.1652-1G>T, r.spl c.1672C>T, p.Q558X	c.797A>G, p.N266S
DSC2			c.2603C>T, p.S868F	c.536A>G, p.D179G c.2596G>C, p.A866P	
JUP			c.1372G>A, p.A458T		

Table 3.6: Final classification of the resulting variants as pathogenic based on *in silico* analysis.

Note that only variants classified as VUS3 and VUS2 were considered pathogenic or potentially pathogenic for further genoype-phenotype correlation studies,

whereas 5 classified as VUS1 and 2 classified as VUS0 were excluded from this study. Therefore only 37 of the 193 variants initially identified in disease-causing genes were considered pathogenic or potentially pathogenic variants. Forty-two index AC cases (42%) carried one or more of these 37 variants and were finally considered as positively genotyped.

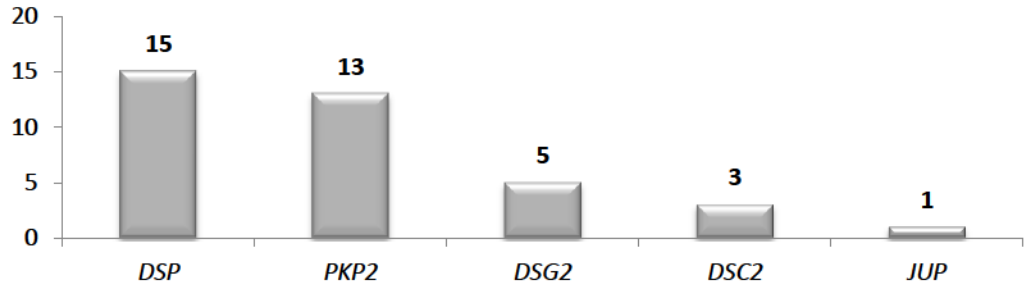

Figure 3.5: Distribution among genes of the 37 resulting variants found in our AC cohort after the IV- filtering step.

3.3 CNVs Analysis

The entire cohort of 99 patients with AC was screened for large intragenic rearrangements in desmosomal genes by MLPA.

Five of the 99 probands (6%) displayed an aberrant MLPA profile, indicating a copy number alteration of the gene due to the lack of one exon or eventually the entire gene. These 5 index AC cases carrying deletions of one or more exons/genes were considered as positively genotyped, increasing from 42 to 46 (46%) the number of positively genotyped patients (see 3.2.4).

Specifically, a 120kb- heterozygous deletion on chr12 containing the entire *PKP2* gene was detected in 2 patients (#36, #42). Bioinformatics means estimated this deletion as a gene dosage reduction ranging from 0.3 to 0.6. The deletion of *PKP2* gene in these 2 patients was further confirmed by relative qPCR for 3 *PKP2* exons (Figure 3.6 and Figure 3.7) and resulted absent in 500 healthy subjects.

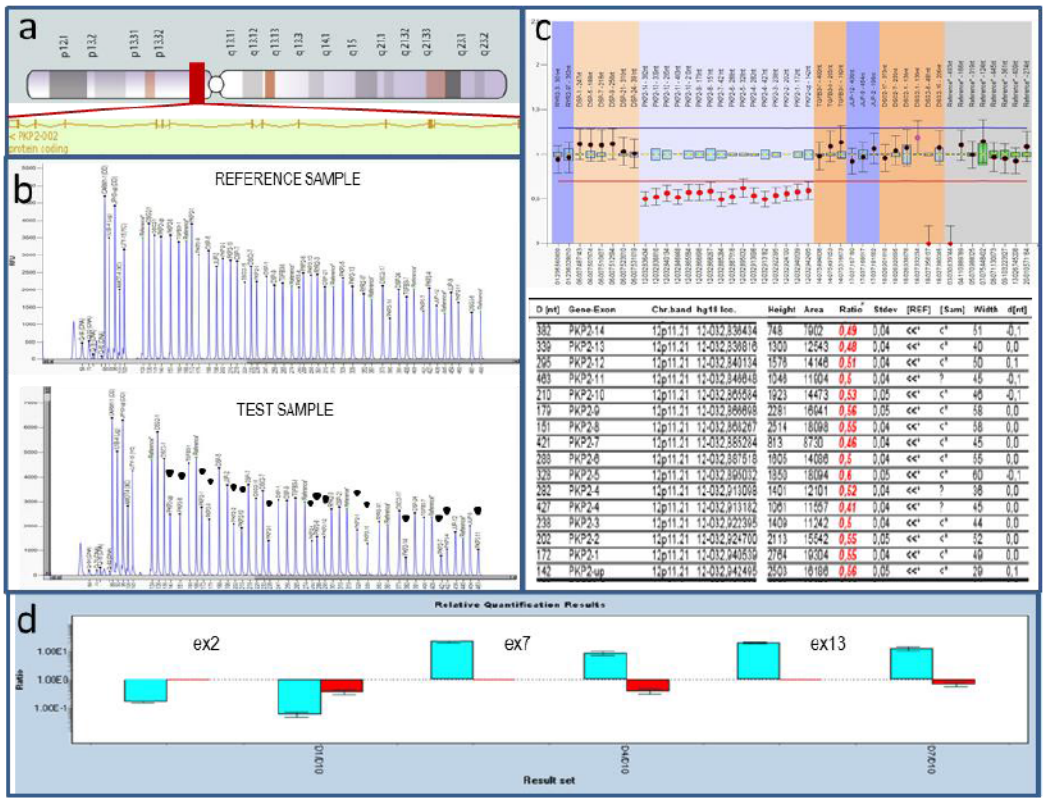

Figure 3.6: MLPA analysis of sample #36 a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

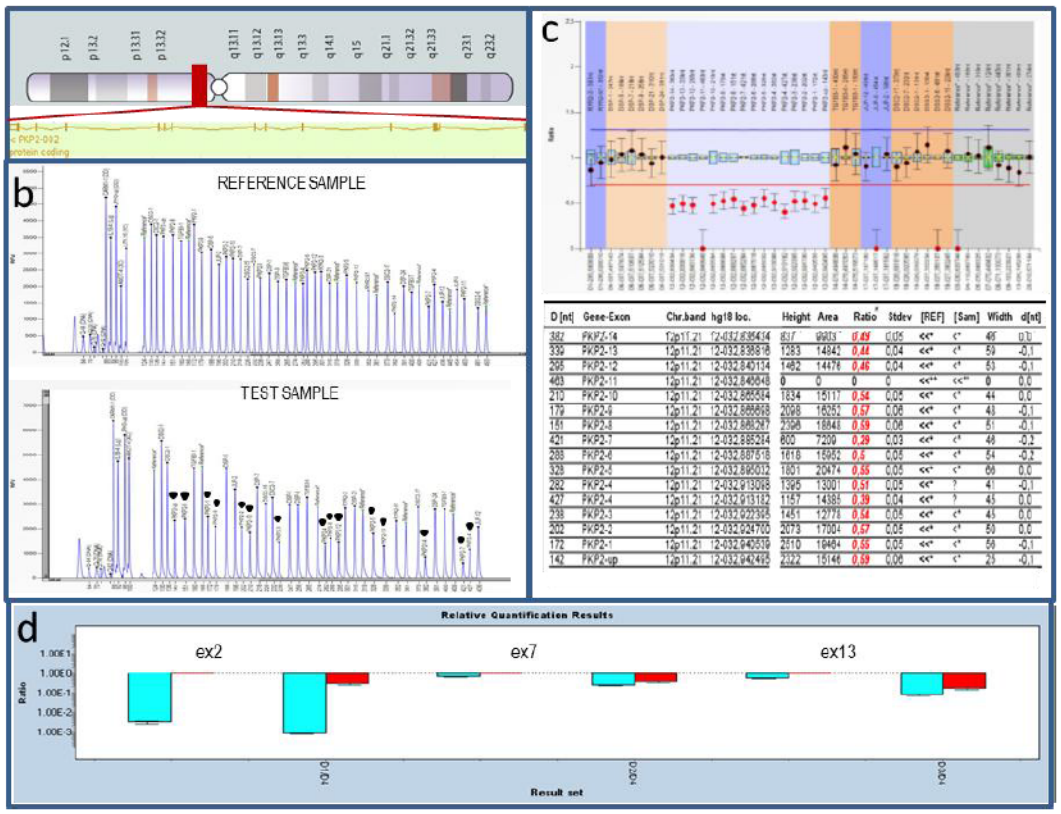

Figure 3.7: MLPA analysis of sample #42 a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

Another patient (#39) carried a large heterozygous deletion of at least 482kb on chr18 which comprise both *DSC2* and *DSG2* genes. Bioinformatics evaluation estimated the gene dosage reduction ratios of 0.35 and 0.51 respectively. qPCR performed for 3 *DSC2* and 3 *DSG2* exons confirmed the deletion of both genes, and its absence in 500 healthy subjects.

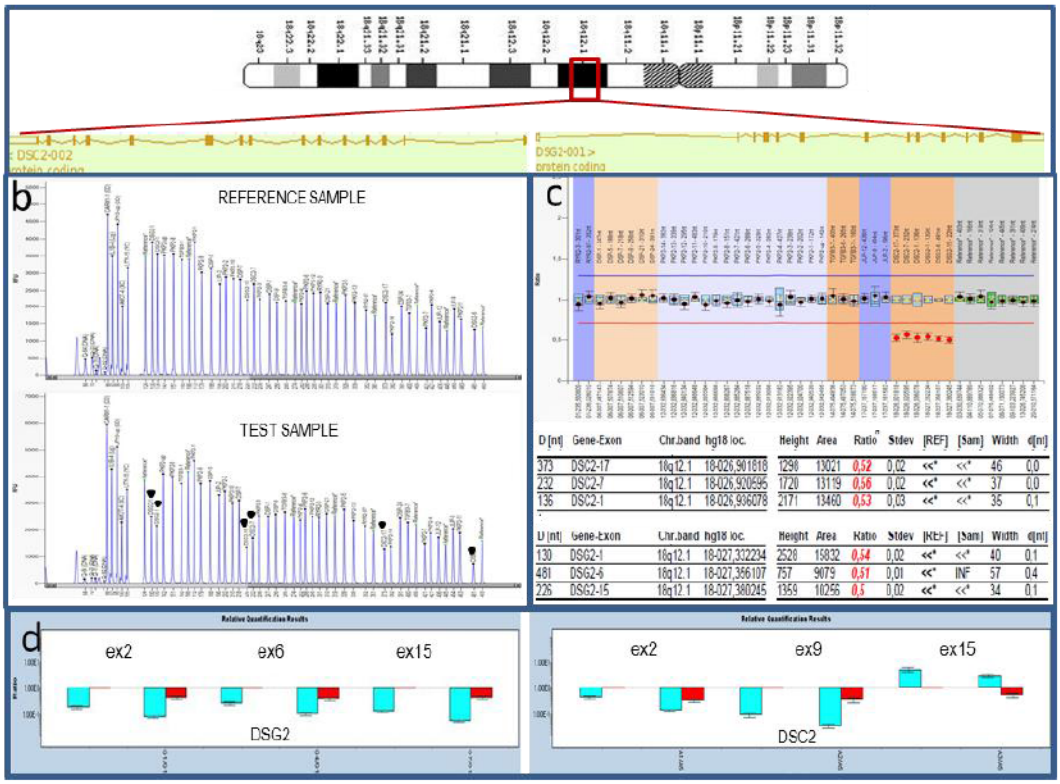

Figure 3.8. MLPA analysis of sample #39 a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

Finally, a single-exon heterozygous deletion was detected in 2 patients (#8 and #38); the deletion of only exon 4 of *PKP2* gene (c.1286_1491del) is predicted to shift the reading frame of the protein translation and create a premature stop codon, ending up with a 350 amino acid-long peptide instead of a 837 amino acid protein (Figure 3.9 and 3.10). A ratio of 0.3 and 0.6 was estimated in these two patients by bioinformatics analysis. qPCR analysis for the specific exon showed a clear-cut reduction compared to 500 healthy control individuals, whereas the other two exon of the same patients did not show quantitative shifts.

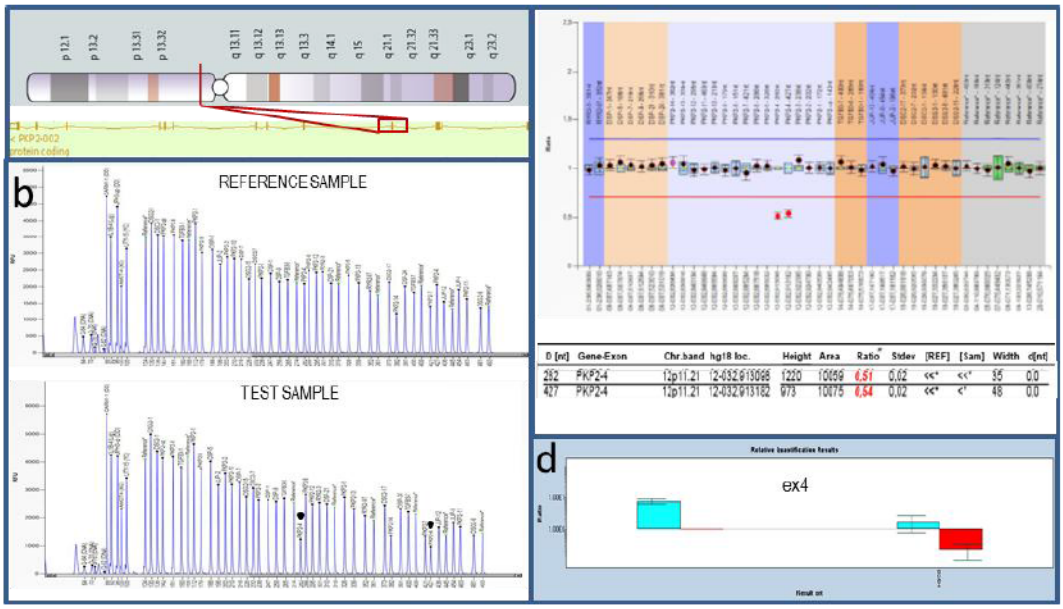

Figure 3.9: MLPA analysis of sample #8 a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

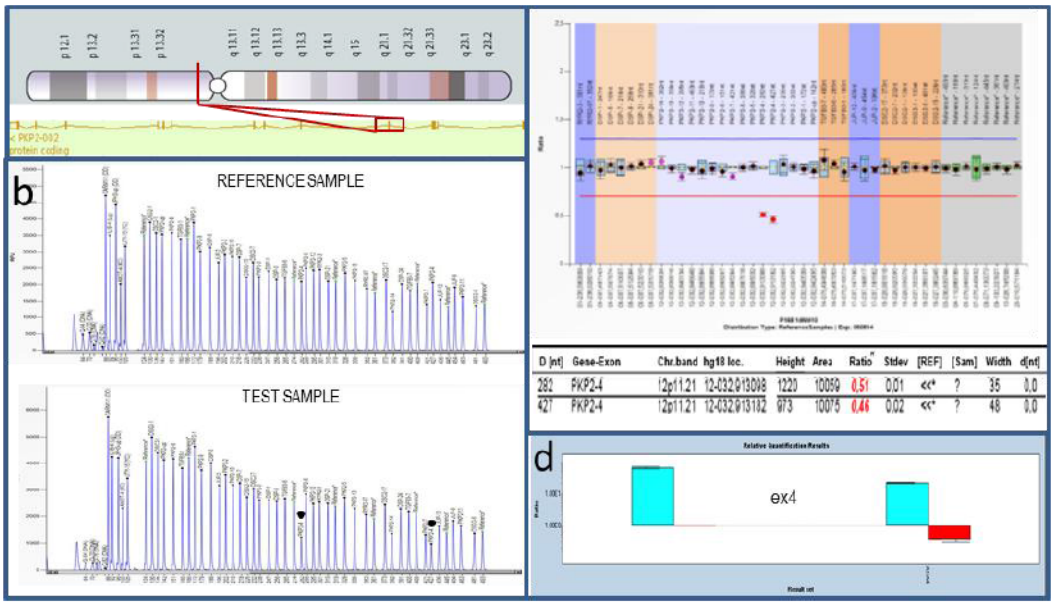

Figure 3.10: MLPA analysis of sample #38 a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

In order to characterize the effect of this exon-specific deletion at the transcript level, cDNA of the patients carrying the deletion was amplified with specific primers. Agarose gel electrophoresis clearly showed two different fragments at

approximately 387 bp and 222 bp, representing the normal transcript containing exon 4 (longest fragment) and the transcript lacking exon 4 of *PKP2* (shortest fragment). Direct sequencing of these two cDNA fragments demonstrated the absence of exon 4 in one allele of both AC patients. (Figure 3.11).

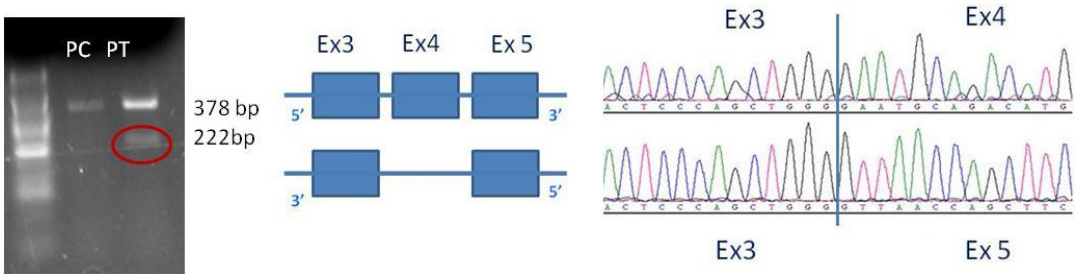

Figure 3.11: Agarose gel run showing the different PCR products and cDNA sequence of the 222bp PCR product showing the lack of ex4 in the transcript. PC: positive control sample, PT: patient sample.

3.4 Variant analysis in the study population

Major AC-related genes were screened in 99 index cases for point mutations by DHPLC and direct sequencing, and for large deletions or duplications by MLPA and qPCR. Subsequent stringent variants filtering led to the identification of 36 point mutations and 3 CNVs (Table 3.7) classified as pathogenic mutations.

