Fakultät für Medizin
I. Medizinische Klinik

Whole exome sequencing to elucidate the molecular basis of cardiac disease

Elisa Mastantuono

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Stefan Engelhardt

Betreuer: Prof. Dr. Thomas Meitinger

Prüfer der Dissertation:

1. Prof. Dr. Karl-Ludwig Laugwitz
2. Prof. Dr. Heribert Schunkert

Die Dissertation wurde am 21.11.2017 bei der Fakultät für Medizin der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 06.03.2018 angenommen.

Acknowledgements

First, I would like to thank my doctoral advisor, Prof. Thomas Meitinger, for giving me the opportunity to work in his Institute and for the hours and the efforts that he spent improving my scientific background. Under his supervision, I have assimilated essential concepts that are certainly not contained in books: think out of the box, fight for your ideas, push the limits, critical thinking, scientific rigour, and many more…it was a real Genomics Boot Camp. Thank you.

I am particularly grateful to Dr. Holger Prokisch for supporting my thesis and for all the efforts he made to improve it, as well as, for the atmosphere that he was able to create in the group.

I am sincerely thankful to Prof. Alessandra Moretti and Prof. Karl Laugwitz for the great support during my PhD studies and the possibility to collaborate in many interesting topics.

Next, I would like to thank Riccardo and Thomas for their great help during these months. Special thanks also to Bader, Arcangela, Caterina, Eliska, Laura, Robert, Peter and Milena for their support.
My sincere thanks also to all my colleagues at the Institute of Human Genetics for the great working atmosphere.

Last but not least I would like to thank my friends and family for their unconditional support and love.

Abstract

Cardiac disorders are a leading cause of death and genetic findings can impact such life-threatening conditions in terms of prevention and intervention. Cardiac disorders are genetically highly heterogeneous, with hundreds of genes and different models of inheritance involved. The identification of the underlying mutation is crucial as it does not only provide the patient with a molecular diagnosis but also directs family screening and can effectively guide therapy.

To tackle the genetic complexity of cardiac disorders, molecular genetic testing needs to be comprehensive and un-targeted with parallel analysis of a large number of genes. Whole exome sequencing (WES) allows the identification of genetic defects in both novel and known disease-associated genes.

In this study, I have applied different approaches of WES analysis to investigate the molecular basis of dominant Mendelian disorders (arrhythmogenic right ventricular cardiomyopathy, ARVC, and long QT syndrome, LQTS), recessive Mendelian disorders (mitochondrial cardiomyopathy) and complex disorders (hypoplastic left heart syndrome).

ARVC is responsible for 20% of sudden cardiac deaths among young individuals and 50% of the cases await a genetic diagnosis. In this study, I have identified two novel candidates (*CDH2* and *MYH10*) that encode for proteins involved in cell adhesion junctions in cardiomyocytes. I have also established an association of CNVs in LQTS and the de novo variants in *CALM* genes (*CALM3* in this study) in malignant LQTS forms.

For recessive forms of mitochondrial cardiomyopathy, I have investigated the genotypic and phenotypic spectrum of nearly 100 nuclear mitochondrial genes associated with cardiac manifestations. This list is constantly growing thanks to the identification of novel genes by WES. Among them, I focused on *ACAD9*, a gene discovered a few years ago and already recognised as one of the most frequent causes of mitochondrial cardiomyopathy. In a genotyped cohort of 67 patients, a positive effect of riboflavin was observed on the survival of the patients. This represents a step toward personalised treatment of patients with *ACAD9* deficiency.

Hypoplastic left heart syndrome (HLHS) is an extremely severe congenital heart defect and little is known about its genetic basis. I have applied a WES analysis in a cohort of 78 HLHS trios. De novo analysis and pathway analysis, together with functional studies in hiPSCs, suggested an important role of defective primary cilium/ autophagy/cell-cycle axis in HLHS pathogenesis.

WES allows the identification not only of mutations related to the disease under investigation but also of pathogenic or likely pathogenic variants in genes associated with other diseases, the so-called “incidental findings” (IFs). In the minimum list of 56 actionable genes for which IFs should be reported to the patient, more than half (31/56) is associated with cardiac disorders. In a cohort of 855 cases sequenced by WES, I have investigated the presence of IFs in those genes, identifying pathogenic or likely pathogenic variants in 4/855 cases.

Although WES application drastically increased the identification of Mendelian disease genes as reflected by the almost 2000 new entries in OMIM since 2008, its genetic yield remains around 50% (depending on the disease). Future perspectives are represented by the use of complementary approaches of WES (or genome sequencing in the near future) and RNA-seq.

Zusammenfassung

Herzerkrankungen gehören zu den häufigsten Todesursachen, wobei genetische Untersuchungen diese lebensbedrohlichen Zustände mittels Prävention und Intervention beeinflussen können. Herzerkrankungen sind genetisch sehr heterogen und können von Mutationen in hunderten von Genen hervorgerufen werden als auch unterschiedlichen Vererbungsmustern folgen. Die Identifizierung der zugrundeliegenden Mutation ist von entscheidender Bedeutung, da sie nicht nur eine molekulare Diagnose für den Patienten liefert, sondern auch das Screening der Familie ermöglicht und die Therapie effektiv steuern kann.

Um die genetische Komplexität von Herzerkrankungen zu erfassen, müssen molekulargenetische Tests umfassend und zielgerichtet sein, wobei eine große Anzahl von Genen parallel analysiert werden muss. Die vollständige Exomsequenzierung (WES) ermöglicht die Identifizierung von genetischen Defekten sowohl in neuen als auch in bekannten krankheitsassoziierten Genen.

In dieser Arbeit habe ich verschiedene Ansätze der WES-Analyse zur Untersuchung der molekularen Grundlagen von dominanten Mendelschen Störungen (arrhythmogene rechtsventrikuläre Kardiomyopathie, ARVC und Long QT Syndrom, LQTS), rezessiven Mendelschen Störungen (mitochondriale Kardiomyopathie) und komplexen Herzsyndromen (hypoplastic links hearz syndrome) angewandt.

ARVC ist für 20% der Fälle von plötzlichem Herztod bei Jugendlichen verantwortlich und 50% der Fälle warten auf eine genetische Diagnose. In dieser Studie habe ich zwei neue Kandidatengene (CDH2 und MYH10) identifiziert, die für Proteine kodieren, die an der Bildung von Zellkontakten in Kardiomyozyten beteiligt sind. In dieser Studie wurde weiterhin eine Assoziation von CNVs mit LQTS und den de novo Varianten in CALM-Genen (CALM3 in dieser Studie) mit malignen LQTS-Formen etabliert.

Für rezessive Formen der mitochondrialen Kardiomyopathie habe ich das genotypische und phänotypische Spektrum von Mutationen in nahezu 100 nukleären mitochondrialen Genen untersucht, die mit kardialen Manifestationen assoziiert sind. Diese Liste wächst stetig aufgrund der kontinuierlichen Identifizierung neuer krankheits-assoziierter Gene durch WES. Unter den krankheits-assoziierten Genen habe ich mich besonders auf ACAD9 konzentriert, ein Gen in dem Mutationen vor einigen Jahren entdeckt wurden und nun als eine der häufigsten Ursachen für die mitochondriale Kardiomyopathie identifiziert werden. In einer genotypisierten Kohorte von 67 Patienten wurde eine positive Wirkung von

Riboflavin auf das Überleben der Patienten beobachtet. Dies ist ein erster Schritt zu einer personalisierten Behandlung von Patienten mit ACAD9-Mangel.

Das hypoplastische Linksherzsyndrom (HLHS) ist ein extrem schwerer angeborener Herzfehler jedoch ist wenig über seine genetische Basis bekannt. Ich habe eine WES-Analyse in einer Kohorte von 78 HLHS-Trios angewendet. Eine Kombination aus de-novo-Analysen, Pathway-Analysen und funktionellen Studien in HiPSCs, schlugen eine wichtige Rolle des defekten primären Cilium / Autophagie / Zell-Achse in HLHS-Pathogenese vor.

WES ermöglicht nicht nur die Identifizierung von Mutationen, die mit der zu untersuchenden Krankheit in Zusammenhang stehen, sondern auch von pathogenen oder wahrscheinlich pathogenen Varianten in Genen, die mit anderen Krankheiten assoziiert sind, den sogenannten "Zufallsbefunden" (IFs). In der Mindestliste von 56 handlungsfähigen Genen, für die IFs dem Patienten berichtet werden sollten, sind mehr als die Hälfte (31/56) mit Herzerkrankungen assoziiert. In einer Kohorte von 855 Fällen, die mittels WES sequenziert wurden, habe ich das Vorhandensein von IFs in diesen Genen untersucht, wobei pathogene oder wahrscheinlich pathogene Varianten in 4/855 Fällen identifiziert wurden.

Obwohl die Anwendung von WES die genetische Diagnose für Mendelschen Krankheit drastisch erhöhte, wie die fast 2000 neuen Einträge bei OMIM seit 2008 zeigen, bleibt ihre genetische Ausbeute bei etwa 50% (abhängig von der Krankheit). Zukünftige Perspektiven werden durch die Verwendung von komplementären Ansätzen von WES (oder Genomsequenzierung in naher Zukunft) und RNA-seq repräsentiert.

Table of contents

I. Introduction

1.1 Genetic basis of cardiac disease

During the past two decades, genetic discoveries have elucidated the molecular basis of many cardiac disorders, including hypertrophic and dilated cardiomyopathy (HCM and DCM), arrhythmogenic ventricular cardiomyopathy (ARVC), inherited arrhythmias and congenital heart defects (CHDs). These findings highlighted the key role of genetic testing in the care of families with inherited cardiac diseases, as reflected in international guidelines recommending genetic testing to improve the diagnosis and management of patients and at-risk family members[1].

A critical aspect of hereditary cardiac conditions is represented by their blurred boundaries with overlapping phenotypes. Cardiac disorders are genetically highly heterogeneous, with hundreds of genes and different model of inheritance involved. To tackle this genetic complexity, molecular genetic testing needs to be comprehensive and un-targeted, with parallel analysis of a large number of genes. Before the advent of Next-Generation Sequencing (NGS), based on massive parallel sequencing of millions of molecules at the same time, the simultaneous analysis of a high number of genes was not feasible, and the genetic testing was based on the sequencing of single genes (first generation sequencing). NGS technologies have completely revolutionized the molecular diagnosis of genetic conditions. First applied only in research projects in 2009, whole exome sequencing (WES) is currently used in the routine diagnosis of Mendelian disease. WES allows the identification of genetic defects in the coding regions of the genome, for a total of more than 20000 genes sequenced for each experiment.

Cardiovascular Mendelian diseases span multiple conditions, for example long QT syndrome, familial hypercholesterolemia, hypertrophic and familial dilated cardiomyopathies, and Marfan syndrome[1]. The yield of genetic testing with next-generation sequencing technologies is variable across these conditions, ranging from 80% in long QT syndrome to as low as 20%-40% in dilated cardiomyopathy[2]. Characterizing the underlying genetic cause of these cases is of high values, guiding for example family screening and fertility planning. Moreover, the identification of the genetic cause can lead to a tremendous impact on the outcome of the patient, in terms of prevention of sudden cardiac death (SCD). In this scenario, WES discoveries in cardiac disease can effectively be seen as a first step towards “precision medicine” interventions[3]. For example, the discovery of the causal gene in long QT syndrome

allows precision targeting of therapy and identification of variants in transthyretin (TTR) or Fabry's disease can allow targeted therapies (including RNA silencing, isoform stabilizers, and enzyme replacement)[4].

1.1.1 Mendelian disorders

Mendelian, or monogenic, diseases are diseases caused by mutations in single genes. They can be divided based on their inheritance pattern. Mendelian inheritance patterns are determined by the location and number of copies (one or two) of the gene involved.

1.1.1.1 Autosomal dominant disorders

Dominant inheritance means that although in autosomal chromosomes there are two copies of each gene, a mutation in just one copy (heterozygous) is sufficient to cause disease. The chance of passing the abnormal copy of the gene to a child (recurrence risk) is 50%. The majority of congenital cardiac disorders are inherited in an autosomal dominant pattern with several family members affected. The non-syndromic forms are mainly represented by such familial cardiomyopathies and congenital arrhythmias. The clinically used term "familial" as an aetiological class for cardiomyopathy typically implies an underlying pathogenic variant in sarcomeric genes, (hypertrophic, dilated, restrictive and left ventricular non-compaction cardiomyopathy) cytoskeletal genes (hypertrophic cardiomyopathy, dilated and restrictive cardiomyopathy) or desmosomal genes (dilated and arrhythmogenic right ventricular cardiomyopathy). Inherited arrhythmias (such as long QT syndrome, Brugada syndrome and cathecolaminergic polymorphic ventricular tachycardia) are known as channelopathies since the underlying genetic defects affect ion channel subunits or proteins that regulate them.

1.1.1.1.1 Arrhythmogenic right ventricular cardiomyopathy (ARVC)

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is an autosomal dominant disease, characterized by fibro-fatty replacement of the right ventricular myocardium that triggers ventricular arrhythmia and can cause sudden cardiac death (SCD), mainly in young individuals[5] (Figure 1). Despite its low prevalence, it is responsible for approximately 20% of SCD cases among young individuals (< 30 years), rarely presenting before the age of 12 or after the age of 60[6]. The estimated prevalence ranges from 1 in 2000 to 1 in 5000, with Caucasians and athletes being more frequently affected[7].

ARVC is primarily caused by mutations in genes encoding proteins of the desmosome and the area composita (junctional structure between cardiomyocytes, composed of both desmosome and adherens junction proteins)[8]. Genes associated with ARVC phenotype include plakophilin 2 (*PKP2*)[9], desmoplakin (*DSP*)[10], desmocollin 2 (*DSC2*)[11], desmoglein 2 (*DSG2*)[12] and junction plakoglobin (*JUP*)[13], with *PKP2* representing the main contributor to ARVC pathogenesis[14–16] Other genes that have been implicated in ARVC include transmembrane protein 43 (*TMEM43*)[17], cardiac ryanodine receptor (*RYR2*)[18,19] transforming growth factor beta 3 (*TGFB3*)[20]. More genes are involved (*SCN5A*[21], *DES*[22], *TTN*[23], *PLN*[24], *LMNA*[25], *CTNNA3*[26]*)* with only a few cases described. Still, 50% of ARVC cases are unsolved[27].

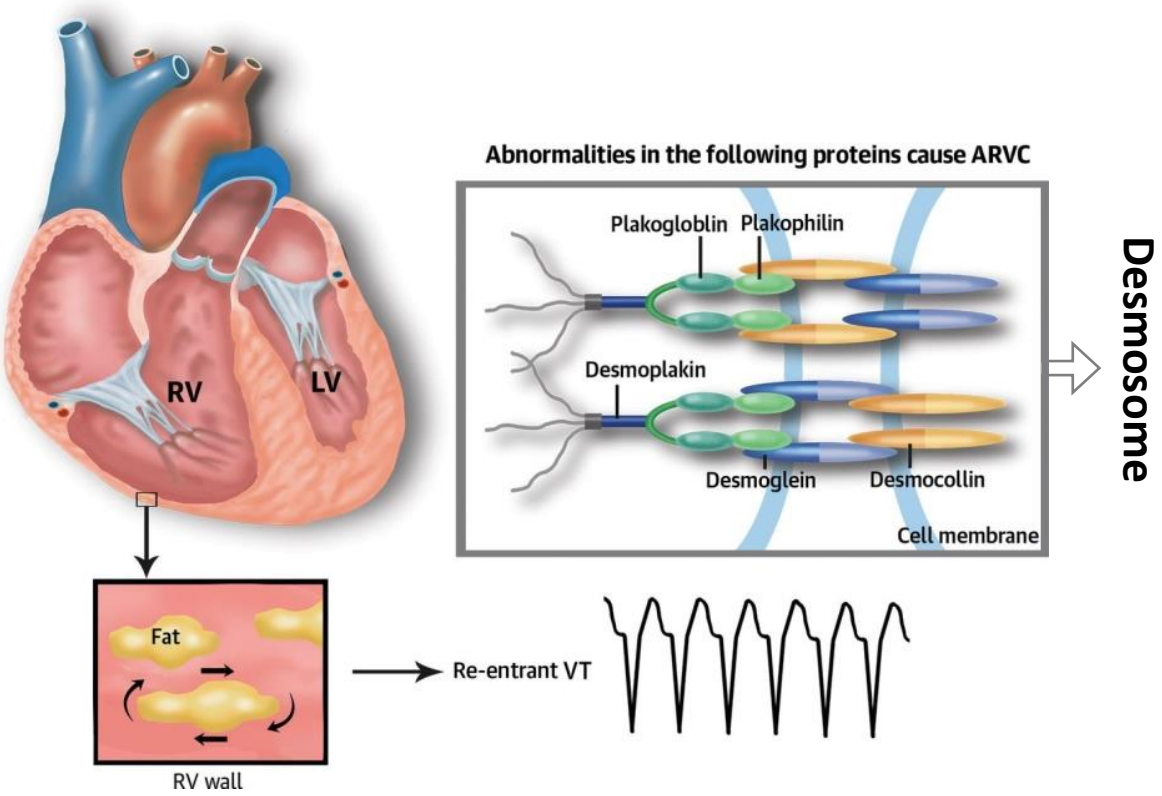

Figure 1. Arrhythmogenic right ventricular cardiomyopathy (ARVC) is characterized by fibro-fatty replacement of the right ventricular myocardium, triggering ventricular arrhythmias (for example, re-entrant ventricular tachycardia as shown in the figure). Mutations are mostly localized in genes encoding proteins of the desmosome. [Modified from [28]]

1.1.1.1.2 Long QT syndrome (LQTS)

Long QT syndrome (LQTS) is a leading cause of SCD among young individuals [29]. Clinically, LQTS is represented typically by a prolongation of the QT interval on the electrocardiography (ECG) and by syncopal events or cardiac arrest, mainly triggered by emotional or exercise stress (Figure 2). SCD can also occur at rest or during sleep. In the typical form, LQTS shows a prevalence of 1:2000 and autosomal dominant pattern of inheritance[30].

Sixteen genes have been identified as responsible for LQTS[31] (Figure 2). The three main genes, *KCNQ1* (LQT1), *KCNH2* (LQT2), and *SCN5A* (LQT3), account for approximately 75% of LQTS cases, while overall the minor genes are related with an additional 5%[31]. *KCNE1* (LQT5) and *KCNE2* (LQT6) encode K+ channel accessory subunits of the protein channels encoded by *KCNQ1* and *KCNH2*[32,33]. In the group of Na+ channels, *AKAP9* variants have been implicated in LQT11 [31]. *CAV3*[34], *SCN4B*[35], and *SNTA1*[35] are considered additional LQTS genes (LQT9, LQT10, and LQT12), clinically showing a LQT3-type phenotype. LQT13 has been linked with heterozygous variants in *KCNJ5* detected in a large four-generation Chinese family[31]. *ANKB*, *KCNJ2*, and *CACNA1C* have been associated with clinical syndromes, also named as LQT4, LQT7, and LQT8[31].

Mutations in genes encoding calmodulin (*CALM1*, *CALM2* and *CALM3*) have been found in association with the most malignant form of LQTS, which causes recurrent cardiac arrest due to ventricular fibrillation manifesting in infancy[36,37]. WES revealed de novo mutations in either *CALM1*, *CALM2* and *CALM3* in unrelated infants[36,38] . Recently, WES allowed the identification of a novel recessive form of LQTS related to biallelic frameshift variants in *TRDN*[39].

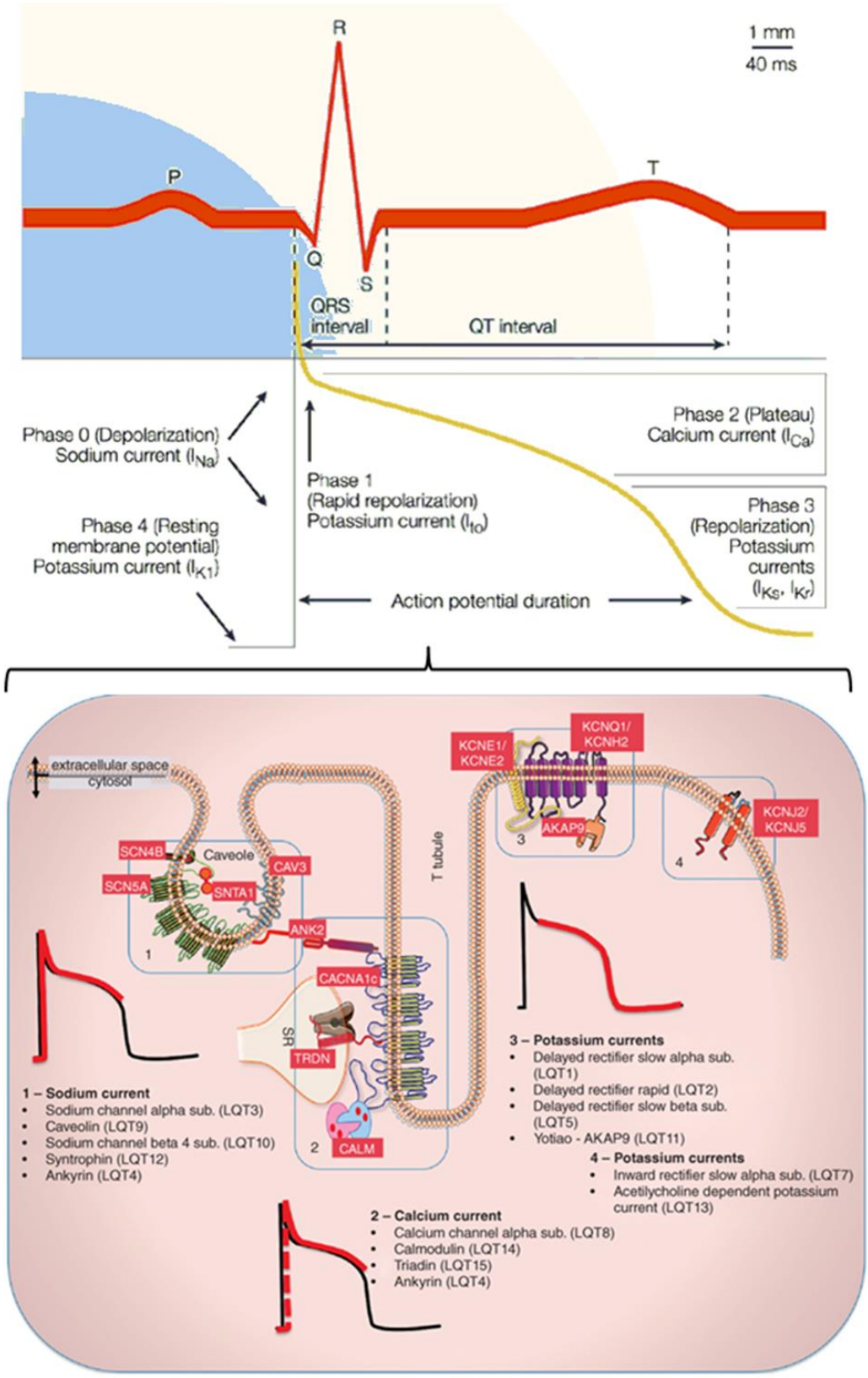

Figure 2. Long QT syndrome (LQTS) is defined by the prolongation of QT interval on the electrocardiogram and it caused by alterations of the cardiac action potential due to mutations in genes encoding cardiac ion channels. [Modified from [40,41]]

1.1.1.2 Autosomal Recessive Disorders

Two mutant alleles of a disease-associated gene are required to cause an autosomal recessive disease. In other words, autosomal recessive diseases are caused either by a homozygous mutation or by two compound heterozygous mutations in the disease-associated gene. In these families, the recurrence risk is 25%. Inborn errors of metabolism represent an important cause of cardiomyopathy in children. They are clinically heterogeneous disorders and often can be inherited in an autosomal recessive manner. The term inborn error of metabolism refers to diseases caused by defects in intermediary metabolism or energy production processes, such as fatty acid oxidation disorders, organic acidemias, storage disorders and mitochondrial disorders.

1.1.1.2.1 Mitochondrial cardiomyopathy

Mitochondrial disorders consist in a heterogeneous group of inborn errors of metabolism showing a wide spectrum of clinical manifestations, with approximately 300 disease-associated genes[42].

Mitochondria exploit the energy stored in fats, carbohydrates, and proteins producing ATP in a process defined as oxidative phosphorylation (OXPHOS). OXPHOS requires four multi-subunit complexes (complex I - complex IV) and two mobile electron carriers, coenzyme Q10 (CoQ10) and cytochrome c. The respiratory chain produces a transmembrane proton gradient that is used by complex V to synthesize ATP[43].

As a consequence of energy deficiency, high-energy demand tissues such as muscle, brain and liver can be impaired, with usually resulting multi-organ disease.

With an estimated prevalence of 1 in 5000 live births, mitochondrial disorders are among the largest groups of inborn errors of metabolism[44]. The onset of mitochondrial diseases peaks in early childhood (first 3 years of life) with a second broad peak towards the second and the fourth decade of life (adult-onset diseases)[44].

Mitochondrial diseases with onset in childhood are usually more severe, in particular when the clinical presentation occurs in infancy or if it involves the heart (mortality of 71% vs. 26% without cardiac phenotype)[45]. Cardiomyopathies have been described in 20–40% of children with mitochondrial diseases[45,46]. Hypertrophic cardiomyopathy is the most common form of cardiomyopathy in mitochondrial disease, but also dilated cardiomyopathy and left ventricular non-compaction appear quite often[47,48]. Conduction system and bradyarrhythmias are the most frequent arrhythmogenic manifestations[49,50].

The advent of NGS drastically improved the diagnostic yield for mitochondrial diseases, allowing the analysis of both the mitochondrial genome (mtDNA) and nuclear genomes. In particular, WES has identified a constantly growing list of new disease genes, leading to the discovery of several disease mechanisms[51–53].

Defects of the mitochondrial energy metabolism can be related to mutations in different mitochondrial pathways and functions, such as OXPHOS system and subunits, transport processes of proteins and substrates across the mitochondrial membranes, quality control systems, essential cofactors, organelle, and membrane integrity (Figure 32).

Mitochondrial genetics is heterogeneous, and the mitochondrial proteome (around 1500 proteins) is controlled by two genomes, the mtDNA and nuclear genome, with the latter accounting for up to 80% of mitochondrial disease in children[54]. In children, the most common mode of inheritance of mitochondrial disease is autosomal recessive, but dominant and X-linked forms are also possible.

1.1.2 Complex disorders

Complex disorders are influenced by a combination of multiple genes and environmental factors.

1.1.2.1 Congenital heart defects (CHDs)

Congenital heart defects (CHDs) are the most common type of birth defect and they are associated with a significant mortality[55]. They affect around 1% of all live births[56].

Genetic factors are known to be involved in the development of CHDs, but the nature of these genetic defects is complex and heterogeneous[55,57–59].

Among the genetic causes of CHDs, more than 110 genes have been associated with human syndromes[60]. Thirty of them have been so far linked with syndromes that include left-sided heart malformations. These loci include those implicated in hypoplastic left heart syndrome, aortic stenosis and coarctation of aorta[55].

Extra-cardiac features are frequently present and they suggest an early pleiotropic role for the underlying molecular pathways in organ development. Left-sided heart malformations without extra-cardiac manifestations (apparently isolated or non-syndromic) seem to have a more complex origin[55]. Familial clustering of cases and an increased risk in first-degree relatives strongly imply a single gene inheritance[61,62]. However, the severe forms (e.g. hypoplastic left heart syndrome, HLHS) are frequently manifesting in sporadic cases, thus suggesting a role for de novo mutations and multi-locus variation[60,63,64].

1.1.2.1.1 Hypoplastic left heart syndrome (HLHS)

Hypoplastic left heart syndrome (HLHS) is characterised by malformation of the structures of the left side of the heart, which encompasses the mitral valve, left ventricle, aortic valve, and aortic arch[65] (Figure 3). Hypoplastic left heart syndrome has a birth incidence of approximately 1/5000 live births and it accounts for 1.4–3.8% of CHDs[66–68] and 23% of cardiac deaths occurring in the first week of life. Males are more frequently affected (3:2 males/females ratio)[69].

The probability of recurrence in a subsequent pregnancy (sibling recurrence risk) is approximately 2–3%[70,71]. A higher risk of recurrence of CHDs have been observed in families with HLHS cases[70]. Based on population data, it was found that 13.5% of first degree relatives of children with HLHS had symptomatic CHD themselves[72]. These findings suggest

a genetic aetiology for several cases of HLHS. However, large studies assessing recurrence risk are still missing.

Inheritance patterns for HLHS have been difficult to determine because of the severity of the phenotype, but evidence suggests more than one pattern of inheritance[65]. Familial recurrence has been reported in consanguineous families with multiple affected siblings, suggesting an autosomal recessive pattern of inheritance[73]. Autosomal dominant inheritance in non-syndromic left-sided defects and an increased incidence of valvular heart defects in parents of a child with HLHS have been reported[74]. HLHS has been described as part of at least 32 syndromes, some of which have a known genetic aetiology, for example, Holt Oram syndrome and Rubinstein Taybi syndrome[65].

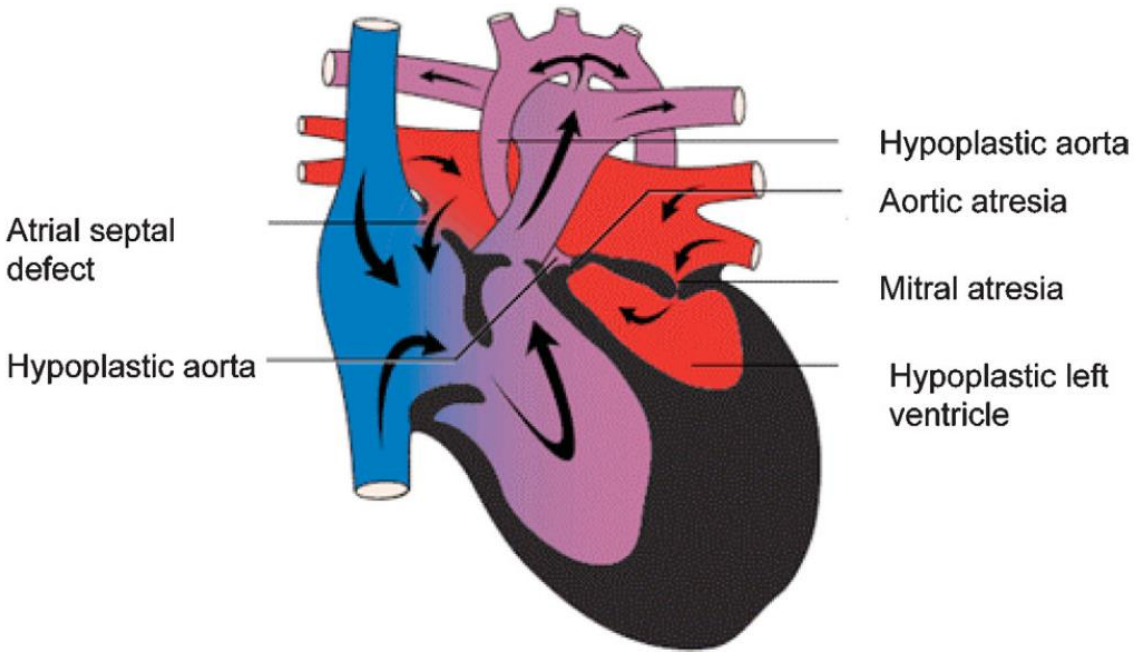

Figure 3. HLHS is defined as underdevelopment of the structures of the left side of the heart, which includes the mitral valve, left ventricle, aortic valve, and aorta. [Taken from [65]]

1.2 Whole exome sequencing (WES)

First described in 2009, whole exome sequencing (WES) is a next-generation sequencing (NGS) technique that consists in the target sequencing of the exome, the portion of the genome including all the protein-coding regions, the exons[75](Figure 4). The term exon derives from "EXpressed regiON," regions that get translated (or expressed) as proteins, compared to the intron, or "INTRagenic regiON" which is not expressed in the final protein[76]. The exome is formed by circa 230,000 exons, representing about 1,5% of the human genome, or approximately 50 million base pairs. It has been estimated that exons harbor about 85% of known mutations with large effects on disease-related traits. Exonic mutations were shown to cause the majority of monogenic diseases[77], with missense and nonsense mutations alone accounting for approximately 60% of disease mutation[78]. WES greatly increased the identification of novel Mendelian disease genes[79] as reflected by the almost 4000 new phenotypic entries in OMIM since 2008[80] (current total: 6107) and an average of 300 genes discovered by WES every year (current total: 3839). Moreover, WES has become the gold standard in molecular diagnostics of Mendelian disorders[81].

Large-scale WES projects[82,83] provided crucial information on variant frequencies in different populations. This process allowed re-evaluating variants previously assumed to be disease-causing and exploring gene function in general[84].

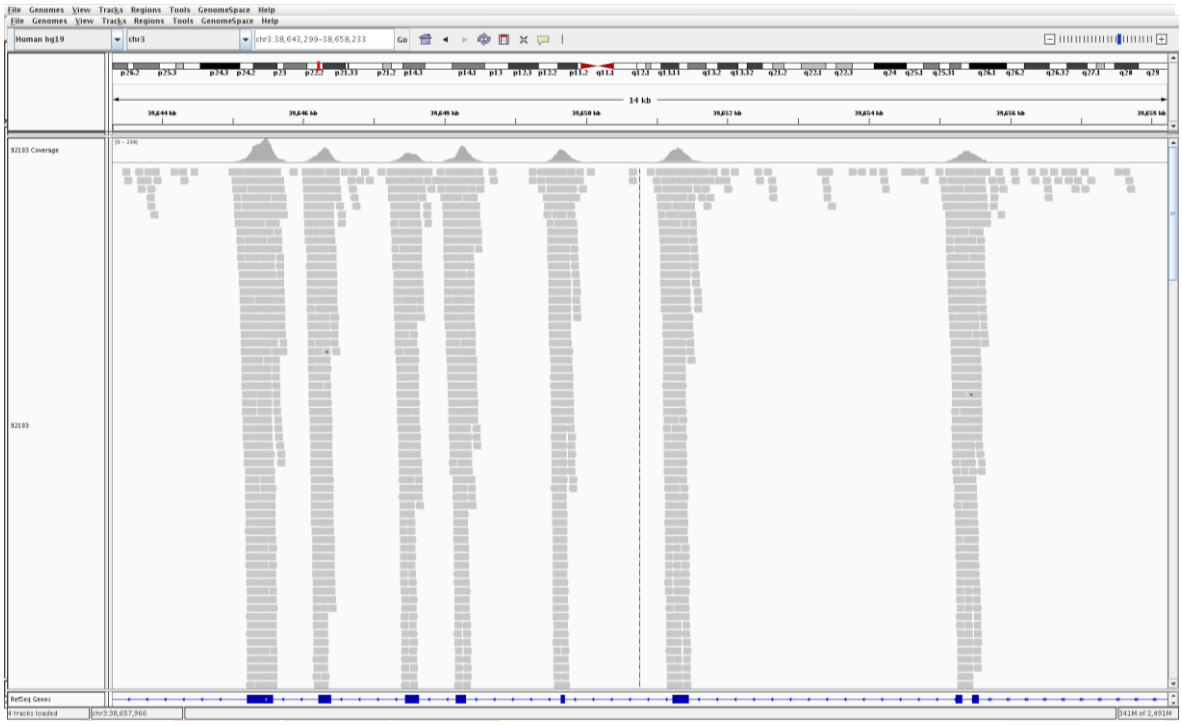

Figure 4. Integrative Genomic Viewer (IGV) visualisation of 7 exons in *SCN5A*. Whole exome sequencing technique is based on the target analysis of the selected coding region of the genome, the exome. In the figure, for each exon present in the reference (blue rectangles) corresponds an enrichment of the exon in the sample, showed by the coverage tracks and the aligned short reads below them (grey bars). Details of these aspects will be provided below.

1.2.1 DNA Sequencing

Since the time DNA was discovered, researchers have sought that if the genetic code could be sequenced or "read", the origins of life itself may be revealed. Although this concept might not be entirely true, the efforts made have certainly revolutionized the biological field[85].

The process of "reading the DNA" is called DNA sequencing and it consists in determining the order of the four different nucleotides (Adenosine, Guanine, Cytosine and Thymine) of a DNA molecule. After the 1962 Nobel recognition for the discovery of the structure of the DNA, it took less than 20 years (1980) for the Nobel Prize in Chemistry, half jointly awarded to Walter Gilbert and Frederick Sanger, "for their contributions concerning the determination of base sequences in nucleic acids"[86]. The Maxam-Gilbert method used chemical breakage, radioisotope and labelling plus gel electrophoresis to sequence the DNA, while the Sanger method utilized radiolabelled di-deoxynucleotide triphosphates (ddNTPs) and gel electrophoresis. Sanger's method had been the sole method for DNA sequencing for three decades, as a result of its lesser technical complexity and a lesser amount of toxic chemicals used, compared to the Maxam-Gilbert method[86].

In 1977 Sanger published the details of the sequencing technique called "chain termination method", subsequently known as Sanger Sequencing[87]. Briefly, this sequencing technology

requires a specific primer to start the read at a specific location along the DNA template, and record the different labels for each nucleotide within the sequence. Sanger sequencing allowed the analysis of the nucleotide sequence of a given portion of DNA. After a series of technical improvements, the Sanger method has reached the maximal capacity to read through 1000–1200 base pairs (bp)[85].

The discovery of thermostable DNA polymerases led to the development of the polymerase chain reaction (PCR) and subsequently improved methods for DNA sequencing, called thermal cycle sequencing[86].

With the advent of automated sequencing in the 1980s, the idea of analysing the entire genome began to be proposed. The "Human Genome Project" (HGP) effectively started in 1990 and took 13 years and $3 billion to be completed[85]. Aiming to sequence the entire genome, a new approach called shotgun sequencing was developed. In this approach, genomic DNA is enzymatically or mechanically fragmented and cloned into sequencing vectors in which cloned DNA fragments can be sequenced individually. The complete

sequence can be then reconstructed by reassembly of sequence fragments based on partial sequence overlaps. Shotgun sequencing was an important step of the HGP and made sequencing the entire human genome feasible. The core concept of massive parallel sequencing used in next-generation sequencing (NGS) was adapted from this approach[86].

1.2.2 Next-generation sequencing (NGS)

The second complete individual genome (belonging to James D. Watson) was sequenced using next-generation sequencing technology (NGS)[85].

Next-Generation Sequencing, or Second-Generation Sequencing, technologies refer to high-throughput massively parallel sequencing of thousands to millions of DNA fragments simultaneously. This massively parallel analysis is achieved by the miniaturization of the amount of individual sequencing reactions, which influences the size of the instruments and the cost per reaction. The read length (the actual number of continuous sequenced bases) for NGS is much shorter than that obtained by Sanger sequencing. NGS provides 50–500 continuous base pair reads and therefore sequencing results are defined as short reads. NGS technologies reduced dramatically sequencing costs to $10^{-5}$$ per base pair (bp) compared to automated Sanger sequencing (First-Generation Sequencing)[88] (Figure 5). Since his introduction in the mid-2000s, NGS technology has drastically revolutionized our understanding of many genomic regions involved in the pathogenesis of human diseases[85].

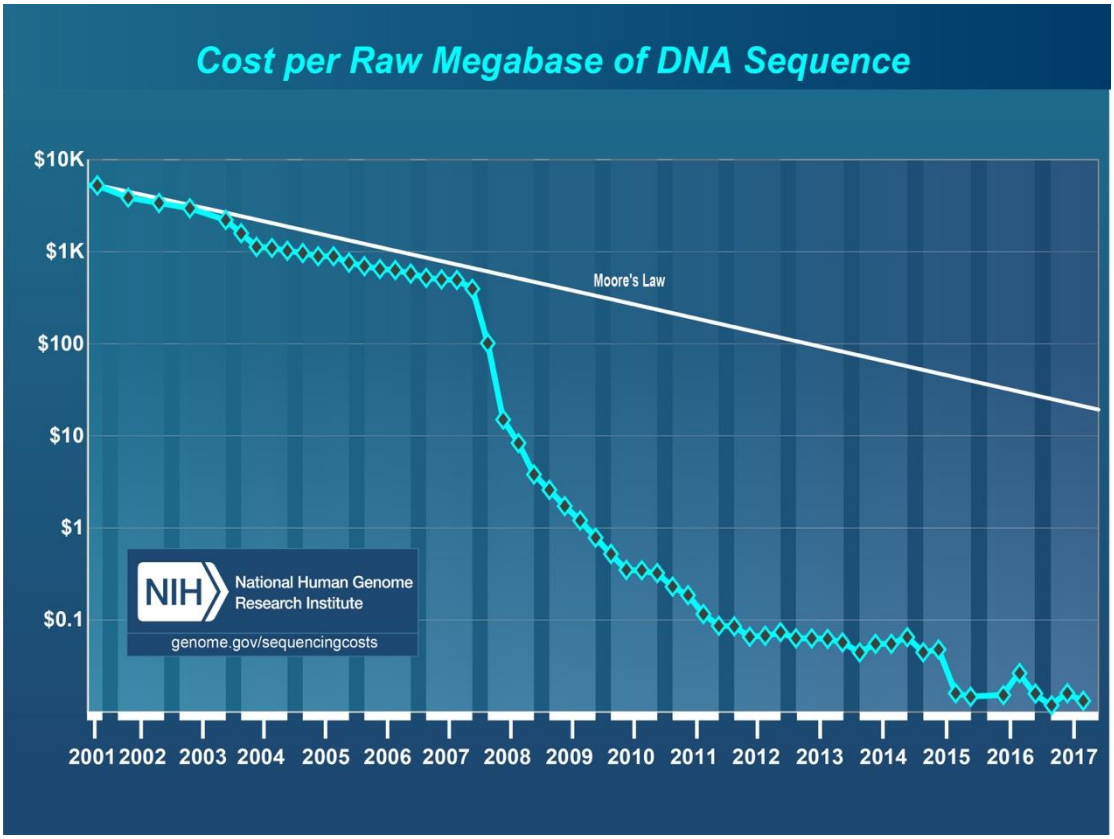

Figure 5. Graph of sequencing costs per raw Megabase of DNA sequence. Moore's law (hypothetical computer industry that predicts the doubling of technology every 2 years) is illustrated in the graph. The abrupt drop in sequencing costs in 2008 represents the change from Sanger-based sequencing to the next-generation technologies. [Taken from [89]]

NGS platforms

Short read sequencing techniques are divided into two categories: sequencing by ligation (SBL) and sequencing by synthesis (SBS). In SBL, a probe sequence that is bound to a fluorophore hybridizes to a DNA fragment is ligated to an adjacent oligonucleotide. The emission spectrum of the fluorophore corresponds to the bases complementary to specific positions in the probe. In SBS, a polymerase is used and a signal, such as a fluorophore or a change in ionic concentration, identifies the incorporation of a nucleotide into an elongating strand. In most SBL and SBS approaches, DNA is clonally amplified on a solid surface. Having thousands of identical copies of a DNA fragment in a defined area, the so-called cluster, allows the distinction of the signal from background noise[90]. A sequencing platform can collect information from millions of reaction centres simultaneously, thus enabling sequencing millions of DNA molecules at the same time.

Between 2005 and 2007, three companies produced distinct NGS platforms. One used SBL sequencing, Roche (454 sequencing), whereas the others adopted the SBS sequencing method, Illumina (Solexa technology) and LifeTechnologies (ABI SOLiD sequencing).

The Illumina SBS-based systems now account for the largest market share for sequencing instruments compared to other platforms. Illumina dominates the short-read sequencing industry thanks to its technology, a high level of cross-platform compatibility and its wide range of platforms. The suite of instruments available ranges from the low-throughput MiSeq to the ultra-high-throughput NovaSeq, which is capable of sequencing more than 2000 human genomes to 30× coverage (number of reads) per year. Further diversification is derived from the various options available for runtime, read structure and read length (up to 300 bp). Illumina NGS workflows include the following basic steps.

Pre-sequencing steps

a) Library preparation. A key procedural difference in NGS methods in comparison to Sanger sequencing is the construction of the sequencing library. The ease and rapidity of NGS library construction are striking: starting from genomic DNA, few preparatory molecular biology steps are performed. First, the DNA is divided into small fragments, following the so-called "fragmentation" process. Then synthetic adapters are ligated to the library fragment ends ("ligation") (Figure 6). These common adapters support subsequent PCR amplification cycles that produce a library ready for quantification, dilution and sequencing. This process requires approximately two days. Adapters can contain the so-called "index sequences" (4-10 bp) that allow distinguishing different samples when pooled and sequenced together. This process is named "barcoding"[91].

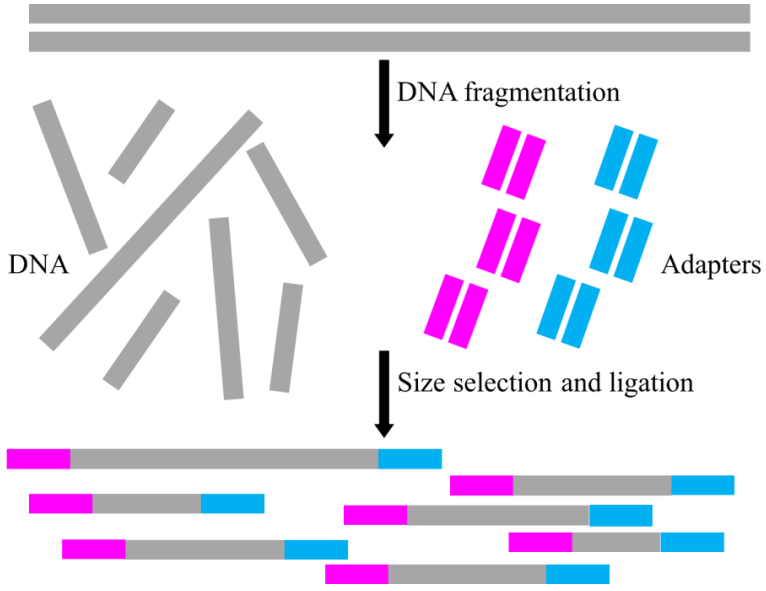

Figure 6. Schematic illustration of the Illumina Library Preparation step. Fragmentation and ligation processes: the DNA is divided into small fragments and synthetic adapters are ligated to the library fragment ends. [Figure modified from Illumina, Inc.[92]]

b) Exon capture. The enrichment of selected areas of the genome by hybridization to known sequences (known as probes) is called capture. The capture was initially developed for the research market and, in the case of WES, was adapted to allow genome-wide coverage and commercial viability. Commercial enrichment kits are offered from Agilent (SureSelect Human AllExon Kit), Illumina (TruSeq Exome Enrichment Kit), and Roche (Nimblegen SeqCap EZ Exome), differing in the bait type (DNA or RNA) and bait length, as well as the captured regions. The sample preparation protocols share basic steps, but important differences can be seen in the design of the oligonucleotide probes (for the selection of target genomic regions, sequence characteristics and lengths of the probes) and the exome capture phases[93]. With the advancement of the kit, the efficiency of the capture, defined by the coverage (number of reads in a certain genomic position), has improved. Accordingly, the list of exons well "captured" and subsequently covered has become higher and higher. During my PhD studies, for example, were used different versions of the enrichment kit: from Agilent SureSelect Human All Exon Kit version 2 in 2014 till version 6 in 2017. In Figure 7 are shown the differences between kit version 2 and kit version 6 for the exons contained in *KCNH2*, a candidate gene for long QT syndrome.

In this work, the library preparation started with 3 μg of high-quality genomic DNA and subsequently samples were sheared to generate small DNA fragments with a median length of 250bp. The Covaris E220 focused-ultrasonicator was used for the shearing process. After DNA shearing, libraries were prepared and the Illumina Paired-End Sequencing Library protocol was followed, adding Illumina compatible adapters and indices. Processes were fully automated by use of the Bravo and Bravo BenchCel 4 R liquid handling platform of Agilent Technologies. Genomic libraries were mixed in solution with an excess of 120 bp long, biotinylated RNA oligonucleotides, termed probes or baits, which are complementary in sequence to the targeted exons. These RNA baits annealed to their targets and were pulled down with streptavidin-coated magnetic beads. The caught sequences were then PCR-amplified with universal primers. Quality and quantity measurements were performed by means of Agilent 2100 Bioanalyzer and the Agilent DNA 1000 kit (Figure 8).

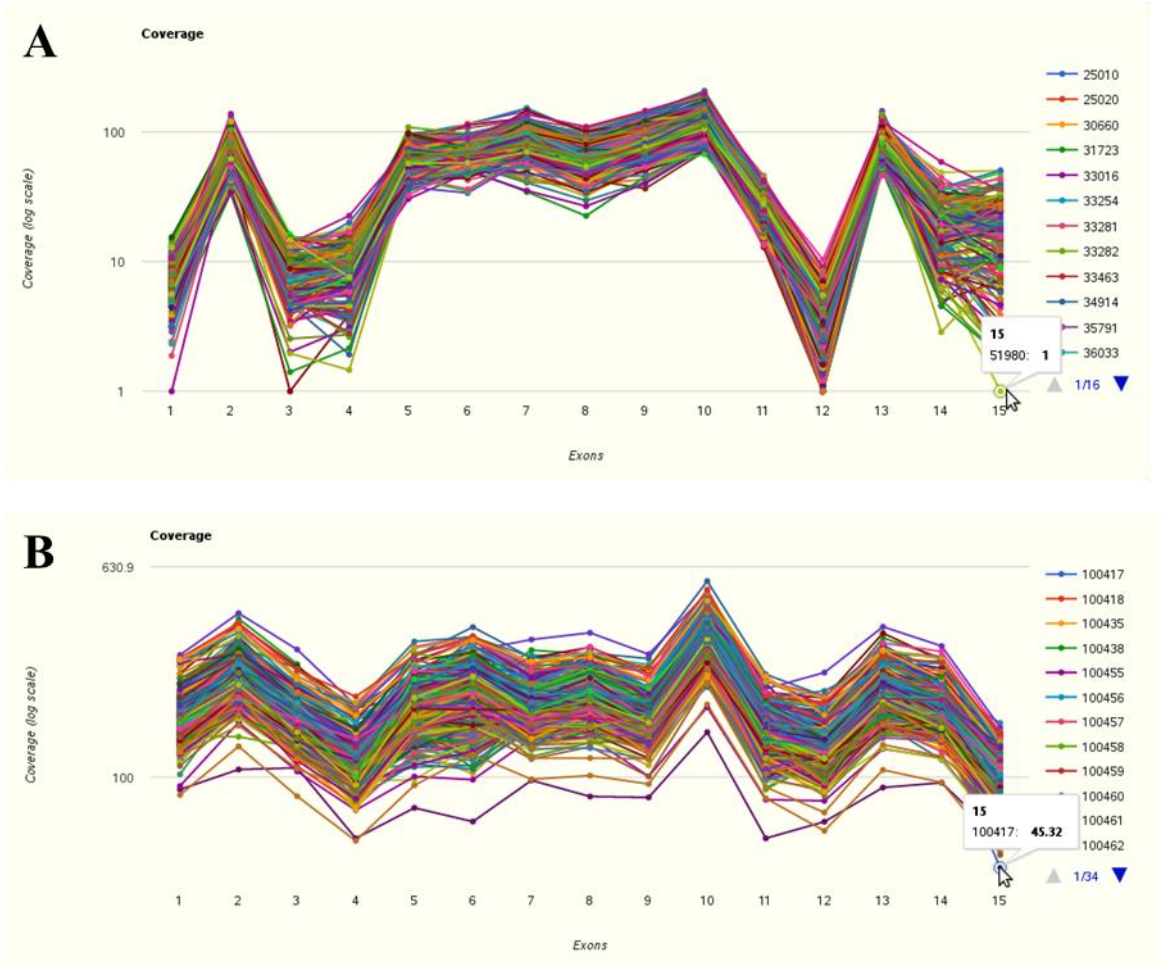

Figure 7. Coverage using two different enrichment kits in *KCNH2*, containing 15 exons. A) Agilent SureSelect Kit version 2; B) Agilent SureSelect Kit version 6. The coverage values indicate the number of reads. Coverage > 20 is generally required in diagnostics. The coloured lines represent the samples sequenced. Version 6 guaranteed a coverage > 40 in all the samples considered, with an average of coverage far above 100 (maximum value 620), whereas with version 2 some exons are not appropriately covered (<10) and 100 is the maximum value reached.

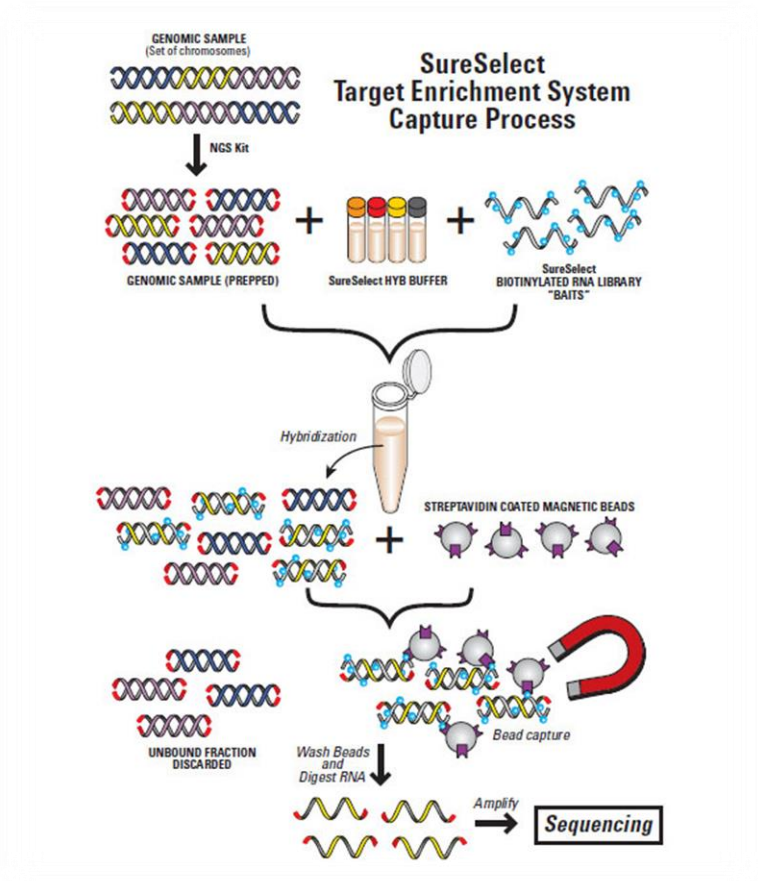

Figure 8. SureSelect target enrichment system capture process. The genomic library is mixed in solution with capture probes. Specific probes bind complementary exomic sequences and are captured via magnetic beads. [Taken from [94]]

c) Attachment to Flow Cell. A flow cell is a glass slide composed by one (MiSeq) or eight lanes (higher throughput instruments). A lane is defined as a channel containing a lawn of oligonucleotides. Using a machine named cBot the prepared DNA fragments of the samples to be sequenced can be attached to the flow cell. As mentioned above, the oligonucleotides attached to the flow cell are complementary to the adapters ligated to the DNA fragments in the library preparation step[95].

d) Cluster Generation. The sequencing process is based on the detection of a fluorescent signal, which is emitted when a labelled dNTP binds to a fragment. Since the signal from a single incorporation process would be too weak to be detected, the fragments are amplified. This process is done by bridge amplification (Figure 9). In this process, the single fragments are copied multiple times in order to produce dense clusters constituted of ≈1,000 fragments with identical sequence information. This process takes place inside the cBot. Once the fragments are amplified the flow cell can be transferred to the sequencer[95].

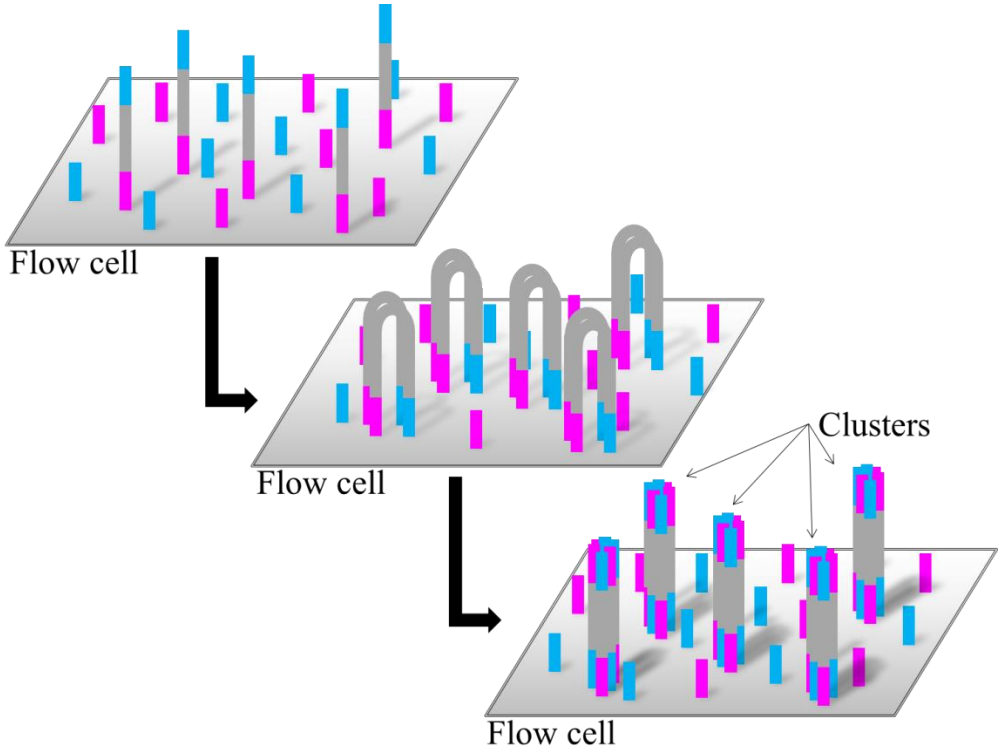

Figure 9. Schematic representation of cluster generation. Single-stranded fragments bind to the flow cell. Clusters are then generated by bridged amplification. [Figure modified from Illumina, Inc. [96]]

Sequencing steps

When the sequencing process is started, a universal sequencing adapter is hybridized to the single-stranded fragments. Sequencing is then performed in cycles and, in each cycle, the complementary sequence of the fragment is extended by one base (incorporation phase). In each cycle DNA polymerases and modified dNTPs are washed through the flow cells and the polymerase extends the appropriate dNTPs to the growing sequences. The modification of dNTPs includes a reversible terminator with four different removable fluorophores, one for each type of dNTP. In this way, only one dNTP binds the growing sequence every cycle and the incorporated dNTP can be detected. The surplus of polymerases and dNTPs is washed away and the incorporated bases are identified by laser-induced excitation of the fluorophores and imaging of the signal. Then, the terminators and fluorophores are removed (cleavage phase) and a new cycle starts[95].

Based on the fluorescent signals of each cycle, the Illumina software decodes the actual sequence of each fragment that was loaded on the flow cell, assigning the according base to each cluster (Figure 10). This process is named “base calling”.

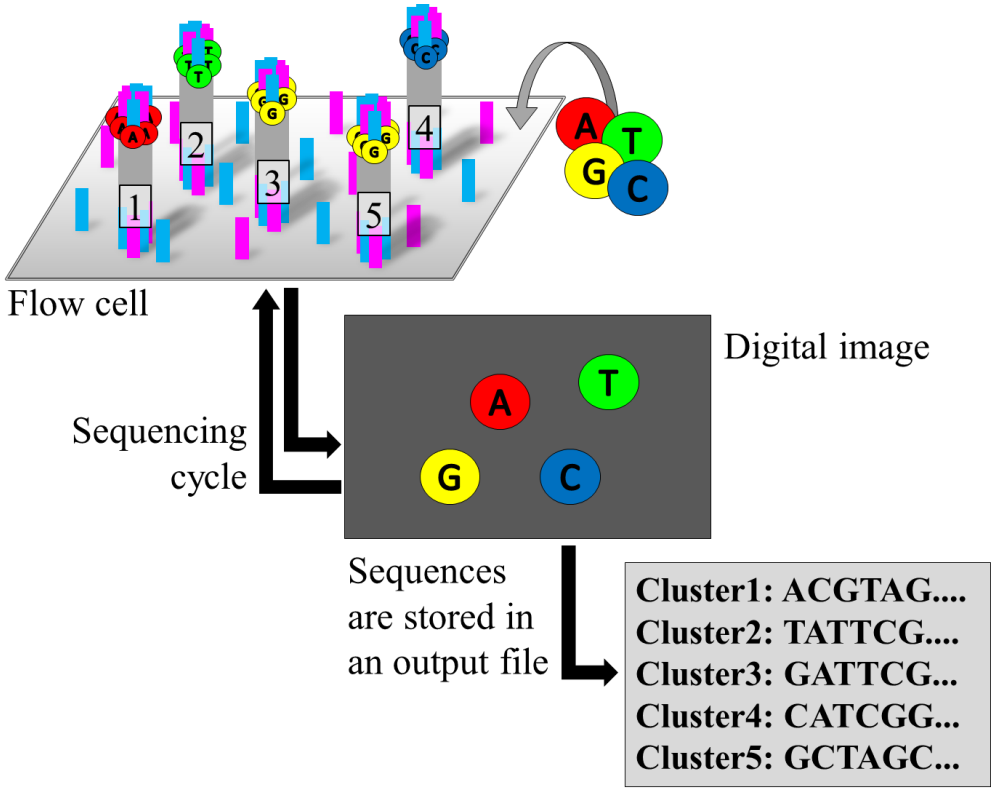

Figure 10. Schematic illustration of the Illumina Sequencing by Synthesis technology. In each cycle, an according fluorescently labelled dNTP is incorporated into the growing sequence and the respective emitted fluorescing signal is detected by a camera. [Figure modified from Illumina, Inc. [96]**]**

Early versions of Illumina technology (2008) were able to perform 35 cycles[97], in other words, sequence 35 bp per fragment. This is related to the fact that not all fragments of a cluster incorporate a dNTP in each cycle, and some can incorporate nothing or more than one. The higher the number of cycles (i.e. the longer the fragment is sequenced) the higher the fraction of signals which are not in "frame". This increasing noise reduces the quality of the base call, a phenomenon that is called "dephasing". Nowadays, the continuously improved chemistry, reagents and machinery can achieve read lengths between 150 and 300 bp. Furthermore, Illumina system supports paired-end sequencing (Figure 11). In brief, after the given number of bases is read, a sequence is obtained and it is called read 1 (R1); if the sequencing is performed in paired-end mode, the machine automatically regenerates the clusters and starts reading the fragment of the opposite strand. This process will run for the same number of cycles, obtaining another sequence named read 2 (R2).

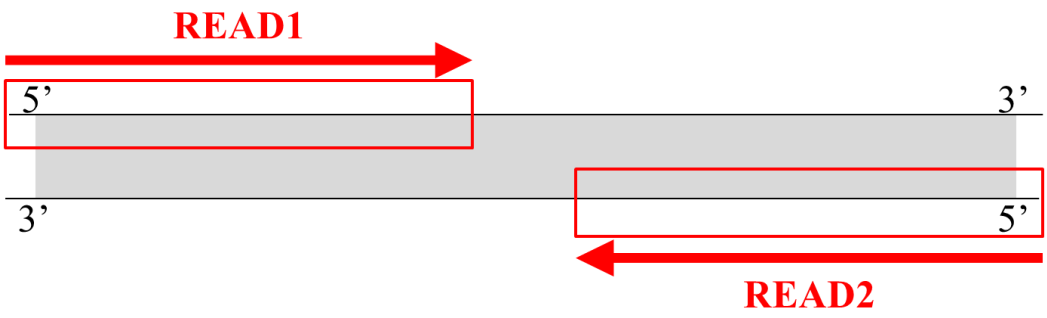

Figure 11. Pair-ended sequencing: each DNA fragment is read from both ends.

The data are stored in FASTQ files, one for the reads from R1 and one for the reads from R2. Each entry of the file contains a unique header, which allows linking the sequences of R1 and R2, the sequence itself, and a quality string that reflects the optical parameters of the sequence. Each nucleotide of the sequence has an associated quality, which is named base quality (BQ) and is represented using the Phred-scale. Phred-scale is a logarithmic scale that represents the error probability, where 10, 20, 30 correspond to the probabilities of error of 0.1, 0.01, 0.001. Therefore, the higher is the number, the lower is the error probability. On average an Illumina run produces qualities between 30 and 40[95].

Sequencing fragments from both ends offer a consistent advantage in data analysis. To illustrate the results of this type of sequencing, we can assume that a fragment of 450bp is sequenced paired-end 150+150bp; this means that we have read 150 bp from the 5’→3’ direction and 150bp from the 3’→5’ of the original fragment, so we know 300bp out of the 450bp of the original fragment. The technological advancement improved not only the read length, but also the number of fragments that can be sequenced. Currently, Illumina systems with NovaSeq can produce up to 2400-3000 Gb of sequence in less than three days. In this study, the samples were subjugated to clustering and sequencing as 100 bp paired-end runs on HiSeq2000, Hiseq2500 and Hiseq4000 systems. The HiSeq 3000/HiSeq 4000 Systems were the first Illumina sequencers to use patterned flow cells for diverse genomic applications. Patterned flow cells use distinct nanowells for cluster generation to make more efficient use of the flow cell surface area. This advanced flow cell design contributed to increase data output, reduce costs, and faster run times (Figure 12). The NovaSeq Series unites the latest high-performance imaging with the next generation of Illumina patterned flow cell technology. The updated NovaSeq flow cell design further reduces the spacing between nanowells, significantly increasing cluster density and data output.