Index case	DSP	PKP2	DSG2	DSC2	JUP	N variants
62 [SD]		c.175 C>T, p.Q59X				1
68 [SD]	c.939+1G>A, r.spl					1
70 [SD]		c.2009 del, p.N670TfsX14				1
2				c.536 A>G, p.D179G [H]		1
1		c.1643 del, p.G548VfsX15				1
16		c.2111 G>A, p.G704E				1
4				c.536 A>G, p.D179G [H]		1
20		c.83 del, p.S29AfsX10				1
5		c.175 C>T, p.Q59X				1
57	c.7461_7464 del, p.D2489MfsX17				c.1372 G>A, p.A458T	2
7	c.6701 A>G, p.N2234S	c.109 A>T, p.K37X				2
8		Deletion ex4				1
89 [CT]		c.1643 del, p.G548VfsX15				1
59 [SD]				c.2603 C>T, p.S868F		1
13	c.3774 C>A, p.D1258E					1
14		c.1820 G>A, p.C607Y				1
15	c.5178 C>A, p.N1726K					1
74 [SD]			c.722 C>T, p.A241V			1
75 [SD]		c.2299+1 G>T, r.spl				1
23		c.147_150del, p.T50SfsX61				1
24	c.88 G>A, p.V30M		c.378+2 T>G, r.spl			2
25	c.2203 G>A, p.G735S					1
76 [SD]	c.448 C>T, p.R150X					1
40		c.2489+1 G>A, r.spl				1
80 [CT]			c.1672 C>T, p.Q558X [H]			1
28		c.1583 C>T, p.T528M				1
29	c.3297_3298 insTTGT, p.C1100LfsX11					1
30		c.147_150 del, p.T50SfsX61				1
82 [CT]	c.5511 dup, p.R1838SfsX19			c.2596 G>C, p.A866P		2
61 [SD]			c.1652-1 G>T, r.spl			1
83 [CT]	c.2956 C>T, p.Q986X					1

32		c.1841 T>C, p.L614P				1
36	c.88G>A, p.V30M	Deletion ex1-14				2
38		Deletion ex4				1
42		Deletion ex1-14				1
39			Deletion ex1-15	Deletion ex1-16		2
49		c.147_150del, p.T50SfsX61				1
90 [SD]	c.897 C>G, p.S299R					1
98 [SD]	c.897 C>G, p. S299R					1
91 [SD]	c.423-1 G>A, r.spl					1
92 [SD]		c.2009 del, p.N670TfsX14				1
93 [SD]	c.4973 C>T , S1658F; c.7039 A>G, I2347V					2
95 [CT]	c.88G>A, p.V30M					1
96 [CT]			c.797 A>G, p.N266S			1
97 [CT]	c.4961 T>C, L1654P	c.2119 C>T, Q707X				1
44		c.1643 del, p.G548VfsX15				
Single Mutation Carriers	11	20	4	3	0	
Multiple Mutation Carriers						8

Table 3.7: Positive genotyped subjects. [SD] sudden death, [CT] cardiac transplantation [H] homozygous

All 39 pathogenic variants were detected in 46 subjects of our cohort (46%). Forty-two of the 46 mutation carriers harbor one or more point mutations, 3 probands carried larger deletions of one or more genes and 1 patient showed both a point mutation in *DSP* and a complete deletion of the *PKP2*-containing allele.

Of these 46 positively genotyped patients, 20 (20%) were single mutation carriers for *PKP2*, 11 (11%) were single mutation carriers *DSP* and 4 probands (2%) carried a single nucleotide variant in *DSG2* gene, one of whom was homozygote. Finally, 2 out of 3 (3%) single mutation *DSC2* carriers were also homozygotes and no single mutation carriers were detected for *JUP*.

Eight of the 46 index patients (8%) carried multiple mutations, 7 were digenic heterozygotes, showing one missense variant in one gene coupled with a “radical mutation” (frameshift, nonsense, splice, deletions) in a second gene, and 1 only was a compound heterozygote for 2 *DSP* missense variants.

Of the 46 positively genotyped index cases, 13 were cases coming of the SCD Veneto Registry and 7 were AC patients who underwent cardiac transplantation.

Figure 3.12 summarizes relative proportion of "radical" mutations, comprising variants with a more drastic predicted effect i.e. frameshift, nonsense, splice site and insertions/deletions, versus missense mutations detected in all five desmosomal genes. Note that, no mutations were detected in the extra-desmosomal genes analyzed and that in *PKP2* missense mutations can be rarely considered as pathogenic. Indeed, 11 of the 15 (73%) *PKP2* mutations were radical, identified in 16 AC probands whereas 4 of the 15 (27%) were missense *PKP2* mutations carried by 4 index cases.

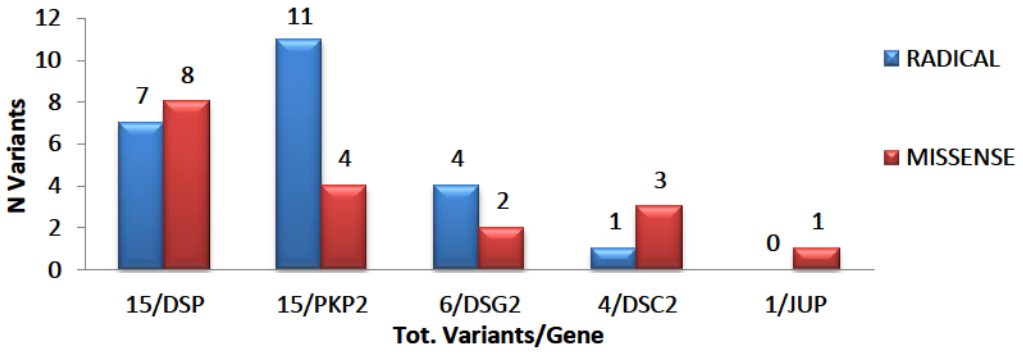

Figure 3.12: Relative proportion of "radical" and missense pathogenic putative variants identified in 5 desmosomal encoding genes.

3.5 Genotype Phenotype correlation between desmosomal genes mutation carriers and non-mutation carriers

Table 3.8 summarizes clinical and genetic data of index AC patients based on the current diagnostic criteria.

INDEX CASES N=99		**MUT+**	**MUT-**
N (%)		46 (46%)	53 (53%)
Age (y)		38±17 (12-78)	38±16 (13-77)
Autopsy diagnosis N=26		***MUT+***	***MUT-***
N (%)		13 (50%)	13(50%)
LD N=13		3 (23%)	10 (77%)
	Age (y)	16±3 (range 12-20)*	24±7 (range 17-40)
BIV N=13		10 (77%)	3 (23%)
	Age (y)	32±12 (range 15-40)	21±4 (range 17-26)
Clinical diagnosis N=73		***MUT+***	***MUT-***
N (%)		33(45%)	40(55%)
Male gender, n (%)		22 (30%)	29 (40%)
Age (y)		37±17 (12-78)	43±16 (15-77)
2010 AC criteria	Definite N=63	30 (41%)	33 (45%)
	Borderline N=3	1	2
	Possible N=7	2	5

Table 3.8: Comparison between desmosomal genes mutation carriers and non mutation carriers. Mut+ : positively genotyped, Mut - : negative genotyped, LD- left dominant pattern of AC, BIV-AC with biventricular involvement. Note * $P<0.05$ is considered statistically significant.

In 99 index AC cases, including both SCD and clinically diagnosed AC cases, 46 resulted positively genotyped patients however no difference in age was evident between groups.

Focusing the attention on SCD subjects, all 26 SCD subjects were males (100%), half of them had predominant or even isolated left ventricular involvement (50%) and 13 of 26 SCD cases were positively genotyped (50%). However, only 3 of the 13 AC patients with a left dominant pattern were positively genotyped (23%). Comparing positively genotyped patients was evident that only 3 had a left dominant AC pattern (23%) and all 3 of them died at youger age (16y±3) compared with patients who exhibited a BIV involevement (p 0.03).

The gender incidence was assessed only on clinically affected patients since SCD cases were all males, showing that female to male ratio of genotyped positive AC patients was maintained 1:3.
Comparing mutation carriers and non mutation carriers fulfilling the established AC diagnostic criteria, 33 were found positively genotyped (45%): 30 classified as definite, 1 classified as borderline and 2 classified as possible. Interestingly, 33 of the 40 non-mutation carriers were classified also as definite.
No significant difference was observed in terms of clinical characteristics according to the task force criteria.

When we divided clinically affected living subjects from SCD cases we observed a higher risk of SCD (5 of 11, 45%) and a significant difference in the mean age of the disease presentation among *DSP* mutation carriers (p 0.04). *DSG2* mutation carriers died also suddenly at a younger age however the sample number is too low to extrapolate statistically significant information. Although the incidence of *PKP2* mutations is higher among AC probands (20, 44%), the risk of SCD is lower (4 of the 20, 20%).

Regarding the distribution of gene mutations in 33 clinically diagnosed AC subjects we observed that all *DSP* (6 of 6, 100%) and 86% (6 of 7) of multiple mutation carriers exhibited definite 2010 AC criteria. Interestingly, we also observed that the vast majority of *PKP2* mutations found in single- *PKP2* carriers were “radical” (11 of 16, 69%), however not all of these *PKP2* “radical” mutation carriers exhibited defined clinical diagnostic AC criteria. Specifically, 4 *PKP2* mutation carriers were clinically classified as possible based on 2010 AC criteria; two of these subjects (about 40y old, male) presented a *PKP2* frameshift [p.G458VfsX15, p.T50SfsX61 respectively] showing a LD pattern of the disease, 1 (44y-old female) displayed a complete deletion of the *PKP2* allele in heterozygosis with absence of any clinical signs of the disease and 1 (15-y-old, male) carried a simple missense mutation [p.T528M] in heterozygosis.

Genotyped positive [MUT+] N=46			***DSP***	***PKP2***	***DSG2***	***DSC2***	***Multiple***
N (%)			11 (24%)	20 (44%)	4 (9%)	3 (6%)	8 (17%)
Male, n (%)			8 (17%)	14 (30%)	2 (4%)	2 (4%)	5 (11%)
Age (y)			35±15 (12-60)	41±18 (12-78)	33±21 (13-63)	36±22 (13-56)	36±18 (15-66)
Autopsy diagnosis N=13			5	4	2	1	1
			3BIV: 2LD	4BIV	1BIV: 1LD	1BIV	1BIV
Age (y)			26±12 (12 -37)*	35±18 (18-60)	19±8 (13-25)	40	15
Clinical diagnosis N=33			6	16	2	2	7
Age			43±14 (18-60)	41±18 (12-78)	47±23 (31-63)	35±30 (13-56)	39±17 (15-66)
	Definite		6	12	2	2	6
	Borderline						1
	Possible			4			
		I Global and/or Regional Dysfunction and Structural Alterations	5M,1m	7M,3m	2M	2M	3M,3m
		II Fibrofatty replacement on endomyocardial biopsy, n (%)	1M				
		III Repolarization abnormalities	2M,1m	7M,1m	1M	2M	3M,1m
		IV Depolarization/ Conduction Abnormalities	2M,2m	1M,7m	1M	1M,1m	3m
		V Arrhythmias	1M,1m	3M,8m	2m	1m	2M, 4m
		VI Family History	1M	2M,2m	1m		1M

Table 3.9 Comparison between the different groups of desmosomal genes mutation carriers. Mut+: positively genotyped. Note * $P<0.05$ is considered statistically significant.

3.6 Next Generation Sequencing

3.6.1 Whole Exome Sequencing

3.6.1.1 WES raw data analysis

Paired-end Whole Exome Sequencing of 12 genotyped subjects on the Illumina HiSeq2000 produced a total of 819.240.262 reads of 100 bp in length, on average 68.270.020 per sample, consisting of a total output of 81.924.026.200 bp.
Mapping by CLC software reduced the total number of reads to 800.428.775 after alignment. Further reduction of the number of reads to 581.800.573 was obtained by considering the number of unique aligned reads and allowing a maximum of 2 mismatches or gaps. Table 3.10 displays mapped and unmapped reads per sample. The number of reads aligned varies among the different samples within each run, ranging from 12 k to 66.8 k, while the number of reads on target varies from a minimum of 48% to a maximum of 57%, depending on the sample.

Sample ID	Mapped Reads	Unmapped Reads
1	27.408.565	4.184.911
2	12.865.444	1.892.534
3	66.790.928	10.216.939
4	57.333.247	9.864.033
5	38.984.184	5.953.774
6	42.198.051	6.919.797
7	46.321.008	24.505.096
8	19.095.462	10.016.630
9	66.340.765	35.528.169
10	71.037.969	37.290.933
11	71.650.527	39.139.587
12	61.774.423	33.115.799

Table 3.10: Results of the alignment per sample

After alignment of sequencing data against the human genome reference, the sequencing reads and coverage of the target regions were observed by uploading BAM files on IGV software. (Robinson et al., 2011). The samples showed a pattern of well covered regions against uncovered or low covered regions, all

through the genomic regions of interest. Overall mean coverage, describing the mean depth of sequenced targets, varied slightly among the samples, from 9.2X to 54.3X. A coverage above 20X, which is considered the cut-off for reliable variant call showing that samples are sequenced sufficiently in terms of number reads, was achieved only for 6 samples. The 20-fold coverage of targets was also different among samples, and varied from 10,20 % to 85.85 % (Table 3.11).

Sample ID	Mean Coverage	% of Target with Coverage ≥ 20 X
1	20,4 X	38,65 %
2	9,2 X	11,07 %
3	51,6 X	81,03 %
4	41,5 X	75,08 %
5	25,1 X	48,38 %
6	27,4 X	53,52 %
7	32,3 X	65,45 %
8	10,4 X	10,20 %
9	49,6 X	85,85 %
10	54,3 X	85,42 %
11	54,2 X	84,89 %
12	46,5 X	81,06 %

Table 3.11: The mean coverage obtained per sample, together with the fraction of the target region showing a coverage >20X

3.6.1.2 SNP Calling analysis- sensitivity and specificity validation

We investigated the differences between the three callers in sensitivity and specificity of SNP calling based on the analysis of 6 samples. GATK identified a total of 1.975.223 variants, CLC identified a total of 1.435.455 variants, and CASAVA identified a total of 260.433 variants (Table 3.12). The last software calls a lower number of variants because SNP calling is automatically limited merely to the enriched target regions (most exonic), and discards the insufficiently covered target regions. On the contrary CLC and CASAVA perform calls also on low covered exonic regions and intronic regions.

N calls	GATK	CLC	CASAVA
Total calls	1.975.223	1.435.455	260.433
Sample 12	238.186	195.672	42.691
Sample 2	132.475	99.600	29.337
Sample 4	382.445	289.124	48.956
Sample 5	430.718	320.178	48.492
Sample 6	381.801	254.541	44.268
Sample 7	409.598	276.340	46.689

Table 3.12: Comparison of the SNP calls across six samples performed by GATK, CLC and CASAVA.

In order to test whether WES could be used as a reliable tool in a clinical setting, we evaluated the analytical sensitivity and specificity of WES for the detection of a series of previously characterized nucleotide changes, by comparing WES data with data obtained by Sanger sequencing of 5 genes.

Initial analysis with the CLC software was able to automatically detect 68 out of the 90 expected nucleotide changes (76%).

22 changes were missed during the initial analysis. The 22 variants were manually searched among the aligned reads, this alternative strategy allowed the detection of 13 among the 'missed' variants. Nine variants previously identified by conventional sequencing remained undetected after both automatic and manual SNP calling. (Table 3.13).

Gene	cDNA	Predicted Protein change	Ex/Int	Coverage	Distance from probes
DSC2	c.70 -154 G>A	-	Int 1	1	154 bp
DSP	c.-1_1insA		Int 1	0	
DSP	c.913A>T	p.I305F	Ex 7	0	356 bp
DSC2	c.2326 A>G	p.I776V	Ex 15	2	
PKP2	c.76G>A	p.D26N	Ex 1	1	
PKP2	c.2489+13_2489+14ins C		Int 12	11	203 bp
DSG2	c.828+16C>A		Int 7	4	62 bp
DSP	c.1141-44C>T		Int 10	1	44 bp
DSP	c.1420-93C>T		Int 12	1	93 bp

Table 3.13: 9 Variant missed by SNP calling, Ex: exon, Int: Intron

Six of the 9 variants were located too distant from the nearest probes to be detected (from 44 to 356 bp). The remaining 3 variants were not called because they were located in uncovered (coverage 0) or low covered (coverage 1 and 2) targeted regions. CLC software automatically discards every variant showing a coverage below 5X assigning it as a false call. The overall sensitivity of WES was evaluated around 90%. The specificity was calculated as >99.9%.

3.6.1.3 SNP Calling analysis- 150 genes associated with arrhythmic inherited cardiomyopathies

All successfully mapped sequence reads of the 12 samples were further analyzed to detect sequence variants, including single nucleotide changes and small insertions and deletions, within a set of 150 genes involved in different inherited cardiomyopathies. Variants were manually called, with the generation of a list of 1003 known or novel nucleotide changes detected in the 12 patients. First step filtering consisted of nonsynonymous and novel variants against the 1000 Genomes project and EVS database resulting in the identification of 100 potentially pathogenic variants. The validation process by Sanger sequencing of these variants successfully confirmed only 3 of them (3%) in 3 patients. Variants were located respectively in the troponin2 gene (*TNNT2*), in the myosin binding protein C3 gene (*MYBPC3*), and in *LMNA* (Table 3.14).

Gene	cDNA	Predicted Protein change	Ex	dbSNP ID	MAF	Type	Reference
TNNT2	c.113C>T	p.A38V	6	rs56816490	NA	M	Millat et al., 2011
MYBPC3	c.2198G>A	p.R733H	23			M	Novel
LMNA	c.949G>A	p.E317K	6	rs56816490	NA	M	Arbustini et al., 2002

Table 3.14: Variants detected by WES and Sanger sequencing, EX: exon, M: missense

The 97 variants that have proven false positives were located in exome regions with low coverage (<5X), or far from the probes.