However, the revolution of WES is not only linked to the improvement in the sequencing technology, but also to the development of bioinformatics pipelines and the frequency information given by large-scale WES projects[98].

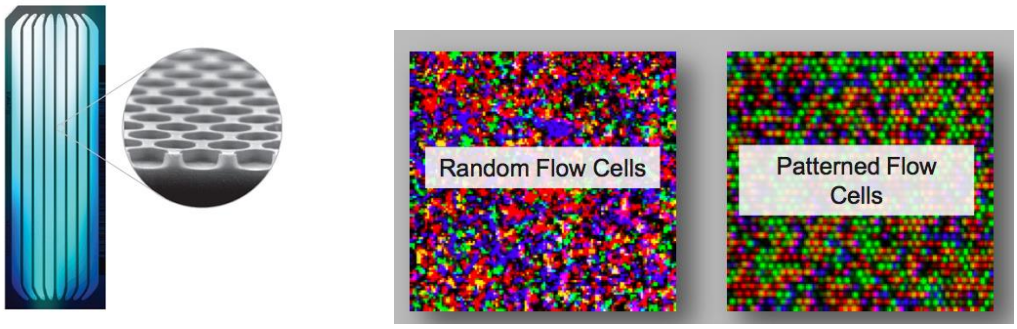

Figure 12. Patterned Sequencing: patterned flow cells contain billions of nanowells at fixed locations across both surfaces of the flow cell. Precise nanowell positioning eliminates the necessity to map cluster sites and reduces of hours the time of each sequencing run. Higher cluster density allows to use more data per flow cell, dropping the cost of the sequencing run. [Taken from Illumina, Inc. [96] and [99]]

Post-sequencing steps (data analysis)

The process of detecting Single Nucleotide Variants (SNVs) and small insertions and deletions (indels) from raw NGS data can be divided into three parts:

a. Alignment: the short NGS reads are aligned to a reference genome (Figure 13).

b. Variant Calling: identification of sites different from the reference sequence.

c. Variant Filtering and Annotation: The called variants can be filtered to remove low-quality variants, and annotated with additional information such as the effect of a variant within a gene[95].

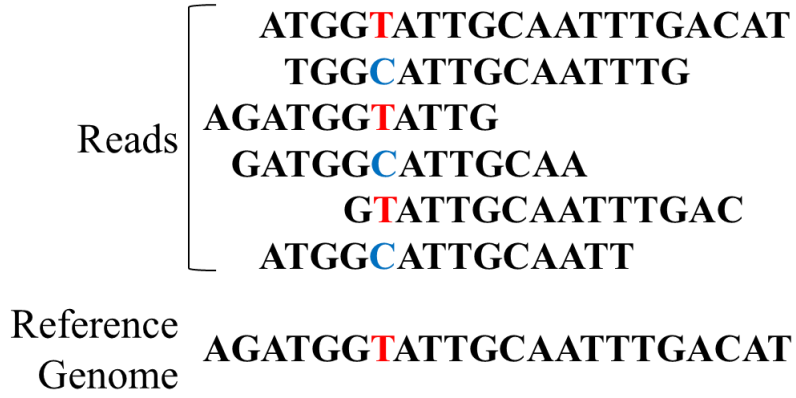

Figure 13. Post sequencing steps: after alignment of the sequenced reads, it is possible to compare them with the reference genome to spot variants.

1.3 Data analysis

As mentioned, the sequencing of a whole exome results in one or more text files in FASTQ format. To make use of this information, the exome of the sequenced individual needs to be constructed. The data is subsequently processed by aligning the short read sequences to a reference genome and then calling single nucleotide variants (SNVs) and small insertions and deletions (indels) resulting from the comparison of the input sequence to the reference sequence. Low-quality variants are filtered out and the remaining high-quality variants are annotated.

1.3.1 Quality scores and compression

Since more and more individuals are sequenced as part of clinical medicine, long-term storage and retrieval of their data is an increasing need.

Each base that is called by a sequencing machine has an associated score. Traditionally, these values have been reported in a format known as “phred score”, which was originally derived from chromatogram traces of early sequencers. The number is expressed as the negative log 10 of the probability of error: q= -10log(p). The most common cut off is Q20, which corresponds to a 1-in-100 chance that a base call is incorrect. Score information represents a large amount of data in an alignment file. To improve compression a smaller set of possible score values has been introduced in the latest years.

1.3.2 Reference genome

The human genome has historically been defined by its reference sequence. The human reference is the product of the Human Genome Project and was derived from the DNA of more than 50 individuals. Clones representing single haplotypes of these individuals were sequenced by a shotgun approach and then patched together in one haploid sequence[3,100,101]. The reference used in this study is the hg19/GRCh37 version of the human genome.

An important aspect to consider is that the GC richness of the genome varies with the first exons generally showing higher GC content than the overall average of around 40%[102]. These areas represent a problem for DNA capture and sequencing, but they also have a functional importance driven by CpG motifs, which are thought to be particularly sensitive to mutation and are clustered in islands near the 5´ end of genes. Another major challenge comes from the repeated sequence, which represents more than 50% of the genome[3,103]. Common types of repeats include segmental duplications, simple repeats, short tandem repeats (recently shown

to have an important role in gene expression)[103], transposon-derived repeats and processed pseudogenes.

1.3.3 Alignment

As anticipated above, the output of sequencing is a large text file (i.e. FASTQ) of short reads along with their quality scores (representing base qualities BQ). To derive a complete picture of a single human genome from these 'raw' reads (namely R1, R2, etc.) it must be compared with the reference genome. Therefore, the alignment process consists in finding where in the genome the sequenced fragments (e.g. R1 and R2) belong to.

Typically, short reads are aligned to a reference genome using an algorithm that searches for the best match. First principles might suggest advantages to de novo assembly (i.e. assembling the genome by overlapping the reads without the aid of a reference sequence). However, de novo assembly, mainly of short reads, is computationally intense and impractical for clinical exome sequencing. However, in WES some regions are not sequenced and the probes designed to capture exons are based already on a reference sequence. Therefore, the majority of human exome sequences are currently aligned to a reference sequence. As mentioned above, the reference sequence itself has been the focus of some concern because it was derived from a pool of individuals and, as such, contains pathologically select variants. In addition, it does not accurately represent longer range haplotypes owing to the switching between reference individuals in some regions[104]. Mapping quality will also be poorer in regions of variation.

Alignment in this study will be performed with Burrows-Wheeler-Alignment (BWA) tool which uses the BWT mapping sequence reads to the reference genome by a backwards search. This search algorithm allows mismatches and gaps, therefore enabling the alignment of longer reads harboring indels. In particular, two FASTQ files (e.g. for R1 and R2) are simultaneously aligned to the reference genome. As anticipated, having the two reads from R1 and R2 together (read pair) is extremely important. For example, it improves alignment quality in a repetitive region; it allows to estimate the insert size and to spot translocations when R1 and R2 align on different chromosomes. Moreover, it makes possible the detection of sequencing errors, when only one read pair shows the problem.

Each read pair is supplied by the alignment software with a value named mapping quality (MQ), which is generally in the phred-score range 0-60. This score reflects several parameters, such as the ambiguity of the alignment with respect to the reference and the status of R1 and R2 being both aligned. After alignment, absolutely identical R1+R2 reads are like

PCR duplicates are removed. They may represent possible errors that could bias the subsequent variant calling.

1.3.4 Variant calling

After alignment, comes the variant calling. Variant calling is the piling-up of all sequenced bases aligning to a certain position and the calculation of the proportion of bases differing from the reference[98]. This step allows calling SNPs, indels, CNVs, and translocations. Generally speaking, aligned reads are piled up and each position of the genome is read several times, from independent fragments[95]. The number of reads covering the same genomic position is named read-depth; while the average of the read-depths over the whole genome or the whole target region is named coverage. From the pile up the variant calling tools detect the genotype at a specific position. Heterozygous variants show a proportion higher than 30% while, variants with a proportion higher than 80% are called as homozygous[98]. However, it can occur that this proportion has a borderline value, therefore need to be manually verified to filter out false positive detections.

Different classes of variation and genomic regions have widely varying call accuracy and reproducibility[105]. Recently, a fully characterized single human's diploid genome has set a reference standard for benchmarking sequencing systems and data processing techniques. The NA12878 genome, available in cell lines from Coriell, has been adopted by a consortium, led by the US National Institute of Standards and Technology, that is called Genome In A Bottle[106,107]. It was derived from 14 data sets from five sequencing technologies and the data is publically available and can be used to test analysis pipelines.

Some variant callers hence employ Bayesian models, such as SAMtools. In contrast, the recently developed GATK HaplotypeCaller first determines regions varying from the reference, so-called active regions and then for each active region, it de novo assembles all possible haplotypes[108].

While evaluating the results of variant calling it is important to consider that the combined error for each base call and alignment can easily exceed 1%. Therefore, statistical methods are being employed to improve the accuracy (Figure 14). While this allows for an overall reduction of false positives, it does not give a confirmation of the quality of the single variant which must be manually investigated and finally confirmed by Sanger sequencing[3,87,109].

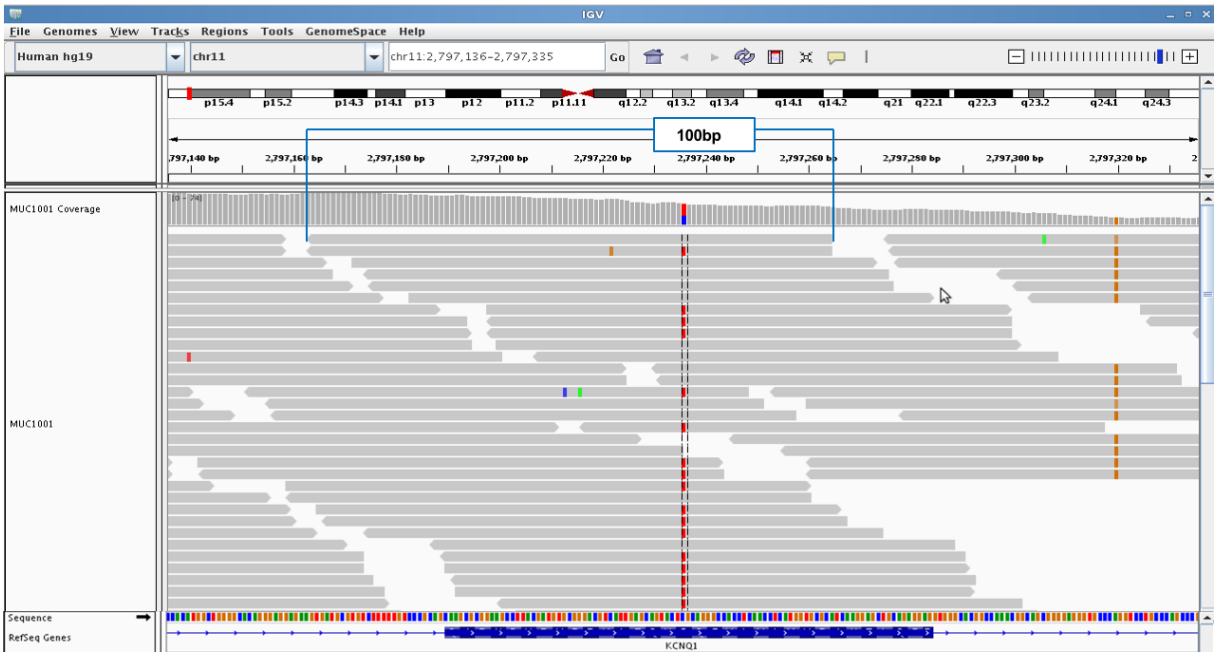

Figure 14. Visualisation of a genomic region in *KCNQ1*. Each read (100 bp) contains bases that are compared to the reference sequence (i.e. Sequence). Grey coloured bars indicate bases that match the reference. Note the presence of false calls (colours within the reads). A heterozygous variant (in the center of the figure and present in 50% of the total reads) is a true call in the exonic region (blue rectangle).

a) Single nucleotide variation (SNV). Overall, single nucleotide variation is called with high sensitivity and specificity for approximately 80% of the genome, approaching 99% concordance with genotyping microarray-based approaches in those regions[107]. This is important because Mendelian diseases are encoded by this class of variation. But it also tells us that we have a problem with the remaining 20% and probably with the diagnosis of repeat expansion disorders.

b) Insertions and deletions (indels)**.** In contrast to single nucleotide variation, calling for small insertions or deletions is less accurate. In one study, the concordance between two platforms for indel calling was only 57% across the genome and 33% for inherited disease risk genes[109]. This is particularly important for clinical genomics, as variations disrupting the reading frame or affecting the structure of the protein are more likely to be more clinically relevant. Careful manual curation is currently the only approach that can resolve these issues. Often it requires playing around with thresholds and scores or to search for candidate regions or search for a second hit. Despite useful, algorithms for calling indels remain less efficient compared to those for calling SNVs. Pindel (widely adopted, including into the Genome Analysis ToolKit GATK framework[108]), local de novo assembly approaches as well as the use of 'known' indel positions increased the number of true detections. However, sensitivity still drops very rapidly to below 50% as the size of the indel increases[95]. Therefore, this class of variation currently cannot be confidently called for clinical purposes and false-negative calls remain a concern[98].

c) Structural variation. In discovery projects, structural variation has been detected through microarrays and sequencing, but algorithms to detect structural variation from short-read sequencing are still limited by the length of the short read[3].

1.3.5 Variant annotation

As a further step, variants are selected considering standard thresholds for read depth and quality. Furthermore, using tools like ANNOVAR or customized tools, they are annotated with scores that predict the functional effect on the protein[98]. Based on a gene definition file derived for example from the University of California Santa Cruz (UCSC) annotation database, the variant is annotated in regard to its genomic location as intergenic, 5'- UTR, exonic, splice site, intronic, or 3'-UTR variant[95]. For exonic variants, the consequence of the variant on the protein sequence is predicted based on the mRNA sequence as synonymous, non-synonymous, frameshift, stop-gain, stop-loss, etc[110,111]. However, these annotations need to be also verified with manual curation, especially in case of gene overlaps and splice isoforms existing for a specific gene.

Minor allele frequencies and pathogenicity scores from public databases are annotated[98]. The minor allele frequency (MAF) from ExAC[112] (Exome Aggregation Consortium Database), gnomAD (Genome Aggregation Database)[113] and 1000 genomes[82] are given for each variant. Lines of evidence are added by the prediction scores. CADD[114], MutationTaster[115,116], PolyPhen-2[117], and SIFT[118] areamong the most used the prediction tools and were used in this study.

1.4 Purpose of this work

Cardiac disorders are genetically highly heterogeneous, with hundreds of genes and different models of inheritance involved. WES is the primary discovery tool in human genetics and it has a potential for most of the clinical fields. Indeed, its un-targeted parallel analysis of 20000 genes and the continuous advancement in bioinformatics pipelines, makes WES suitable for a broad range of genetic conditions.

The aim of this study was to explore the use of WES in gene discovery of genetically heterogeneous cardiac disorders. In this study, different approaches of WES analysis have been applied in:

- Family studies with a dominant pattern of inheritance to identify novel candidates. For this purpose, I selected two large pedigrees showing a clear dominant model of inheritance with multiple affected family members affected by arrhythmogenic right ventricular cardiomyopathy (ARVC) and idiopathic ventricular fibrillation (IVF).

- Family studies with a typical recessive model of inheritance, such as mitochondrial disorders, to expand the clinical spectrum of cardiac manifestations in such disorders. Indeed, mitochondrial disorders are often presenting with a cardiac manifestation and half of the 300 genes so far discovered have been detected by WES.

- A cohort of 78 trios with HLHS, using de novo, pathway and CNV analyses to elucidate the genetic basis of the complex cardiac disease.

- A cohort of 26 families with LQTS and no mutation in the three major genes (*KCNQ1*, *KCNH2* and *SCN5A*) to evaluate the genetic yield of WES.

- A cohort of 855 patients to quantify the detection of incidental findings (IFs).

II. Methods

2.1 Internal pipeline

With the pipeline of the Institute of Human Genetics, using the latest version of the enrichment kit, around of 80000 SNVs and 7000 indel are normally detected in each sample. In order to filter and to identify candidate variants among this large amount of data, many bioinformatics processes and manual curation of filters are required. For clinical filtering of variants, I had to modulate and to adapt the filters according to the project or question and that often required also to step back into the data analysis itself. During my PhD studies, my work was focused on filtering and interpretation of variants. The exome data have been queried via a custom web interface thus allowing the filtering steps. The exome sequencing internal pipeline was developed in 2008 by the group of Dr. Strom and since then it has constantly been updated based on feedback from physicians and novel methods developed.

This pipeline (Figure 15) started with FASTQ files. After alignment (BWA tool) variants were called, filtered and annotated using SAMtools and customized Perl scripts[95]. The annotated variants were then inserted into an internal database, including sample information (e.g. gender, pedigree name and disease). In addition, in the database are stored a variant table with characteristics on the variant (e.g. position and type of the variant) and a variant to sample table in order to link variants to samples[95].

The database can be queried via an internal web interface allowing filtering for certain characteristics of the variants (e.g. variant frequency or variant type) and showing basic statistics (e.g. coverage or duplicate rate)[95].

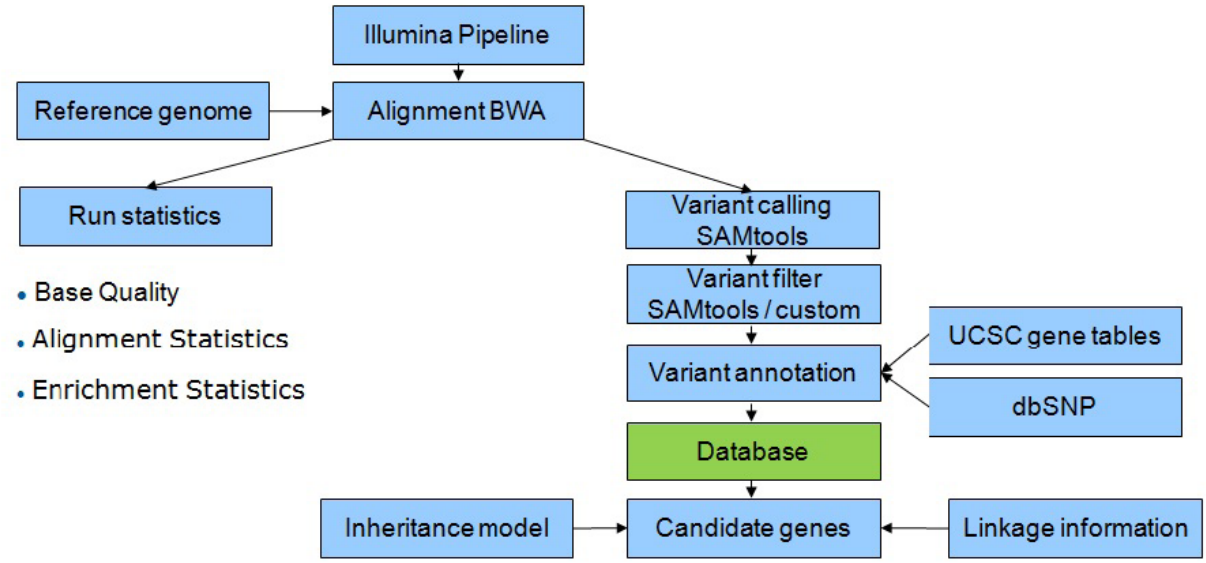

Figure 15. Internal pipeline: an overview of the pre-existing exome sequencing data analysis pipeline is reported here as it was initially developed by Eck, 2014 and Wieland, 2015 [95,119,120]**.**

2.2 Variant filtering

Variant calling usually detects for each exome more than 60000 genetic variants. The majority of those variants are unlikely to be relevant for the disease under investigation and therefore they need to be filtered out to identify the final disease-causing variant. Using the internal database, I applied different strategies of analysis to identify the causative variants contained in this study.

2.2.1 Internal database

The internal database contains currently almost 15000 exomes, from hundreds of different international projects, sequenced on Illumina HiSeq2000, HiSeq2500, and Hiseq4000 machines. Samples were sequenced to identify pathogenic variants and disease-associated genes in rare and common diseases. Samples carry a variety of medical disorders, mainly tumors, developmental disorders, neurological disorders, mitochondrial diseases and cardiovascular disease (Figure 16). Of note, all samples diagnosed with a different disease were by default used as controls. The existence of an internal database represents a valid support for variant prioritization, in addition to the publicly available exome databases. Moreover, it offers the possibility to see genotype-phenotype correlations because some phenotypic information is available for each sample such as the primary medical condition for which the sequencing was ordered.

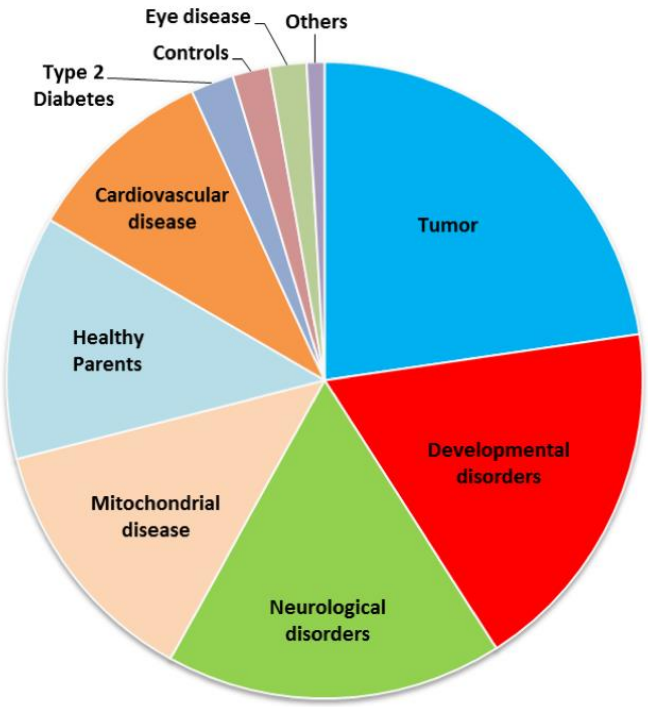

Figure 16. Distribution of samples among disease groups in the internal exome database.

2.2.2 Mode of inheritance and strategies of analysis

Significant genetic support is mandatory for assigning causality to gene variants identified using WES[109]. The use of an inheritance model or expected segregating variant frequency threshold is helpful to filter out and reduce the number of variants that have to be analysed. In an autosomal dominant inheritance model, candidates are heterozygous variants seen in all affected individuals in a family (Figure 17). Using the internal database, I performed dominant pedigree analysis on ARVC families and LQTS families. In diseases with autosomal recessive inheritance, the pattern is usually due to compound heterozygous in all of the affected individuals in a family (when parents are non-consanguineous) and only less frequently, homozygous (Figure 17). Unaffected parents are heterozygous for one allele, while other unaffected family members are either heterozygous or homozygous for the reference allele. I applied a recessive type of analysis for suspected mitochondrial disorders patients. Another scenario is represented by the de novo variants. In this case, the allele in the affected individual is not present in either parent. However, the possibility to perform this filtering depends on the availability of exomes from the parents (i.e. trio sequencing) (Figure 17). This type of analysis is particularly powerful because it reduces a lot the number of variants to analyse (0-5 variants) for each proband in the trio. In this work, I used this strategy of analysis for HLHS trios (Figure 18).

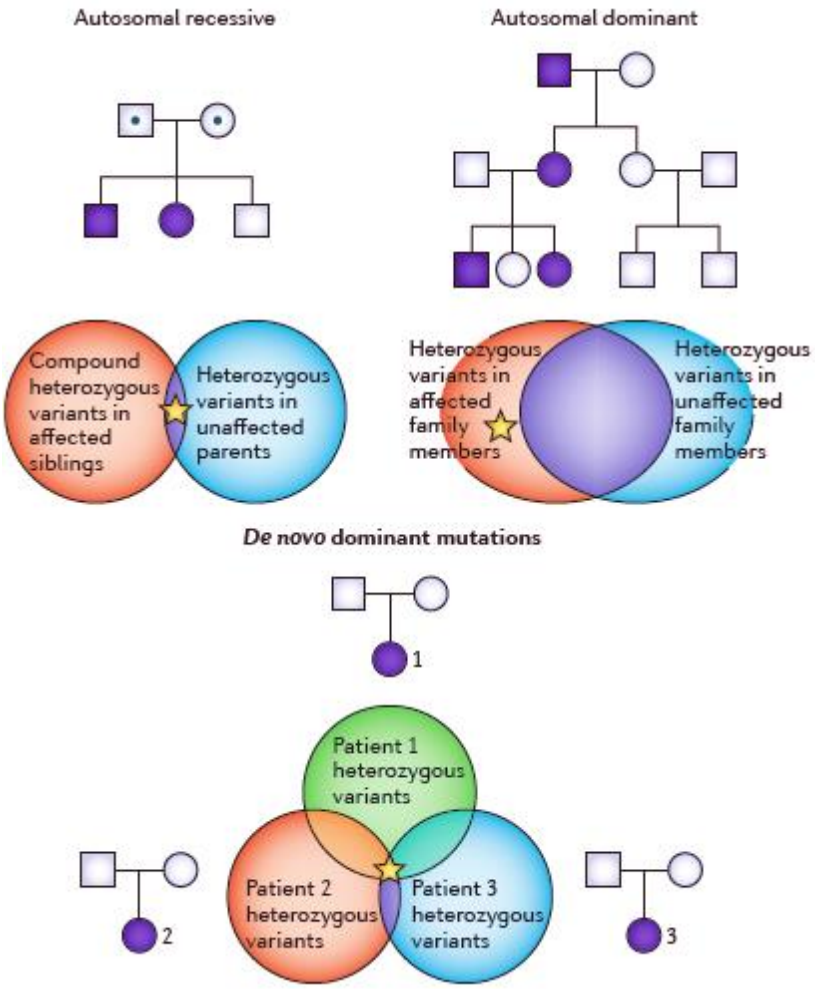

Figure 17. Gene identification approaches for the identification of disease-causing variants. Purple symbols indicate affected individuals. A dot in the centre of a symbol indicates carriers of the disease-causing variant. Stars underline the area where it is expected to contain the disease-causing gene [Modified from [121]]

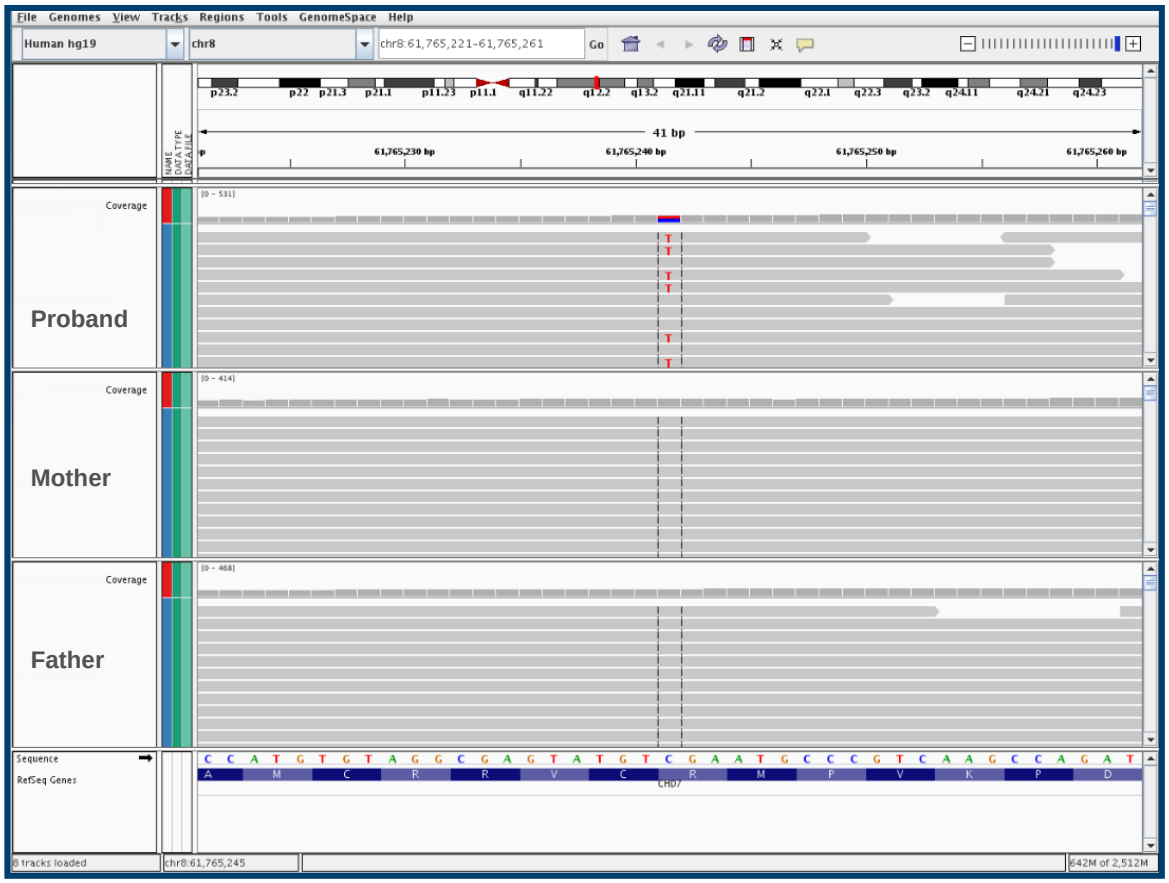

Figure 18. Visualization with Integrative Genomics Viewer (IGV) of a de novo heterozygous variant in an HLHS trio.

2.2.3 Variant class and prediction scores

Further selection can be carried out using annotations for the variant class. Indeed, exonic missense, nonsense, stop-loss, frameshift and splice site variants may all affect protein function. Therefore, the variants identified with WES are normally reduced to those most likely affecting the protein structure. The so-called loss of function (LOF): nonsense, start-loss, stop-loss and splice-site variants as well as frameshift insertions and deletions (indel) have quite well defined effects on the protein and are present in low numbers[98]. At first, the analysis starts with a search for LOF variants. The output of the LOF variants need to be manually evaluated, since many times the annotation of splice, nonsense or frameshift can be referred to minor isoforms of the gene. For this work, I generally selected annotation for the canonical form of a transcript (using UCSC and Ensembl database). As a second step, non-synonymous missense variants were evaluated. Several in silico prediction tools are available to identify those amino acid changes that most likely affect protein structure. In this work, I filtered the missense variants considering the predictions of SIFT[122], Polyphen2[122], and CADD[123]). In SIFT values below 0.05 indicated damaging predictions and in Polyphen2

values above 0.5 indicated possibly or probably damaging prediction. Using the CADD score, values above 15 referred to high conserved residues and higher is the value of the score higher is the predicted damaging impact on the protein function. It is important to mention that these tools alone have to be considered carefully because they represent a prediction and their accuracies are far to be perfect (around 80%[124]). As a general concept, the concordance of ≥ 3 tools (SIFT, Polyphen2 and CADD) confers a better indication of the effect on the protein structure. However, a good exercise it is verifying the conservations among different species directly in UCSC; also scores like GERP and Phylop can be helpful. Genes differ also in the amount of potentially disruptive genetic variation they can tolerate. This important gene feature is indexed as intolerance scores. A good example is the probability Loss of Function intolerance score (pLI) used in ExAC[112]. The pLI indicates the probability that a gene is intolerant to a LoF mutation. This analysis is based on high-quality exome sequence data for 60706 individuals of diverse ethnicities. The ExAC consortium used a mutational model to compare the observed number of rare variants per gene to an expected number, and then quantified deviation from expectation. The pLI score specifically reflects the probability that a given gene falls into the haploinsufficient category, therefore is extremely intolerant of loss-of-function variation. Genes with high pLI scores (pLI ≥ 0.9) are extremely LoF intolerant, whereby genes with low pLI scores (pLI ≤ 0.1) are LoF tolerant. However, the prediction of the effect of a variant on the protein structure and its function is only one of the criteria that need to be considered when assessing disease-relevant variants.