3.6.2 NGS- Targeted Resequencing

Targeted resequencing has been validated (Sikkema-Raddatz et al., 2013) showing both a sensitivity and specificity around 100% and is routinely used as a stand-alone diagnostic test at the Cardiogenetics Laboratory of the University of Groningen.

Four AC patients from the Cardiogenetics Laboratory of the University of Groningen and one subject from the Cardiovascular Laboratory of the University of Padua, which did not show potentially pathogenic variants in other desmosomal genes, were subjected to targeted-NGS for 55 genes associated with different hereditary cardiomyopathies.

3.6.2.1 Raw data analysis

150 bp paired-end sequencing of the libraries on the Illumina MiSeq produced on average 110.185.963 reads per sample, in table 3.15 are depicted the total number of reads obtained per sample and mapped bases.

Sample	Total Numer of Reads	Alligned Bases
1	103.336.675	8964.222
2	118.658.699	9171.513
3	119.677.954	9720.766
4	99.070.524	7425.109

Table 3.15: Total number of reads obtained per sample and mapped bases.

The total number of reads produced and the number of mapped reads are comparable among the samples. The mean depth of coverage over all samples was around 300X, in accordance with the theoretical coverage of 245X. The homogeneity of the distribution of reads between samples and the mean target coverage demonstrated the high performance of the sequence capture approach.

3.6.2.2 SNP Calling analysis

SNP calling performed on the sample from the Cardiovascular Genetics laboratory of Padua identified a total of 947 variants, and 197 mapped on the target panel regions. A first variant filtering performed by Cartagenia Bench Lab removed 2 variants with low quality that passed over the previous quality filters.

The resulting 195 variants were subjected to a series of filtering processes developed by the Cardiogenetics laboratory of Groningen, that take into consideration the presence and the frequency of the variants in several public databases (1000 Genomes Projects, EVS), and an in house variant database, radically reducing the amount of potentially pathogenic variants to 3 (table 3.16).

Gene	cDNA	Predicted Protein change	Ex	dbSNP ID	MAF	Type	Reference
TTN	c.102737 G>A	p.R34246H	358	rs372716177	T=0.0004/2	M	Millat et al., 2011
TTN	c.17279 C>T	p.T5760M	60			M	Novel
LAMA4	c.1552-4 G>A		int13	rs368746644	NA	Sp	Arbustini et al., 2002

Table 3.16: Output of the Cartagenia filtering pipeline, Ex:exon, M: missense, Sp: splicing

The three variants detected in the patient are located in extra desmosomal gene that have never been associated with AC (Table 3.16). 2 variants were located in *TTN* and one in the laminin-α-4 gene (*LAMA4*). The three novel variants were subsequently confirmed by Sanger sequencing.

3.7 Genotype Phenotype correlation

Family members of 9 AC index probands were available both for clinical and genetic testing.

3.7.1 Family #1

Patient #1 (male, aged 12) met a definite diagnosis of AC according to the revised 2010 Task Force criteria. Desmosomal genes analysis revealed a single nucleotide deletion c.1643del in exon 7 of *PKP2* predicted to shift the reading frame and to introduce a premature stop codon (p.G548VfsX15) (Figure 3.13). This variant has been described as pathogenic in different studies (Gerull et al., 2004; Dalal et al., 2006; Fressart et al., 2010; Bauce et al., 2011; Rigato et al., 2013; Alcalde et al., 2014), and it resulted absent in the 1000 Genomes Project and EVS databases. Analysis extended to available family members showed that this mutation was inherited by the probands' father (aged 55) and was also carried by his 14-year old sister both previously asymptomatic.

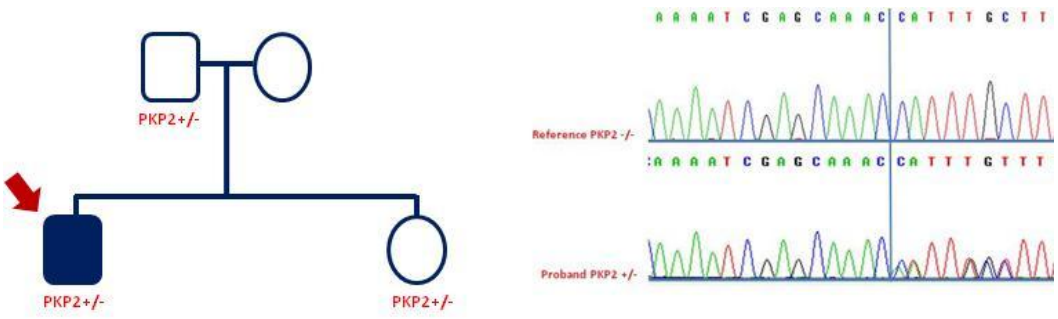

Figure 3.13: Pedigree of the proband (arrow) and her family for *PKP2* mutation p.G548VfsX15. Mutation status is indicated as follows: white = unaffected; black = affected; +/+ = homozygous for the aforementioned mutation; +/- = heterozygous for the aforementioned mutation.

3.7.2 Family #2

Patient #2 is a male of age 13 years affected by a classic form of AC who did not show family history of the disease. Screening of desmosomal genes identified the missense mutation c.536A>G, p.D179G in exon 5 of *DSC2* in homozygosis. The mutation involves a highly conserved amino acid located in the first cadherin domain, was predicted to be deleterious by SIFT and PolyPhen-2. It was absent in the 1000 Genomes Project and EVS databases, however it had been detected in

the healthy control population with an allele frequency of 2.7% (De Bortoli et al., 2010). The patient also carries a *PKP2* variant (c.1012A>G, p.T338A, rs139851304, MAF C= C=0.001/0), affecting a weakly conserved amino acid and predicted as tolerated by Polyphen-2 and SIFT.

The *DSC2* variants was carried in heterozygosis by his unaffected parents, while the *PKP2* polymorphism was carried by his father (Figure 3.14).

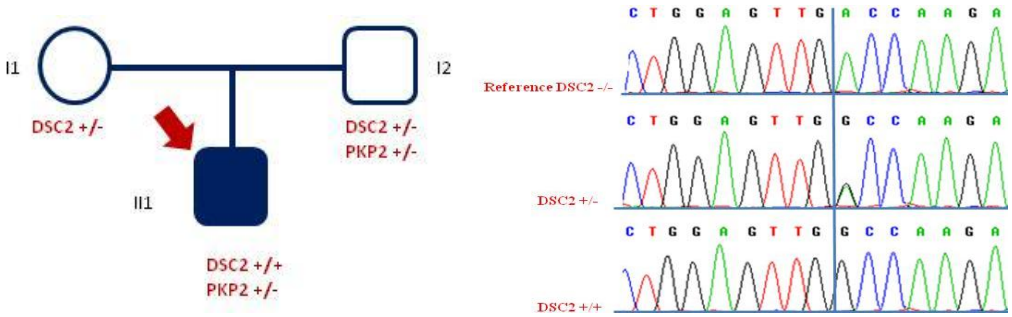

Figure 3.14: Pedigree of the proband (arrow) and his parents for *DSC2* mutation p.D179G. Mutation status is indicated as follows: white = unaffected; black = affected; +/+ = homozygous for the aforementioned mutation; +/- = heterozygous for the aforementioned mutation.

3.7.3 Family #4

Patient #4 (female, aged 56) was affected by a classic form of AC. She was found to carry the homozygous mutation p.D179G in exon 5 of *DSC2*. Cascade screening was performed on all available family members, who were clinically unaffected and resulted to carry the *DSC2* variants in heterozygosis (Figure 3.15).

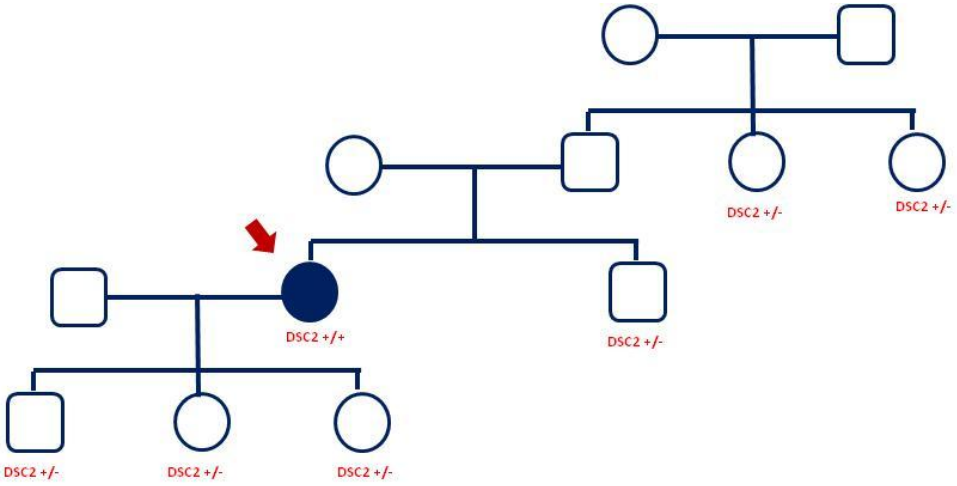

Figure 3.15: Pedigree of the proband (arrow) and her family for DSC2 mutation p.D179G. Mutation status is indicated as follows: white = unaffected; black = affected; +/+ = homozygous for the aforementioned mutation; +/- = heterozygous for the aforementioned mutation.

Data on families #2 and #4 support the presence of an autosomal recessive form of AC with a homozygous variant on *DSC2* as a cause of AC, without any sign of cardiocutaneous syndrome.

3.7.4 Family #8

Patient #8 (male, aged 15) was affected by a severe form of AC. Screening for desmosomal genes mutations by DHPLC and Sanger sequencing did not identifiy any pathogenic variants. Successive MLPA analysis detected the presence of a novel deletion on chr 18 including both *DSG2* and *DSC2* genes (Figure 3.16). This deletion was never observed in the 1000 control chromosomes and it was not detected in the 1000 Genomes Projects and EVS databases.

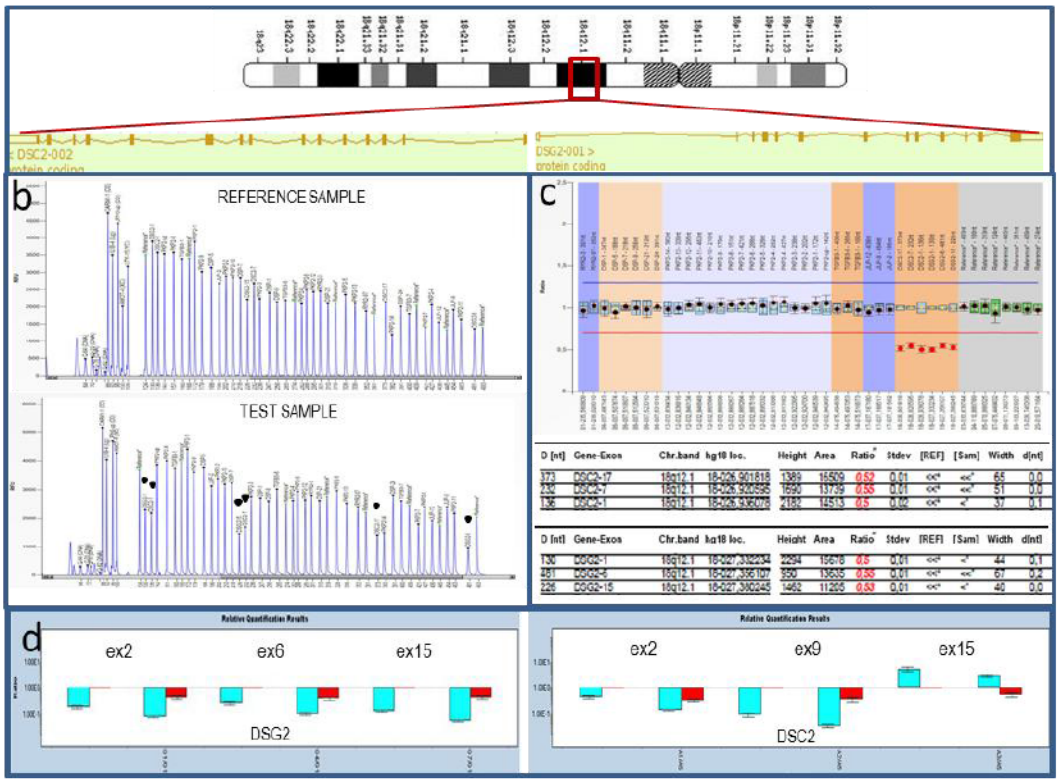

Figure 3.16: MLPA analysis of the probands' mother, indicating the presence of the deletion a) Schematic representation of the deleted chromosomal region b) Raw data of MLPA runs c) Histograms plotting the DQ of the fragments corresponding to every exon analyzed and layout of Coffalyzer results d) qPCR validation of MLPA results, in red the ratio between the test sample and the reference sample.

The genetic study was extended the probands' parents, showing that the deletion was carried by his mother, who also met a definite diagnosis of AC according to the revised 2010 Task Force criteria (Figure 3.17).

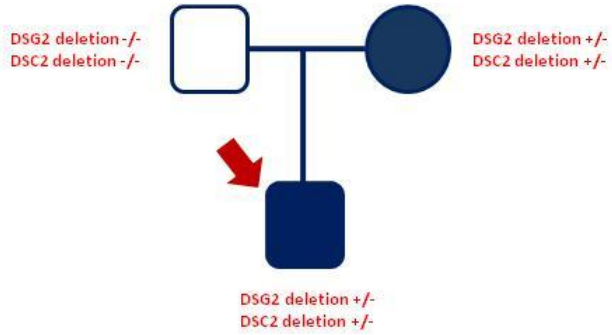

Figure 3.17: Pedigree of the proband (arrow) and his family for *DSC2* and *DSG2* deletion.

3.7.5 Family #9

The proband #9 (male, age 15) met a definite diagnosis of AC according to the revised 2010 Task Force criteria, and did not show family history of the disease. Genetic testing revealed the presence of 2 missense variants in two different genes. One variant involving a highly conserved amino acid in the N-terminal region of *DSG2* (c.875G>A, p.R292H), and one in *DSC2* (c.2194 T>G, p. L732V) (Figure 3.18). The *DSG2* variant were considered intolerant by Polyphen2 and SIFT, while the *DSC2* variant, also affecting a highly conserved amino acid, was predicted to be benign. Although both variants were excluded in the variants filtering process due to high allelic frequency (MAF> 0.02%), cascade screening on the proband's unaffected relatives showed that the mother was a carrier of the *DSC2* variant and the father carried the *DSG2* variant. These data suggest an autosomal recessive mode of transmission. This case is similar to another study recently describing the p.R292H mutation as pathogenic when present in the homozygous state (Jiménez-Jáimez et al., 2014).

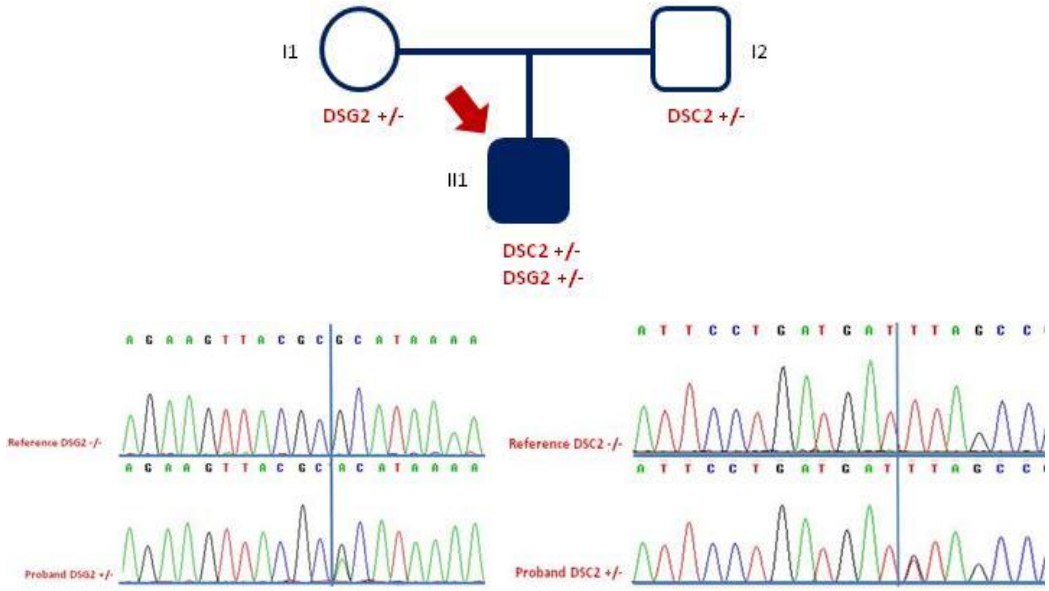

Figure 3.18: Pedigree of the proband #9 (arrow) and his parents for DSG2 mutation p.R292H and for DSC2 mutation p. L732V. Mutation status is indicated as follows: white = unaffected; black = affected; +/+ = homozygous for the aforementioned mutation; +/- = heterozygous for the aforementioned mutation.

3.7.6 Family #58

The proband #58 (male, aged 77) was affected by a biventricular form of AC and showed family history of the disease. DHPLC analysis followed by Sanger sequencing of abnormal elution profiles did not identify any mutation in desmosomal genes. Successive WES detected the heterozygous point mutation c.949G>A, p.E317K in the exon 6 of the *LMNA* gene (NM_170707.2) (Figure 3.19).

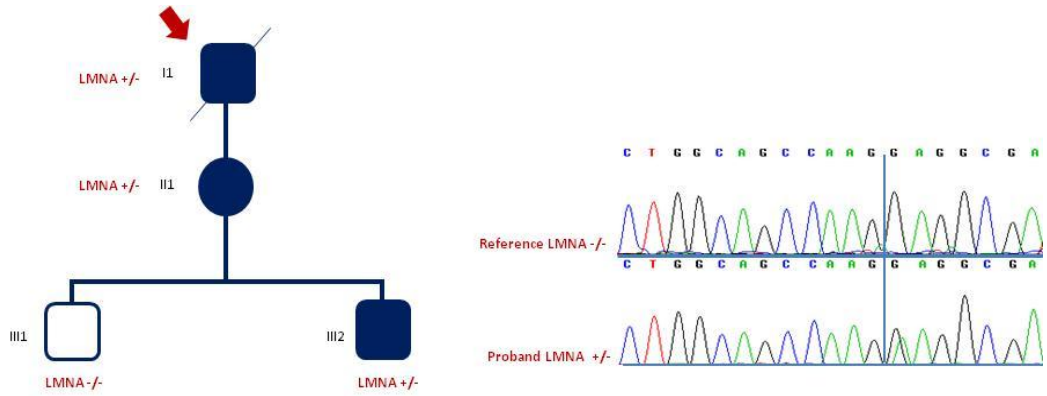

Figure 3.19: Pedigree of the proband (arrow) and his family for the *LMNA* mutation p.E317K.

The variant, located in the coil 2 rod domain, results in an amino acid substitution altering completely the physico-chemical properties of the amino acid. The mutation affects a highly conserved amino acid, and is predicted to be deleterious by Polyphen and SIFT.

This variant has been described in a case of familial autosomal dominant DCM with atrio ventricular block, in the proband the *LMNA* expression of the myocyte nuclei was reduced or absent (Arbustini et al., 2002). This variant is present in ClinVar database and classified as "likely pathogenic", it is also in dbSNP database (rs56816490) but not detected in the 1000 genomes or EVS dataset. The genetic study was extended to available family members. Three subjects, the probands daughter (aged 50) and one of their sons (aged 21), also clinically affected, resulted to carry the mutation. While the clinically unaffected son (aged 24) did not carry the mutation.

3.7.7 Family #59

The proband #59 was a 40 years-old competitive athlete who died suddenly, and received an autoptic diagnosis of AC. Sanger sequencing identified 2 heterozygous variants: one in *DSG2*, c.2137 G> A, p.E713K (rs79241126, MAF A=0.0260/130), and one in *DSC2*, c.2603 C>T, p.S868F (rs141873745, MAF A=0.0002/1).

Only the *DSC2* mutation is predicted as deleterious by Polyphen-2 and SIFT. Cascade screening showed that the brother and the sister were carriers for only one of the two variants.

WES confirmed the 2 variants and identified an additional variant c.68C>T, p.A23V located in the cardiac troponin T gene (*TNNT2*), whose alterations are known to significantly contribute to dilated cardiomyopathy rather than AC (Figure 3.20).

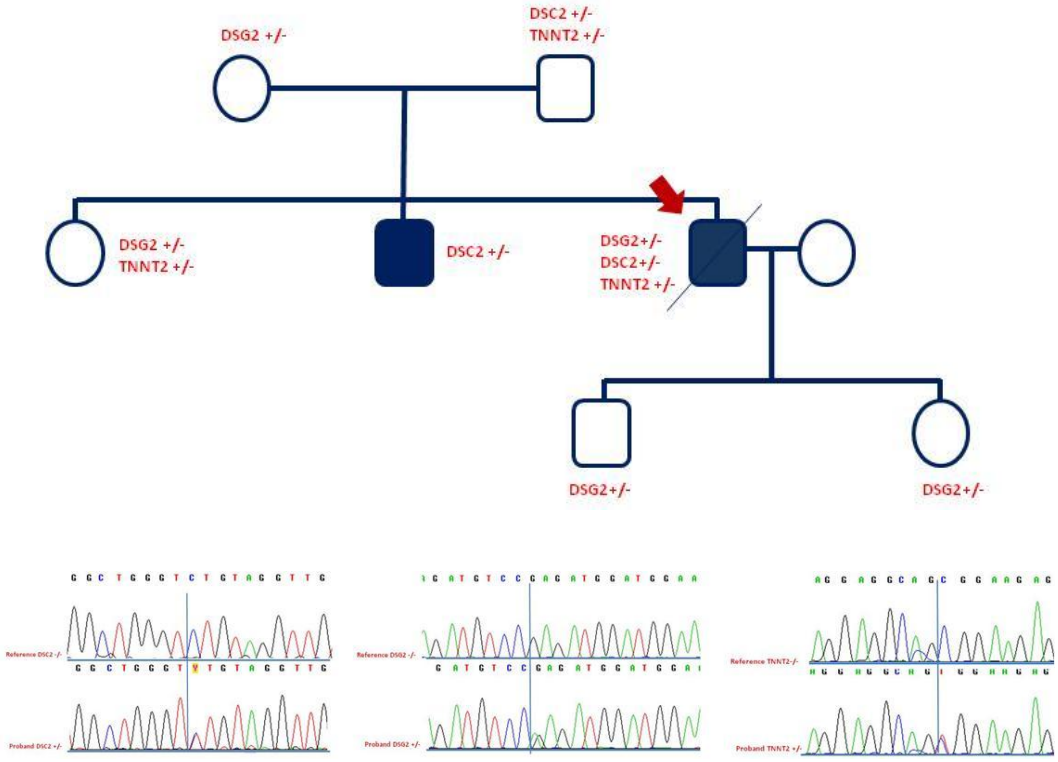

Figure 3.20: Pedigree of the proband #59 and his family for the *DSG2*, *DSC2*, *TNNT2* mutations.

3.7.8 Family #74

The proband #74 was a 13 years-old competitive athlete who died suddenly, and had an autoptic diagnosis of AC. DHPLC analysis followed by Sanger sequencing identified 2 heterozygous variants: one in *DSG2* and one in *PKP2*. The *DSG2* variant c.722 C>T, p.A241V was novel and was absent both in the 1000 alleles of the control population and in the 1000 Genomes and EVS dataset. Moreover it is located in the cadherin2 domain, it is predicted as deleterious by Polyphen and SIFT. The proband also carried the c.1576A>G, p.T526A *PKP2* variant, predicted as tolerated by *in silico* analysis tools and affecting a weakly conserved amino acid. Probands' 11-year old sister is a carrier of the only *DSG2* variant (Figure 3.21). No deeper co-segregation studies in the family were possible since the two children had foster parents.

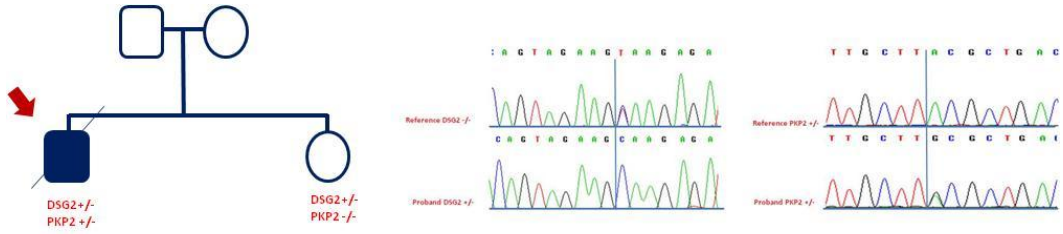

Figure 3.21: Pedigree of the proband #74 and his family for the *DSG2* and *PKP2* mutations.

3.7.9 Family #76

The proband #76 was a 20 years old male who died suddenly and the autopsy showed evidences of left-dominant AC.

Genetic analysis identified a novel heterozygous mutation (c.448C>T, p.R150X) in exon 4 leading to the premature truncation of *DSP*.

The same mutation was shared by the father (aged 54) and the sister (aged 19) (Figure 3.22). Interestingly, the father appeared to be the first mutation carrier in the family, suggesting a dominant de novo mutation (Pilichou et al., 2014).

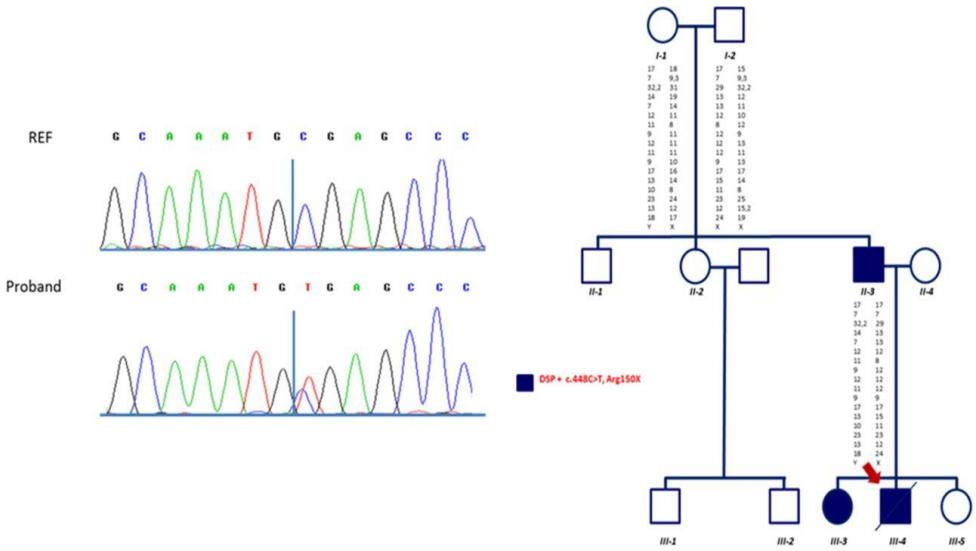

Figure 3.22: Pedigree of the proband (arrow) and his family for the *DSP* mutation p.R150X (from Pilichou et al., 2014).

3.7.10 Family #87

A fourteen year-old boy was transplanted at the Cardiology of Tehran and referred to the Laboratory of Cardiovascular Genetics for genetic screening of biventricular AC. Family history of sudden cardiac death was present in two brothers of the proband. ECG recording showed atrial fibrillation and conduction abnormalities (RBBB type), whereas Echo analysis showed biventricular enlargement. Targeted resequencing of 55 genes carried out in collaboration with the University of Groningen, identified 2 different missense variants in *TTN* (c.102737G>A, p.R34246H, and c.17279C>T, p.T5760M) and one variant in *LAMA4* (c.1552-4G>A), confirmed by Sanger sequencing (Figure 3.23). The variant c.1552-4G>A in *LAMA4* is predicted to affect the splicing process (MaxEnt: -5.3%; NNSPLICE: 0.7%; HSF: 0.1%). The *TTN* variant p.T5760M affects a weakly conserved amino acid, while the p.R34246H variant affects an highly conserved amino acid and is classified as a variant of uncertain significance in ClinVar (update Apr 19, 2012).

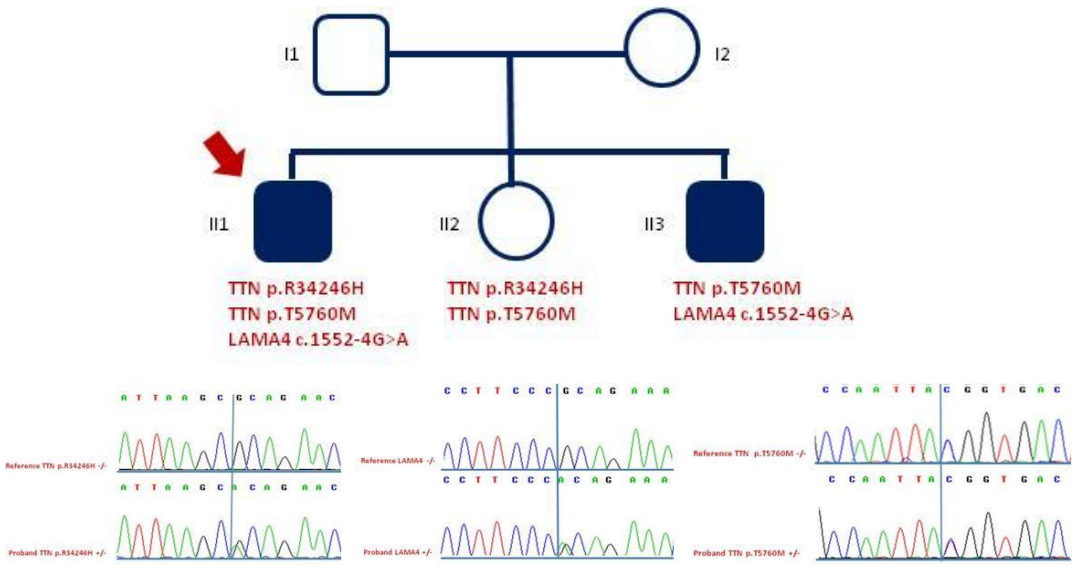

Figure 3.23: Pedigree of the proband (arrow) and his family for the LAMA4 and TTN mutations.

4. DISCUSSION

Arrhythmogenic Cardiomyopathy is an inherited heart muscle disease that may result in arrhythmias and sudden death, especially in the young and athletes (Thiene et al., 1988; Corrado et al., 1990). The main pathological feature of AC is the progressive loss of myocardial cells, which are replaced by fibrous and adipose tissue (Nava et al., 1988; Thiene et al., 1988). Major and minor criteria were established and revised by an international Task Force, to be specific for AC diagnosis (Marcus et al., 2010). AC criteria consider cardiac morphology and function, tissue characterization, electric rhythm conduction and the presence of arrhythmic events as well as family history, including the identification of pathogenic mutations as a major diagnostic criterion.

Although the first genetic sightings were accomplished on the recessive form known as Naxos disease, AC is nowadays most commonly considered as an autosomal dominant trait with variable expressivity and age-dependent penetrance (Nava et al., 1988; McKenna et al., 1994; Sen-Chowdhry et al., 2004; van Tintelen et al., 2006). Desmosomal point mutations are detected in nearly half of the AC probands (Basso et al., 2012; Kapplinger et al., 2011; den Haan et al., 2009). However, recent studies highlighted the presence of desmosomal gene variants in the general population, intensifying the debate regarding the pathogenic effects of missense and "radical" mutations. At present "radical" mutations are considered highly pathogenic due to abnormal protein length advocating a haploinsufficiency mechanism, whereas missense mutations are cautiously interpreted using complex bioinformatics algorithms. Moreover, the identification of compound and digenic heterozygous carriers, associated with a more severe form of the disease, increased the genetic complexity of the disease suggesting a recessive inheritance pattern (Xu et al., 2011; Bauce et al. 2010; Rigato et al., 2013). Further rare CNVs other that point mutations have been also associated with AC, by three independent studies describing different large deletions encompassing *PKP2* gene in Dutch and Italian AC patients (Cox et al., 2011; Roberts et al., 2013; Li Mura et al., 2013).

The aim of the present study was to prospectively assess the prevalence, age/gender relation and underlying genetic mutations in a large series of clinically affected and SCD AC cases.

4.1 Comprehensive genetic analysis in 99 index cases

Ninety-nine index cases referred for molecular genetic screening to the Cardiovascular Genetics Laboratory of Padua; 73/99 probands had a clinical diagnosis of AC based on the 2010 Task Force criteria (Marcus et al., 2010), of which 14 underwent cardiac transplantation, and 26 index cases were sudden death victims recorded in the Veneto Registry of Cardio-Cerebro-Vascular Pathology. Genetic screening carried out in the entire cohort for 5 desmosomal genes (*DSP, PKP2, DSG2, DSC2, JUP*), and in 46 of the 99 probands for 3 additional non desmosomal genes (*CTNNA3, DES, PLN*), identified 193 different nucleotide variants underlying the complex genetic status of the disease. Of note, the majority of the nucleotide variants resided in *DSP*.

The first genetic analysis excluded all intronic, UTR and synonymous variants reducing the total number of variants to 81. The 81 detected variants were prioritized based on the frequency of variants in the general population, the conservation of the nucleotide and the amino acid conservation among species, as well as other *in silico* prediction analytical methods. Finally a nucleotide variant was classified as pathogenic once it was predicted to alter the structure or the function of the encoded protein, its presence was assessed in a large healthy ethnically matched control population and when literature or family information implied correlation with the disease.

The second variant filtering was based on frequency data provided by available public databases such as 1000 Genomes Project and Exome Variant Server and an internal control group composed of 500 unrelated healthy athletes who were screened as controls to assess frequency of novel variants in the Italian population. Minor allele frequency cut-off was established at <0.02% given the prevalence of this rare disease as 1:5000 (0.0002, 0.02%). Twenty-nine variants above this threshold were excluded due to high frequency in the general population whereas 52 unique variants advanced to the next variant filtering step. However one *DSP* variant excluded because of frequency beyond the established threshold, was

reclassified as a mutation due to *in vitro* function evidences proving its association with the disease (Yang et al., 2006). Of note, all 53 variants were located in desmosomal encoding genes: *PKP2* (19, 36%), followed by *DSP* (19, 36%), *DSG2* (8, 15%), *DSC2* (3, 6%), *JUP* (4, 7%), in agreement with previous reported data considering AC as a desmosomal disease (Basso et al., 2012). No nucleotide variants were found in our cohort in extra desmosomal genes. Indeed extra desmosomal genes mutations in AC are rare and usually reported as founder mutations (p.S358L in *TMEM43*, p.S13F in *DES*, p.R14del in *PLN* ((Merner et al. 2008; Bergman et al., 2007; van der Zwaag et al., 2012). Fifty-four positive-mutation carriers were identified from this filtering step, of which 11% were single mutation carriers for *DSP*, 18% for *PKP2*, 4% for *DSG2* and 3 % for *DSC2* and *JUP* respectively. Fifteen of these 54 genotype-positive AC cases were multiple mutation carriers (15%).

The following variant filter was based on evolutionary conservation of amino acids and led to the exclusion of 8 more missense variants, weakly or not conserved among species, reducing the amount of putative pathogenic variants to 44.

The resulting variants from the evolutionary conservation- filtering step were underwent a final *in silico* evaluation regarding predicted variant pathogenicity, which excluded from the study 4 more variants classified as VUS1 (unlikely to be pathogenic) and 1 classified as VUS0 (not pathogenic) and reduced the putative pathogenic variants to 37. A further reduction in 42 index AC cases (42%) who carried one or more of these 37 variants were finally considered as positively genotyped.