2.2.4 Variant quality, mapping quality and coverage filters

During a variant search, I often applied different filters for variant quality, coverage and especially variant frequency. Filtering is a process based on a delicate balance between specificity and sensitivity, in order to detect the true variants with the minimum number of false positives detected.

For variant quality, the general cut-off adopted in this study was ≤ 30. With lower values are considered variants in candidate genes, especially if the coverage is good, and variant visualisation with IGV is then a recommended step to check the variant and exclude a false positive result. For mapping quality, the cut-off used was 50 and also, in this case, the visualisation with IGV can discriminate between variants aligned and mapped correctly. The coverage is one of the most important parameters to check if the variant is a true call, and I adopted a filter of 15 (number of reads that cover the variant). The percentage of reads gives the estimation of the heterozygosity and homozygosity status of the variant. When the

coverage is very low, this percentage cannot be guaranteed. In this case, only Sanger validation can confirm the WES result[98].

2.2.5 Variant frequency

Each individual harbors around 12,000-14,000 predicted protein-altering variants[125,126], therefore filtering out the pathogenic variants from the benign ones is essential. One of the primary criteria for predicting if a variant is likely to have a functional effect on the encoded protein is the minor allele frequency (MAF) of the variant (Figure 19). A low MAF in or absence from exome databases (such as ExAC) of controls is recognised as a necessary, but not sufficient, criteria for variant pathogenicity[126–128]. The use of ExAC, which contain the population allele frequencies of 10 million genomic variants through the analysis of exome sequencing data from over 60,000 humans, provides valid frequency estimates improving the filtering capacity of variant discovery in Mendelian disorders[126]. In practice, there is great debate about which allele frequency should be selected, with MAF cutoffs of 1% and 0.1% considered as conservative thresholds for recessive and dominant diseases respectively[126]. Population genetics, however, suggested that disease-causing variants must be much rarer than these cut-offs[126,129]. In this study, in order to avoid over-filtering, the MAF threshold was selected on a per-study basis, using the carrier frequency for a related disease and following the Hardy-Weidenberg law. For example, I have used a specific disease-frequency filter, such as 0.0001 for ARVC and HLHS, based on their prevalence of 1:5000. However, to avoid to apply too stringent filters, for each analysis were tested also default cut-offs (autosomal dominant search MAF < 0.1% and recessive search on MAF <1%). Indeed, with a stringent filter, less false positives are detected, but there is a higher risk to miss possibly pathogenic variants. In the internal database, the frequency of the selected variant is filtered in the internal exome cohort (currently 15000 exomes). Then, I compared the frequencies of the variants using publicly available databases, such as ExAC and gnomAD. The in-house cohort is smaller, but it allows for correction of systematic errors of the exome analysis pipeline and individual level information like bi-allelic variants in a given gene[98].

Large sequencing datasets like ExAC a gnomAD can be used to estimate the incidence and prevalence of carriers, given the genetic spectrum of the disease[98]. In this study, I have estimated the incidence of *ACAD9* deficiency in the European population. The incidence of *ACAD9* deficiency in the European population was calculated by computing the minor allele frequency of LOF variants contained in ExAC and GnomAD and the causal *ACAD9* variants present in this cohort. Then the total allele frequency was used to estimate the prevalence of

ACAD9 deficiency. The incidence was calculated in the European population (5.1 millions of individuals).

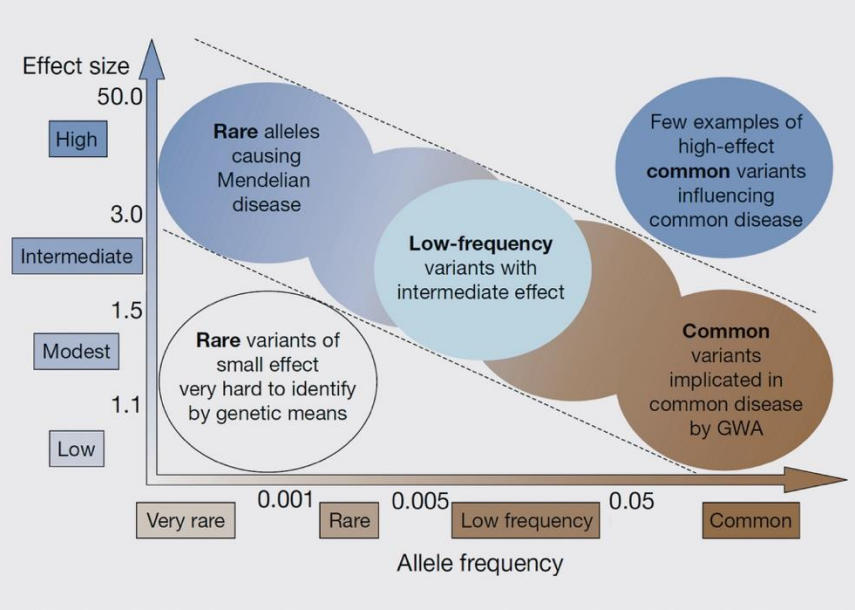

Figure 19. Allele frequency and effect size. [Taken from [130]]

2.2.6 Database of known variants

The process of disease-causing variants classification often includes independent observations of patients with similar clinical phenotypes and carrying the same variant. For rare variants, the discovery of two additional individuals with the same symptoms and carrying the same genetic variant are sufficient for a statistical support of the genetic finding. Sources of information are usually published studies and public databases that need to be searched. Several efforts aiming to specifically aggregate relevant clinical data exist and include databases such as Decipher[131], HGMD[132] or ClinVar[133]. In the internal database, HGMD and ClinVar reports are annotated for each variant. For HGMD, the variant needs to be classified checking the source (publication with the given patient); indeed, HGMD database is a comprehensive database for human mutations, but there is not a classification of the variants that are reported as DM (disease-causing mutation), DM? (possibly disease-causing) or (FP=functional polymorphism). Moreover, the variants are not systematically reviewed. Only variants present in literature are present in the database and not the ones inserted from laboratories for genetic testing. ClinVar was created more recently and it contains the classification of the variants in benign, likely benign, uncertain significance, likely pathogenic and pathogenic. Given the ClinVar annotation, this feature helped a lot the variant filtering process. In addition, ClinVar reports include the date of the submission of

the variant and the system of classification used. Each variant is also systematically checked and reviewed. The results obtained by diagnostic laboratories are also included. Decipher is not included in the database annotated for each variant, so I interrogated the database manually. Variants reported in public databases can contain false positives and lack of pathological variants; therefore, I investigated the results with caution and evaluated carefully the information present in the report or in literature.

Variants present in the internal database are also annotated with the OMIM[134] link. For each sample analysed, I also used OMIM search present in the database, using clinical keywords related to the phenotype under investigation (such as "cardiomyopathy"). I also used the Mouse Genome Informatics Database (MGI)[135] report to add supporting evidence to the genetic findings.

2.2.7 Pathway analysis

To determine whether a set of variants contributes to functions relating to a phenotype of interest, genes can be assessed for the enrichment of GO functions. In this study, I have performed pathways analysis using DAVID with the default parameters (cellular function, biological processes, molecular functions annotations and KEGG pathways annotation) and I have further applied Benjamini and Bonferroni correction.

2.2.8 Variant visualization

False-positive variant calls represent a common issue, and therefore any candidates should be independently verified. To address this point, I have applied the Integrative Genomics Viewer (IGV)[136], a visualization tool that allows manual exploration of variants on desktop computers (Figure 20).

IGV shows different colours and levels of transparency to point out events in the alignment data or to visually deemphasize others. Different characteristics can be used to select the reads' colour composition and to group, sort and filter the reads. Important features included are sample identifier, strand, read group, mapping quality, base call at a particular position, pair distance and orientation and custom tags. The coverage track and the % of reads to visualise heterozygosity and homozygosity of the variants are helpful parameters to quickly analyse the variants.

Individual read bases that match the reference genome are displayed in the same colour as the read (generally grey), while mismatches are colour-coded by the called base and are assigned a transparency value proportional to the base call quality (phred) score[137]. This is

the default colouring scheme for reading bases in order to underline high-quality mismatches (Figure 20). A number of options are available for paired-end alignments to help elucidate structural variants, such as deletions, insertions and rearrangements. To emphasize possible inter-chromosomal rearrangements, alignments that fall on a different chromosome are coloured with the same colour of the mate chromosome.

The CNVs search is associated with a high number of false positive calls. This search is still not optimal and the high number of false positive requires manual investigation with IGV. To give an order of magnitude, I manually investigated with IGV more than 1500 de novo CNVs in the HLHS cohort, but only around 50 were considered as possible.

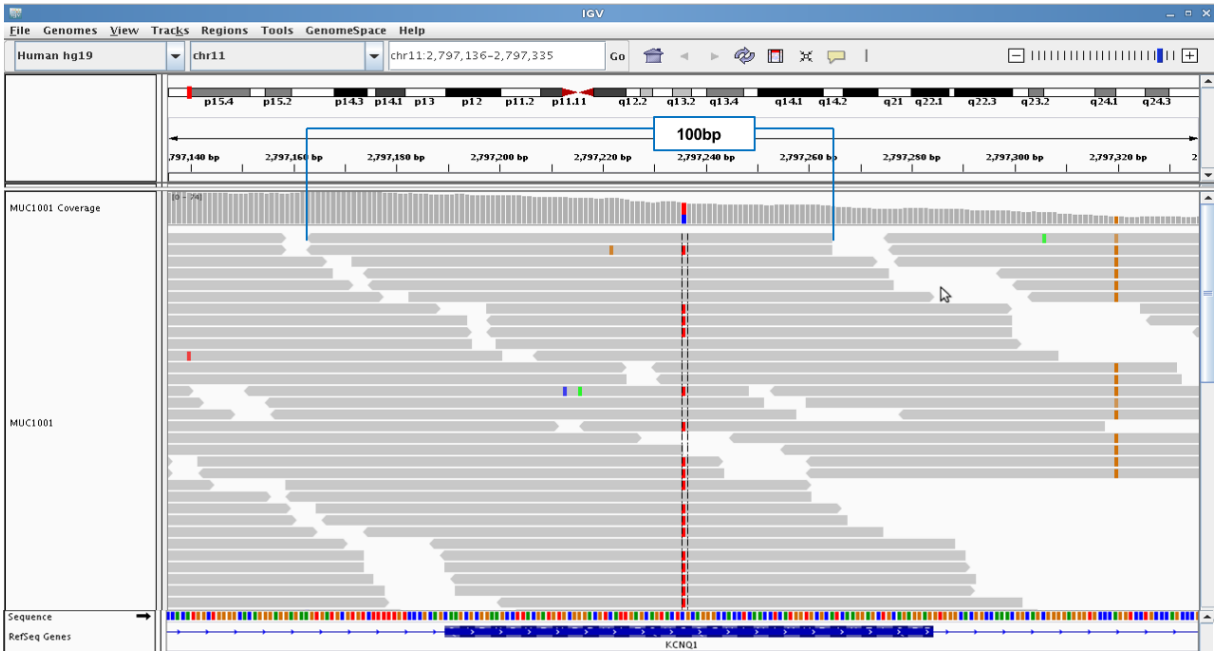

Figure 20. Visualization with IGV of a heterozygous variant.

2.2.9 Coverage

As anticipated in the variant calling section, deep sequencing refers to the general concept of aiming for a high number of replicate reads spanning each region of a sequence. The number of reads covering the same genomic position is named read-depth, while the average of the read-depths over the whole genome or the whole target region is named coverage[95].

Sufficient coverage and read depth are required for variant calling. A minimum read depth of 20x-40x is assumed to give sufficient power to detect heterozygous variants[95]. Read coverage of a specific genomic position is considered insufficient when the number of reads mapped is less than 10. Also, the uniformity of coverage (the percentage of a genomic region that is covered with at least 20 reads) is an important parameter to take into account.

The internal database shows that coverage is highly reproducible across a gene (Figure 21).

This tool is particularly useful to check the coverage of the samples in known disease-associated genes. In the internal database, coverage tool and candidate gene list search are combined in order to give information of the coverage of each gene present in the list (for example: mitochondrial disease-associated gene list with around 300 genes).

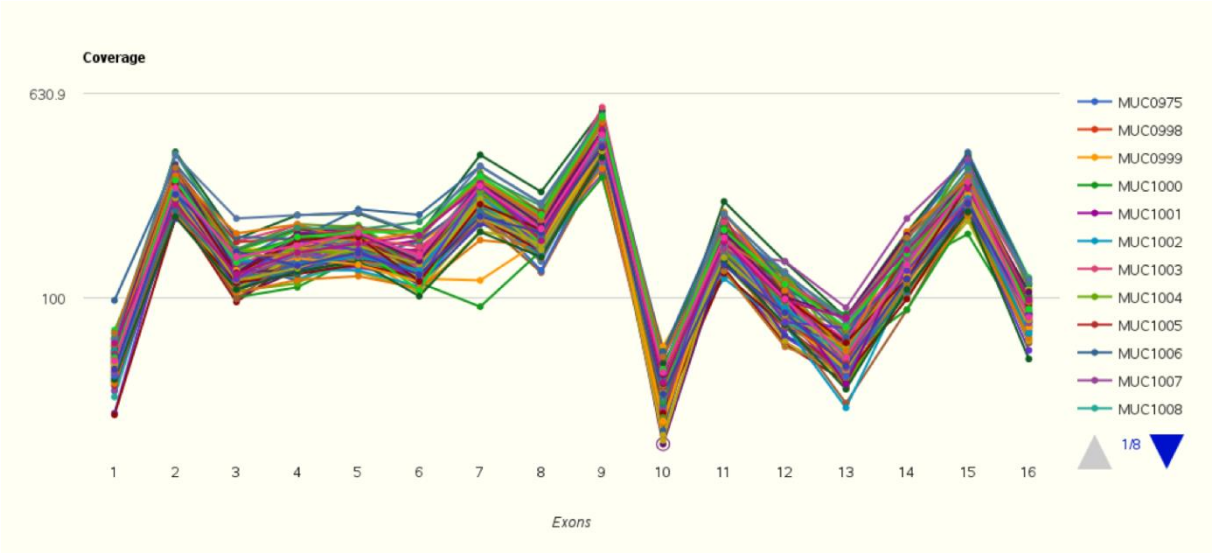

Figure 21. Coverage in *KCNQ1*. All the 16 exons have a coverage > 30. Coloured lines represent different samples.

2.2.10 Copy number variants (CNVs)

Copy number variants (CNVs) can range from 1 bp to several megabases[138,139]. In short-read sequencing, the detection of these structural variants is particularly controversial. Different strategies for detecting CNV are available: read depth, split-read, paired-end, assembly, and a combination approach. Among those, the read depth approach is considered the most powerful for WES data, since the target size is around 100 bp[98,139]. For read depth analysis, the normalized read depth of a chromosomal position is statistically compared to an expected read depth in this region[98]. The statistical power improves with the size of the region, making the evaluation of small CNVs more difficult[98]. In this study, I applied ExomeDepth, which uses a Hidden Markov Model to compare the read depth of a certain region in a sample to the read depth of around ten control samples[98]. ExomeDepth consists in a CNV calling algorithm that applies the robust beta-binomial model for the read depth data in order to build an optimized reference set. As a result, it maximizes the statistical power to detect even small CNVs in the presence of technical variability across the samples. Furthermore, because ExomeDepth assumes that the CNV call is not present in the reference sample, it is best suited to call rare CNVs.

The ExomeDepth calling results in a list of CNVs identified by genomic coordinates (start and end of the putative CNV). Furthermore, its output includes the type of the CNV (deletion or duplication) and a Bayes Factor assessing the evidence in favour of (or against) the CNV. In order to identify putative rare de novo CNVs in HLHS trios, I have queried the internal database to show only CNVs identified in the affected subjects and not in the unaffected parents (de novo), containing ≥ 3 exons and with a genotype frequency cut-off of 0.0001 in the internal database. ExomeDepth has shown the highest sensitivity when compared to other tools but also has a considerable high false positive rate[139]. Therefore, the resulting CNVs were inspected visually using IGV and only the plausible CNVs (labelled as "possible") were considered for further analyses. In HLHS project, among more than 1500 CNVs, only around 50 were labelled as possible after manual inspection.

As additional step, those CNVs were verified in DGV (Database for Genomic Variants http://dgv.tcag.ca/gb2/gbrowse/dgv2_hg19/) as a dataset of CNVs in healthy individuals. Finally, Decipher (https://decipher.sanger.ac.uk/disorders#syndromes/) and CHDWiki (http://homes.esat.kuleuven.be/~bioiuser/chdwiki/index.php/CHD:Map) were used as reference sets for known microdeletions and duplications respectively associated with CHDs and syndromic disorders.

2.3 Variant pathogenicity

2.3.1 Variant classification

The assessment of pathogenicity of genetic variants represents one the biggest challenge in clinical genomics[140]. It is well-known that the majority of genetic variants contained in the human population are unlikely to be disease-causing[140]. For example, a genome sequence includes about 3.5 million genetic variants, of which 0.6 million are rare or novel. The interpretation of the clinical relevance of these genetic variants represents the main barrier to further development of clinical genomics[140].

The evaluation of pathogenicity of a genetic variant is even more complicated by the fact that laboratories use different criteria for assessing genetic evidence[141]. The decision of how to take in account each type of evidence is challenging. To overcome this issue, the American College of Medical Genetics and Genomics (ACMG) published in 2015 guidelines for the evaluation of variants in genes associated with Mendelian diseases (Figure 22). [142,143]

Following the standard criteria, the variants are classified in benign, likely benign, uncertain significance (VUS), likely pathogenic and pathogenic. Supportive evidence of pathogenicity and benignity of the variants are reported in Figure 22.

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 22. Evidence framework of the American College of Medical Genetics and Genomics (ACMG) 2015 standards and guidelines for interpretation of sequenced variants. The table includes the criteria by the type of evidence for a benign (left side) or pathogenic (right side) assessment. BS, benign strong; BP, benign supporting; FH, family history; LOF, loss of function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong. [Taken from [142]]

2.3.2 Variant interpretation

The accurate filtering of variants can massively reduce the amount of candidate variants. In some circumstances, this is in the only step required to identify a disease-causing variant in a known disease gene. On the other hand, in most of the cases, additional evaluation is needed. Even when dealing with sporadic cases, it is generally helpful to analyse more than one individual of a family because this allows the identification of variants segregating with the disease within the pedigree. Novel genes that have not previously been implicated in disease or even genes with an unknown function challenge the search. Convincing evidence for a new genotype-phenotype correlation needs likely pathogenic variants in the same gene in

unrelated patients sharing a common distinct phenotype. Especially for VUS variants but not only, functional studies are required.

2.3.3 Variant validation

In a molecular diagnostics setting, a validation of the putative causative variant identified by WES is necessary[144]. For this scope, Sanger sequencing is used[87]. Segregation of the variant in the family needs to be confirmed. Indeed, the lack of segregation is a sufficient criteria to exclude the variant.

Sanger validation

DNA sequences were amplified by PCR using the Qiagen Taq DNA Polymerase Kit (Qiagen, Hilden, Germany) and a PeqStar thermal cycler (PeqLab Biotechnology, Erlangen, Germany). PCR reactions were performed in a 25 µL PCR reaction with 25 ng gDNA, 0.5 U Taq DNA Polymerase, 1x PCR Buffer, 0.2 mM dNTPs, 1x Q-Solution, and 0.2 µM of each, forward and reverse primers, as follows[98]:

Step	Temperature [°C]	Time
Heat Lid	110	
Denature	95	5 min
Start Cycle (40x)		
Denature	95	30 s
Anneal	primer T_m - 5	30 s
Extend	72	1 min/kb
End Cycle		
Extend	72	10 min

1 µL of the PCR product was subjected to agarose gel electrophoresis to access quality and product size. The remainder was purified using the MultiScreen® PCRµ96 Filter Plate (Merck Millipore, USA) according to the manufacturer's protocol. The purified PCR product was used for cycle sequencing using the ABI BigDye Terminator v.3.1 Cycle Sequencing kit (Life Technologies, Carlsbad, CA, USA) following these conditions[98]:

Sequencing reaction:

Component	5 µL reaction	Final concentration
BigDye Terminator v.3.1 Ready Reaction mix	1 µL	-
BigDye Terminator 5x Sequencing Buffer	1 µL	1x
10 µM forward or reverse primer	1 µL	2 µM
PCR product	1 µL	-
Ultrapure H_2O	1 µL	-

Sequencing program:

Step	Temperature [°C]	Time
Heat Lid	110	
Denature	96	1 min
Start Cycle (25x)		
Denature	96	10 s
Anneal	50	5 s
Extend	60	1 min 30 sec
End cycle		

The sequencing reaction was precipitated with 25 µL 100% ethanol for 15 min in the dark followed by centrifugation at 3000 g for 30 min at RT. The washing step included 125 µL 70% ethanol, centrifuged at 2000 g for 10 min, and left to dry at RT in the dark. Then the pellet was resuspended in 25 µL ultrapure H_2O, transferred to a microtiter plate, and located into the automated ABI 3730 sequencer. The obtained sequences were analysed using Staden Package (http://staden.sourceforge.net)[98].

2.3.4 Incidental findings

The American College of Medical Genetics and Genomics (ACMG) has recommended the identification and return of "incidental findings" (IFs) from a minimum list of 56 actionable genes, representing conditions with well-known guidelines for prevention and treatment. Using the internal pipeline, I have prioritized nonsynonymous SNVs, splicing and insertion/deletions (indels) detected in the 56 ACMG genes[145]. To reduce false positives, I filtered out variants according to standard thresholds (variant quality <30 and depth <10). I adopted, as internal frequency cut-off, a minor allele frequency of 1%. Variants reported as pathogenic or likely pathogenic in ClinVar[133] and variants reported as disease-causing (DM) in HGMD[146] were further investigated. Novel disruptive variants (nonsense, frameshift and canonical splice), were also evaluated. For variant classification, I followed stringent criteria for reporting IFs, given the fact that the a priori probability of pathogenicity is lower than in disease-targeted testing[141]. At this phase, I applied an additional disease-specific frequency filter for each gene using the internal database, 1000 Genomes and ExAC[112]. As a further step, I reviewed the primary literature for the remaining variants and only variants putatively producing the phenotype of interest were considered.

2.4 Data collection

Genomic DNA of patients belongs to the DNA collection of IHG (Helmholtz Zentrum München, Munich, Germany). For all samples, informed consent was obtained.

I collected and reviewed all the clinical data obtained by the collaborators in order to clarify the genotype-phenotype correlations. For the *ACAD9* cohort, I contributed to the creation of a Survey Monkey circulated among the collaborators. I designed the clinical questionnaire and investigated the clinical aspects of the patients, focusing on the cardiac manifestations, therapy and clinical outcome of the patients.

III. Results

3.1 Genetic basis of ARVC

3.1.1 Defects in *CDH2*-encoded N-cadherin in ARVC

In 1996 a family of northern European ancestry with ARVC was referred to the Department of Medicine at the J Floor Old Groote Schuur Hospital in Cape Town (Prof. Mayosi) for genetic investigation[147]. Five members in the family were defined as definite ARVC (patients II:4, III:2, III:3, III:4, and III:5), while one member was classified as borderline ARVC (patient II:7). Micro-array genotyping of 25 microsatellite markers was initially performed to identify putative linkage to the ARVC loci [148]. The analysis resulted in the identification of a region of putative linkage on chromosome 10p12-p14.22[147]. However, at that time candidate screening of several genes within the putative linkage region did not identify a disease-causing mutation. The original phenotyping of this family had been performed at Wentworth Hospital in Durban between 1990 and 2000 [149]. The clinical status of all the family members were reassessed considering the revised diagnostic criteria for ARVC in 2010 (Table 1). Patients III:6 and III:8, previously labelled as affected,[148] did not fulfilled the revised criteria for the diagnosis of ARVC[147].

The phenotypic reclassification of two family members from affected to unaffected status leaded to a re-analysis of the previously performed genetic linkage analysis. With the advent of NGS technologies, WES was used to identify the genetic cause of ARVC in this family[147].

3.1.1.1 ARVC family 1

A 3-generation family of northern European descent from the KwaZulu-Natal province of South Africa was enrolled in the ARVC Registry of South Africa (Prof. Bongani Mayosi group). For thirteen members was possible to obtain clinical and genetic information and 6 were defined as clinically affected. Clinical features of the affected individuals are summarized in Table 1 and the pedigree is presented in Figure 23[147].

The proband (III:2) was a male who presented at 16 years with palpitations associated with ventricular tachycardia. The ECG at rest had an epsilon wave in lead V1 and inverted T waves in V1-3. The echocardiogram and right ventriculogram showed a severe dilatation of the right ventricle with reduced systolic function. Despite attempts with several antiarrhythmic medications, he experienced sudden cardiac death at the age of 22 years. The

sister (III:3) had several episodes of ventricular tachycardia with left bundle branch block morphology, necessitating cardioversion and amiodarone treatment. Similarly to the brother, she also died suddenly at the age of 24 years. No clinical data were available for the parents (mother II:2 and father II:1). A cousin of the index case (III:4) had chest pain and palpitations at the age of 23 years. The Holter ECG showed several ventricular ectopies and at the echocardiography global dilatation of the right ventricle was noticed. On the cardiac magnetic imaging emerged areas of fat in the right ventricular free wall and the endomyocardial biopsy confirmed these findings. The electrophysiological study documented inducible non-sustained monomorphic ventricular tachycardia with a left bundle branch morphology and right axis deviation. The brother of III:4 (III:5) had structural and ECG abnormalities fulfilling the diagnostic criteria for ARVC. The mother (II:4) of subjects III:4, III:6, and III:7 showed ECG abnormalities with T wave inversion in V1-3, which, is compatible with a diagnosis of ARVC with partial penetrance. Subject II:7 showed clinically palpitations, T wave inversion in V1-3 at the ECG, and mild right ventricular dilatation on cardiac magnetic resonance imaging. The pedigree structure indicates autosomal dominant inheritance of ARVC with variable penetrance (Figure 23)[147].

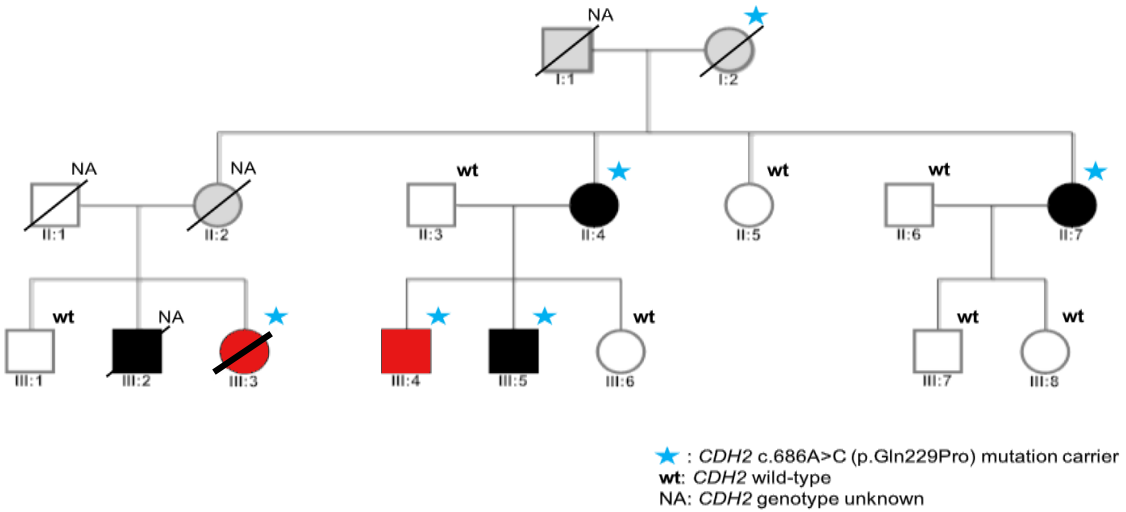

Figure 23. Pedigree of ARVC family 1 depicting the cosegregation of the *CDH2* c.686A>C (p.Gln229Pro) mutation in the South African family with arrhythmogenic right ventricular cardiomyopathy. Filled black symbols: clinically affected subjects; filled red symbols: affected individuals sequenced by WES; open symbols: unaffected subjects; grey-shaded symbols denote subjects with phenotype unknown or insufficient phenotypic data: *CDH2* c.686A>C (p.Gln229Pro) indicates mutation carrier; NA: *CDH2* genotype unknown; and wt: CDH2 wild-type. [Modified from[147]]

Table 1. Clinical characteristics of ARVC affected subjects in ARVC family 1

ID	Sex	Age at diagnosis (years)	Symptoms	Cardiac imaging	Tissue characteristics	ECG repolarization abnormalities	Depolarization/conduction abnormalities	Arrhythmias	Diagnostic criteria	Outcome
II:4	F	44	Palpitations, chest pain, presyncope	Echo: normal MRI not done	Myocardial biopsy not done	TWI V1-V3	ECG: no epsilon wave SAECG not done	24-hour Holter ECG not done No history of VT	DEFINITE ARVC 2 major	Alive
II:7	F	44	Palpitations, chest pain	Echo: normal MRI: normal	Myocardial biopsy not done	TWI V1- V3	ECG: no epsilon wave SAECG: no late potentials	24-hour Holter ECG not done EST normal	DEFINITE ARVC 2 minor	Alive
III:2	M	16	Palpitations	RVA: dilated RV with hypokinetic outflow tract	Myocardial biopsy not done	TWI V1-V3	ECG: epsilon wave SAECG: not done	24-hour Holter ECG not done VT with LBBB morphology	DEFINITE ARVC 4 major, 1 minor	Died age 22
III:3	F	13	Palpitations	Echo: dilated RV with hypokinesia	Myocardial biopsy not done	Information not available	Information not available	24-hour Holter ECG not done VT with LBBB morphology	DEFINITE ARVC 2 major, 1 minor	Died age 24
III:4	M	23	Palpitations, chest pain	RVA: bulging hypokinetic RV inferior wall and sub-tricuspid dyskinesia	Fatty infiltration (not quantified), scattered chronic inflammatory cells	TWI V1-V5	ECG: no epsilon wave SAECG: no late potentials	EPS: NSVT LBBB morphology right axis PVC 9341/24h (no meds 1995) PVC 12620/24h (Sotalol 1997)	DEFINITE ARVC 3 major, 1 minor	Alive
III:5	M	21	Palpitations, chest pain	RVA: anterior hypokinesia	Fibrofatty replacement (not quantified)	No TWI	ECG: no epsilon wave SAECG: no late potentials	PVC 2035/24h	DEFINITE ARVC 2 major, 1 minor	Alive

ACM, arrhythmogenic cardiomyopathy; ARVC, arrhythmogenic right ventricular cardiomyopathy; EPS, electrophysiological study; Echo, echocardiogram; EST, exercise stress test; F, female; LBBB, left bundle branch block; MRI, magnetic resonance imaging; M, male; NSVT, non-sustained ventricular tachycardia; RV, right ventricle; PVC, premature ventricular contractions; RVA, right ventricular akinesia; SAECG, signal-averaged electrocardiogram (ECG); TWI, T-wave inversion; VT, ventricular. [Modified from[147]]

3.1.1.2 WES identified *CDH2* heterozygous variants in ARVC family 1

WES has been performed in 2 affected individuals (patients III:3 and III:4; Figure 23), as distantly related as possible (Figure 23 and 24), considering that the smallest region of overlapping haplotypes reduces the total number of variants to analyse. III:3 and III:4 are two first cousins so they share a rare allele that is identical by descent in approximately one-eighth of the genome.

In both samples, exome sequencing yielded around 17 Gb of mappable sequences and 98% of the targeted sequences were covered at least 20-fold with a mean coverage of at least 199x (Table 2).

Table 2. Sample statistics of individual III:3 and III:4

	Individual III:3	Individual III:4
Total variants called	Off-target variants are not called 76.5% reads on target	Off-target variants are not called 76.0% reads on target
Total reads detected	169469462	174409354
Total reads mapped % mapped	167670685 98,94	172615468 98,97
Seq (Gb)	16.950	17.440
Duplicates	12%	15%
Total variants on target	69813 SNV 6587 INDEL 374 PINDEL	66528 SNV 5574 INDEL 383 PINDEL
Avg Coverage % >= 1X % >= 4X % >= 8X % >= 20X	199.8 X 98,54 96,96 95,15 92,16	204.5 X 98,13 96,6 94,9 92,06
Transition/transversion ratio	2,22	2,13
Variants present in the 1000Genomes database	58433	57707
Variants present in the dbSNP database	60354	59603
Variants present in ExAC database	49561	49637
Novel variants	7132	7616
Missense variants	9482	9391
Nonsense variants	69	81
Synonymous variants	10606	10564
Frameshift variants	285	271
Inframe insertion/deletion	209	210
Splice variants	133	134
5’ UTR variants	2927	2879
3’ UTR variants	3442	3252
Intronic variants	35869	35731

I queried the internal database to filter the variants following a dominant inheritance pattern (variant allele 1, so heterozygous variants) to consider only the non-synonymous variants (SNVs and indels) in common between the two samples sequenced (Figure 24). I applied standard quality filters (coverage > 20, SNV quality > 30 and Mapping quality > 50) and a specific disease-frequency filter (MAF ≤ 0.0001). I selected a MAF ≤ 0.0001 as a cut-off considering an estimated prevalence of ARVC of 1:5000. Only 13 heterozygous variants survived after applying those parameters (Table 3) and also the manual inspection with IGV. Among these, the CDH2 variant (c.686A>C; p.Gln229Pro) was the only one simultaneously predicted to impact the protein by all in silico tools that were used (Polyphen 2, SIFT, CADD, Mutation Taster) and not present in internal and publicly available exome databases (1000 genomes and ExAC)[147]. Using the list of 13 variants, Maryam Fish (Prof. Mayosi group) performed segregation analysis in the ARVC family and only 3 variants (in *CDH2*, *NEK10*, and *ADD1*) were found to be also present in all affected family members. The CDH2-encoded N-cadherin is a cell-cell adhesion protein expressed in the heart. Cardiac alterations were documented in previously published CDH2 knockout and over-expression in animal models[150]. Considering the biological plausibility, this variant was considered likely pathogenic (Figure 24).

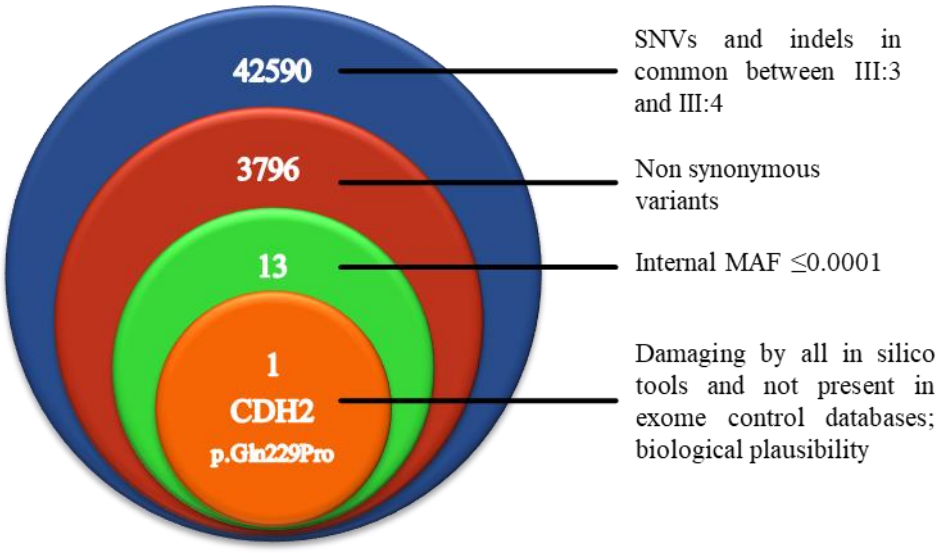

Figure 24. Schematic representation of variant exome filtering in subjects III:3 and III:4. [Modified from[147]]

Table 3. List of non-synonymous variants (n=13) identified through WES in both affected individuals (III:3 and III:4) surviving internal quality and frequency filters

Nucleotide change	Amino acid change	Type	Gene	Protein function	Polyphen2	SIFT	CADD	Mutation Taster	Internal database	1000 Genomes	ExAC EUR	ExAC AFR
c.268G>A	p.Val90Ile	missense	*MYO16*	Unconventional Myosin	0.113	np	0.023	Polymorphism	0,0001	0	0,0002	0
c.217T>C	p.Cys73Arg	missense	*KRTAP16-1*	Keratin-Associated Protein	**0.879**	**0,02**	**22,3**	**Disease Causing**	0,0001	0,0006	0	0.004
c.1151C>T	p.Pro384Leu	missense	*KRT31*	Keratin	0.001	**0,01**	**23,4**	**Disease Causing**	0,00006	0,0007	0	0,0025
c.3568A>G	p.Ile1190Val	missense	*CLTC*	Clathrin	0.001	0,24	**14,71**	**Disease Causing**	0	0	0	0
c.606G>C	p.Met202Ile	missense	*SCN4A*	Voltage-gated Sodium Channel	0.009	1	12,07	Polymorphism	0	0	0	0
c.971_973 delATA	p.Asn325del	indel	*ANKRD62*	Ankyrin Repeat Domain-Containing Protein	np	np	np	Polymorphism	0,00006	0	0	0.003
c.826C>G	p.Leu276Val	missense	*ESCO1*	N-Acetyltransferase	0,01	0,3	0,0003	Polymorphism	0	0.003	0,00005	0.008
c.686A>C	p.Gln229Pro	missense	*CDH2*	Cadherin	**0.785**	**0,03**	**16,6**	**Disease Causing**	0	0	0	0
c.1333A>G	p.Ile445Val	missense	*GALNT2*	Polypeptide N-Acetylgalactosaminyl transferase	0.001	0,76	11,97	**Disease Causing**	0,00006	0	0,0002	0,0002
c.1379T>G	p.Met460Arg	missense	*KLF11*	Zinc finger transcription factor	0.483	0,09	**26,2**	**Disease Causing**	0	0	0	0
c.1889G>A	p.Arg630Lys	missense	*NEK10*	Serine/Threonine Kinase	0.034	0,35	**22,4**	Polymorphism	0	0	0	0
c.1903C>G	p.Gln635Glu	missense	*ADD1*	Adducin	0.006	1	0.213	Polymorphism	0,00006	0	0,0001	0,0001
c.1181G>A	p.Arg394His	missense	*FSCN3*	Fascin	0,99	0,09	**28,3**	**Disease Causing**	0,00006	0	0,00005	0,0001

SIFT: values below 0.05 indicate damaging predictions – Polyphen2: values over 0.5 indicate possibly or probably damaging predictions – CADD: scores > 15 indicate damaging predictions. Minor Allele Frequency (MAF) of the variant in the different exome databases (internal database, 1000 Genomes, ExAC) is reported. EUR denotes the European non-Finnish and AFR denotes the African sub-populations of ExAC. [Modified from[147]]

3.1.2 Deficits in RhoA/cytoskeletal pathway in ARVC

I applied whole exome sequencing analysis in a severe ARVC patient lacking mutations in known ARVC-associated genes in order to identify a novel candidate.

3.1.2.1 ARVC family 2

The index patient of family 2 (Figure 25) was a young boy with a severe form of ARVC clinically evaluated at the Papworth Hospital NHS Foundation Trust in Cambridge. He presented with ventricular flutter that led to SCD. He showed 2 major and 2 minor criteria for the diagnosis of ARVC. The family history was negative for syncopal events, cardiac events or SCD (Table 4).

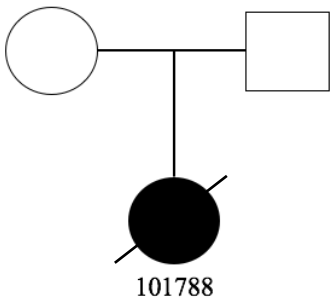

Figure 25. Pedigree of ARVC family 2.

Table 4. Clinical characteristics of the patients and ARVC diagnostic criteria.

Clinical features	
Sex	Male
Clinical Feature At Presentation	Ventricular Flutter
Familial History	None
Deoparization/Conduction Deficits	Major
Repolarization Abnormalities	Minor
Global/Regional Dysfunction And/Or Structural Alterations	Major
Arrhythmias	Minor
Total Diagnostic Criteria	2 Major, 2 Minor

3.1.2.2 WES identified a *MYH10* heterozygous variant in ARVC family 2

Exome sequencing yielded 13 Gb of mappable sequences and 98% of the targeted sequences were covered at least 20-fold with a mean coverage of at least 139x. Considering a dominant effect, I have quested the index subject's WES data for protein-altering sequence variations with a MAF ≤ 0.0001 (based on ARVC prevalence) in the internal database. Among 211 heterozygous variants, I selected the loss of function variants (frameshift, nonsense and canonical splice) and I further performed a phenotype-based search. This filtering allowed the identification of a nonsense variant in *MYH10* (c.1729C>T, p.R577*) (Figure 26)[151]. The variant was not present among 10900 internal exomes and 60706 exomes in the ExAC database. For *MYH10* the probability of loss of function intolerance (pLI), that defines if a gene falls into the haploinsufficient category, reached the maximum value (Table 5)[151]. In other words, a LOF variant has a high probability to be disease-causing in this gene. The amino acid change *MYH10* p.R577* involved a highly conserved amino acid residue (CADD score 38). MYH10 gene encodes the non-muscle myosin IIB (NMIIB) of the actomyosin cytoskeleton, which has been reported to regulate both actin dynamics as well as accumulation of active RhoA-GTP at the adherens junctions[152]. To add functional evidence, induced pluripotent stem cells (iPSCs) were generated (Prof. Alessandra Moretti and Prof. Karl Laugwitz group).

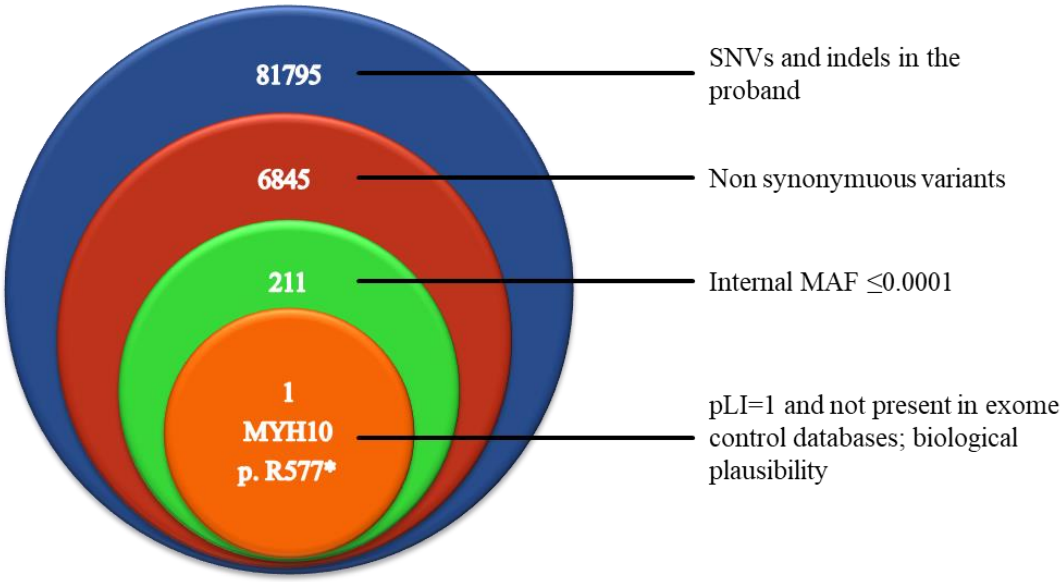

Figure 26 Schematic representation of variant exome filtering in ARVC family 2 proband.

Table 5. List of ARVC-associated genes and their probabilities of loss of function intolerance (pLI).

Gene	Locus	ExAC pLI	ExAC n LOF expected	ExAC n LOF observed	Reference
PKP2	12p11	0	28.8	19	(Gerull *et al.*, 2004)
DSP	6p24.3	1	87.7	14	(Rampazzo *et al.*, 2002)
DSG2	18q12.1	0	26.1	10	(Pilichou *et al.*, 2006)
DSC2	18q12.1	0	25.9	10	(Syrris *et al.*, 2006)
JUP	17q21.2	0.04	19.1	6	(McKoy *et al.*, 2000)
CTNNA3	10q 22.2	0	68	11.7	(van Hengel *et al.*, 2013)
CDH2	18q12.1	0.9	29.1	5	(Mayosi *et al.*, 2017)
DES	2q35	0	16.7	7	(Klauke *et al.*, 2010)
TMEM43	3p25.1	0	17.2	13	(Merner *et al.*, 2008)
LMNA	1q22	0.99	19.1	1	(Quarta *et al.*, 2012)
TGFB3	14q24.3	0.92	12.3	1	(Beffagna *et al.*, 2005)
RYR2	1q43	1	164.5	27	(Tiso *et al.*, 2001)
PLN	6q22.1	0.11	1.2	1	(van der Zwaag *et al.*, 2012)
SCN5A	3p21	1	55.5	9	(Erkapic *et al.*, 2008)
TTN	2q32.1-q32.3	0	894	247	(Taylor *et al.*, 2011)
MYH10	17p13.1	1	79.9	10	

ExAC pLI score: probability of loss of function intolerance according to ExAC; ExAC n LOF expected: number of loss of function variants expected to be present in the gene in the ExAC cohort; EXAC n LOF observed: number of loss of function variants observed in the gene in the ExAC cohort. [Modified from[151]**]**

3.1.3 Deficit in *DSP* in left-dominant arrhythmogenic cardiomyopathy

Idiopathic ventricular fibrillation (IVF) is a rare cause of sudden cardiac death (SCD). Patients with IVF present with a sudden onset of ventricular fibrillation (VF) of unknown origin that is not identified even after extensive diagnostic testing.

In a 3-generation family of European ancestry multiple episodes of SCD below the age of 40 were reported. The family has been monitored since early 1990's and for more than 20 years the clinical diagnosis was indicated as IVF, due to the unusual and complex clinical manifestations. With the advent of whole exome sequencing, the family was investigated with the aim of identifying the underlying genetic cause.

3.1.3.1 ARVC family 3

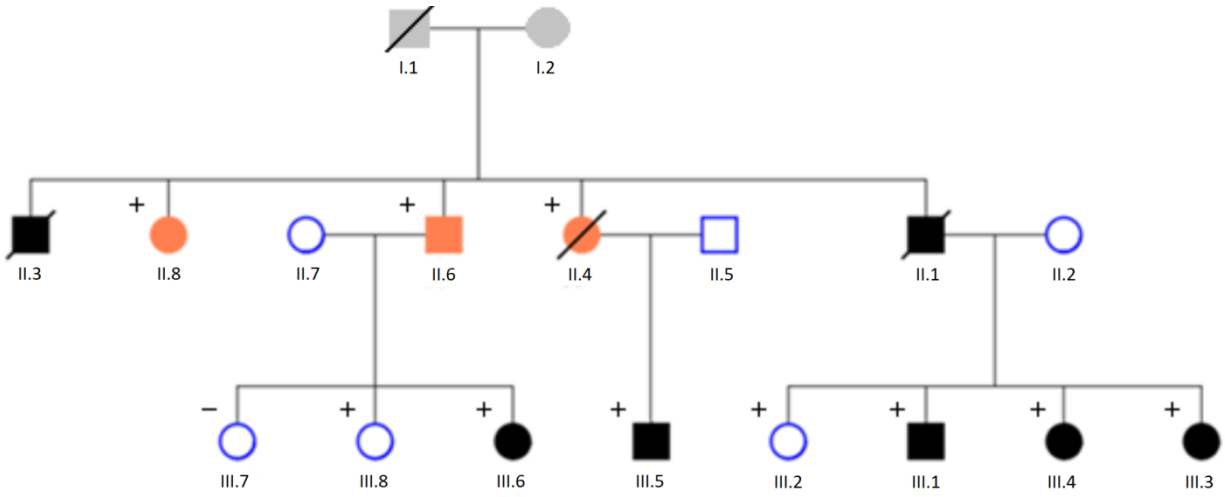

Figure 27. Pedigree of the ARVC family 3. It shows 3-generation of a family of European ancestry from Germany with positive family history for SCD below the age of 40. Filled black symbols: clinically affected subjects; filled red symbols: affected individuals sequenced by WES; open symbols: unaffected subjects; grey-shaded symbols denote subjects with phenotype unknown or insufficient phenotypic data; + indicates mutation carrier of Ser2168Argfs*18 in *DSP*.

Family 3 consists in a 3-generation family of European ancestry from Germany (Figure 27). Clinical evaluations were performed at the Department of Cardiology of the LMU Munich University Hospital (Prof. Stefan Kääb). In this family, three (II.1, II.3, II.4) out of five siblings experienced sudden unexpected deaths. In 1990, a male sibling (II.3) was found dead in bed at the age of 32 years. Two years later (1992), a brother (II.1) experienced SCD while playing a musical instrument. From his clinical history, 3 syncopal events just 3 months before the SCD and palpitations from the age of 24 emerged. The basal 12-lead ECG showed inverted T waves in V4-6. The autopsy demonstrated an enlargement of ventricles and pericardium myocarditis was suspected. In 2002, a sister (II.4) died suddenly while driving her car at the age of 38. Her

clinical history documented palpitations since the age of 25. Her basal 12-lead ECG showed inverted T waves in V1-6 and complex ventricular extra-systoles. The echocardiogram was reported as normal. Another brother (II.6), experienced 2 syncopal events and an episode of chest pain with cardiac enzyme elevation at the age of 38. In 2003, he experienced an additional episode of chest pain and coronary spasm after Ach infusion. Considering the dramatic family history of SCD, he was treated with implantable cardioverter defibrillator (ICD). During the same year, in another sister (II.8) an electrophysiology study was performed and it documented an induced ventricular tachycardia (VT). Accordingly, the patient was ICD implanted. Her resting 12-lead ECG showed inverted T waves in V3-6, II-III and aVF. Complex ventricular extra-systoles appeared during stress test, while the echocardiogram and myocardial scintigraphy showed LV hypokinesia. In 2006, individual III.1 (son of II.1) at age 25 suffered from chest pain with cardiac enzyme elevation. His basal 12-lead ECG showed inverted T waves in V3-6 and complex ventricular extra-systoles during stress-test. One year later, he experienced another episode of chest pain and therefore he was ICD implanted. His sister (III.4) presented with chest pain at age 18 and she experienced a syncopal event at age 25. Another sister (III.3) presented with palpitations at age 21. In both cases, cardiological clinical investigations were referred as normal and they are still carefully monitored for an eventual ICD implantation. A young female, III.6 (daughter of II.6), presented with palpitations at the age of 19. Her resting 12-lead ECG showed inverted T waves in V1-3 while cardiac MRI documented subepicardial left enhancement of the left ventricle. She experienced multiple episodes of palpitations at age 25 and was ICD implanted. Individual III.5 (son of II.4) presented with chest pain during adolescence. To summarise, the pedigree indicates autosomal dominant inheritance with high penetrance.

3.1.3.2 WES identified *DSP* heterozygous variants in ARVC family 3

Three (II.4, II.6 and II.8) individuals with the most severe phenotype were selected for whole exome analysis. The DNA of individuals II.3 and II.4 was not available. The choice of 3 samples to analyse instead of 2 (as performed in ARVC family 1) was related to the fact that the individuals were siblings, so they share approximately 50% of SNVs. Exome sequencing yielded 11 Gb of mappable sequences and 99% of the targeted sequences were covered at least 20-fold with a mean coverage of at least 133x in all 3 samples. I have performed the dominant analysis using the same quality (SNV >30, Mapping >50) and frequency filters (MAF ≤

0.0001) in 1, 2 and 3 siblings, respectively. With 1 sibling 110 variants were obtained, with 2 siblings 55 variants and with 3 siblings 24 variants (1/2*1/2=1/4 of variants to analyse).

The bioinformatics filters were based on a dominant inheritance pattern and a MAF ≤ 0.0001 based on the disease prevalence. The 24 heterozygous variants in common among all 3 affected family members were inspected by IGV. The complete list of variants detected is reported in (Table 6). Among these, 2 variants were loss of function: a nonsense variant in *CCDC181* and a frameshift variant in *DSP*. They were both absent among exome internal controls and ExAC controls. For *DSP*, the probability of loss of function intolerance (pLI) in ExAC browser reached the maximum value (1), while for *CCDC181* it was zero. In addition, considering a phenotype-based search, the variant Ser2168Argfs*18 in *DSP* emerged as the only variant possibly related to the phenotype under investigation (Figure 28). Indeed, *DSP* is a known ARVC-associated gene that can also present a left ventricular involvement[153]. Therefore, the variant Ser2168Argfs*18 was considered likely pathogenic for the observed phenotype. To support this finding, I further performed a segregation analysis in the family and the *DSP* variant was confirmed in all the affected family members (Figure 27). Of note, it was also present in two individuals rated as unaffected (III.8 and III.2), but considering the young age of the individuals (below 20), a later manifestation cannot be excluded and closed follow-up visits are required before assessing the unaffected status.

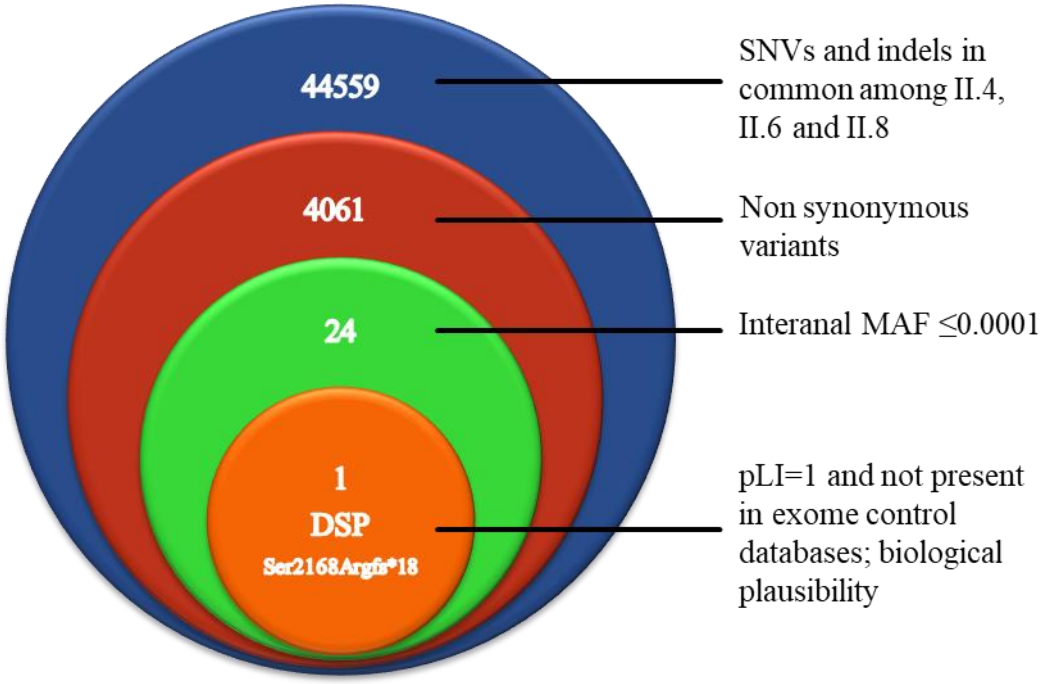

Figure 28. Schematic representation of variant exome filtering in II.4, II.6 and II.8.

Table 6. List of non-synonymous variants (n=24) identified through WES in 3 samples (II:4, II:6, II:8) surviviving the internal quality and frequency filters.

Gene	Function	nucleotide	Aminoacid	ExAC pLI	pph2	Sift	CADD	Internal database	1000 genomes	ExAC EUR	ExAC AFR
DSP	frameshift	6502_6505delAGTC	Ser2168Argfs*18	1				0	0		
IGSF9B	missense	2861 C>T	Arg954Gln	1	0,001	0,04	27,3	1	0	59986--2--0	11408--8--0
SRCAP	missense	1477 A>G	Ser493Gly	1	0.034	0,11	16,46	0	0	55532--1--0	7644--0--0
NPLOC4	missense	545 G>A	Ala182Val	1	0.114	0,3	26,1	0	0	54972--0--0	7496--0--0
LRP1B	missense	1430 A>G	Val477Ala	1	0	0,75	11,26	0	0	62453--7--0	11932--0--0
APC	missense	681 C>G	Asp227Glu	1	0,005	1	6,457	0	0		
PHF3	missense	1807 C>T	His603Tyr	1	0.016	0,01	0,05	0	0		
PSIP1	missense	995 C>A	Gly332Val	0,95	0,004	0,12	20,6	0	0	55632--4--0	7623--0--0
SPARC	missense	208 C>T	Gly70Ser	0,89	0,387	0,61	27	0	0		
DLX3	missense	148 A>T	Tyr50Asn	0,01	0,484	0,2	25,1	0	0	55837--5--0	7652--0--0
IQCH	missense	1654 A>G	Thr552Ala	0	0,661	0,29	24,8	0	0		
CHRNE	missense	1466 G>A	Pro489Leu	0	0,435	0,08	17,94	1	0	53984--1--0	7211--0--0
MAB21L3	missense	545 A>G	Gln182Arg	0	0,159	1	0,002	0	0		
ADAMTSL4	missense	196 C>T	Arg66Cys	0	0.855	0	29,4	1	0	62458--14--0	11798--1--0
IFI16	missense	1843 G>A	Val615Ile	0	0.003	0,15	0,075	1	0	63185--5--0	12016--0--0
CCDC181	nonsense	1048 T>A	Arg350Stop	0		1	35	0	0		
NEB	missense	1940 C>T	Arg647Gln	0	0.564	0,02	27,9	1	0		
GALNT5	missense	1673 G>A	Arg558Gln	0	0,958	0,05	29,1	0	0		
EVC	missense	1272 G>T	Gln424His	0	0,789	0,14	25,4	0	0		
PCDH18	missense	2416 C>T	Val806Met	0	0.048	0,13	21,8	0	0		
SRD5A1	missense	485 C>T	Met162Thr	0	0.221	0	24,5	0	0		
TDRD7	missense	2092 G>C	Glu698Gln	0	0,903	0,08	23,9	1	0	55641--2--0	7648--0--0
NECAB2	missense	985 G>A	Asp329Asn		0,999	0,2	23,2	0	0,000199681	63347--2--0	12011--0—0
LEKR1	missense	1273 A>G	Lys425Glu		0,938	0,58	28,2	0	0		

SIFT: values below 0.05 indicate damaging predictions – Polyphen2: values over 0.5 indicate possibly or probably damaging predictions – CADD: scores > 15 indicate damaging predictions. Allele frequency of each variant in the different exome databases (IHG Helmholtz, 1000 Genomes, ExAC) is reported. EUR denotes the European non-Finnish and AFR denotes the African sub-populations of ExAC.

3.2 Genetic causes of mitochondrial cardiomyopathy

Among the cardiac conditions with a recessive mode of inheritance, mitochondrial disorders represent an important cause of cardiomyopathy in an isolated form or as a part of multisystem involvement. NGS techniques are the state of the art tool to diagnose mitochondrial disorders, enabling rapid, cost-effective genome-wide screening with a diagnostic yield of greater than 60%. Of the approximate 300 mitochondrial disease-associated genes described to date, around half of them have been discovered with NGS technologies in the past 7 years. These findings have facilitated our understanding of the pathophysiology of mitochondrial diseases. Despite the scientific progress discussed thus far, it is important to underline the fact that most of these disorders do not have a treatment. Nevertheless, there are several mitochondrial defects, especially related to cofactor metabolism, that can be corrected by specific treatment strategies[154], highlighting the need for molecular diagnostics.

Interestingly, some of the most frequent mutated genes with a cardiac manifestation have been discovered in the last years thanks to the application of NGS technologies (*ACAD9, AGK, ELAC2*) and this gene list is constantly growing. During my PhD studies, a comprehensive and detailed collection and analysis of cardiac manifestations in the genes implicated in mitochondrial disorders was still lacking. For example, a recent review incorporated around 30 nuclear genes[155]. In order to improve the analysis of mitochondrial exomes and to update the candidate gene list used in the internal database, I started to search in literature and in publicly available databases the list of mitochondrial disease-associated genes to investigate the cardiac clinical manifestations in mitochondrial disorders. Moreover, for each gene I annotated the number of patients with cardiac phenotype reported. Surprisingly, I identified the presence of a cardiac phenotype in 90 mitochondrial genes and a list of reported 1880 patients. This list represents a minimal number, given the fact that many patients are not published or included in publicly available sources.

I noticed that for some genes the cardiac phenotype represented a frequent and key clinical feature (major symptom) in the clinical spectrum, for other genes it appeared more sporadically (minor symptom). Therefore, I divided the list of 90 genes into two groups (Figure 31). Concerning the number of patients per gene defect, there was a long tail distribution (Table 7 and Figure 29) with disorders like Friedreich ataxia for which more than 1200 patients have a reported cardiac involvement, Barth syndrome with more than 200 patients, and *ACAD9* with over 50 patients. On the other hand, there are around 70 disease entities with less than 10 published cases with cardiac involvement. Usually, not all patients with a certain gene defect

develop a cardiac phenotype, but the mitochondrial dysfunction predisposes patients to cardiac abnormalities. For most gene defects associated with a cardiac phenotype, NGS diagnostics extended the clinical spectrum adding patients with extracardiac features. For each of the 90 genes I also investigated in literature the presence of specific forms of cardiac manifestations, such as hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmias, congenital heart defects (CHDs), left ventricle non-compaction (LVNC) and restrictive cardiomyopathy (Figure 30).

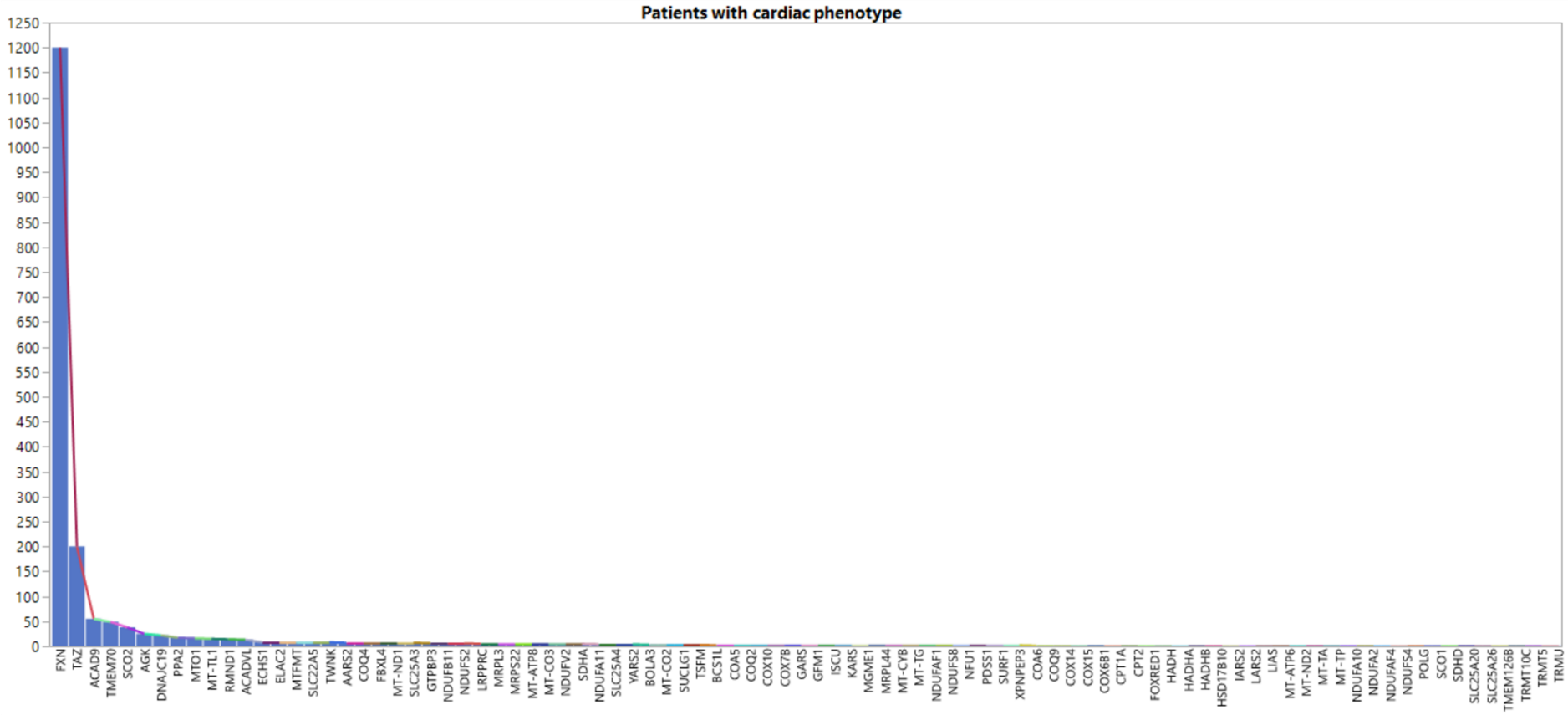

Figure 29. Number of patients with cardiac phenotype for each of the 90 mitochondrial genes.

Table 7. Number of patients with cardiac phenotype for each of the 90 mitochondrial genes.