A parallel CNVs analysis, carried out both by MLPA and qPCR in all 99 probands identified 3 additional large genomic rearrangements involving desmosomal genes, that could not have been detected by DHPLC and Sanger sequencing: a heterozygous deletion of 120 kb located on chr 12 comprising the entire *PKP2* gene in 2 patients, a heterozygous deletion of at least 482 kb on chr 18 encompassing both *DSC2* and *DSG2* genes in 1 patient, and a heterozygous deletion involving only exon 4 of *PKP2* in other 2 patients. CNVs investigation allowed the identification of an additional 4% of desmosomal mutation carriers, increasing the diagnostic yield of AC genetic testing from 42% to 46% in our

population. We considered as mutation carriers all 46 probands (46%) who carried one or more of the 37 identified point variants or of the 3 large deletions.

4.2 Genotype complexity in AC

Forty-six (46%) index AC cases harbored at least one desmosome point mutation or a large deletion in desmosomal encoding genes; 20 of the 46 genotype positive AC probands were single *PKP2* mutation carriers (43%), 11 single *DSP* mutation carriers, 4 single *DSG2* mutation carriers and 3 were single *DSC2* mutation carriers. Note that 3 homozygous carriers have been identified for 1 nonsense *DSG2* mutation and 2 *DSC2* missense mutations. Moreover, 8 of 46 genotype positive AC probands carried multiple mutations and most of them were digenic heterozygotes (7 of 8, 88%).

4.3 Recessive inheritance pattern in AC

Cascade genetic family screening has the power to unveil clinical features and improve characterization of the detected variants. However the low number of relatives available for clinical evaluation or the small size of the families often made difficult a final assessment of the role and/or the contribution of the different mutations identified. Herein, we report our findings in 3 AC families supporting the idea that AC might be present as a recessive form without associated palmoplantar keratoderma and wooly hair, in concordance to previous reported studies (Alcalde et al., 2014; Awad et al., 2006; Sato et al., 2011; Rasmussen et al., 2013).

In two families we detected the presence of the same homozygous founder *DSC2* mutation (p.D179G) (Lorenzon et al., submitted), while in another family the only affected proband was found to carry a *DSC2* mutation (p.L732V) associated with a *DSG2* mutation (p.R292H). These data together with a high incidence of digenic heterozygotes (8 in 99, 8%) in our cohort, suggest an autosomal recessive mode of transmission, confirming the complexity of AC genetics and highlighting the need of cascade clinical and genetic screening in larger populations in order to better understand the genetic basis of the disease. We cannot exclude the possibility that AC cases hosting 1 AC-associated mutation do not carry a compound mutation in yet unknown AC susceptibility gene.

4.4 Prediction of missense mutations pathogenicity

While the pathogenicity prediction value of features such as the frequency of the altered allele in the population and the phylogenetic conservation of the amino acids is well established, the utility of *in silico* analysis remains questionable, even though it represents the first tool enabling the prioritization of the increasing amount of VUSs emerging from the new sequencing technologies. *In silico* analysis are based on different statistical algorithms combining the outcomes of series of predictors that often consider MAF, conservation data, predictions about the possible impact of an amino acid substitution on the structure and function of a protein, the difference in the physicochemical properties of the amino acids, while they do not consider for instance any possible interaction of the protein inside the cell. Herein we identified 53 desmosomal gene variants after MAF-Filtering in 54 positive-mutation carriers, of which 11% *DSP* single mutation carriers, 18% *PKP2* single mutation carriers, 4% *DSG2* single mutation carriers, 3 % *DSC2* single mutation carriers, 3% *JUP* single mutation carriers and 15% multiple mutation carriers. Stringent analysis, limited only to carriers of variants resulted from the four filtering steps reduced the number of variants to consider pathogenic to 37, the number of mutation carriers to 42 index cases of which 11% *DSP* single mutation carriers, 20% *PKP2* single mutation carriers, 4% *DSG2* single mutation carriers, 3 % for *DSC2* single mutation carriers and 8% multiple mutation carriers. These results showed a significant reduction in multiple mutation carriers and that *JUP* variants alone cannot cause the development of the disease. However, the exclusion of variants during the 4-filtering procedure due to statistical evaluation is still complicated since different algorithms can sometimes provide discordant outputs, suggesting the need of more than one *in silico* tool comparison and validation of predictions by co-segregation analysis and/or functional studies to finally verify the real pathogenicity of mutations. This raise the question whether exclusion of genetic variants during the 4-filtering fails to detect a large portion of mutation carriers.

4.5 Prediction of "radical" mutations pathogenicity

PKP2 mutations are the most common AC-related genetic variants, representing 13/37=35% of the total AC variants identified in this study and responsible for 20/46=43% of our genotype positive AC cohort.

Eleven of the 15 *PKP2* mutations identified in 16 AC index cases were "radical" mutation (frameshift, nonsense, splice sites). The pathogenicity of missense and "radical" mutations in cardiac diseases and specifically in AC is a matter of intense debate. Recent studies have shown that stop-codon mutations in *PKP2* are more pathogenic because they alter protein length and that truncating PKP2 proteins may lead to haploinsufficiency due to protein instability (Kapplinger et al., 2011; Andreasen et al., 2013; Joshi-Mukherjee et al., 2008; Herman et al., 2012; Rasmussen et al., 2014). Even though this study showed a clear-cut prevalence of "radical" *PKP2* mutations compared to missense mutation, we found no difference in the age onset (41±19 vs 40±19) or in a higher risk of SCD events in *PKP2* mutation carriers.

Actually our data supports the idea that a heterozygous gene deletion might be insufficient to determine alone the disease. Specifically, 3 *PKP2* "radical" mutation carriers were identified and clinically classified as possible based on 2010 AC criteria; two of these subjects (about 40y old, males) presented a *PKP2* frameshift [p.G458VfsX15, p.T50SfsX61 respectively] showing a LD pattern of the disease and 1 (44y-old female) displayed a complete deletion of the *PKP2* containing allele in heterozygosis with absence of any clinical signs of the disease. In contrast we observed that another AC carrier of the entire *PKP2* allele deletion and of an additional *DSP* missense mutation (p.V30M) fulfilled definite diagnostic criteria, as observed also in all the compound/digenic mutation carriers in this study. Further a young boy with the deletion of the allele containing both *DSG2* and *DSC2* genes was also classified as definite. Thus, in agreement with previous studies the pathogenicity of "radical" mutations as well as missense mutations should be accomplished only by co-segregation studies and clinical analysis in the families.

4.6 Genotype-phenotype correlation

Previous studies on AC index cases consistently demonstrated a male predominance, suggesting that men develop a more severe disease phenotype, making them more likely to come to medical attention (Cox et al., 2011; Dalal et al., 2006; Nava et al., 2000; Rigato et al., 2013). These authors hypothesized a direct sex-hormone influence on the disease pathobiology and development of arrhythmogenic myocardial substrate, advocating the protective effect of estrogens to explain why females with desmosomal gene mutations tend to develop a less severe phenotype and fewer arrhythmic complications compared with males.

Our cohort included 77 (78%) males and only 22 (22%) females presenting a 4:1 ratio, of which 26 SCD cases were all males. Excluding this bias from our cohort of SCD males, the sex ratio males:females in the clinically diagnosed group was 3:1, as previously reported elsewhere.

Although our findings confirmed a male predominance with a ratio 3:1 among AC index cases, a 2:1 male to female ratio was observed among desmosomal mutation gene carriers with definite criteria suggesting that genotype positive females have the same probability of developing specific AC features and the fact that these are not translated to SCD events needs further investigation.

Comparing mutation carriers and non mutation carriers fulfilling the established AC diagnostic criteria, 46 were found positively genotyped (46%): 30 classified as definite, 1 classified as borderline and 2 classified as possible. Interestingly, 33 of the 40 non-mutation carriers were also classified as definite.

Sorting genotyped positive patients for the different desmosomal gene mutations we observed no significant differences in the average age of the initial presentation (around 38y) of the disease. Although multiple mutation carriers and *DSP* single mutation carriers showed definite diagnostic criteria for AC, no significant correlations were also observed in terms of clinical characteristics and in the average age at diagnosis between the different single mutation carriers and the multiple mutation carriers. These data do not reflect previous studies reporting that multiple mutation carriers correlation with an earlier onset of symptoms and a more severe disease (Xu et al., 2010). However our results might be biased due to high strigency in variants filtering which relocated multiple mutation carriers in

the group of single mutation carriers (7 multple mutation carriers after 4-step filtering became single mutation carriers), and to the small number of multiple mutation carriers.

Separating clinically affected living subjects from SCD cases made evident a higher risk of SCD (5 of 11, 45%) among *DSP* mutation carriers at a significantly younger age (p 0.04). In contrast given the higher incidence of *PKP2* mutations among AC probands (20, 44%), the risk of SCD in *PKP2* mutation carriers is lower (4 of the 20, 20%).

4.7 Next generation sequencing in AC

4.7.1 Whole exome sequencing and troubleshooting

Twelve of the 99 genotyped patients underwent paired-end WES in 2 successive runs, obtaining 68.270.020 reads per sample on average. To elaborate these data we used 3 different pipelines: we used CLC software as the main mapping tool due to its ability to recognize faster the highest number of unique hits and it did not require indexing of the reference. However a huge variability was observed in the number of reads aligned against the human genome reference and the number of reads on target for each sample within every sequencing run. Mean coverage variability was also observed among different samples (inter-variability) and within the same sample (intra-variability) in terms of well covered regions versus uncovered or low covered regions of interest.

Accordingly, we obtained in nearly 80% of the exome a mean coverage above 20X only for 6 of the 12 samples sequenced by WES; mean coverage of 20X is considered the minimal coverage value for reliable variant call, showing that sequencing did not produce enough reads (Sims et al., 2014).

Inter- and intra- variability observed among samples, even though the sample procedure was applied, depended mainly on the quality of the original sample. Many of our samples were obtained from post mortem sampling resulting in tissue and DNA degradation which was translated in this process as poor enrichment of target regions. We tried to overcome this difficulty by quality control of the starting DNA sample and libraries, however when the samples are pooled together calculation and pipetting cannot be perfectly accurate and this result in unbalanced number of reads produced for each sample during multiplex sequencing. These

technical problems create an enormous effect on the yield and quality of sequence reads obtained by NGS.

Next, we used 3 pipelines also for SNP calling comparing them in terms of sensitivity and specificity of the calls. In 6 samples, GATK and CLC called respectively a total of 1.975.223 and 1.435.455 variants, while CASAVA called only about 1/6 of them. In fact CASAVA automatically calls variants located exclusively in the exonic target regions that reached an appropriate coverage >20X, thus not focusing on the low covered exonic regions or the intronic regions. CLC software instead was demonstrated to be user-friendly and flexibile, able to call a great amount of variants even in non coding regions.

4.7.2 Whole exome sequencing: sensitivity and sensibility

In the clinical setting a higher specificity and sensitivity of diagnostic tools is required (Gargis et al., 2012; Rehm et al., 2013). To this regard we compared WES data to the ones deriving from conventional sequencing.

SNP calling by CLC software, automatically detected 68 out of the 90 variants (76%), missing 22 nucleotide changes, that were manually called looking through the alignment data. Thirteen of these 22 missing variants were manually detected, exhibiting 90% sensitivity and >99.9% specificity of WES. The rest 9 variants missed by automatic and manual SNP calling, 6 were intronic variants located from 44 to 356 bp to the nearest probe and thus of difficult detection, while 3 exonic missense variants were located in target uncovered regions or with a coverage below 5X. Note that, CLC software automatically discards all calls with low quality parameters.

4.7.3 Whole exome sequencing: clinical application and efficacy

WES data of the 12 sample were further interrogated for variants located in 150 additional genes associated with different from AC inherited cardiomyopathies. The pipeline used for variants filtering was the same as previously described in conventional sequencing, resulting in the identification of more than 100 putative pathogenic variants. However, conventional sequencing methods confirmed only 3 of the 100 extra desmosomal variants detected in 3 different patients, of which 1 in *TNNT2*, 1 in *MYBPC3*, and 1 in *LMNA* genes respectively. In this way 3 more patients were positively genotyped allowing a differential diagnosis with respect

to AC. For instance, *LMNA* mutation carrier was an AC-labeled proband who died shortly after cardiac transplantation. Cascade family screening enabled the study of his family identified 2 more mutation carriers who exhibited a DCM-like pattern. Indeed this variant has been previously described in a case of familial autosomal dominant DCM with atrio ventricular block (Arbustini et al., 2002).

The rest 97 false positive WES calls were related not only to a low sequence coverage (<10x in 2 samples) but also to systematic errors due to repetitive elements, homopolymer stretches and small insertions or deletions resulting in an aberrant alignment process. The presence of false positives implies that this pilot WES experiments requires improvements both at the enrichment level and the set-up of the bioinformatic pipeline, and highlights the value of confirmatory Sanger sequencing especially at this stage of the progress of NGS technology.

Although, our NGS procedure still presents some limitations to overcome, the data obtained from these two pilot WES experiments demonstrated the utility of sequencing simultaneously the entire exome of every patient since a higher throughput of data can be analyzed at a lower cost with respect to conventional sequencing methods and WES data can be stored and re-analyzed after time.

4.7.4 Targeted resequencing vs WES strategy

Last, we evaluated another NGS strategy (Targeted Resequencing-TR) in order to extend out analysis from 5 desmosomal genes to a specific set of 55 inherited cardiomyopathy-related genes. Targeted resequencing panel was previously validated by the Cardiogenetics Laboratory of the University of Groningen and at present is used as a routine diagnostic test.

Four AC patients from the Cardiogenetics Laboratory of Groningen and 1 negative-genotyped AC patient from the Cardiovascular Laboratory of Padua, were sequenced with a paired-end strategy producing on average 110.185.963 reads per sample to be mapped on 55 genes, corresponding to approximately 1.6 fold the reads produced by WES which were expected to map all over the exome. No intra- or inter- variability was observed in terms of total number of reads produced and mapped, while the mean coverage was around 300X, confirming the good quality of the enrichment. Variant calling and filtering was performed by an in-house pipeline and led to the detection of 3 variants in 2 extra desmosomal

genes (p.R34246H and p.T5760M in *TTN*; c.1552-4G>A in *LAMA4*) in the patient coming from Padua.

This work highlighted the need for improvement of the WES protocol and allowed a differential diagnosis of Dilated Cardiomyopathy based on genetics evidences.

5. CONCLUSIONS

Comprehensive mutation screening and filtering in a large cohort of unrelated consecutive index AC cases identified 37 putative pathogenic mutations exclusively in desmosomal-related genes in 46% of cases.

In order to provide a more stringent analysis, we limited our study to the only variant carriers resulting from the 4-filtering steps. The characterization of identified variants as pathogenic mutations or benign variants is essential for the understanding of their role in the disease, and this process is even more critical with the use of genetics as a diagnostic criterion. According to the results of the present study, compound/digenic heterozygosity (8%) and homozygosity (1%) in AC-desmosomal gene mutations suggest a different more complex mode of inheritance. This highlights the importance of screening the entire panel of desmosomal genes, even after a single mutation has been identified, and have significant implications for genetic counseling.

Further CNVs investigation allowed the identification of an additional 4% of CNVs carriers, increasing the diagnostic yield of AC genetic testing to 46% in our population. These data emphasize the diagnostic impact of performing additional CNVs analysis in AC genetic testing and the importance of comprehensively analyze the entire AC cohort.

Radical mutations in our cohort showed a lower risk of SCD event, suggesting that pathogenicity should be assessed carefully by studying the family co-segregation with the disease as well.

Clinical and familiar correlation is essential to elucidate the role of both genetic point/radical variants and CNVs, for genetic counselling and management.

Finally, our pilot NGS experiments highlighted the potential of WES and Targeted resequencing in the clinical setting in terms of time, cost and effectiveness, by enabling a differential diagnosis in 4 probands with extra desmosomal mutations. However the application of NGS to the molecular diagnosis of AC raises new challenges in distinguishing pathogenic from non pathogenic vairiants and overall in the patients' management and treatment.

6. REFERENCES

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061-1073.

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012;30:224-6.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; 7:7.20.

Affymetrix ENCODE Transcriptome Project; Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature. 2009; 457:1028-32.

Agrawal N , Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Treviño L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.Science 2011; 333: 1154-1155.

Alcalde M, Campuzano O, Berne P, Garcia-Pavia P, Doltra A, Arbelo E, Sarquella-Brugada G, Iglesias A, Alonso-Pulpon L, Brugada J Brugada R. Stop-gain mutations in PKP2 are associated with a later age of onset of arrhythmogenic right ventricular cardiomyopathy. PLoS ONE 2014; 9:e100560.

Alcalde M, Campuzano O, Sarquella-Brugada G, Arbelo E, Allegue C, Partemi S, Iglesias A, Oliva A, Brugada J, Brugada R. Clinical interpretation of genetic variants in arrhythmogenic right ventricular cardiomyopathy. Clin Res Cardiol. 2014 [Epub ahead of print]

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Müller F; FANTOM Consortium, Forrest AR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507:455-61.

Andreasen C, Nielsen JB, Refsgaard L, Holst AG, Christensen AH, Andreasen L, Sajadieh A, Haunsø S, Svendsen JH, Olesen MS.New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur J Hum Genet. 2013; 21:918-928.

Ankala A, da Silva C, Gualandi F, Ferlini A, Bean LJ, Collins C, Tanner AK, Hegde MR. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. Ann Neurol. 2014. [Epub ahead of print]

Arbustini E, Pilotto A, Repetto A, Grasso M, Negri A, Diegoli M, Campana C, Scelsi L, Baldini E, Gavazzi A, Tavazzi L. Autosomal dominant dilated cardiomyopathy with atrioventricular block: a lamin A/C defect-related disease. J Am Coll Cardiol. 2002; 39:981-990.