Gene	N	Gene	N	Gene	N	Gene	N	Gene	N	Gene	N
NDUFA2	1	TRMU	1	LIAS	1	COQ2	2	MT-CO3	5	MTFMT	8
NDUFA10	1	LARS2	1	SLC25A26	1	PDSS1	2	MT-ATP8	5	ECHS1	8
NDUFS4	1	IARS2	1	NDUFS8	2	ISCU	2	LRPPRC	5	SLC22A5	8
MT-ND2	1	MT-TP	1	NDUFAF1	2	NFU1	2	MRPS22	5	ACADVL	13
NDUFAF4	1	COQ9	1	BCS1L	2	COX10	2	MRPL3	5	RMND1	14
FOXRED1	1	COX15	1	MT-CYB	2	GFM1	2	NDUFB11	6	MT-TL1	15
TMEM126B	1	SCO1	1	COX7B	2	MT-CO2	3	NDUFS2	6	MTO1	16
SDHD	1	COA6	1	SURF1	2	TSFM	3	GTPBP3	6	PPA2	17
COX6B1	1	CPT1A	1	COA5	2	SUCLG1	3	MT-ND1	7	DNAJC19	22
COX14	1	CPT2	1	XPNPEP3	2	BOLA3	3	FBXL4	7	AGK	25
MT-ATP6	1	HADH	1	MGME1	2	NDUFA11	4	AARS2	7	SCO2	38
POLG	1	HADHA	1	MRPL44	2	SLC25A4	4	COQ4	7	TMEM70	49
HSD17B10	1	HADHB	1	GARS	2	YARS2	4	SLC25A3	7	ACAD9	55
TRMT5	1	SLC25A20	1	KARS	2	NDUFV2	5	TWNK	8	TAZ	200
TRMT10C	1	MT-TA	1	MT-TG	2	SDHA	5	ELAC2	8	FXN	1200

In the cases investigated, hypertrophic cardiomyopathy (HCM) was the most frequent manifestation (Figure 30). HCM is characterized by progressive myocardial thickening, diastolic and systolic ventricular dysfunction, histopathologic changes, such as myocyte disarray and fibrosis, and arrhythmias which may cause SCD. HCM was reported to be associated with 77 mitochondrial disease genes out of 100 with a cardiac phenotype.

Dilated cardiomyopathy (DCM) is characterized by progressive myocardial dilatation and thinning. Both diastolic, as well as systolic ventricular dysfunction, can occur and DCM is often associated with the occurrence of cardiac conduction system diseases, arrhythmias and sudden arrhythmic death. In this study, DCM was often found as a consequence of a progressed HCM and in total, I identified it in association with 26 mitochondrial disease genes (Figure 30).

Arrhythmias were found in mitochondrial patients as cardiac conduction defects typically involving sinus node dysfunction, atrioventricular block, ventricular conduction delay, or ventricular pre-excitation. Progression in intraventricular conduction defect is unpredictable and responsible for SCD. I identified 16 mitochondrial disease genes in association with arrhythmias (Figure 30).

Congenital heart defects (CHDs) have been rarely recognized to be linked to mitochondrial disorders. I discovered CHDs in patients with mitochondrial disorders due to 15 different genetic defects (Figure 30). For the first time, mitochondrial gene defects have been considered in the pathophysiology of CHDs, however, currently the pathomechanism remains largely unknown. The CHDs include patent ductus arteriosus (PDA), ventricular and septal defects (VSD and ASD) or more complex CHD defects (tetralogy of Fallot, transpositions of great arteries). Among the 15 genes, considering 3 of the most frequent causes of cardiac disorder, CHDs have been reported in 10% of *ACAD9* and *TAZ* cases and 30% of *TMEM70* cases[156,157]. This figure was clearly above the 1% of cases with CHDs identified in the general population (Figure 31). The observation of CHD is present in only a fraction of patients with the clinical presentation of mitochondrial disorders.

Left ventricular non-compaction (LVNC) is a morphological abnormality of excessive trabeculation of the left ventricular myocardium, which is often complicated by ventricular dysfunction and arrhythmias. In this study, LVNC was typical for *TAZ* and *DNAJC19* defects and reported in a total of 10 mitochondrial disease genes (Figure 31).

Restrictive cardiomyopathy (RCM) was found as a very rare manifestation of mitochondrial diseases.

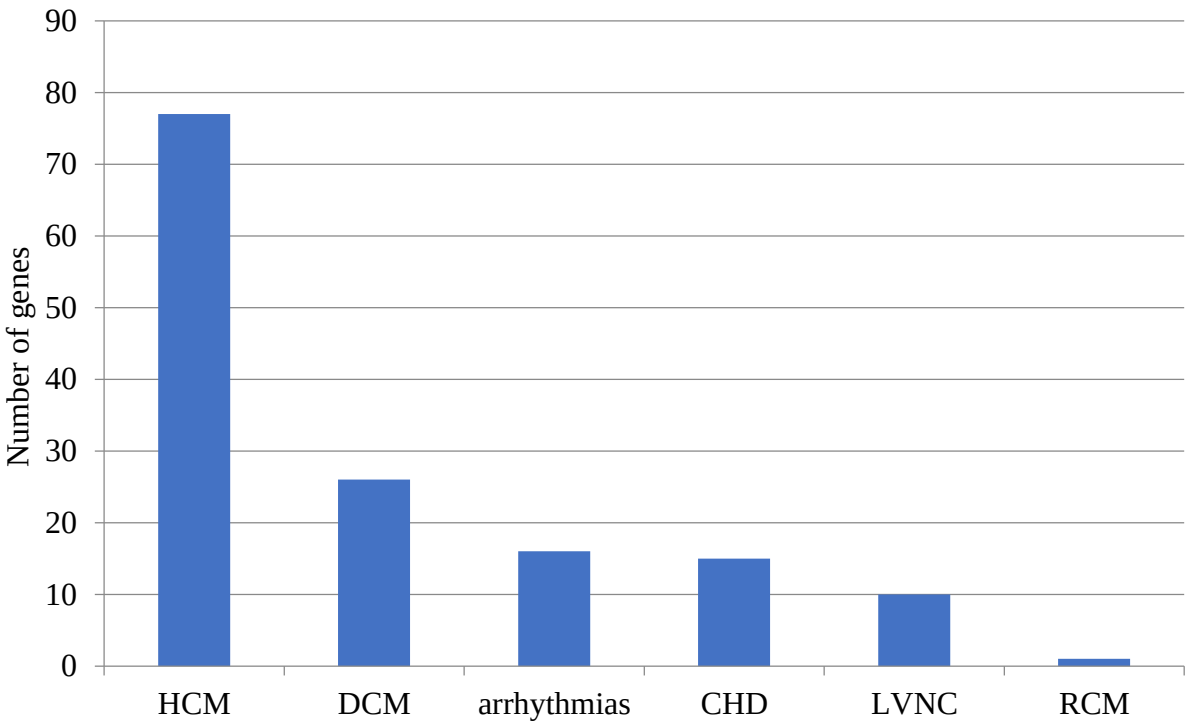

Figure 30. Hypertrophic cardiomyopathy (HCM) is the most frequent manifestation in mitochondrial disorders, followed by dilated cardiomyopathy (DCM), arrhythmias, congenital heart disease (CHD), left ventricular non-compaction (LVNC) and restrictive cardiomyopathy (RCM).

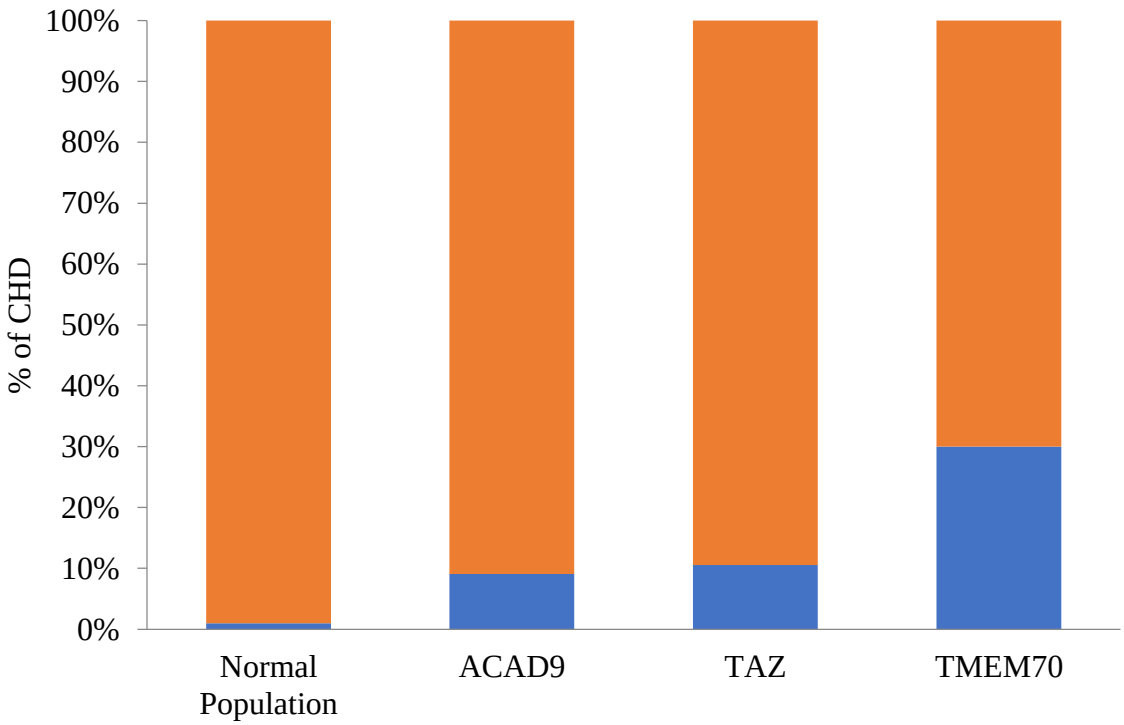

Figure 31. Congenital heart defects (CHDs) are underdiagnosed in mitochondrial disorders. I observed the occurrence of CHDs in 30% of *TMEM70* cases and 10% of *ACAD9* and *TAZ* cases, thus clearly above the 1% of CHDs cases occurring in the general population.

To break down the complexity, I grouped the genes, based on a recent publication of Wortmann et al.[52], into (1) disorders of oxidative phosphorylation (OXPHOS) subunits and their assembly factors, (2) defects of mitochondrial DNA, RNA and protein synthesis, (3) defects in the substrate-generating upstream reactions of OXPHOS, (4) defects in cofactor metabolism, (5) defects in mitochondrial homeostasis, and (6) defects in relevant inhibitors (Figure 32).

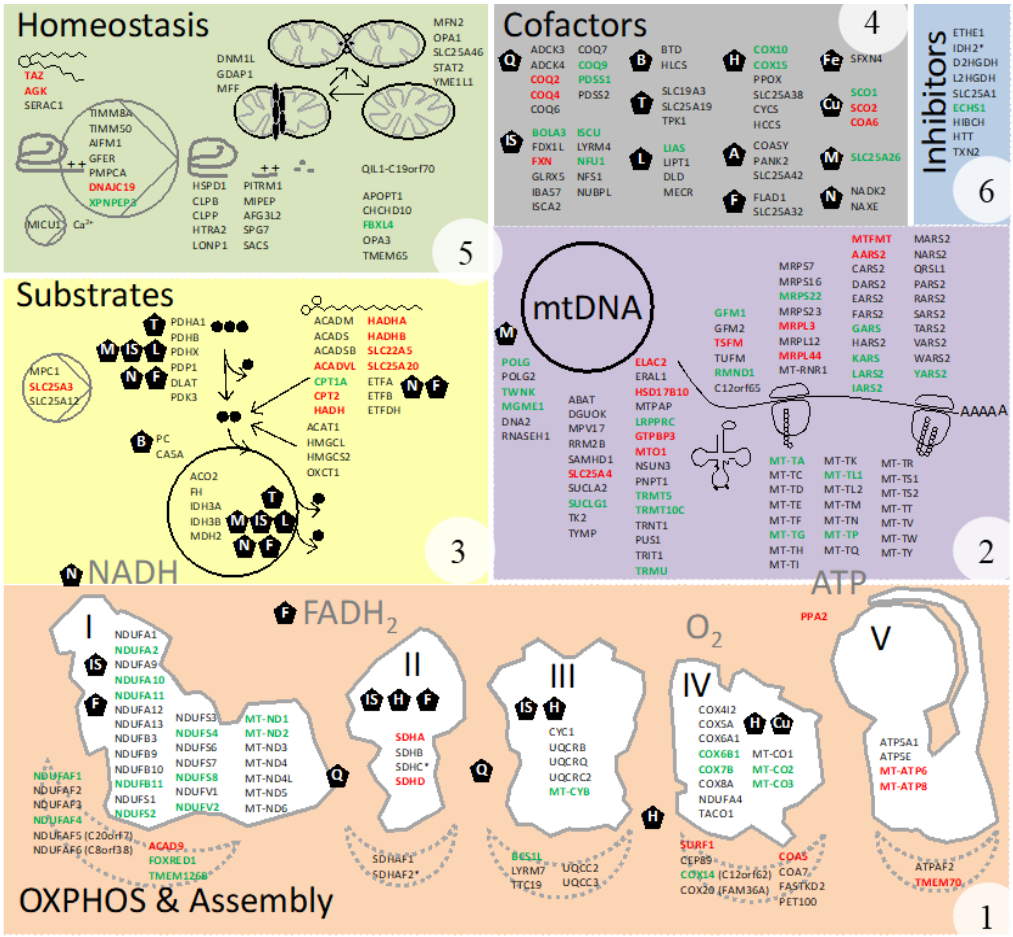

Figure 32. Summary of mitochondrial genes with cardiac manifestations: (1) disorders of oxidative phosphorylation (OXPHOS) subunits and their assembly factors; (2) defects of mitochondrial DNA, RNA and protein synthesis; (3) defects in the substrate-generating upstream reactions of OXPHOS; (4) defects in cofactor metabolism; (5) defects in mitochondrial homeostasis; and (6) defects in relevant inhibitors. Red= Major symptom; Green= Minor symptom. [Modified from [52]]

1- OXPHOS subunits and their assembling factors

Complex I deficiency is the most common biochemical phenotype observed in individuals with mitochondrial disease. One third of childhood-onset cases are related to this OXPHOS defect. Complex I is divided in 44 subunits and 37 are encoded by nuclear DNA. Genetic defects have been described in structural genes, assembly factors, and other factors involved in mitochondrial translation (Figure 32).

On a clinical level, complex I deficiency is diverse, with patients presenting with Leigh syndrome, with fatal infantile lactic acidosis (FILA) or leukoencephalopathy, lactic acidosis, MELAS. Hypertrophic cardiomyopathy is also a common feature and it can appear isolated or associated with multisystem disease. Isolated HCM has been related to mutations in nuclear-encoded subunits (*NDUFS2, NDUFV2*) and assembly factors (*ACAD9, NDUFAF1*). Cases with dilated cardiomyopathy, left ventricle non-compaction or conduction defects have been described [158].

At least 100 patients with complex I deficiency and cardiac abnormalities have been so far published., with *ACAD9* deficiency representing the most frequent genetic cause observed.

2- Defects of mitochondrial DNA, RNA and protein synthesis

Defects in mitochondrial protein translation are associated with multiple OXPHOS deficiencies and severe mitochondrial disease. In this subgroup are contained genes encoding elongation factors, aminoacyl-tRNA synthetases, tRNA-modifying enzymes, a mitochondrial peptide release factor, and an RNase that processes mitochondrial RNA. Mitochondrial disorders due to protein translation defects show neurological symptoms, hypertrophic cardiomyopathy and multisystem disorder. Cardiac abnormalities have been related to defects in mitochondrial ribosomal proteins (*MRPS22, MRPL3, MRPL44*), mitochondrial tyrosine (*YARS2*) or alanine (*AARS2*) tRNA aminoacylation, and other enzymes involved in mitochondrial RNA metabolism[159,160]. *ELAC2* mutations are typically presenting with early severe forms of hypertrophic cardiomyopathy together with a multisystem disease, but also isolated forms of cardiomyopathy have been described.

Post-transcriptional modifications of mitochondrial tRNAs are required for their stability and function. In this subgroup, cardiomyopathy has been associated with mutations in *MTO1, GTPBP3* and *TRMT5*. Hypertrophic cardiomyopathy plus arrhythmias, together with developmental delay, encephalopathy, hypotonia and lactic acidosis, represent the typical clinical spectrum of variants identified in *MTO1*[161]. Mutations in *GTPBP3* have been associated with

early forms of hypertrophic or dilated cardiomyopathy and conduction defects[162] with encephalopathy, lactic acidemia, hypoglycemia, and hyperammonemia. *TRMT5* deficiency includes, among the clinical features, hypertrophic cardiomyopathy, exercise intolerance, global developmental delay, hypotonia, peripheral neuropathy, renal tubulopathy, and lactic acidosis[163].

3– Defects in the substrate-generating upstream reactions of OXPHOS

Deficiencies of the mitochondrial phosphate carrier *SLC25A3* have been associated with hypertrophic cardiomyopathy accompanied by myopathy and lactic acidosis[164].

Loss of function mutations of the carnitine transport SLC22A5 have been linked with cardiomyopathy and SCD. Interestingly, hypertrophy cardiomyopathy and her exercise tolerance were strongly responsive to l-carnitine supplementation[165].

Biallelic variants in *PPA2*, the nuclear-encoded mitochondrial inorganic pyrophosphatase, were detected in 17 patients presenting with seizures, lactic acidosis, cardiac arrhythmia with specific sensitivity to alcohol, leading to SCD[166].

4- Defects in relevant cofactors

Among the cofactors, some are necessary for the respiratory chain enzymes like coenzyme Q, iron-sulphur clusters, riboflavin and haem.

Primary CoQ10 deficiency is a heterogeneous condition and the clinical presentations is inclusive of encephalomyopathy, myopathy, cerebellar ataxia, nephrotic syndrome, and severe infantile multisystem mitochondrial disease. Hypertrophic cardiomyopathy have been related to mutations in *COQ2, COQ4* and *COQ9*[167,168] and supplementation with Q10 has shown beneficial effects in these patients.

Fatal infantile cardioencephalomyopathy due to cytochrome c oxidase (COX) deficiency-1 is caused by biallelic variants in the *SCO2* gene. *SCO2* is a mitochondrial copper-binding protein involved in the biogenesis of the Cu(A) site in the cytochrome c oxidase (CcO) subunit Cox2. It is involved in cellular copper homeostasis. Mutations in *SCO1* and *SCO2* are associated with distinct clinical phenotypes linked to tissue-specific cytochrome c oxidase deficiency. *SCO2* is highly expressed in muscular tissue, whereas *SCO1* in liver tissue. Accordingly, *SCO1* deficiency is mostly related to hepatic liver failure, while *SCO2* deficiency is linked with severe cardiac failure. The onset of the cardiomyopathy is either in utero or in the first days of life[169]. Copper-histidine supplementation in cell culture and the treatment in one patient has shown an improvement of the cardiac manifestation[170].

5- Defects in mitochondrial homeostasis

Mitochondrial homeostasis involves several essential aspects of mitochondrial biogenesis, lipid synthesis, protein import, fission and fusion, quality control and targeted degradation.

Barth syndrome is due to mutations in the X-linked *TAZ* gene, which codes for Tafazzin, a phospholipid transacylase involved in the remodelling of cardiolipin. Barth syndrome is characterized by cardiomyopathy, skeletal myopathy, developmental delay, neutropenia and increased urinary levels of 3-methylglutaconic acid. Cardiomyopathy is the presenting manifestation in more than 70% of affected males and usually appears in infancy. Interestingly, in these patients, left ventricular non-compaction and dilated cardiomyopathies are more frequent cardiological findings than hypertrophic cardiomyopathy[171]. Patients can manifest with supraventricular and ventricular arrhythmias, sometimes related to SCD[172]. Recently, cardiac-specific loss of succinate dehydrogenase (complex II) activity has been described as a key event in the pathogenesis of cardiomyopathy in a mouse model of Barth syndrome[173]. Often it requires time until the cardiac phenotype is recognized as Barth syndrome.

Sengers syndrome is caused by the deficiency of the acylglycerol kinase (AGK) which is also involved in the mitochondrial protein import. The clinical spectrum is characterized by the presence of hypertrophic cardiomyopathy, cataracts, myopathy, exercise intolerance, and lactic acidosis[174].

3-methylglutaconic aciduria associated with *DNAJC19* mutations (DCMA syndrome) is mainly associated with dilated cardiomyopathy or left ventricular non-compaction, and non-progressive cerebellar ataxia[175].

6- Defects in relevant inhibitors

Mitochondrial short-chain enoyl-CoA hydratase-1 deficiency (ECHS1) is an autosomal recessive inborn error of metabolism caused by compound heterozygous mutation in the *ECHS1* gene and characterized by severe developmental delay, increased lactic acid and cardiomyopathy due to the accumulation of toxic metabolites. Usually, these patients are affected by hypertrophic cardiomyopathy, and more rarely, by the dilated cardiomyoapathy[176].

3.2.1 Variants in *ACAD9* associated with mitochondrial cardiomyopathy

Among the 100 genes associated with mitochondrial cardiomyopathy, *ACAD9* is one of the most frequent. This gene was discovered by WES in 2010 and since then 67 patients have been described in the literature. In vitro experiments done in the Institute of Human Genetics (PhD work of Birgit Repp) have underlined a positive effect of riboflavin on mitochondrial respiration. Therefore, I investigated the clinical spectrum of *ACAD9* patients focusing on the cardiac manifestations and the clinical effect of riboflavin treatment. For these reasons, a Survey Monkey was created and circulated among the collaborators. I used the clinical data collected with the Survey to investigate the effect of riboflavin on the outcome of the patients, considering also clinical features and age of onset (Table 8). I used this data to study the effect of riboflavin on the survival of *ACAD9* patients and computed surviving curves (Figure 33). In addition, I used the list of *ACAD9* pathogenic variants to calculate the incidence of *ACAD9* deficiency in the European population. I compute the minor allele frequency of LOF variants contained in ExAC and GnomAD and the minor allele frequency (taken from ExAC and GnomAD) of the causal *ACAD9* variants present in this cohort (Table 11).

3.2.1.1 *ACAD9* cohort

67 individuals from 47 families, of which 26 newly discovered (Table 9 and10) were enrolled. 34 subjects were still alive at a median age of 14 years, with a range of 24 days – 44 years. On the other hand, the median age of patients deceased was 3 months, with a range of 1 day – 44 years[250].

The majority of the patients had a clinical presentation in the first year of life. This subgroup showed a poor survival, with more than 50% not surviving the first two years[250]. Patients with a later presentation (up to 1 year of age) had a very good survival (more than 90% surviving 10 years). In one individual (Ind. 18) fetal cardiomegaly was identified, and two other cases were reported with fetal rhythm alterations. These three individuals died early after birth. Clinical characteristics of ACAD9 deficiency were cardiomyopathy (86%), exercise intolerance (74%) and muscular weakness (79%)[250]. The clinical findings are contained in Table 8 and Figure 33 A, B.

Table 8. Clinical features of the *ACAD9* cohort. [Modified from[150]]

	n	n available	%
Prenatal findings			
Cardiomegaly	1	59	2
Rhythm abnormalities	2	59	3
Decreased child movements	1	59	2
Oligohydramnios	4	59	7
Intrauterine growth failure	6	59	10
Neonatal course			
Lactic acidosis	21	62	34
Cardiomyopathy	15	62	24
Rhythm abnormalities	4	53	8
Respiratory failure necessitating artificial ventilation	6	53	11
Severe liver dysfunction/failure	2	53	4
Severe renal dysfuntion/failure	2	53	4
Most frequent clinical findings			
Cardiomyopathy at presentation/during course	42/54	63/63	67/86
Muscular weakness at presentation/during course	18/34	45/46	40/74
Exercise intolerance at presentation/during course	19/31	46/44	41/70
Neurological findings			
Severe intellectural disability (clinical impression)	1	48	2
Mild intellectual disability (clinical impression)	13	45	29
Severe developmental delay (clinical impression)	4	49	8
Mild developmental delay (clinical impression)	20	48	42
Optic atrophy, retinits pigmentosa	0	67	0
Neuroradiological findings			
MRI: basal ganglia alterations	4	23	17
MRI: leukoencephalopathy	5	20	25
MRI: global brain atrophy	1	19	5
MRI: isolated cerebellar atrophy	1	19	5
MRS: lactate peak (any location)	2	16	13
Activities of daily living			
Age adequate behaviour	27	36	75
Attending/finished regular school	26	38	68
Able to sit independently	31	39	79
Able to walk independently	30	38	79
Able to eat and drink independently	30	38	79
Able to perform personal hygiene independently	28	38	74
Able to communicate with words/sentences	32/29	40/39	80/74

Cardiac manifestations

Isolated hypertrophic cardiomyopathy (HCM) was the most prevalent cardiac manifestation (42/54=78%) while isolated dilated cardiomyopathy (7/54=13%) and combined forms were (6/54=11%) less common[250]. Three patients survived an arrhythmogenic event that required resuscitation (3/33=9%), as indicated earlier progressive cardiorespiratory failure (15/31=48%) and sudden cardiac events (10/31=32%) were the main causes leading to death[250].

During the neonatal period, 15 patients (15/62=24%) manifested hypertrophic cardiomyopathy and 12 of them died early in infancy. In addition, four patients showed rhythm abnormalities (4/62=6%) and five (5/6=8%) CHDs[250].

Four cases in this cohort underwent cardiac transplantation. Individual 20 presented with rapidly progressive cardiomyopathy in the first year of life and she had heart transplantation at two years. Unfortunately, four years later she died for heart failure[250].

Individual 21 developed a hypertrophic cardiomyopathy at 18 months and later she manifested neurological features (ataxia and epilepsy). She received a heart transplantation at 9 years of age. She is currently alive (15 years old). Individual 22 (35 years old) manifested a hypertrophic cardiomyopathy in childhood and she received a heart transplantation at 18 years. Individual 30 presented with tachycardia in the first days of life and he developed cardiomyopathy at one month. Despite heart transplantation, he died at three months of age[250].

Riboflavin and other oral vitamin treatment

Of the entire cohort of 67 patients, 20 patients were reported as not treated; data about treatment and/or effect were unavailable for 15 patients[250]. Data on the general clinical effect of riboflavin as reported by the responsible physician were available for 31 patients. For 20 patients (20/31 =65%) physicians reported a good effect, for 11 (35%) no effect. No clinical deterioration or side effects were reported. Detailed data on onset of riboflavin treatment, dosage and duration were only available for a minority of patients and have not been investigated[250].

To analyze the effect of riboflavin treatment, I focused on the patients presenting during the first year of life as these was the largest subgroup and the group with the shortest survival (suggesting the most severe course). For 37 of these 48 patients, data on riboflavin treatment were available (n=17 untreated, n=20 treated). Figure 33 section C and D shows the Kaplan-Meier curve for both groups of patients and indicates a significantly better survival rate for patients with oral riboflavin treatment (deceased n= 7/20) in contrast to untreated patients (deceased n=16/17).

Regarding other food supplements, several patients were reported as taking coenzyme Q10, biotin and L-carnitine with anecdotal positive effects[250].

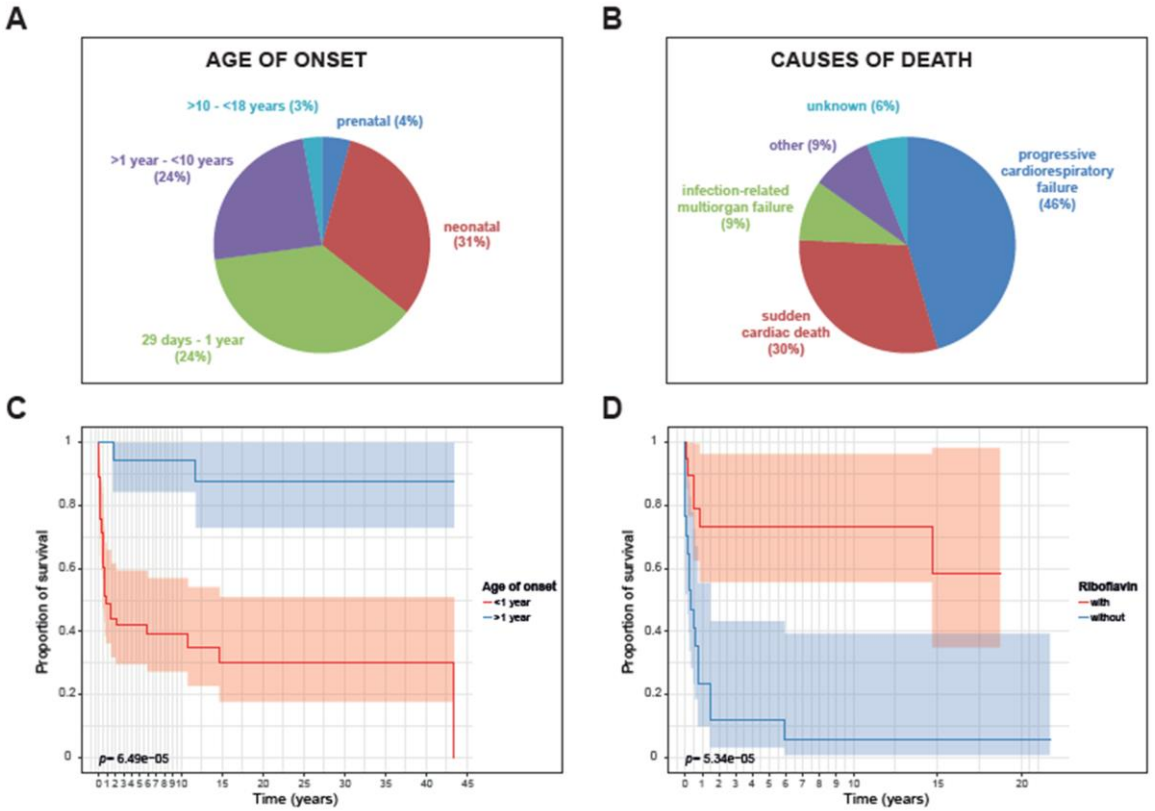

Figure 33. Age of onset, causes of death, survival and effect of riboflavin on the survival of ACAD9 patients. (A) Age of onset of symptoms, (B) Causes of death, (C) Kaplan-Maier survival rates. In red, patients with presentation in the first year. In blue, patients with a later presentation (D) In red, patients with in the first year and treated with riboflavin. In blue, patients of the same age category but untreated with riboflavin. [Modified from[250]]

Table 9. Details of the *ACAD9* cohort: genetic features.