Armstrong DK, McKenna KE, Purkis PE, Green KJ, Eady RA, Leigh IM, and Hughes AE. Haploinsufficiency of desmoplakin causes a striate subtype of palmoplantar keratoderma. Hum Mol Genet 1999; 8:143-148.

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; 81:964-973.

Awad MM, Dalal D, Cho E, Amat-Alarcon N, James C, Tichnell C, Tucker A, Russell SD, Bluemke DA, Dietz HC, Calkins H, Judge DP. DSG2 Mutations Contribute to Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy. Am J Hum Genet 2006, 79:136-142.

Awad MM, Dalal D, Tichnell C, James C, Tucker A, Abraham T, Spevak PJ, Calkins H, Judge DP. Recessive arrhythmogenic right ventricular dysplasia due to novel cryptic splice mutation in PKP2. Hum Mutat. 2006; 27.1157.

Baev V, Milev I, Naydenov M, Vachev T, Apostolova E, Mehterov N, Gozmanva M, Minkov G, Sablok G, Yahubyan G. Insight into small RNA abundance and expression in high- and low-temperature stress response using deep sequencing in Arabidopsis. Plant Physiol Biochem. 2014; 84:105-14.

Bagnall RD, Molloy LK, Kalman JM, Semsarian C. Exome sequencing identifies a mutation in the ACTN2 gene in a family with idiopathic ventricular fibrillation, left ventricular noncompaction, and sudden death. BMC Med Genet. 2014; 15:99.

Bahareh Rabbani, Nejat Mahdieh, Kazuyoshi Hosomichi, Hirofumi Nakaoka and Ituro Inoue. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders, Journal of Human Genetics, 2012; 57:621-632.

Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, Angrist M, Bhak J, Bobe J, Callow MJ, Cano C, Chou MF, Chung WK, Douglas SM, Estep PW, Gore A, Hulick P, Labarga A, Lee JH, Lunshof JE, Kim BC, Kim JI, Li Z, Murray MF, Nilsen GB, Peters BA, Raman AM, Rienhoff HY, Robasky K, Wheeler MT, Vandewege W, Vorhaus DB, Yang JL, Yang L, Aach J, Ashley EA, Drmanac R, Kim SJ, Li JB, Peshkin L, Seidman CE, Seo JS, Zhang K, Rehm HL, Church GM. A public resource facilitating clinical use of genomes. Proc Natl Acad Sci U S A. 2012; 109:11920-7.

Bhaijee F, Pepper DJ, Pitman KT, Bell D.New developments in the molecular pathogenesis of head and neck tumors: a review of tumor-specific fusion oncogenes in mucoepidermoid carcinoma, adenoid cystic carcinoma, and NUT midline carcinoma. Ann Diagn Pathol. 2011;15:69-77.

Barahona-Dussault C, Benito B, Campuzano O, Iglesias A, Leung TL, Robb L, Talajic M, Brugada R Role of genetic testing in arrhythmogenic right ventricular cardiomyopathy/dysplasia. Clin Genet 2010; 77:37–48.

Basso C, Bauce B, Corrado D, Thiene G. Pathophysiology of arrhythmogenic cardiomyopathy. Nature Reviews Cardiology, 2011; 9:223–233.

Basso C, Corrado D, Bauce B, Thiene G. Arrhythmogenic right ventricular cardiomyopathy. Circ Arrhythm Electrophysiol. 2012; 5:1233-1246.

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

Basso C, Wichter T, Danieli GA, Corrado D, Czarnowska E, Fontaine G, McKenna WJ, Nava A, Protonotarios N, Antoniades L, Wlodarska K, D'Alessi F, Thiene G. Arrhythmogenic right ventricular cardiomyopathy: clinical registry and database, evaluation of therapies, pathology registry, DNA banking. Eur Heart J 2004, 25:531-534.

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

Beckmann J, SE, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007; 8:639-646.

Beffagna G, De Bortoli M, Nava A, Salamon M, Lorenzon A, Zaccolo M, Mancuso L, Sigalotti L, Bauce B, Occhi G, Basso C, Lanfranchi G, Towbin JA, Thiene G, Danieli GA, Rampazzo A. 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.

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

Bentley DR. Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 2006;16:545–552.

Bergman JE, Veenstra-Knol HE, van Essen AJ, van Ravenswaaij CM, den Dunnen WF, van den Wijngaard A, van Tintelen JP. Two related Dutch families with a clinically variable presentation of cardioskeletal myopathy caused by a novel S13F mutation in the desmin gene. Eur J Med Genet. 2007; 50:355-66.

Bisgaard AM, Rackauskaite G, Thelle T, Kirchhoff M, Bryndorf T. Twins with mental retardation and an interstitial deletion 7q34q36.2 leading to the diagnosis of long QT syndrome. Am J Med Genet A. 2006; 140:644-648.

Biswas A, Rao VR, Seth S, Maulik SK. Next generation sequencing in cardiomyopathy: towards personalized genomics and medicine. Mol Biol Rep. 2014; 41:4881-4888.

Blue GM, Kirk EP, Giannoulatou E, Dunwoodie SL, Ho JW, Hilton DC, White SM, Sholler GF, Harvey RP, Winlaw DS. Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. J Am Coll Cardiol. 2014; 64:2498-506.

Bomar L, Maltz M, Colston S, Graf J. Directed culturing of microorganisms using metatranscriptomics. MBio. 2011; 2:e00012-11.

Bormann Chung CA, Boyd VL, McKernan KJ, Fu Y, Monighetti C, Peckham HE, Barker M.Whole methylome analysis by ultra-deep sequencing using two-base encoding. PLoS One. 2010;5:e9320.

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: 892-895.

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

Campuzano O, Allegue C, Sarquella-Brugada G, Coll M, Mates J, Alcalde M, Ferrer-Costa C, Iglesias A, Brugada J, Brugada R. The role of clinical, genetic and segregation evaluation in sudden infant death. Forensic Sci Int. 2014; 242:9-15.

Campuzano O, Sanchez-Molero O, Allegue C, Coll M, Mademont-Soler I, Selga E, Ferrer-Costa C, Mates J, Iglesias A, Sarquella-Brugada G, Cesar S, Brugada J, Castellà J, Medallo J, Brugada R. Post-mortem genetic analysis in juvenile cases of sudden cardiac death. Forensic Sci Int. 2014; 245C:30-37.

Cantsilieris S, Baird PN, White SJ. Molecular methods for genotyping complex copy number polymorphisms. Genomics. 2013; 101:86-93.

Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat. Genet. 2007; 39:S16–S2.

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet. 2011; 43:956-63.

Chen SN, Gurha P, Lombardi R, Ruggiero A, Willerson JT, Marian AJ. The hippo pathway is activated and is a causal mechanism for adipogenesis in arrhythmogenic cardiomyopathy. Circ Res. 2014; 114:454-468.

Chen X, Bonne S, Hatzfeld M, van Roy F, and 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; 277:10512-10522.

Chen W, Cheng YM, Zhang SW, Pan Q2.Supervised method for periodontitis phenotypes prediction based on microbial composition using 16S rRNA sequences. Int J Comput Biol Drug Des. 2014;7:214-24.

Chidgey M, Brakebusch C, Gustafsson E, Cruchley A, Hail C, Kirk S, Merritt A, North A, Tselepis C, Hewitt J, Byrne C, Fassler R, Garrod DJ. Mice lacking desmocollin 1 show epidermal fragility accompanied by barrier defects and abnormal differentiation. Cell Biol. 2001; 155:821-832.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009; 106:19096-101.

Choquet K, La Piana R, Brais B. A novel frameshift mutation in FGF14 causes an autosomal dominant episodic ataxia. Neurogenetics. 2015. [Epub ahead of print]
Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 2008; 5: 613–619.

Coonar AS, Protonotarios N, Tsatsopoulou A, Needham EW, Houlston RS, Cliff S, Otter MI, Murday VA, Mattu RK, McKenna WJ. Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21. Circulation. 1998; 97:2049-2058.

Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Arch Pathol Lab Med. 2013; 137:415-433.

Corrado D, Fontaine G, Marcus FI, McKenna WJ, Nava A, Thiene G, and Wichter T. Arrhythmogenic right ventricular dysplasia/cardiomyopathy: need for an international registry. Study Group on Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy of the Working Groups on Myocardial and Pericardial Disease and Arrhythmias of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the World Heart Federation. Circulation 2000; 101:E101-106.

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; 89:588-96.

Cox MG, van der Zwaag PA, van der Werf C, van der Smagt JJ, Noorman M, Bhuiyan ZA, Wiesfeld AC, Volders PG, van Langen IM, Atsma DE, Dooijes D, van den Wijngaard A, Houweling AC, Jongbloed JD, Jordaens L, Cramer MJ, Doevendans PA, de Bakker JM, Wilde AA, van Tintelen JP, Hauer RN. Arrhythmogenic right ventricular dysplasia/cardiomyopathy: pathogenic desmosome mutations in index-patients predict outcome of family screening: Dutch arrhythmogenic right ventricular dysplasia/cardiomyopathy genotype-phenotype follow-up study. Circulation. 2011; 123:2690-2700.

Crotti L, Johnson CN, Graf E, De Ferrari GM, Cuneo BF, Ovadia M, Papagiannis J, Feldkamp MD, Rathi SG, Kunic JD, Pedrazzini M, Wieland T, Lichtner P, Beckmann BM, Clark T, Shaffer C, Benson DW, Kääb S, Meitinger T, Strom TM, Chazin WJ, Schwartz PJ, George AL Jr. Calmodulin mutations associated with recurrent cardiac arrest in infants. Circulation. 2013; 127:1009-1117.

D'Amati G, di Gioia CR, Giordano C, Gallo P. Myocyte transdifferentiation: a possible pathogenetic mechanism for arrhythmogenic right ventricular cardiomyopathy. Arch Pathol Lab Med 2000; 124: 287–290.

Dalal D, Tandri H, Judge DP, Amat N, Macedo R, Jain R, Tichnell C, Daly A, James C, Russell SD, Abraham T, Bluemke DA, Calkins H. Morphologic variants of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy a genetics-magnetic resonance imaging correlation study. J Am Coll Cardiol. 2009; 53:1289-1299.

Daliento L, Turrini P, Nava A, Rizzoli G, Angelini A, Buja G, Scognamiglio R, Thiene G. Arrhythmogenic right ventricular cardiomyopathy in young versus adult patients: similarities and differences. J Am Coll Cardiol. 1995; 25:655-664.

Dames S, Chou LS, Xiao Y, Wayman T, Stocks J, Singleton M, Eilbeck K, Mao R. The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. J Mol Diagn. 2013; 15:526-534.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156-2158.

De Bortoli M, Beffagna G, Bauce B, Lorenzon A, Smaniotto G, Rigato I, Calore M, Li Mura IE, Basso C, Thiene G, Lanfranchi G, Danieli GA, Nava A, Rampazzo A. The p.A897KfsX4 frameshift variation in desmocollin-2 is not a causative mutation in arrhythmogenic right ventricular cardiomyopathy. Eur J Hum Genet. 2010; 18:776-782.

Della Mina E, Ciccne R, Brustia F, Bayindir B, Limongelli I, Vetro A, Iascone M, Pezzoli L, Bellazzi R, Perotti G, De Giorgis V, Lunghi S, Coppola G, Orcesi S, Merli P, Savasta S, Veggiotti P, Zuffardi O. Improving molecular diagnosis in epilepsy by a dedicated high-throughput sequencing platform. Eur J Hum Genet. 2014 [Epub ahead of print]

Delmar M, McKenna WJ. The cardiac desmosome and arrhythmogenic cardiomyopathies. Circ Res. 2010 Sep 17; 107:700-14.

Den Dunnen JT and Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat. 2000; 15:7-12.

den Haan AD, Tan BY, Zikusoka MN, Lladó LI, Jain R, Daly A, Tichnell C, James C, Amat-Alarcon N, Abraham T, Russell SD, Bluemke DA, Calkins H, Dalal D, Judge DP. Comprehensive desmosome mutation analysis in north americans with arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Cardiovasc Genet. 2009; 2:428-35.

Dokuparti MV, Pamuru PR, Thakkar B, Tanjore RR, Nallari P. Etiopathogenesis of arrhythmogenic right ventricular cardiomyopathy. J Hum Genet. 2005; 50:375-381.

Dorn C, Grunert M, Sperling SR. Application of high-throughput sequencing for studying genomic variations in congenital heart disease. Brief Funct Genomics. 2014; 13:51-65.

Driessen M, Kienhuis AS, Pennings JL, Pronk TE, van de Brandhof EJ, Roodbergen M, Spaink HP, van de Water B, van der Ven LT. Exploring the zebrafish embryo as an alternative model for the evaluation of liver toxicity by histopathology and expression profiling. Arch Toxicol. 2013; 87:807-23.

Duan J, Deng HW, Wang YP. Common copy number variation detection from multiple sequenced samples. IEEE Trans Biomed Eng. 2014;61:928-37.

Duan J, Zhang JG, Deng HW, Wang YP. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. PLoS One. 2013; 8:e59128.

Elliott P, O'Mahony C, Syrris P, Evans A, Rivera Sorensen C, Sheppard MN, Carr-White G, Pantazis A, McKenna WJ. Prevalence of desmosomal protein gene mutations in patients with dilated cardiomyopathy. Circ Cardiovasc Genet. 2010; 3:314-22.

ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306:636-40.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57-74.

Ericsen AJ, Starrett GJ, Greene JM, Lauck M, Raveendran M, Deiros DR, Mohns MS, Vince N, Cain BT, Pham NH, Weinfurter JT, Bailey AL, Budde ML, Wiseman RW, Gibbs R, Muzny D, Friedrich TC, Rogers J, O'Connor DH. Whole genome sequencing of SIV-infected macaques identifies candidate loci that may contribute to host control of virus replication. Genome Biol. 2014;15:478.

Fischer SG, Lerman LS. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory, Proc.Natl. Acad. Sci. USA 80 1983; 1579–1583.

Fokkema IF, den Dunnen JT, Taschner PE. LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. Hum Mutat. 2005; 26:63-8.

Fontaine G, Fontaliran F, and Frank R. Arrhythmogenic right ventricular cardiomyopathies: clinical forms and main differential diagnoses. Circulation 1998; 97:1532-1535.

Frank R, Fontaine G, Vedel J, Mialet G, Sol C, Guiraudon G, and Grosgogeat Y. Electrocardiologie de quatre cas de dysplasie ventriculaire droite arythmogene. Arch Mal Coeur Vaiss 1978; 71:963-972.

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 95 proteins. Eur J Cell Biol 2006; 85:69-82.

Fournier PE, Dubourg G, Raoult D. Clinical detection and characterization of bacterial pathogens in the genomics era. Genome Med. 2014;6:114.
Fouse SD, Nagarajan RO, Costello JF. Genome-scale DNA methylation analysis. Epigenomics. 2010;2:105-17.

Fressart V, Duthoit G, Donal E, Probst V, Deharo JC, Chevalier P, Klug D, Dubourg O, Delacretaz E, Cosnay P, Scanu P, Extramiana F, Keller D, Hidden-Lucet F, Simon F, Bessirard V, Roux-Buisson N, Hebert JL, Azarine A, Casset-Senon D, Rouzet F, Lecarpentier Y, Fontaine G, Coirault C, Frank R, Hainque B, Charron P. Desmosomal gene analysis in arrhythmogenic right ventricular dysplasia/cardiomyopathy: spectrum of mutations and clinical impact in practice. Europace. 2010; 12:861-868.

G. Mellars, K. Gomez. Mutation detection by Southern Blotting. Methods Mol Biol. 2011; 688:281-291.

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

Garcia-Gras E, Lombardi R, Giocondo MJ, Willerson JT, Schneider MD, Khoury DS, Marian AJ. Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhythmogenic right ventricular cardiomyopathy. J Clin Invest. 2006; 116:2012-2021.

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

Garcia-Pavia P, Syrris P, Salas C, Evans A, Mirelis JG, Cobo-Marcos M, Vilches C, Bornstein B, Segovia J, Alonso-Pulpon L, Elliott PM. Desmosomal protein gene mutations in patients with idiopathic dilated cardiomyopathy undergoing cardiac transplantation: a clinicopathological study. Heart 2011; 97:1744-1752.

Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, Lu F, Lyon E, Voelkerding KV, Zehnbauer BA, Agarwala R, Bennett SF, Chen B, Chin EL, Compton JG, Das S, Farkas DH, Ferber MJ, Funke BH, Furtado MR, Ganova-Raeva LM, Geigenmüller U, Gunselman SJ, Hegde MR, Johnson PL, Kasarskis A, Kulkarni S, Lenk T, Liu CS, Manion M, Manolio TA, Mardis ER, Merker JD, Rajeevan MS, Reese MG, Rehm HL, Simen BB, Yeakley JM, Zook JM, Lubin IM. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012; 30:1033-6.

Genome 10K Community of Scientists. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered. 2009;100:659-74.

Gerull B, Heuser A, Wichter T, Paul M, Basson CT, McDermott DA, Lerman BB, Markowitz SM, Ellinor PT, MacRae CA, Peters S, Grossmann KS, Michely B, Sasse-Klaassen S, Birchmeier W, Dietz R, Breithardt G, Schulze-Bahr E, Thierfelder L. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. Nat Genet 2004; 36:1162-1164.

Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, Thibodeau P, Bachand I, Bao JY, Tong AH, Lin CH, Millet B, Jaafari N, Joober R, Dion PA, Lok S, Krebs MO, Rouleau GA.Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet. 2011; 43:860-3.

Glod J, Song M, Sharma A, Tyagi R, Rhodes RH, Weissmann DJ, Roychowdhury S, Khan A, Kane MP, Hirshfield K, Ganesan S, DiPaola RS, Rodriguez-Rodriguez L. Next generation sequencing as an aid to diagnosis and treatment of an unusual pediatric brain cancer. J Pers Med. 2014; 4:402-11.

Gonzaga-Jauregui C, Lupski JR, Gibbs RA.Human genome sequencing in health and disease. Annu Rev Med. 2012;63:35-61.
Grantham R. Amino acid difference formula to help explain protein evolution. Science 1974; 185:862–864.

Gréen A, Gréen H, Rehnberg M, Svensson A, Gunnarsson C, Jonasson J. Assessment of HaloPlex Amplification for Sequence Capture and Massively Parallel Sequencing of Arrhythmogenic Right Ventricular Cardiomyopathy-Associated Genes. J Mol Diagn. 2015; 17:31-42.

Green KJ, Sympson CL. Desmosomes: new perspectives on a classic. J Invest Dermatol. 2007; 127: 2499-2515.

Grossmann KS, Grund C, Huelsken J, Behrend M, Erdmann B, Franke WW, and Birchmeier W. Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. J Cell Biol 2004; 167:149-160.

Guilhamon P, Butcher LM, Presneau N, Wilson GA, Feber A, Paul DS, Schütte M, Haybaeck J, Keilholz U, Hoffman J, Ross MT, Flanagan AM, Beck S. Assessment of patient-derived tumour xenografts (PDXs) as a discovery tool for cancer epigenomics. Genome Med. 2014; 6:116.

Gupta P, Bilinska ZT, Sylvius N, Boudreau E, Veinot JP, Labib S, Bolongo PM, Hamza A, Jackson T, Ploski R, Walski M, Grzybowski J, Walczak E, Religa G, Fidzianska A, Tesson F. Genetic and ultrastructural studies in dilated cardiomyopathy patients: a large deletion in the lamin A/C gene is associated with cardiomyocyte nuclear envelope disruption. Basic Res Cardiol. 2010; 105:365-377.

Haywood AF, Merner ND, Hodgkinson KA, Houston J, Syrris P, Booth V, Connors S, Pantazis A, Quarta G, Elliott P, McKenna W, Young TL. Recurrent missense mutations in tmem43 (arvd5) due to founder effects cause arrhythmogenic cardiomyopathies in the UK and Canada. Eur Heart J. 2013; 34:1002-1011.

Hedberg C, Melberg A, Kuhl A, Jenne D, Oldfors A. Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7 is caused by a DES mutation. Eur J Hum Genet. 2012; 20:984-985.

Henikoff S., Henikoff J.G. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992; 89:10915-10919.

Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, Teodorescu DL, Cirino AL,Banner NR, Pennell DJ, Graw S, Merlo M, Di Lenarda A, Sinagra G, Bos JM, Ackerman MJ, Mitchell RN, Murry CE, Lakdawala NK, Ho CY, Barton PJ, Cook SA,Mestroni L, Seidman JG, Seidman CE. Truncations of titin causing dilated cardiomyopathy. N Engl J Med. 2012; 366:619-28.

Hertz CL, Christiansen SL, Ferrero-Miliani L, Fordyce SL, Dahl M, Holst AG, Ottesen GL, Frank-Hansen R, Bundgaard H, Morling N. Next-generation sequencing of 34 genes in sudden unexplained death victims in forensics and in patients with channelopathic cardiac diseases. Int J Legal Med. 2014 [Epub ahead of print]

Heuser A, Plovie ER, Ellinor PT, Grossmann KS, Shin JT, Wichter T, Basson CT, Lerman BB, Sasseklaassen S, Thierfelder L, MacRae CA, and Gerull B. Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet 2006; 79:1081-8.

Horpaopan S, Spier I, Zink AM, Altmüller J, Holzapfel S, Laner A, Vogt S, Uhlhaas S, Heilmann S, Stienen D, Pasternack SM, Keppler K, Adam R, Kayser K, Moebus S, Draaken M, Degenhardt F, Engels H, Hofmann A, Nöthen MM, Steinke V, Perez-Bouza A, Herms S, Holinski-Feder E, Fröhlich H, Thiele H, Hoffmann P, Aretz S. Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis. Int J Cancer. 2015;1 36:E578-89.

Igartua C, Myers RA, Mathias RA, Pino-Yanes M, Eng C, Graves PE, Levin AM, Del-Rio-Navarro BE, Jackson DJ, Livne OE, Rafaels N, Edlund CK, Yang JJ, Huntsman S, Salam MT, Romieu I, Mourad R, Gern JE, Lemanske RF, Wyss A, Hoppin JA, Barnes KC, Burchard EG, Gauderman WJ, Martinez FD, Raby BA, Weiss ST, Williams LK, London SJ, Gilliland FD, Nicolae DL, Ober C. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat Commun. 2015; 6:5965.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. Science 2009; 324: 218–223.

International HapMap Consortium. The International HapMap Project. Nature. 2003;426:789-96.

Iqbal Z, Neveling K, Razzaq A, Shahzad M, Zahoor MY, Qasim M, Gilissen C, Wieskamp N, Kwint MP, Gijsen S, de Brouwer AP, Veltman JA, Riazuddin S, van Bokhoven H.Targeted next generation sequencing reveals a novel intragenic deletion of the TPO gene in a family with intellectual disability. Arch Med Res. 2012; 43:312-316.

Jiménez-Jáimez J, López Moreno E, Barrio López MT, González-Molina M, Alvarez M, Tercedor L. A recessive inheritance pattern contributes to arrhythmogenic biventricular cardiomyopathy. Rev Esp Cardiol (Engl Ed). 2014; 67:772-774.

Joshi-Mukherjee R, Coombs W, Musa H, Oxford E, Taffet S, Delmar M. Characterization of the molecular phenotype of two arrhythmogenic right ventricular cardiomyopathy (ARVC)-related plakophilin-2 (PKP2) mutations. Heart Rhythm. 2008; 5:1715-23.

Jones MA, Rhodenizer D, da Silva C, Huff IJ, Keong L, Bean LJ, Coffee B, Collins C, Tanner AK, He M, Hegde MR. Molecular diagnostic testing for congenital disorders of glycosylation (CDG): detection rate for single gene testing and next generation sequencing panel testing. Mol Genet Metab. 2013; 110:78-85.

Kapplinger JD, Landstrom AP, Salisbury BA, Callis TE, Pollevick GD, Tester DJ, Cox MG, Bhuiyan Z, Bikker H, Wiesfeld AC, Hauer RN, van Tintelen JP, Jongbloed JD, Calkins H, Judge DP, Wilde AA, Ackerman MJ. Distinguishing arrhythmogenic right ventricular cardiomyopathy/dysplasia-associated mutations from background genetic noise. J Am Coll Cardiol. 2011; 57:2317-2327.

Kazazian HH, Boehm CD, Seltzer WK. ACMG recommendation for standards for interpretation of sequence variations. Genet Med 2000; 2:302–303.

Kleber AG, Saffitz JE. Role of the intercalated disc in cardiac propagation and arrhythmogenesis. Front Physiol. 2014; 5:404.

Klimczak D, Pączek L, Jażdżewski K, Kuch M. MicroRNAs: powerful regulators and potential diagnostic tools in cardiovascular disease. Kardiol Pol. 2014 .[Epub ahead of print]

Koopmann TT, Alders M, Jongbloed RJ, Guerrero S, Mannens MM, Wilde AA, Bezzina CR. Long QT syndrome caused by a large duplication in the KCNH2 (HERG) gene undetectable by current polymerase chain reaction-based exon-scanning methodologies. Heart Rhythm. 2006; 3:52-55.

Kougioumtzi A, Tsaparas P, Magklara A.Deep sequencing reveals new aspects of progesterone receptor signaling in breast cancer cells. PLoS One. 2014; 9:e98404.

Krawitz PM, Schiska D, Krüger U, Appelt S, Heinrich V, Parkhomchuk D, Timmermann B, Millan JM, Robinson PN, Mundlos S, Hecht J, Gross M. Screening for single nucleotide variants, small indels and exon deletions with a next-generation sequencing based gene panel approach for Usher syndrome. Mol Genet Genomic Med. 2014; 2:393-401.

Lampa S, Dahlö M, Olason PI, Hagberg J, Spjuth O. Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. Gigascience. 2013; 2:9.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS,

Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860-921.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42:D980-985.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947-2948.

Leung CL, Green KJ, and Liem RK. Plakins: a family of versatile cytolinker proteins. Trends Cell Biol 2002; 12:37-45.

Li D, Liu Y, Maruyama M, Zhu W, Chen H, Zhang W, Reuter S, Lin SF, Haneline LS, Field LJ, Chen PS, Shou W. Restrictive loss of plakoglobin in cardiomyocytes leads to arrhythmogenic cardiomyopathy. Hum Mol Genet. 2011; 20:4582-4596.

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; 31:1134-1144.

Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, Viale A, Wright C, Schweitzer PA, Gao Y, Kim D, Boland J, Hicks B, Kim R, Chhangawala S, Jafari N, Raghavachari N, Gandara J, Garcia-Reyero N, Hendrickson C, Roberson D, Rosenfeld J, Smith T, Underwood JG, Wang M, Zumbo P, Baldwin DA, Grills GS, Mason CE. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. Nat Biotechnol. 2014; 32:915-25.

Li Mura IE, Bauce B, Nava A, Fanciulli M, Vazza G, Mazzotti E, Rigato I, De Bortoli M, Beffagna G, Lorenzon A, Calore M, Dazzo E, Nobile C, Mostacciuolo ML, Corrado D, Basso C, Daliento L, Thiene G, Rampazzo A. Identification of a PKP2 gene deletion in a family with arrhythmogenic right ventricular cardiomyopathy. Eur J Hum Genet. 2013; 21:1226-1231.

Liang WC, Mitsuhashi H, Keduka E, Nonaka I, Noguchi S, Nishino I, Hayashi YK. TMEM43 mutations in Emery-Dreifuss muscular dystrophy-related myopathy. Ann Neurol. 2011; 69:1005-1013.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462: 315–322.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M.Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012; 2012:251364.

Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, Tang WH, Moravec CS, Margulies KB, Cappola TP, Li M; for the MAGNet consortium. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. Genomics. 2014. [Epub ahead of print]

Lombardi R, da Graca Cabreira-Hansen M, Bell A, Fromm RR, Willerson JT, Marian AJ. Nuclear plakoglobin is essential for differentiation of cardiac progenitor cells to adipocytes in arrhythmogenic right ventricular cardiomyopathy. Circ Res. 2011; 109:1342-1353.

Lombardi R, Dong J, Rodriguez G, Bell A, Leung TK, Schwartz RJ, Willerson JT, Brugada R, Marian AJ. Genetic fate mapping identifies second heart field progenitor cells as a source of adipocytes in arrhythmogenic right ventricular cardiomyopathy. Circ Res. 2009; 104:1076–1084.

Lopes LR, Zekavati A, Syrris P, Hubank M, Giambartolomei C, Dalageorgou C, Jenkins S, McKenna W; Uk10k Consortium, Plagnol V, Elliott PM. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. J Med Genet. 2013; 50: 228-239.

Lopez-Ayala JM, Ortiz-Genga M, Gomez-Milanes I, Lopez-Cuenca D, Ruiz-Espejo F, Sanchez-Munoz JJ, Oliva-Sandoval MJ, Monserrat L, Gimeno JR.A mutation in the Z-line Cypher/ZASP protein is associated with arrhythmogenic right ventricular cardiomyopathy. Clin Genet. 2014 [Epub ahead of print]

Loporcaro CG, Tester DJ, Maleszewski JJ, Kruisselbrink T, Ackerman MJ. Confirmation of cause and manner of death via a comprehensive cardiac autopsy including whole exome next-generation sequencing. Arch Pathol Lab Med. 2014; 138:1083-109.

Lorenzon A, Pilichou K, Rigato I, Vazza G, De Bortoli M, Calore M, Occhi G, Carturan E, Lazzarini E, Cason M, Mazzotti E, Poloni G, Mostacciuolo ML, Daliento L, Thiene G, Corrado D, Basso C, Bauce B, Rampazzo A. Homozygous founder mutation in Desmocollin-2 causes severe forms of arrhythmogenic cardiomyopathy. Paper submitted

Lwande O, Venter M, Lutomiah J, Michuki G, Rumberia C, Gakuya F, Obanda V, Tigoi C, Odhiambo C, Nindo F, Symekher S, Sang R. Whole genome phylogenetic investigation of a West Nile virus strain isolated from a tick sampled from livestock in north eastern Kenya. Parasit Vectors. 2014; 7:542.

Mallat Z, Tedgui A, Fontaliran F, Frank R, Durigon M, and Fontaine G. Evidence of apoptosis in Marcus FI, Fontaine GH, Guiraudon G, Frank R, Laurenceau JL, Malergue C, and Grosgogeat Y. Right ventricular dysplasia: a report of 24 adult cases. Circulation 1982; 65:384-398.

Manson-Bahr D, Ball R, Gundem G, Sethia K, Mills R, Rochester M, Goody V, Anderson E, O'Meara S, Flather M, Keeling M, Yazbek-Hanna M, Hurst R, Curley H, Clark J, Brewer DS, McDermott U, Cooper C.Mutation detection in formalin-fixed prostate cancer biopsies taken at the time of diagnosis using next-generation DNA sequencing. J Clin Pathol. 2015 Jan 13. pii: jclinpath-2014-202754.

Marcacci M, Ancora M, Mangone I, Teodori L, Di Sabatino D, De Massis F, Camma' C, Savini G, Lorusso A. Whole genome sequence analysis of the arctic-lineage strain responsible for distemper in Italian wolves and dogs through a fast and robust next generation sequencing protocol. J Virol Methods. 2014;202:64-8.
Marcus FI, McKenna WJ, Sherrill D, Basso C, Bauce B, Bluemke DA, Calkins H, Corrado D, Cox MG, Daubert JP, Fontaine G, Gear K, Hauer R, Nava A, Picard MH, Protonotarios N, Saffitz JE, Sanborn DM, Steinberg JS, Tandri H, Thiene G, Towbin JA, Tsatsopoulou A, Wichter T, Zareba W. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. Circulation. 2010;121:1533-41.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376-80.

Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot X, Turner SD, Sebra R, Kasarskis A, Peto T, Crook D, Sifri CD.Klebsiella pneumoniae carbapenemase (KPC) producing K. pneumoniae at a Single Institution: Insights into Endemicity from Whole Genome Sequencing. Antimicrob Agents Chemother. 2015 [Epub ahead of print]

Matsumoto T, Wu J, Antonio BA, Sasaki T.Development in rice genome research based on accurate genome sequence. Int J Plant Genomics. 2008; 2008:348621.

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PIW, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitc J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat. Genet. 2008; 40:1166–1174.

McKenna WJ, Thiene G, Nava A, Fontaliran F, Blomstrom-Lundqvist C, Fontaine G, Camerini F. Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Diseaseof the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. Br Heart J 1994; 71:215–218.

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

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome research 2010; 20:816-825.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 2005; 33: 5868–5877.

Menezes MJ, Guo Y, Zhang J, Riley LG, Cooper ST, Thorburn DR, Li J, Dong D, Li Z, Glessner J, Davis RL, Sue CM, Alexander SI, Arbuckle S, Kirwan P, Keating BJ, Xu X, Hakonarson H, Christodoulou J. Mutation in mitochondrial ribosomal protein S7 (MRPS7) causes congenital sensorineural deafness, progressive hepatic and renal failure, and lactic acidemia. Hum Mol Genet. 2015 [Epub ahead of print]

Merner ND, Hodgkinson KA, Haywood AF, Connors S, French VM, Drenckhahn JD, Kupprion C, Ramadanova K, Thierfelder L, McKenna W, Gallagher B, Morris-Larkin L, Bassett AS, Parfrey PS, Young TL. 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; 82:809-821.

Merritt AJ, Berika MY, Zhai W, Kirk SE, Ji B, Hardman MJ, and Garrod DR. Suprabasal desmoglein 3 expression in the epidermis of transgenic mice results in hyperproliferation and abnormal differentiation. Mol Cell Biol 2002; 22:5846-58.

Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11:31–46.

Mook OR, Haagmans MA, Soucy JF, van de Meerakker JB, Baas F, Jakobs ME, Hofman N, Christiaans I, Lekanne Deprez RH, Mannens MM. Targeted sequence capture and GS-FLX Titanium sequencing of 23 hypertrophic and dilated cardiomyopathy genes: implementation into diagnostics. J Med Genet. 2013; 50:614-626.

Morgan S, Shoai M, Fratta P, Sidle K, Orrell R, Sweeney MG, Shatunov A, Sproviero W, Jones A, Al-Chalabi A, Malaspina A, Houlden H, Hardy J, Pittman A. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. Neurobiol Aging. 2014; S0197-4580(14)00833-1.
Mullis K.B. The unusual origin of the polymerase chain reaction. Sci Am 1990 262:56–61, 64–65

Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. Nat Rev Genet. 2013; 14:321-332.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008; 320: 1344–1349.

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

Narula N, Tester DJ, Paulmichl A, Maleszewski JJ, Ackerman MJ. Post-mortem Whole Exome Sequencing with Gene-Specific Analysis for Autopsy-Negative Sudden Unexplained Death in the Young: A Case Series. Pediatr Cardiol. 2014. [Epub ahead of print]

Nava A, Bauce B, Basso C, Muriago M, Rampazzo A, Villanova C, Daliento L, Buja G, Corrado D, Danieli GA, and Thiene G. Clinical profile and long-term follow-up of 37 families with arrhythmogenic right ventricular cardiomyopathy. J Am Coll Cardiol 2000; 36:2226-2233.

Nava A, Thiene G, Canciani B, Scognamiglio R, Daliento L, Buja G, Martini B, Stritoni P, and Fasoli G. Familial occurrence of right ventricular dysplasia: a study involving nine families. J Am Coll Cardiol 1988; 12:1222-1228.

Neff T, Armstrong SA. Chromatin maps, histone modifications and leukemia. Leukemia. 2009;23:1243-51.

Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet. 2012; 13:667-672.

Newell-Price J, Clark AJ, King P. DNA methylation and silencing of gene expression. Trends Endocrinol Metab 2000; 11:142–148.

Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. Genome Res. 2001; 11:863-874.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010; 42:30-5.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461:272-6.