ID, family, sex	Country	Ref.	cDNA	Protein	Variant type	MAF	cDNA	Protein	Variant type	MAF
1, 1, F	Italy	Pat.1, #35834, [177]	c.130T>A	p.Phe44Ile	missense	np	c.797G>A	p.Arg266Gln	missense	1.219e-5
2, 1, M	Italy	Pat.2, #52935, [177]	c.130T>A	p.Phe44Ile	missense	np	c.797G>A	p.Arg266Gln	missense	1.219e-5
3, 2, F	Italy	Pat.3, #52933, [177]	c.797G>A	p.Arg266Gln	missense	1.219e-5	c.1249C>T	p.Arg417Cys	missense	np
4, 3, F	UK	Pat.4, #52674, [177]	c.976G>C	p.Ala326Pro	missense	5.279e-5	c.1594C>T	p.Arg532Trp	missense	4.089e-6
5, 4, F	Netherlands	Pat.1, CB, VII:11, [178]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
6, 4, F	Netherlands	Pat.2, #49591, MJ, VII:6, [178]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
7, 4, M	Netherlands	Pat.3, JJ, VII:8, [178]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
8, 5, F	Netherlands	Pat.4, CV, [178]	c.380G>A	p.Arg127Gln	missense	np	c.1405C>T	p.Arg469Trp	missense	0.0003754
9, 6, F	Canada	Pat.1, #69842, [179]	c.1553G>A	p.Arg518His	missense	8.66e-5	c.1553G>A	p.Arg518His	missense	8.66e-5
10, 7, F	Thailand/ Germany	Pat.2 (twin sister), [179]	c.187G>T	p.Glu63X	nonsense	np	c.1237G>A	p.Glu413Lys	missense	1.625e-5
11, 7, F	Thailand/ Germany	Pat.3 (twin sister), [179]	c.187G>T	p.Glu63X	nonsense	np	c.1237G>A	p.Glu413Lys	missense	1.625e-5
12, 8, M	Turkey	Pat.1, #59029, 72545, [180]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
13, 8, F	Turkey	Pat.2, #59033, [180]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
14, 8, F	Turkey	Pat.3, #59036, [180]	c.1594C>T	p.Arg532Trp	missense	4.089e-6	c.1594C>T	p.Arg532Trp	missense	4.089e-6
15, 9, M	Not available	[181]	c.260T>A	p.Ile87Asn	missense	np	c.976G>A	p.Ala326Pro	missense	np
16, 10, M	Italy	[182]	c.1240C>T	p.Arg414Cys	missense	1.219e-5	c.1240C>T	p.Arg414Cys	missense	1.219e-5
17, 11, F	Turkey	[183]	c.659C>T	p.Ala220Val	missense	np	c.659C>T	p.Ala220Val	missense	np
18, 12, M	France	[184]	c.1030-1G>T	acceptor splice	splice	np	c.1249C>T	p.Arg417Cys	missense	np
19, 13, F	West Africa	P2, [185]	c.976G>A	p.Ala326Pro	missense	5.279e-5	c.1595G>A	p.Arg532Gln	missense	8.174e-6
20, 14, F	Portugal	P3, [185]	c.358delT	p.Phe120fs	frameshift	0.000109	c.1594C>T	p.Arg532Trp	missense	4.089e-6
21, 15, F	France	P4, [185]	c.976G>C	p.Ala326Pro	missense	5.279e-5	c.1595G>A	p.Arg532Gln	missense	8.174e-6
22, 16, F	France	P5, [185]	c.151-2A>G	acceptor splice	splice site	np	c.1298G>A	p.Arg433Gln	missense	4.062e-6
23, 17, M	French Caribbean	P6, [185]	c.1237G>A	p.Glu413Lys	missense	1.625e-5	c.1552C>T	p.Arg518Cys	missense	8.66e-5
24, 18, F	French Caribbean	P7, [185]	c.1552C>T	p.Arg518Cys	missense	8.66e-5	c.1564-6_1569del	splice site	splice	np
25, 19, F	French Caribbean	P8, [185]	c.1A>G	p.Met1?	start lost	1.277e-5	c.796C>T	P.Arg266Trp	missense	8.123e-6
26, 20, F	Japan	Pt090, [186]	c.1150G>A	p.Val384Met	missense	4.065e-6	c.1817T>A	p.Leu606His	missense	np
27, 21, F	Japan	Pt025, [186]	c.811T>G	p.Cys271Gly	missense	np	c.1766-2A>G	splice site	splice site	np
28, 22, M	not available	[187]	c.187G>T	p.Glu63X	nonsense	np	c.941T>C	p.Leu314Pro	missense	np
29, 23, F	Poland	P1, [188]	c.514G>A	p.Gly172Arg	missense	1.804e-5	c.803C>T	p.Ser268Phe	missense	8.124e-6
30, 24, M	Poland	P2, [188]	c.1552C>T	p.Arg518Cys	missense	8.66e-5	c.1553G>A	p.Arg518His	missense	8.66e-5

ID, family, sex	Country	Ref.	cDNA	Protein	Variant type	MAF	cDNA	Protein	Variant type	MAF
31, 25, M	Poland	P3, [188]	c.728C>G	p.Thr243Arg	missense	8.121e-6	c.1552C>T	p.Arg518Cys	missense	8.66e-5
32, 26, F	Morocco	I-1, [189]	c.1636G>C	p.Val546Leu	missense	np	c.1636G>C	p.Val546Leu	missense	np
33, 26, F	Morocco	I-2, [189]	c.1636G>C	p.Val546Leu	missense	np	c.1636G>C	p.Val546Leu	missense	np
34, 26, M	Morocco	I-3, [189]	c.1636G>C	p.Val546Leu	missense	np	c.1636G>C	p.Val546Leu	missense	np
35, 26, F	Morocco	I,4, [189]	c.1636G>C	p.Val546Leu	missense	np	c.1636G>C	p.Val546Leu	missense	np
36, 27, M	Belgium	II-1, [189]	c.509C>T	p.Ala170Val	missense	1.218e-5	c.1687C>G	p.His563Asp	missense	np
37, 27, F	Belgium	II-2, [189]	c.509C>T	p.Ala170Val	missense	1-218e-5	c.1687C>G	p.His563Asp	missense	np
38, 28, F	Congo	III-3, [189]	c.1240C>A	p.Arg414Ser	missense	4.062e-6	c.1650_1672dup	p.Leu558fs	frameshift	np
39, 28, F	Congo	III-6, [189]	c.1240C>A	p.Arg414Ser	missense	4.062e-6	c.1650_1672dup	p.Leu558fs	frameshift	np
40, 28, F	Congo	III-7, [189]	c.1240C>A	p.Arg414Ser	missense	4.062e-6	c.1650_1672dup	p.Leu558fs	frameshift	np
41, 29, F	Tunisia	[190]	c.1240C>T	p.Arg414Cys	missense	1.219e-5	c.1240C>T	p.Arg414Cys	missense	1.219e-5
42, 30, M	Austria	This study	c.1690G>A	p.Glu564Lys	missense	4.074e-6	c.1832A>G	p.Tyr611Cys	missense	8.121e-6
43, 31, M	Pakistan	This study	c.1553G>A	p.Arg518His	missense	1.083e-5	c.1553G>A	p.Arg518His	missense	1.083e-5
44, 32, F	Asia	This study	c.293T>C	p.Leu98Ser	missense	np	c.293T>C	p.Leu98Ser	missense	np
45, 33, F	Sri Lanka	This study	c.1253A>G	p.Asp418Gly	missense	np	c.1253A>G	p.Asp418Gly	missense	np
46, 33, M	Sri Lanka	This study	c.1253A>G	p.Asp418Gly	missense	np	c.1253A>G	p.Asp418Gly	missense	np
47, 33, M	Sri Lanka	This study	c.1253A>G	p.Asp418Gly	missense	np	c.1253A>G	p.Asp418Gly	missense	np
48, 34, F	Italy	This study	c.857T>C	p.Leu286Pro	missense	np	c.1240C>T	p.Arg414Cys	missense	1.219e-5
49, 35, F	France	This study	c.976G>C	p.Ala326Pro	missense	5.279e-5	c.1651A>G	p.Ser551Gly	missense	np
50, 36, M	Bahrain	This study	c.1684G>A	p.Asp562Asn	missense	4.34e-5	c.1684G>A	p.Asp562Asn	missense	4.34e-5
51, 37, F	not available	This study	c.1805C>T	p.Ser602Phe	missense	np	c.1805C>T	p.Ser602Phe	missense	np
52, 37, M	not available	This study	c.1805C>T	p.Ser602Phe	missense	np	c.1805C>T	p.Ser602Phe	missense	np
53, 38, M	Nigeria	This study	c.868G>A	p.Gly290Arg	missense	3.228e-5	c.1237G>A	p.Glu413Lys	missense	1.625e-5
54, 38, M	Nigeria	This study	c.868G>A	p.Gly290Arg	missense	3.228e-5	c.1237G>A	p.Glu413Lys	missense	1.625e-5
55, 38, M	Nigeria	This study	c.868G>A	p.Gly290Arg	missense	3.228e-5	c.1237G>A	p.Glu413Lys	missense	1.625e-5
56, 39, F	UK	This study	c.976G>C	p.Ala326Pro	missense	5.279e-5	c.1594C>T	p.Arg532Trp	missense	4.089e-6
57, 40, M	UK	This study	c.665T>A	p.Ile222Asn	missense	np	c.1249C>T	p.Arg417Cys	missense	np
58, 40, M	UK	This study	c.665T>A	p.Ile222Asn	missense	np	c.1249C>T	p.Arg417Cys	missense	np
59, 41, M	UK	This study	c.1150G>A	p.Val384Met	missense	4.065e-6	c.1168G>A	p.Ala390Thr	missense	1.444e-5
60, 41, M	UK	This study	c.1150G>A	p.Val384Met	missense	4.065e-6	c.1168G>A	p.Ala390Thr	missense	1.444e-5
61, 42, F	UK	This study	c.1552C>T	p.Arg518Cys	missense	8.66e-5	c.1715G>A	p.Cys572Tyr	missense	np
62, 43, F	Belgium	This study	c.976G>C	p.Ala326Pro	missense	5.279e-5	c.1552C>T	p.Arg518Cys	missense	8.66e-5
63, 44, M	Italy	This study	c.1240C>T	p.Arg414Cys	missense	1.219e-5	c.1646G>A	p.Arg549Gln	missense	8.18e-6
64, 45, F	Germany/ Poland	This study	c.569C>T	p.Ala190Val	missense	4.074e-6	c.1405C>T	p.Arg469Trp	missense	0.0003754
65, 46, M	not available	This study	c.1240C>T	p.Arg414Cys	missense	1.219e-5	c.1240C>T	p.Arg414Cys	missense	1.219e-5
66, 46, M	not available	This study	c.1240C>T	p.Arg414Cys	missense	1.219e-5	c.1240C>T	p.Arg414Cys	missense	1.219e-5
67, 47, M	not available	This study	c.555-2A>G	splice site	splice site	np	c.1168G>A	p.Ala390Thr	missense	1.444e-5

MAF=Minor allele frequency reported in gnomAD, np=not present. [Modified from[250]]

Table 10. Detail of the *ACAD9* cohort: clinical features.

ID, Family, Sex	Ref.	Age of onset	Age of onset*	Current age in day (* if dead)	Prenatal findings	Cardio-myopathy	Arrhythmias	Cardiological features	Neurological features	Effect of Riboflavin (rated by physician)	Effect of Riboflavin (cells)
1, 1, F	Pat.1, #35834, [177]	neonatal	1	49*		PS	NA	HCM	NA	untreated	positive effect
2, 1, M	Pat.2, #52935, [177]	neonatal	1	3600		PS	NA	HCM	no	positive effect	positive effect
3, 2, F	Pat.3, #52933, [177]	early infancy	2	4320*		PS	NA	HCM	no	positive effect	positive effect
4, 3, F	Pat.4, #52674, [177]	1 w	1	750*		DS	No	DCM	mild ID	positive effect	positive effect
5, 4, F	Pat.1, CB, VII:11, [178]	4 y	3	15480		no	No	No	no	positive effect	NA
6, 4, F	Pat.2, #49591, MJ, VII:6, [178]	4 y	3	12240		no	No	No	mild ID	positive effect	positive effect
7, 4, M	Pat.3, JJ, VII:8, [178]	4 y	3	14760		no	No	No	no	positive effect	NA
8, 5, F	Pat.4, CV, [178]	early childhood	3	10800		DS	No	HCM	mild DD	positive effect	NA
9, 6, F	Pat.1, #69842, [179]	8 m	2	6480		PS	NA	HCM	no	positive effect	positive effect
10, 7, F	Pat.2 (twin sister), [179]	4 m	2	180*		PS	NA	HCM	NA	NA	NA
11, 7, F	Pat.3 (twin sister), [179]	6 m	2	240*		PS	NA	HCM	NA	NA	NA
12, 8, M	Pat.1, #59029, 72545, [180]	6 m	2	5370*	Oligohydramnios	DS	No	HCM, DCM	mild DD	positive effect	positive effect

ID, Family, Sex	Ref.	Age of onset	Age of onset*	Current age in day (* if dead)	Prenatal findings	Cardio-myopathy	Arrhythmias	Cardiological features	Neurological features	Effect of Riboflavin (rated by physician)	Effect of Riboflavin (cells)
13, 8, F	Pat.2, #59033, [180]	4.5 m	2	2160	Oligohydramnios	DS	No	HCM	mild DD	positive effect	NA
14, 8, F	Pat.3, #59036, [180]	4 m	2	3240	Oligohydramnios	DS	No	HCM, DCM	mild DD	positive effect	NA
15, 9, M	[181]	neonatal	1	28*		PS	NA	NA	NA	NA	NA
16, 10, M	[182]	1 y	2	6840		no	NA	No	mild ID, mild DD	positive effect	NA
17, 11,F	[183]	neonatal	1	180*		DS	No	HCM	NA	no effect	NA
18, 12, M	[184]	prenatal	0	1*		PS	PS	HCM	NA	untreated	Positive effect
19, 13, F	P2, [185]	neonatal	1	70*	IGR, fetal rhythm abnormalities	PS	PS	HCM	no	untreated	NA
20, 14, F	P3, [185]	1 y	2	2160*		PS	PS	HCM, DCM	NA	untreated	positive effect
21, 15, F	P4, [185]	18 m	3	5400		PS	No	HCM	no	untreated	no effect
22, 16, F	P5, [185]	9 y	3	12600		PS	No	HCM	no	NA	no effect
23, 17, M	P6, [185]	4 m	2	7920		PS	No	HCM	no	untreated	no effect
24, 18, F	P7, [185]	15 m	3	630*		PS	NA	HCM	no	untreated	NA
25, 19, F	P8, [185]	1 y	2	540*		DS	No	HCM	no	untreated	NA
26, 20, F	Pt090, [186]	2 w	1	14	IGR	PS	NA	HCM	NA	NA	NA
27, 21, F	Pt025, [186]	14 y	4	5040		PS	NA	HCM	NA	NA	NA
28, 22, M	[187]	neonatal	1	1*	IGR, fetal rhythm abnormalities	PS	PS	HCM, DCM	NA	NA	NA
29, 23, F	P1, [188]	2 d	1	3180		PS	DS	HCM	mild DD	no effect	NA
30, 24, M	P2, [188]	1 m	1	90*	32 Hbd: preterm uterus contractions	PS	DS	HCM, DCM	mild DD	untreated	NA
31, 25, M	P3, [188]	2 y	3	3240		no	No	no	no	NA	NA

ID, Family, Sex	Ref.	Age of onset	Age of onset*	Current age in day (* if dead)	Prenatal findings	Cardio-myopathy	Arrhythmias	Cardiological features	Neurological features	Effect of Riboflavin (rated by physician)	Effect of Riboflavin (cells)
32, 26, F	I-1, [189]	neonatal	1	150*	prematurely 26.7w due to pre-eclampsia	NA	NA	NA	NA	NA	NA
33, 26, F	I-2, [189]	2 m	2	315*		DS	DS	HCM	severe DD	no effect	NA
34, 26, M	I-3, [189]	15 d	1	270*		PS	NA	HCM	mild ID, severe DD	untreated	NA
35, 26, F	I,4, [189]	15 m	3	3240		NA	NA	NA	mild ID, mild DD	positive effect	NA
36, 27, M	II-1, [189]	12 y	4	10440		PS	NA	HCM	mild DD	NA	NA
37, 27, F	II-2, [189]	8 y	3	9360		PS	NA	HCM	mild ID, mild DD	untreated	NA
38, 28, F	III-3, [189]	2 d	1	9*	born after preclampsia of the mother, 29w	NA	NA	NA	NA	untreated	NA
39, 28, F	III-6, [189]	neonatal	1	2*		PS	NA	HCM	NA	untreated	NA
40, 28, F	III-7, [189]	neonatal	1	180*		PS	NA	HCM, DCM	mild ID	no effect	NA
41, 29, F	[190]	6 y	3	12240		no	NA	No	no	NA	NA
42, 30, M	This study	prenatal	0	2*	lissencephalopathy agenesis of the corpus callosum, IGR	PS	NA	DCM	NA	untreated	no effect
43, 31, M	This study	3 m	2	120*		PS	No	HCM	NA	untreated	NA
44, 32, F	This study	7 y	3	5040		no	NA	No	no	no effect	no effect
45, 33, F	This study	4 m	2	540*		PS	NA	DCM	mild ID, severe DD	untreated	NA
46, 33, M	This study	9 m	2	3960		DS	No		no	positive effect	NA

ID, Family, Sex	Ref.	Age of onset	Age of onset*	Current age in day (* if dead)	Prenatal findings	Cardio-myopathy	Arrhythmias	Cardiological features	Neurological features	Effect of Riboflavin (rated by physician)	Effect of Riboflavin (cells)
47, 33, M	This study	7m	2	720		DS	No	HCM	mild ID, mild DD	positive effect	NA
48, 34, F	This study	neonatal	1	2880		DS	No	HCM, DCM	mild ID, mild DD	no effect	Positive effect
49, 35, F	This study	1 y	2	3960*		PS	No	HCM	mild ID, mild DD	NA	NA
50, 36, M	This study	neonatal	1	1080	hydronephrosis left, IGR	no	No	no	mild DD	NA	NA
51, 37, F	This study	2 m	2	180*		PS	NA	HCM, DCM	NA	untreated	NA
52, 37, M	This study	10 y	3	4680		PS	No	HCM	mild ID, mild DD	positive effect	NA
53, 38, M	This study	9 m	2	270*		PS	No	HCM	no	untreated	NA
54, 38, M	This study	9 m	2	3210		PS	PS	HCM	severe ID, mild DD	positive effect	NA
55, 38, M	This study	prenatal	0	280		PS	PS	HCM	mild ID, mild DD	positive effect	NA
56, 39, F	This study	neonatal	1	1*	oligoidramnios, IGR, decreased child movements	PS	NA	HCM	NA	untreated	NA
57, 40, M	This study	neonatal	1	28*		no	No	no	NA	no effect	NA
58, 40, M	This study	6 m	2	210*		NA	NA	NA	no	untreated	NA
59, 41, M	This study	2 y	3	13680		PS	No	HCM	mild DD	no effect	NA
60, 41, M	This study	9 y	3	15840		DS	NA	NA	no	no effect	NA
61, 42, F	This study	2 m	2	NA		PS	NA	HCM	no	NA	NA
62, 43, F	This	2 y	3	1080		PS	NA	HCM	no	untreated	NA

ID, Family, Sex	Ref.	Age of onset	Age of onset*	Current age in day (* if dead)	Prenatal findings	Cardio-myopathy	Arrhythmias	Cardiological features	Neurological features	Effect of Riboflavin (rated by physician)	Effect of Riboflavin (cells)
	study										
63, 44, M	This study	neonatal	1	3240		PS	No	HCM	mild DD	positive effect	NA
64, 45, F	This study	1 y	2	6000		PS	No	HCM	mild DD	positive effect	NA
65, 46, M	This study	infancy	2	15840*		PS	NA	HCM, DCM	no	NA	NA
66, 46, M	This study	neonatal	1	5760		PS	No	HCM	no	no effect	NA
67, 47, M	This study	neonatal	1	60*		PS	NA	HCM	severe DD	no effect	NA

***0 = prenatal, 1 = neonatal (birth -28 days), 2 =>1-12 months, 3 = >12m -10y, 4 = adolescence, 5 = adult, 6 = na. DCM= Dilated Cardiomyopathy, DD=developmental delay, DS=developing symptom, HCM=Hypertrophic cardiomyopathy, ID=intellectual disability; PS= presenting symptom; NA= not available. [Modified from[250]]**

3.2.1.2 Pathogenic bi-allelic variants in *ACAD9*

Figure 34 and Table 9 show the 15 previously unreported variants (bold in Figure 34) and 34 known variants in *ACAD9* found in the cohort. Of these 40 were missense, one frame shift, one nonsense, six splice site and one start codon mutation. No individual carried two variants predicted to lead to a loss of protein function[250].

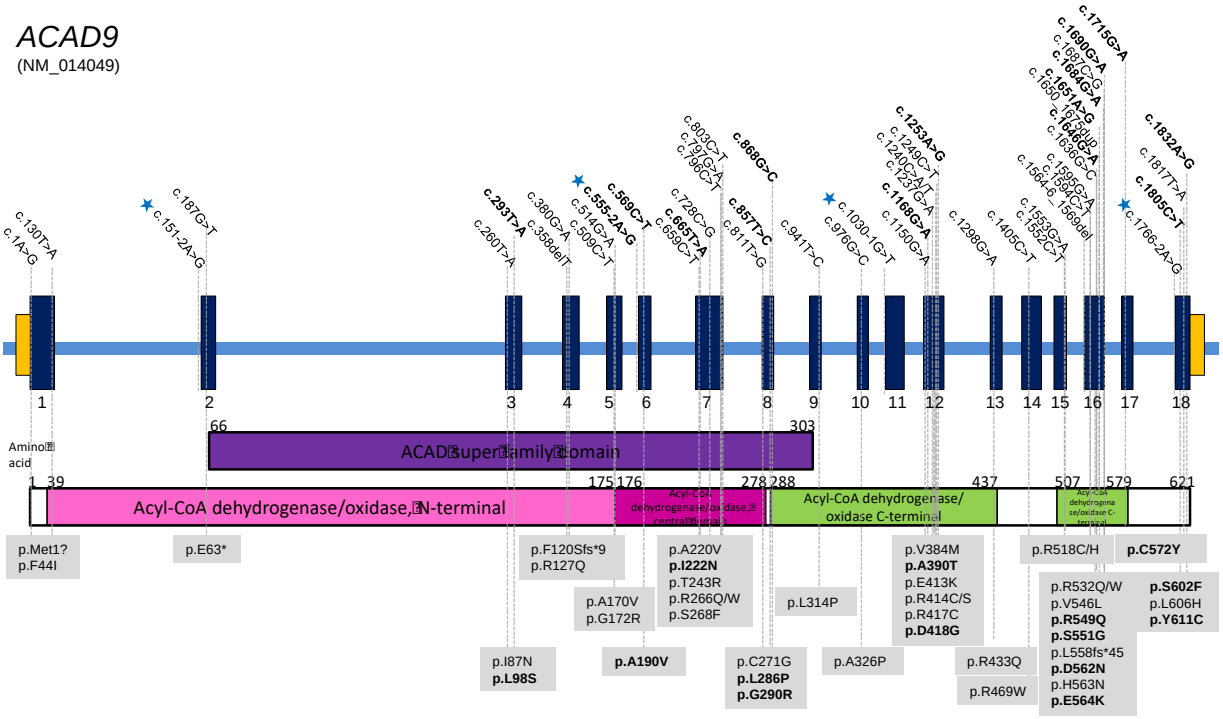

Figure 34. *ACAD9* mutation status, gene structure and conservation of affected amino acid residues. Gene structure of *ACAD9* with localization of mutations in 67 patients. Blue asterisks indicate splice site mutations. Newly identified mutations are shown in bold. [Modified from[250]]

Based on the prevalence of deleterious *ACAD9* alleles in the normal population (GnomAD, www.gnomad.broadinstitute.org,) I estimated that approximately 27 children with *ACAD9* deficiency will be born each year in Europe (Table 11). It was estimated by computing the minor allele frequency of LOF variants contained in ExAC and GnomAD and the causal *ACAD9* variants present in this cohort. Then the total allele frequency was used to estimate the prevalence of *ACAD9* deficiency. The incidence was then calculated in the European population (5.1 millions of individuals)[250].

Table 11. Estimated incidence of *ACAD9* deficiency in the European population.

		Formula	
ACAD9 LoFs* in ExAC *ACAD*9 LoFs in GnomAD	(A)	Allele count - n homozygous carrier/ 1/2 x allele number	0.000361172 0.000689169
Causal ACAD9 missense and LoF variants (this paper)	(B)	Allele count- n homozygous carrier/ 1/2 x allele number	0.001246382
Total allele frequency (missense + LoF) in ExAC, GnomAD and this paper	(C)	$(A+B)^2$	0.0000053236
Estimated prevalence	(D)	1/C	1:187842
Estimated incidence#	(E)	$5.1*10^6/D$	27

*LoF: unequivocal, splice site, and frameshift variants in the canonical *ACAD9* transcript. #European population: 5.1 million births per year. [Modified from[250]]

3.2.2 Variants in *ELAC2* and isolated mitochondrial cardiomyopathy

I investigated two unrelated cases presenting with severe isolated forms of heart failure with onset infancy of unknown origin. WES was applied aiming to find a genetic cause. In both cases, a mitochondrial disorder was not suspected before performing the genetic testing.

3.2.2.1 *ELAC2* cohort

The *ELAC2* cohort is composed of two families, genetic and clinical findings are reported in Table 12 and Figure 35. Individual #88033 (Department of Pediatric Cardiology at DeutschesHerzZentrum in Munich), a boy, was the first child of healthy, unrelated parents from Germany. The patient died at the age of 4 months due to a severe progressive heart failure. Individual #86493 (Department of Pediatric Cardiology at Klinikum Grosshadern in Munich), a girl, was the first child of healthy, unrelated parents from Germany. Early in infancy, she developed a severe dilative cardiomyopathy. Subsequently, she showed signs of muscular weakness, dystrophy and mild developmental delay with language impairment. Further investigations did not reveal a specific form of congenital myopathy. Therefore, these clinical aspects were considered to be consequences of the severe heart failure. The cardiomyopathy progressively worsened and the patient required heart transplantation at the age of 5 years. At the last follow up (6 years old) the cardiac situation remained stable under immunosuppressive treatment (tacrolimus) and the general clinical condition was considered satisfactory.

3.2.2.2 Pathogenic bi-allelic variants in *ELAC2*

Based on a recessive model of inheritance, I searched for genes carrying rare homozygous and compound heterozygous nonsynonymous variants, using the default cut-off for recessive disorders (MAF≤ 0.01). In individual #88033 22 bi-allelic variants in 14 genes were identified. Among those, two genes were present in OMIM (*ELAC2*, and *LRP2*). In individual #88033 8 bi-allelic variants distributed in 8 genes were detected and 4 were listed in OMIM (*ABCC12, ELAC2, CD109, TNRC18*). A second phenotype-based search (OMIM search term “cardiomyopathy”) was performed and *ELAC2* was the only gene which had previously been linked to cardiomyopathy.

These findings not only underline the presence of severe isolated forms of cardiomyopathy in *ELAC2* cases, thus expanding its phenotypic spectrum, but also arise important aspects such

as the screening of mitochondrial gene for isolated forms of pediatric cardiomyopathy and the successful effect of heart transplantation in mitochondrial cases.

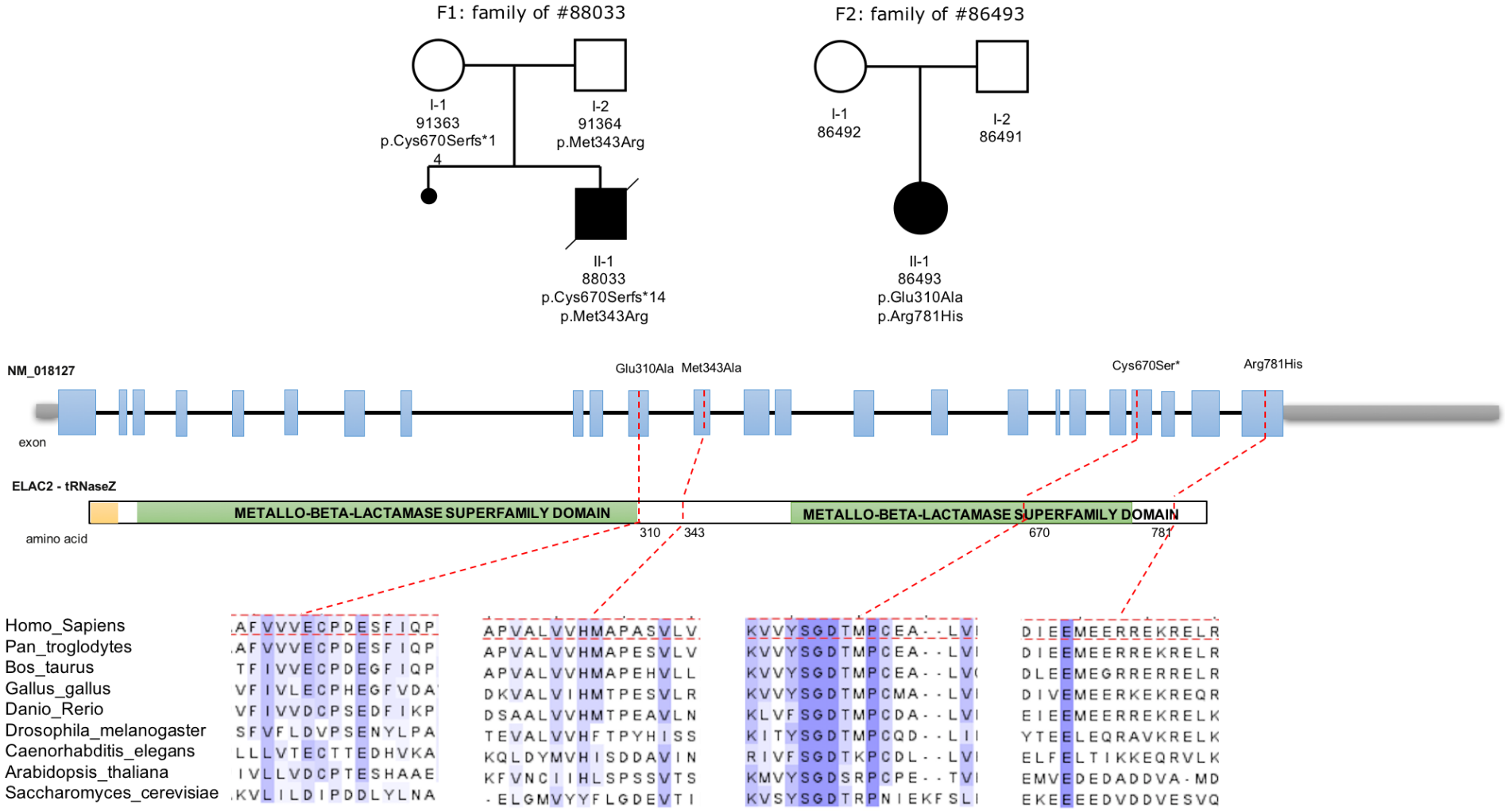

Figure 35. *ELAC2* mutation status and gene structure. Pedigrees of two families with mutations in *ELAC2*. Gene structure of *ELAC2* with known protein domains of the gene product and localization and conservation of amino acid residues affected by mutations. Intronic regions are drawn to scale.

Table 12. Genetic and clinical findings in individuals with *ELAC2* mutations.

ID	*ELAC2* mutations	Age at onset	Course	HCM	Other features
#88033[a]	c.[2009delG(;)1028T>G], p.[Cys670Serfs*14(;) Met343Arg]	3 months	death at 4 months	yes	Massive cardiomegaly with acute cardiac failure
#86493[a]	c.[929A>C];[2342G>A], p.[Glu310Ala]; [Arg781His]	4 months	Alive at 6 years	DCM	Mild psychomotor delay, muscular hypotonia, severe and progressive heart failure (heart transplantated at age 5)
#61525	c.[631C>T; 1559C>T], p.[Arg211*; Thr520Ile]	4 months	death at 6 months	yes	Intrauterine growth retardation, lactic acidosis, myocardial damage and necrosis associated with acute cardiac failure
#57415	c.[631C>T; 1559C>T], p.[Arg211*; Thr520Ile]	3 months	Alive at 3 years	yes	Psychomotor and growth retardation, muscular hypotonia, microcephaly, dysphagia, lactic acidosis, sensorineural hearing impairment, hyperintensities in basal ganglia at the age of 3 months
#61982	c.[460T>C; 460T>C], p.[Phe154Leu; Phe154Leu	2 months	Death at 11 months	yes	Intrauterine growth retardation, lactic acidosis, cardiac failure, normal muscle biopsy findings
#36355	c.[1267C>T; 1267C>T], p.[Leu423Phe; Leu423Phe]	5 months	Alive at 13 years	yes	Mild psychomotor delay, muscular hypotonia
#65937	c.[1267C>T; 1267C>T], p.[Leu423Phe; Leu423Phe]	5 months	Death at 11 months	yes;later	Psycomotor retardation, muscolar hypotonia, cardiac failure, COX-deficient fibers

Abbreviations: HCM, hypertrophic cardiomyopathy; DCM, dilated cardiomyopathy. *ELAC2* mutations: cDNA (NM_018127.6) and Protein (NP_060597). [a] novel case present in this study (in red).

3.3 Genetic basis of HLHS

Considering the extremely severe phenotype and the negative family history, WES was performed to investigate with a de novo analysis the genetic basis of more than 70 trios with hypoplastic left heart syndrome (HLHS). Before initiating exome sequencing, I have searched the literature and created a candidate gene list of potential congenital heart defects (CHDs) genes (Table 18). The list was in total formed by 459 genes, dividing them in: 190 syndromic (i.e. genes already associated in humans with syndromes including CHDs), 114 non-syndromic (i.e. genes already associated in humans with isolated CHDs) and 155 candidates from animal models (i.e. genes that in animal models showed to be associated with CHDs). The aim of the study was to identify novel genes that could help understanding the pathomechanism of such a severe disorder. I also investigated the presence of de novo CNVs because they are a known cause of HLHS. Furthermore, I computed a GO enrichment analysis to identify novel pathways potentially involved in the dysregulation of heart development. After a de novo analysis, I have performed a recessive analysis in HLHS cases, using the candidate gene list to filter the results. Considering the complex inheritance of the disorders, I have also analysed all rare inherited variants present in the probands to evaluate an additional role played by the genetic background. The lists of genes identified by recessive and all the rare inherited variants were used to compute a GO enrichment analysis.