Nijman IJ, van Montfrans JM, Hoogstraat M, Boes ML, van de Corput L, Renner ED, van Zon P, van Lieshout S, Elferink MG, van der Burg M, Vermont CL, van der Zwaag B, Janson E, Cuppen E, Ploos van Amstel JK, van Gijn ME. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. J Allergy Clin Immunol. 2014; 133:529-534.

Nollau P, Wagener C. Methods for detection of point mutations: performance and quality assessment. IFCC Scientific Division, Committee on Molecular Biology Techniques. Clin Chem. 1997; 43:1114-1128.

North AJ, Bardsley WG, Hyam J, Bornslaeger EA, Cordingley HC, Trinnaman B, Hatzfeld M, Green KJ, Magee AI, Garrod DR. Molecular map of the desmosomal plaque. J Cell Sci. 1999; 112:4325-4336.

Nuber UA, Schäfer S, Schmidt A, Koch PJ, and Franke WW. The widespread human desmocollin Dsc2 and tissue-specific patterns of synthesis of various desmocollin subtypes. Eur J Cell Biol 1995; 66:69-74.

Oefner PJ, Underhill PA. Comparative DNA sequencing by denaturing high-performance liquid chromatography (dHPLC). Am J Hum Genet 1995; 57:A266.

Ohno S, Nagaoka I, Fukuyama M, Kimura H, Itoh H, Makiyama T, Shimizu A, Horie MAge-dependent clinical and genetic characteristics in Japanese patients with arrhythmogenic right ventricular cardiomyopathy/dysplasia. Circ J 2013; 77:1534–1542.

Otten E, Asimaki A, Maass A, van Langen IM, van der Wal A, de Jonge N, van den Berg MP, Saffitz JE, Wilde AA, Jongbloed JD, van Tintelen JP. Desmin mutations as a cause of right ventricular heart failure affect the intercalated disks. Heart Rhythm. 2010; 7:1058-1064.

Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata.Nucleic Acids Res. 2012; 40:D571-9.

Pacurar DI, Pacurar ML, Pacurar AM, Gutierrez L, Bellini C. A novel viable allele of Arabidopsis CULLIN1 identified in a screen for superroot2 suppressors by next generation sequencing-assisted mapping. PLoS One. 2014; 9:e100846.

Patel D, Gemel J, Xu Q, Simon AR, Lin X, Matiukas A, Beyer EC, Veenstra RD. Atrial fibrillation-associated connexin40 mutants make hemichannels and synergistically form gap junction channels with novel properties. FEBS Lett. 2014; 588:1458-1464.

Perazzolo Marra M, Thiene G, Rizzo S, De Lazzari M, Carturan E, Tona F, Caforio AL, Cacciavillani L, Marcolongo R, Tarantini G, Corbetti F, Iliceto S, Basso C. Cardiac magnetic resonance features of biopsy-proven endomyocardial diseases. JACC Cardiovasc Imaging. 2014; 7:309-312.

Pilichou K, Mancini M, Rigato I, Lazzarini E, Giorgi B, Carturan E, Bauce B, d'Amati G, Marra MP, Basso C. Nonischemic left ventricular scar: sporadic or familial? Screen the genes, scan the mutation carriers. Circulation. 2014; 130:e180-182.

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

Pilichou K, Remme CA, Basso C, Campian ME, Rizzo S, Barnett P, Scicluna BP, Bauce B, van den Hoff MJ, de Bakker JM, Tan HL, Valente M, Nava A, Wilde AA, Moorman AF, Thiene G, Bezzina CR. Myocyte necrosis underlies progressive myocardial dystrophy in mouse dsg2-related arrhythmogenic right ventricular cardiomyopathy. J Exp Med. 2009; 206:1787-1802.

Pylro VS, Roesch LF, Morais DK, Clark IM, Hirsch PR, Tótola MR. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods. 2014;107:30-7.

Posch MG, Posch MJ, Geier C, Erdmann B, Mueller W, Richter A, Ruppert V, Pankuweit S, Maisch B, Perrot A, Buttgereit J, Dietz R, Haverkamp W, Ozcelik C. A missense variant in desmoglein-2 predisposes to dilated cardiomyopathy. Mol Genet Metab. 2008;95:74-80.

Priori SG, Napolitano C, Tiso N, Memmi M, Vignati G, Bloise R, Sorrentino V, Danieli GA. Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. Circulation 2001; 103, 196-200.

Protonotarios N, Tsatsopoulou A, Patsourakos P, Alexopoulos D, Gezerlis P, Simitsis S, and Scampardonis G. Cardiac abnormalities in familial palmoplantar keratosis. Br Heart J 1986; 56:321-326.

Quarta G, Syrris P, Ashworth M, Jenkins S, Zuborne Alapi K, Morgan J, Muir A, Pantazis A, McKenna WJ, Elliott PM. Mutations in the Lamin A/C gene mimic arrhythmogenic right ventricular cardiomyopathy. Eur Heart J. 2012; 33:1128-1136.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev. 2006; 20:3407–3425.

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

Rasmussen TB, Palmfeldt J, Nissen PH, Magnoni R, Dalager S, Jensen UB, Kim WY, Heickendorff L, Mølgaard H, Jensen HK, Baandrup UT, Bross P, Mogensen J. Mutated desmoglein-2 proteins are incorporated into desmosomes and exhibit dominant-negative effects in arrhythmogenic right ventricular cardiomyopathy. Hum Mutat. 2013; 34:697-705.

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013; 15:733-47.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med 2008: 10: 294–300.

Rigato I, Bauce B, Rampazzo A, Zorzi A, Pilichou K, Mazzotti E, Migliore F, Marra MP, Lorenzon A, De Bortoli M, Calore M, Nava A, Daliento L, Gregori D, Iliceto S, Thiene G, Basso C, Corrado D. Compound and digenic heterozygosity predicts lifetime arrhythmic outcome and sudden cardiac death in desmosomal gene-related arrhythmogenic right ventricular cardiomyopathy. Circ Cardiovasc Genet. 2013; 6:533-542.

Rizzo S, Lodder EM, Verkerk AO, Wolswinkel R, Beekman L, Pilichou K, Basso C, Remme CA, Thiene G, Bezzina CR. Intercalated disc abnormalities, reduced Na(+) current density, and conduction slowing in desmoglein-2 mutant mice prior to cardiomyopathic changes. Cardiovasc Res. 2012; 95:409-418.

Roberts JD, Herkert JC, Rutberg J, Nikkel SM, Wiesfeld AC, Dooijes D, Gow RM, van Tintelen JP, Gollob MH. Detection of genomic deletions of PKP2 in arrhythmogenic right ventricular cardiomyopathy. Clin Genet. 2013; 83:452-456.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative Genomics Viewer. Nature Biotechnology 2011; 29, 24–26.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 201; 29:24-6.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 475:348-52.

Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132:365-386.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell. 2006; 127:1193-207.

Saffitz JE.The pathobiology of arrhythmogenic cardiomyopathy. Annu Rev Pathol 2011; 6:299–321.

Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors Proc Natl Acad Sci U S A. 1977; 74: 5463–5467.

Sato T, Nishio H, Suzuki K.Sudden death during exercise in a juvenile with arrhythmogenic right ventricular cardiomyopathy and desmoglein-2 gene substitution: a case report. Leg Med (Tokyo). 2011; 13:298-300.

Schäfer S, Koch PJ, and Franke WW. Identification of the ubiquitous human desmoglein, Dsg2, and the expression catalogue of the desmoglein subfamily of desmosomal cadherins. Exp Cell Res 1994; 211:391-399.

Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. 2015 [Epub ahead of print]

Schottmann G, Jungbluth H, Schara U, Knierim E, Morales Gonzalez S, Gill E, Seifert F, Norwood F, Deshpande C, von Au K, Schuelke M, Senderek. Recessive truncating IGHMBP2 mutations presenting as axonal sensorimotor neuropathy. Neurology. 2015 [Epub ahead of print]

Schroyen M, Tuggle CK. Current transcriptomics in pig immunity research. Mamm Genome. 2014. [Epub ahead of print]

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005; 309:1728-32.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002; 30:e57.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. Science. 2008;322:1849-51.

Sen-Chowdhry S, Prasad SK, Syrris P, Wage R, Ward D, Merrifield R, Smith GC, Firmin DN, Pennell DJ, McKenna WJ. Cardiovascular magnetic resonance in arrhythmogenic right ventricular cardiomyopathy revisited: comparison with task force criteria and genotype. J Am Coll Cardiol. 2006; 48:2132-40.

Sen-Chowdhry S, Syrris P, Prasad SK, Hughes SE, Merrifield R, Ward D, Pennell DJ, McKenna WJ. Left-dominant arrhythmogenic cardiomyopathy: an under-recognized clinical entity. J Am Coll Cardiol 2008; 52:2175–2187.

Sen-Chowdhry, M. D. Lowe, S. C. Sporton, and W. J. McKenna. Arrhythmogenic right ventricular cardiomyopathy: clinical presentation, diagnosis, and management. Am.J.Med. 2004; 117:685-695.

Sharma N, Bhalla PL, Singh MB.Transcriptome-wide profiling and expression analysis of transcription factor families in a liverwort, Marchantia polymorpha. BMC Genomics. 2013; 14:915.

Shi Y, Tyson GW, DeLong EF. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature. 2009;459:266-9.

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 1993; 16:325-332.

Shin J, Ming GL, Song H. Decoding neural transcriptomes and epigenomes via high-throughput sequencing. Nat Neurosci. 2014; 17:1463-75.

Sidransky D. Nucleic acid-based methods for the detection of cancer. Science 1997; 278:1054–1059.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15:121-32.

Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev 2007; 21: 601–614.

Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, Thomas G, Valle D, Hoover-Fong JE, Goldstein DB. Whole genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet. 2010; 6:e1000991.

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 1975; 98: 503-517.

Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN: The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. 2014; 133:1-9.

Stransky N , Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, Carter SL, Voet D, Cortés ML, Auclair D, Berger MF, Saksena G, Guiducci C, Onofrio RC, Parkin M, Romkes M, Weissfeld JL, Seethala RR, Wang L, Rangel-Escareño C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Winckler W, Ardlie K, Gabriel SB, Meyerson M, Lander ES, Getz G, Golub TR, Garraway LA, Grandis JR. The mutational landscape of head and neck squamous cell carcinoma. Science 2011; 333: 1157-1160.

Stewart FJ, Dmytrenko O, Delong EF, Cavanaugh CM.Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of solemya velum. Front Microbiol. 2011; 2:134.

Syrris P, Ward D, Evans A, Asimaki A, Gandjbakhch E, Sen-Chowdhry S, McKenna WJ. Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. Am J Hum Genet 2006; 79:978-984.

Tabone T, Sallmann G, Chiotis M, Law M, Cotton R. Chemical cleavage of mismatch (CCM) to locate base mismatches in heteroduplex DNA. Nat Protoc. 2006; 1:2297-2304.

Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS. Tiny RNAs associated with transcription start sites in animals. Nat Genet. 2009; 41:572-8.

Taib N, Mangot JF, Domaizon I, Bronner G, Debroas D.Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. PLoS One. 2013; 8:e58950.

Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc 2012; 7: 617–636.

Tandri H, Saranathan M, Rodriguez ER, Martinez C, Bomma C, Nasir K, Rosen B, Lima JA, Calkins H, Bluemke DA. Noninvasive detection of myocardial fibrosis in arrhythmogenic right ventricular cardiomyopathy using delayed-enhancement magnetic resonance imaging. J Am Coll Cardiol. 2005; 45:98-103.

Tang S,Wang J,Zhang VW,Li FY,Landsverk M,Cui H,Truong CK,Wang G,Chen LC,Graham B,Scaglia F,Schmitt ES,Craigen WJ,Wong LJ.Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. Hum Mutat. 2013; 34:882-893.

Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J MedGenet 2006; 43:295–305.

Taylor M, Graw S, Sinagra G, Barnes C, Slavov D, Brun F, Pinamonti B, Salcedo EE, Sauer W, Pyxaras S, Anderson B, Simon B, Bogomolovas J, Labeit S, Granzier H, Mestroni L. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. Circulation. 201; 124:876-885.

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

Thiene G, Corrado D, Nava A, Rossi L, Poletti A, Boffa GM, Daliento L, Pennelli N. Right ventricular cardiomyopathy: is there evidence of an inflammatory aetiology? Eur Heart J. 1991; 12:22-25.

Thiene G, Nava A, Corrado D, Rossi L, Pennelli N. Right ventricular cardiomyopathy and sudden death in young people. N Engl J Med. 1988; 318:129-133.

Toma I, Siegel MO, Keiser J, Yakovleva A, Kim A, Davenport L, Devaney J, Hoffman EP, Alsubail R, Crandall KA, Castro-Nallar E, Pérez-Losada M, Hilton SK, Chawla LS, McCaffrey TA, Simon GL. Single-molecule long-read 16S sequencing to characterize the lung microbiome from mechanically ventilated patients with suspected pneumonia. J Clin Microbiol. 2014; 52:3913-21.

Valente M, Calabrese F, Thiene G, Angelini A, Basso C, Nava A, Rossi L. In vivo evidence of apoptosis in arrhythmogenic right ventricular cardiomyopathy. Am J Pathol. 1998; 152:479-484.

van der Zwaag PA, Jongbloed JD, van den Berg MP, van der Smagt JJ, Jongbloed R, Bikker H, Hofstra RM, van Tintelen JP. A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. Hum Mutat. 2009; 30:1278-1283.

van der Zwaag PA, van Rijsingen IA, Asimaki A, Jongbloed JD, van Veldhuisen DJ, Wiesfeld AC, Cox MG, van Lochem LT, de Boer RA, Hofstra RM, Christiaans I, van Spaendonck-Zwarts KY, Lekanne dit Deprez RH, Judge DP, Calkins H, Suurmeijer AJ, Hauer RN, Saffitz JE, Wilde AA, van den Berg MP, van Tintelen JP. 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; 14:1199-1207.

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

van Spaendonck-Zwarts KY, van Rijsingen IA, van den Berg MP, Lekanne Deprez RH, Post JG, van Mil AM, Asselbergs FW, Christiaans I, van Langen IM, Wilde AA, de Boer RA, Jongbloed JD, Pinto YM, van Tintelen JP.Genetic analysis in 418 index patients with idiopathic dilated cardiomyopathy: overview of 10 years' experience. Eur J Heart Fail. 2013; 15:628-636.

van Tintelen JP, Entius MM, Bhuiyan ZA, Jongbloed R, Wiesfeld AC, Wilde AA, van der Smagt J, Boven LG, Mannens MM, van Langen IM, Hofstra RM, Otterspoor LC, Doevendans PA, Rodriguez LM, van Gelder IC, Hauer RN. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation. 2006; 113:1650-1658.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D,

Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. Science. 2001; 291:1304-51.

Verma A.Next-generation sequencing and genetic diagnosis of Charcot-Marie-Tooth disease. Ann Indian Acad Neurol. 2014;17:383-6.

Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, Soto-Ortolaza AI, Cobb SA, Wilhoite GJ, Bacon JA, Behrouz B, Melrose HL, Hentati E, Puschmann A, Evans DM, Conibear E, Wasserman WW, Aasly JO, Burkhard PR, Djaldetti R, Ghika J, Hentati F, Krygowska-Wajs A, Lynch T, Melamed E, Rajput A, Rajput AH, Solida A, Wu RM, Uitti RJ, Wszolek ZK, Vingerhoets F, Farrer MJ. VPS35 mutations in Parkinson disease. Am J Hum Genet. 2011; 89:162-7.

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. Nat Genet. 2010;42:1109-12.

Wang H, Nettleton D, Ying K.Copy number variation detection using next generation sequencing read counts. BMC Bioinformatics. 2014; 15:109.

Wang M, Borris L, Aarestrup FM, Hasman H. Identification of a Pseudomonas aeruginosa co-producing NDM-1, VIM-5 and VIM-6 metallo-β-lactamases in Denmark using whole-genome sequencing. Int J Antimicrob Agents. 2014. [Epub ahead of print]

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

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 2009; 138: 1019–1031.

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

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369:1502-1511.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014; 312:1870-9.

Yang Z, Bowles NE, Scherer SE, Taylor MD, Kearney DL, Ge S, Nadvoretskiy VV, DeFreitas G, Carabello B, Brandon LI, Godsel LM, Green KJ, Saffitz JE, Li H, Danieli GA, Calkins H, Marcus F, Towbin JA Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Res. 2006; 99:646-655.

Yoon S, Xuan Z, Makarov V, Ye K, Sebat J Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res 2009; 19: 1586–1592.

Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 2009; 10:451-481.

Zhang M.,Tavora F., Burke A. Desmosomal protein gene mutations in patients with idiopathic DCM Heart 2011; 97: 2090.

Zhao H, Li Y, Wang S, Yang Y, Wang J, Ruan X, Yang Y, Cai K, Zhang B, Cui P, Yan J, Zhao Y, Wakeland EK, Li Q, Hu S, Fang X. Whole transcriptome RNA-seq analysis: tumorigenesis and metastasis of melanoma. Gene. 2014; 548:234-43.

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics. 2013;14 Suppl 11:S1.

Zhao W, Zhu QY, Zhang JT, Liu H, Wang LJ, Chen ZQ, Guan LP, Huang XS, Yang L, Yu SY. Exome sequencing identifies novel compound heterozygous mutations in SPG11 that cause autosomal recessive hereditary spastic paraplegia. J Neurol Sci. 2013; 335:112-117.

Zhou X, Wang B, Pan Q, Zhang J, Kumar S, Sun X, Liu Z, Pan H, Lin Y, Liu G, Zhan W, Li M, Ren B, Ma X, Ruan H, Cheng C, Wang D, Shi F, Hui Y, Tao Y, Zhang C, Zhu P, Xiang Z, Jiang W, Chang J, Wang H, Cao Z, Jiang Z, Li B, Yang G, Roos C, Garber PA, Bruford MW, Li R, Li M. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. Nat Genet. 2014;46:1303-10.