3.3.1 HLHS cohort

Seventy-eight probands with HLHS with apparently no evident syndromic features and their unaffected parents had been enrolled in Munich (29 trios, Krane group), Utah (46 trios, Prof. Gruber group) and Amsterdam (3 trios, Prof. Bezzina group). A male prevalence is present in this cohort (54 males and 24 females). Demographic information is reported in Table 13.
Forty-nine trios were selected as controls from the KORA populations and sequenced with the same method and analysed through the same pipeline. Twenty of the control-trios were recruited from the Kora Augsburg Diabetes Family Study on type 2 diabetes, 29 from the KORA-gen study.

Table 13. Demographics information of the individuals with hypoplastic left heart syndrome (HLHS).

Category	N	%
Anatomic observation		
AA/MA	44	56.4%
AS/MS	17	21.8%
AA/MS	12	15.4%
AS/MA	2	2.6%
Others	3	3.8%
Major syndromic features		
Yes	5	6.4%
No	73	93.6%
Gender		
Female	24	30.8%%
Male	54	69.2%
Race		
Black	1	1.3%
White	77	98.7%%
Ethnicity		
Hispanic/Latino	0	0.0%
Not Hispanic/Latino	78	100.0%
Total	78	100.0%

AA- aortic atresia, AS- aortic stenosis, MA, mitral atresia, MS- mitral stenosis

3.3.2 De novo analysis

WES performed in 78 HLHS trios and in 49 control trios reached a mean coverage depth across 50 Mb of captured sequence between 68- and 162-fold (median 119-fold). About 96% of nucleotides in the target region have been covered by at least 20-times. By comparing the sequencing data with the reference sequence (human genome assembly hg19), the pipeline detected an average of 23.503 total coding variants was detected per sample (including all synonymous, nonsense, frameshift-inducing insertions or deletions and canonical splice variants within the captured exons). De novo analysis was performed in cases and control trios, using as filters SNV quality > 30, Mapping quality > 50 and MAF ≤ 0.0001 (calculation based on HLHS prevalence).

The list of de novo mutations detected in cases encompassed 122 variants, divided in 32 synonymous and 90 non-synonymous (80 missense, 4 frameshift, 3 nonsense, 2 splice variants and 1 inframe deletion). Among the 90 non-synonymous variants, 87 (accuracy 97%) were positively confirmed by Sanger Sequencing.

The average number of de novo non-synonymous mutations per subject was 1.175 in cases, while in controls it was 0.73. As shown in Table 14, fifty-seven (73%) probands in the case group and thirty-three (67 %) participants in the control group carried at least one de novo variant (p=0.4). Non-synonymous de novo variants were present in 50 (64%) cases and 24

(48%) controls (p=0.1). No differences were observed when considering only the loss of function de novo mutations (frameshift, nonsense and splice variants), identified in 10 (12.8%) cases and 5 (10%) controls.

The distribution of the number of de novo mutations per patient in the different subpopulations is represented in Figure 36A, B. In 69% (20/29) of Munich cases, 61% (28/46) of Utah cases and two out of three (66%) Amsterdam cases at least one de novo mutation was identified

.

Table 14. A and B. Summary of the de novo analysis in the cases and controls under study.

A	**HLHS**
Subjects	78
De novo	122
De novo ratio	1.56
NS de novo	87
NS de novo ratio	1.175
Cases with de novo	57 (73%)*
Cases with NS de novo	50 (64%)§
Cases with LOF de novo	10 (12.8 %)

B	**CONTROLS**
Subjects	49
De novo	55
De novo ratio	1.122
NS de novo	36
NS de novo ratio	0.73
Cases with de novo	32 (67%)
Cases with NS de novo	24 (48%)
Cases with LOF de novo	5 (10%)

NS= non synonymous mutations
LOF= loss of function mutations
* p value = 0.4
§ p value = 0.1

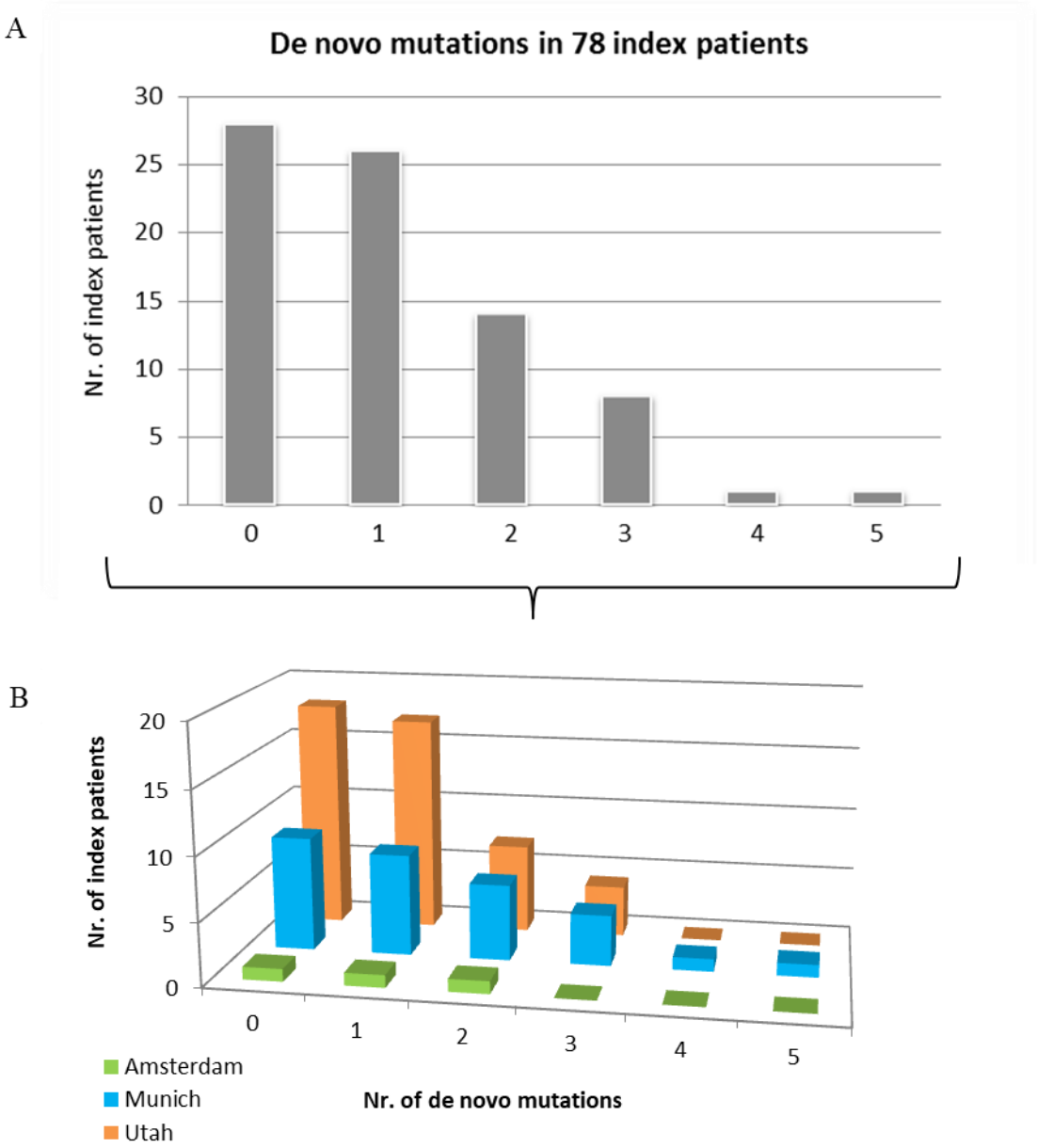

Figure 36. A) Number of non-synonymous de novo mutations identified in 78 probands B) distribution of de novo mutation across the different cohorts.

The 87 non-synonymous de novo mutations identified in cases are distributed over 84 different genes (Table 15). In 4 index patients (5%) de novo mutations were identified in genes associated with syndromes in which cardiac malformations are frequently present: a frameshift mutation in *NOTCH2* (present in HGMD and already described in a case with Hajdu-Cheney syndrome)[191], a nonsense mutation in *CHD7* (present in HGMD and already associated with CHARGE syndrome)[192], a missense mutation with damaging predictions in *CREBBP* (missense mutations in this gene have been previously associated with Rubinstein-Taybi syndrome)[193] and a missense mutation with damaging predictions in *FGFR2* (gene associated with Apert syndrome)[194]. Interestingly, the case with the *CHD7* nonsense variant

(the only case of therapeutic abortion in the case cohort), is also carrying two additional de novo missense mutations predicted to be damaging on *MACF1* and *LRP6.* Both these genes are implicated in Wnt signalling[195].

Looking at de novo mutations occurring more than once in the same gene or family of genes, I have identified 16 probands (21%) carrying a mutation in 9 different gene-families, as detailed in (Table 16). Four of these gene-families (*PTCH* genes, *GLI* genes, *FGFR* genes and *MACF1*), are associated with pathways already implicated in early cardiac development. This provides additional biological evidence for their involvement in the pathogenesis of HLHS. *PTCH* and *GLI* genes have been associated with hedgehog signalling (SHH)[196], *FGFR1* and *FGFR2* are part of the FGF[197,198] signalling and *MACF1* have been associated with Wnt signalling[199]. *MACF1* and *MYRF* are the only two genes that appeared twice in the de novo list, with one proband carrying mutations in both these genes.

Focusing only on potential LOF variants like frameshift, nonsense and splice variants, I detected a total of 10 such mutations distributed on 9 genes and detailed in (Table 17). *NOTCH2*, *CHD7* and *MYRF* appear again in this list. Another of these 10 genes is *DSP*, desmoplakin, where a frameshift mutation is present as the only de novo mutation in one of the cases. *DSP* is a well-known gene, implicated in ARVC (arrhythmogenic right ventricular cardiomyopathy)[15]. Finally, worth of note is the nonsense mutation in the *ETS2* gene, a transcriptional factor part of the E26 transformation-specific family that plays a key role in cell migration, proliferation, and differentiation. In particular, it has been demonstrated that ETS shows a critical role in directing cardiac progenitors during cardiopoiesis in embryonic stem cells[200].

In 10 of the HLHS probands (12.8%), I found a de novo mutation in the genes present in the candidate gene list; those genes are specifically highlighted in (Table 15). Among these, 8/10 (80%) were enriched in the Munich cohort, even though the smaller number of patients in comparison with the Utah cohort. This result can be related to a different clinical approach used in Munich to distinguish between syndromic and non-syndromic cases. In only 1 control I have identified a de novo mutation in the genes present in the candidate gene list (p=0.04).

In summary, in this cohort of non-syndromic HLHS cases, I identified de novo variants in previously reported CHD genes in 15 cases (19%). Specifically, 4 cases (5%) carry mutations related to monogenic syndromes in which CHDs are present (*NOTCH2*, *CREBBP*, *CHD7*, *FGFR2*) and further 7 cases (9%) carry de novo mutations in genes known to be implicated in early cardiac development (*FGFR1*, *GLI1*, *GLI2*, *MACF1*, *ETS2*, *PTCH1*, *PTCH2*). In 5 cases

(6%), CNVs were identified in 2 syndromic genes (*FANCM*, *BBS5)* and 3 in genes related to early cardiac development *(REL*, *USP34*, *NRG1*).

I identified only a single gene, *MYRF*, with de novo frameshift mutations in two unrelated affected individuals. This gene is a transcriptional factor so far known for its role in myelin regulation[201]. However, it is ubiquitously expressed and further studies are needed to elucidate its possible role in cardiac development.

Table 15. De novo non-synonymous mutations (n= 87) identified in 78 index patients.

Sample ID	Gene	Function	AA change	Protein function	Polyphen	Sift	EXAC (EA)
73119	KMT2A	missense	Pro198His	Histone-lysine N-methyltransferase 2A	**0.88**	0.06	
	FGFR1 §	missense	His166Pro	Fibroblast growth factor receptor 1	**0.99**	**0.03**	
69270	BAI2	missense	Gly267Glu	Brain-specific angiogenesis inhibitor 2	0.003	0.07	
	DENND5B	missense	Lys834Arg	DENN domain-containing protein 5B	**1**	**0**	
	SYBU	missense	Pro93Hys	Syntabulin	**0.99**	**0**	
73503	LRIG3	missense	Glu570Lys	Leucine rich repeats domain 3	0.01	np	
	GLI2 §	missense	Ala60Val	Zinc Finger protein	0.01	0.08	
73485	CCDC88C	missense	Met1357Val	Negative regulator of Wnt signalling	0.12	**0.01**	
	PWP2	missense	Arg709Cys	Periodic tryptophan protein homolog	**0.65**	**0.01**	$2.997\ 10^{-05}$
69252	KIAA1107	missense	Lys727Glu	Uncharacterized protein	0.03	0.53	
69255	NOTCH2 *	**frameshift**	Pro2149Arg fs*X	Belongs to the NOTCH family (NOTCH signalling)	np	np	
69675	ANK3	missense	Pro3516Leu	Ankyrin 3	0.4	**0.01**	
	CREBBP *	missense	Phe1544Ser	CREB binding protein	**0.99**	**0**	
	CDK3	missense	Arg157Cys	Cyclin dependent kinase 3	**0.99**	**0**	
70315	AIM1L	missense	Val498Met	Absent in melanoma 1-like protein	np	np	
	MACF1	2 missense	Ala4013Leu	Microtubule-actin cross-linking factor 1 involved in Wnt signalling	**0.99**	**0.01**	
	MYRF	**frameshift**	Ile213Thr fs*59	Myelin regulator factor	np	np	
	NDUFB10	missense	Pro23Ser	NADH dehydrogenase ubiquinone	**0.98**	0.09	
70270	DENND4B	missense	Gly308Arg	DENN domain-containing protein 4B	0.07	0.076	
73491	TRPC7	missense	Thr615Met	Transient receptor potential cation channel subfamily M member 2	**0.99**	np	
73495	CYP17A1	missense	Trp397Arg	Belongs to the cytochrome P450 family	**1**	**0**	
	PAMR1	missense	His504Tyr	Peptidase domain-containing protein assoc. with muscle regeneration	**0.99**	**0**	

76357	FAR1	missense	Arg266Cys	Fatty acyl-CoA reductase 1	0.03	0.18	
66111	MACF1	missense	Ala5268Pro	Microtubule-actin cross-linking factor 1	**0.99**	np	
	LRP6	missense	Arg360Cys	Low-density lipoprotein protein, receptor in Wnt signalling	**0.91**	**0.03**	$2.998\ 10^{-05}$
	ATF6B	missense	Ala408Val	Cyclic AMP-dependent transcription factor ATF-6 beta	0.2	0.56	
	CHD7 *	**nonsense**	Arg2027X	Chromodomain-helicase-DNA-binding protein	np	1	
73498	ETS2 §	**nonsense**	Tyr352X	Binds Ets site and stimulates transcription	np	np	
84476	KIR2DS4	missense	Val45Gly	Killer cell immunoglobulin-like receptor	**0.99**	**0**	
	PPP2R2B	nonsense	Trp163Stop	Protein phosphatase 2, regulatory subunit B	np	np	
	OR2J2	missense	Leu67Ile	Olfactory receptor, family 2	**0.99**	**0**	
84479	NLRC3	missense	Val389Leu	NOD-like receptor family member	0	0.35	
73513	BCL9L	missense	Arg710Gln	B-cell CLL/lymphoma 9-like	0.6	**0.02**	$1.518\ 10^{-05}$
	C2orf54	missense	Arg28Trp	Uncharacterized Protein	**0.99**	**0**	$2.59\ 10^{-05}$
84490	MYRF	frameshift	Met876Val fs*23	Myelin regulatory factor	np	np	
	HDAC5	missense	Arg659Cys	Histone deacetylase 5	**0.99**	**0**	$2.945\ 10^{-05}$
	SLC26A6	missense	Lys3720Arg	Solute carrier family 26	0.06	0.3	
HLHS0157	FOXP4 §	missense	Ala264Ser	Forkhead box transcription factor	0	0.43	
HLHS0163	ZNF469	missense	Pro1853Ala	Zinc-finger protein	0.3	0.1	
	TNIP3	missense	Ser166Pro	TNFAIP3 interacting protein 3	0.002	np	
84496	CAMSAP2	missense	Tyr25Cys	Calmodulin regulated spectrin-associated protein family	**0.97**	**0**	
	OBSL1	missense	Glu542Lys	Cytoskeletal adaptor protein	0.05	0.8	$1.566\ 10^{-05}$
84495	C7orf61	missense	Arg175Gly	Uncharacterized Protein	**0.99**	**0**	$1.613\ 10^{-05}$
HLHS0001	ZNF394	missense	Val501Ile	Zinc Finger Protein 394	0.29	**0**	
HLHS0007	SLU7	missense	Lys302Arg	Pre-mRNA-splicing factor	0.04	0.2	
HLHS0034	EIF2AK3	**indel**	Arg718 -	Transcriptional factor, belongs to the protein kinase superfamily	np	np	
HLHS0040	PTCH1 §	missense	Asn872Ser	Protein patched homolog 1	0.01	1	
HLHS0061	MVP	missense	Arg512Cys	Major Vault Protein	**0.73**	**0**	
	CABLES2	missense	Phe348Leu	Belongs to the cyclin family	**1**	**0**	

HLHS0010	ALG12	missense	His473Tyr	Belongs to the glycosyltransferase family	0.17	**0.02**	
HLHS0064	RAB11FIP3	missense	Ser318Arg	Mediates the interaction with Rab11	**0.99**	0.15	
HLHS0067	AGPAT1	missense	Arg222Cys	Belongs to the 1-acyl-sn-glycerol-3-phosphate acyltransferase family	**0.78**	**0.03**	
	MYPN	missense	Ile378Met	Myopalladin, component of the sarcomere	0.03	np	
HLHS0076	TCHP	missense	Glu378Lys	Trichoplein keratin filament-binding protein with oncostatic activity	**0.95**	0.18	
	OSGIN2	missense	Tyr179Cys	Oxidative stress-induced growth inhibitor	**1**	**0**	
HLHS0085	SDS	**splice**		Serine Dehydratase	np	np	
	RPL23A	missense	Pro66Arg	Belongs to the ribosomal protein L23P family	0.42	**0.06**	$1.525\ 10^{-05}$
	MCMBP	missense	Pro516Leu	Acts as a regulator of DNA replication	**0.99**	0.07	$1.504\ 10^{-05}$
HLHS0088	MRPS9	missense	Arg345Gln	Mammalian mitochondrial ribosomal protein	**0.98**	**0.02**	$1.503\ 10^{-05}$
	FHOD1	missense	Arg981Trp	Formin homolog overexpressed in spleen	**0.94**	0.05	
	ANKRD35	missense	Arg306Leu	Ankyrin repeat domain-containing protein	0.04	**0.04**	
HLHS0091	SLC9A4	missense	Gly187Cys	Solute carrier family 9, subfamily A	**0.56**	np	
	LRRN1	missense	Val303Leu	Leucine-rich repeat neuronal protein	0.03	0.38	
HLHS0013	CYP2A13	missense	Arg265Trp	Belongs to the cytochrome P450 family	**0.91**	**0**	$1.499\ 10^{-05}$
HLHS0097	BAIAP3	missense	Pro1039Leu	p53-target gene encodes a brain-specific angiogenesis inhibitor	0.18	0.42	
	CENPJ	missense	Asn277Lys	Belongs to the centromere protein family	0.07	**0.03**	0.000306
HLHS0100	TADA2B	missense	Tyr166Cys	Coactivates PAX5-dependent transcription	**0.99**	**0**	$1.637\ 10^{-05}$
HLHS0103	NLRP1	missense	Arg1336His	Encodes a member of the Ced-4 family of apoptosis proteins	0.001	0.55	$1.503\ 10^{-05}$
HLHS0106	UCKL1	missense	Ala2Asp	Belongs to the uridine kinase family	0.38	**0**	
HLHS0109	MYO9A	missense	Ala1154Glu	Unconventional Myosin, actin-based motor molecules with ATPase activity	0.07	0.28	
HLHS0115	MFNG	missense	Pro48Ser	Glycosyltransferase involved in inactivation of Notch signalling	0.001	0.49	
HLHS0118	CSRP3	missense	Ala77Thr	Cardiac LIM protein	**0.73**	**0.01**	
HLHS0121	IGFBP1	missense	Ala39Pro	Insulin-like growth factor-binding protein 1	**0.96**	**0.02**	
	FGFR2 *	missense	Gly238Arg	Member of the fibroblast growth factor receptor family	**1**	**0.01**	

	RLF	missense	Pro1054Leu	Zinc finger protein	0.004	0.25	
HLHS0016	CGNL1	missense	Leu162Met	Member of the cingulin family	**0.79**	0.65	
	WDR25	missense	Arg413Gln	WD repeat protein	0.005	**0.03**	
HLHS0124	PTCH2	missense	Arg555Trp	Protein patched homolog 2	**0.75**	**0.04**	$1.512\ 10^{-05}$
	BRCA1	missense	Arg979Cys	Breast cancer type 1 susceptibility protein	**0.46**	0.08	3 10-^{05}
HLHS0136	RTKN2	missense	Ile408Thr	Rhotekin-2, related to lymphopoiesis	**0.99**	**0**	$1.5\ 10^{-05}$
HLHS0145	KAT2A	missense	Thr477Met	Histone acetyltransferase	0.05	0.31	
HLHS0022	DSP	**frameshift**	Leu2461Gln fs*4	Desmoplakin	np	np	
HLHS0028	GLI1 §	missense	Ser378Pro	Zinc finger protein	**1**	**0**	
	AVIL	missense	Arg539Gln	Advillin, Ca2+-regulated actin-binding protein	0.004	0.18	0.0001049
HLHS0031	WDFY3	missense	Trp3154Cys	WD repeat protein	**0.98**	**0**	
	PYROXD1	**splice**		Pyridine nucleotide-disulphide oxidoreductase domain 1	np	np	$6.231\ 10^{-05}$

Samples ID are highlighted in orange for USA cohort, in green for Amsterdam cohort and in light blue for Munich cohort.
§: genes present in the Candidate gene list and associated to cardiac defects only in animal models (colour code red)
* : genes present in the Candidate gene list and related to syndromic CHD (colour code green)
~: genes present in the Candidate gene list and related to non-syndromic CHD (colour code blue)
Polyphen: < 0.5 benign; > 0.5 possibly damaging; >0.8 probably damaging
Sift: > 0.05 benign; ≤ 0.05 damaging
LOF variants are highlighted in bold

Table 16. Genes/families of genes with more than one de novo non-synonymous variant.

Sample ID	Chromosome position	Gene	Protein change	Gene family	Polyphen	Sift
HLHS0040	chr9:98224226 T>C	PTCH1 §	Asn872Ser	Protein patched homolog 1. Receptor for SHH	benign (0.01)	tolerated (1)
HLHS0124	chr1:45294014 G>A	PTCH2	Arg555Trp		possibly damaging (0.75)	deleterious (0.04)
HLHS0028	chr12:57861831 T>C	GLI1 §	Ser378Pro	Zinc finger protein involved in SHH	probably damaging (1)	deleterious (0)
73503	chr2:121684967 C>T	GLI2 §	Ala60Val		benign (0.01)	tolerated (0.08)
73119	chr8:38285563 T>G	FGFR1 §	His166Pro	Fibroblast growth factor receptor	probably damaging (0.99)	deleterious (0.03)
HLHS0121	chr10:123298142 C>G	FGFR2 *	Gly238Arg		probably damaging (1)	deleterious (0.02)
70270	chr1:153914478 C>T	DENND4B	Gly308Arg	DENN domain-containing protein	benign (0.07)	tolerated (0.37)
69270	chr12:31576605 T>C	DENND5B	Lys834Arg		probably damaging (1)	deleterious (0)
70315	chr1:39903674 G>T chr1:39903675 C>T	MACF1	Ala4013Leu	Microtubule-actin cross-linking factor	probably damaging (0.99)	deleterious (0.01)
66111	chr1:39945586 G>C	MACF1	Ala5268Pro		probably damaging (0.99)	deleterious (0.01)
70315	chr11:61537895 delT	MYRF	Ile213Thr fs*59	Myelin regulator factor	-	-
84490	chr11:61548663 delAT	MYRF	Met876Val fs*23		-	-
84490	chr3:48668699 T>C	SLC26A6	Lys3720Arg	Solute carrier family	benign (0.06)	tolerated (0.3)
HLHS0091	chr2:103095600 G>T	SLC9A4	Gly187Cys		benign (0.56)	-
73495	chr10:104591319 A>G	CYP17A1	Trp397Arg	Cytochrome P450	probably damaging (1)	deleterious (0)
HLHS0013	chr19:41597775 C>T	CYP2A13	Arg265Trp		possibly damaging (0.91)	deleterious (0)
HLHS0163	ch16:88499519 C>G	ZNF469	Pro1853Ala	Zinc finger protein	benign (0.03)	tolerated (0.11)
HLHS0011	chr7:99091337 C>T	ZNF394	Val501Ile		benign (0.29)	tolerated (0)

SHH: Sonic hedgehog signalling pathway
§: genes present in the Candidate gene list and associated to cardiac defects only in animal models
*: genes present in the Candidate gene list and related to syndromic CHD

Table 17. Loss of function mutations (LoF) - de novo haploinsufficiency.

Sample ID	Chr. position	Function	Gene	Description	HGMD	OMIM	Mouse
66111	chr8:61765241 C>T	nonsense	CHD7	Chromodomainhelicase DNA binding protein	CM060921	608892: CHARGE syndrome	
HLHS0022	chr6:7584816 del CT	frameshift	DSP	Desmoplakin		125647: ARVD 8	MGI:109611 Homozygous targeted null mutants die by embryonic day E6; rescue by aggregation with wild-type tetraploid morulae increase embryonic survival with noted major defects in heart muscle
HLHS0034	chr2:88874845 del CTT	indel	EIF2AK3	Trascriptional factor			
73498	chr21:40190395 C>A	nonsense	ETS2	Binds specifically Ets site in gene promoters and stimulates transcription			
84490	chr11:61548663 del AT	frameshift	MYRF	Myelin regulator factor			
70315	chr11:61537895 delT	frameshift	MYRF	Myelin regulator factor			
69255	chr1:120458895 del AG	frameshift	NOTCH2	Belongs to the NOTCH family (NOTCH signalling)	CD112312	600275: Hajdu Cheney syndrome 610205: Alagille syndrome 2	
HLHS0031	chr12:21598401 GT->AT	splice	PYROXD1	Pyridine nucleotide-disulphide oxidoreductase			
84476	chr5:146077585 C>T	nonsense	PPP2R2B	Protein phosphatase 2			
HLHS0085	chr12:113831822 AG->AA	splice	SDS	Serine Dehydratase			

HGMD=Human Gene Mutation Database; MGI: Mouse Genome Informatics

Table 18. Candidate gene list (n= 459) divided in non-syndromic, syndromic and animal models.

Non Syndromic (n=114)									
ABCB1	ACP6	ACTA2	ACTC1	ACVR1	ACVR2B	AHSA2	ALDH1A2	ANKRD1	AXIN1
BCL9	BMPR2	CCT4	CFC1	CHD1L	CITED2	CNN1	CRELD1	CRKL	CUL3
CYFIP1	DAPK3	DLC1	DTNA	DUSP1	ETS1	FMO5	FOXH1	GALTN11	GATA4
GATA6	GDF1	GJA5	GJA8	GPC5	HAND1	HAND2	HAS2	HEY2	HOMEZ
HUWE1	IRX4	ISL1	JUN	JUP	KDM5A	KDM5B	KIF3C	LEFTY1	LEFTY2
MAML3	MATR3	MCTP2	MED13L	MED15	MED9	MESP1	MSX1	MTHFR	MTR
MTRR	MYH11	MYH6	MYH7	NEK2	NEXN	NFATC1	NIPA1	NIPA2	NKX2-5
NKX2-6	NODAL	NOS3	NOTCH1	NR2F2	NRP1	NUB1	NUP188	PAPOLG	PDGFRA
PLAGL1	PLXND1	PPM1K	PRKAB2	PTPRE	RAB10	REL	RNF20	ROCK2	SEMA3D
SESN1	SMAD3	SMAD6	SMAD7	SOX7	SREBF1	STX8	SUV420H1	TAB2	TBX15
TBX18	TBX2	TBX20	TDGF1	TLL1	TOP2A	UBE2B	USP34	USP44	VEGFA
WDR5	XPO1	ZFPM2	ZIC3						

Syndromic (n=190)									
ACVRL1	ADAMTS10	AGGF1	ALDH1L2	ALX3	AMER1	ANKRD11	ARL13B	ARL6	ARX
ATIC	ATRX	B3GALTL	BBS1	BBS10	BBS12	BBS2	BBS4	BBS5	BBS7
BBS9	BCOR	BMP4	BRAF	BRCA2	BRIP1	CACNA1C	CBL	CCBE1	CCDC28B
CCDC39	CCDC40	CEP290	CHD7	CNOT1	COL18A1	COL2A1	COL3A1	COL6A1	CREBBP
DHCR24	DHCR7	DLL3	DNAAF1	DNAH11	DNAH2	DNAH5	DNAI1	DNAI2	DPM1
EHMT1	ELN	EP300	ESCO2	EVC	EVC2	EYA1	FAM58A	FANCA	FANCB
FANCC	FANCD2	FANCE	FANCF	FANCG	FANCI	FANCL	FANCM	FBN1	FBN2
FGD1	FGFR2	FLNA	FLNB	FOXC1	FOXC2	FOXF1	FOXL1	FOXL2	FTO
G6PC3	GATA5	GJA1	GPC3	HCCS	HOXA1	HRAS	IFT122	IFT172	IGBP1
INVS	JAG1	KCNJ2	KDM6A	KMT2D	KRAS	KRTAP7-1	LBR	LEMD3	LMNA
LRP2	LRP5	LRRC6	MAP2K1	MAP2K2	MAP4K4	MED12	MEK2	MGAT2	MGP
MID1	MKKS	MKS1	MYCN	NEFL	NF1	NIPBL	NME8	NOTCH2	NPHP3
NRAS	NRGN	NSD1	NSDHL	OFD1	PALB2	PAX3	PEX1	PEX10	PEX11B
PEX12	PEX13	PEX14	PEX16	PEX19	PEX2	PEX26	PEX3	PEX5	PEX6
PEX7	PIFO	PITX2	PQBP1	PTEN	PTPN11	RAB23	RAD51C	RAF1	RAI1
RBM10	RECQL4	RIPK4	ROR2	RPGRIP1L	RPS19	SALL1	SALL4	SCLT1	SEMA3E
SETBP1	SH3PXD2B	SHOC2	SIX6	SKI	SLC29A3	SLC2A10	SLX4	SNX3	SOS1
SOX2	SOX9	SPA17	STRA6	TBX1	TBX3	TBX5	TFAP2B	TGFBR1	TGFBR2
TP63	TRIM32	TTC8	TWIST1	UBR1	VPS33B	WDPCP	WNT3	ZBTB21	ZEB2

Candidates in animal models (n=155)									
ADAM17	ADAP2	AKT1	ALDH1A1	ANAPC10	ANGPT1	ARID3B	ASXL2	ATE1	ATP4A
BCL6	BICC1	BMP1	BMP10	BMP2	BMP5	BMP6	BMP7	BMPR1A	BTC
CDH2	CHD5	CHRD	COL6A2	CSPG2	CSRP1	CYR61	DAND5	DLL1	DMRT2
DVL1	DVL2	ECE1	ECE2	EDN1	EDNRA	EED	EFNB2	EGFR	ENG
EPHB2	EPO	EPOR	ETS2	EZH1	FBLN2	FGF10	FGF8	FGFR1	FOLR1
FOXA2	FOXF2	FOXJ1	FOXK2	FOXP1	FOXP4	FURIN	GJA9	GJC1	GLI1
GLI2	GLI3	GPR161	GSK3B	HBEGF	HDAC1	HDAC5	HDAC9	HES1	HEX
HEY1	HIC2	HOPX	HOXA3	HRG	HSPB7	HSPG2	ID2	IFT20	IFT88
IGFBP4	IHH	ILK	ITGA5	JAZF1	JMJD6	KCTD10	KIF3A	KIF3B	KIFAP3
KLF13	KLF3	LAMA1	LTBP3	MAPK9	MDS1	MEF2A	MEF2C	MEOX2	MESP2
MPZL1	MSX2	MYC	MYH10	MYOCD	NEK8	NFATC2	NFATC3	NFATC4	NOTCH4
NRG1	NTF3	NTN1A	NTRK3	NUMBL	OSR1	PCSK5	PCSK6	PHC1	PICALM
PKD2	PPP3CA	PROX1	PTCH1	RARA	RTTN	RXRA	SEMA3C	SHH	SMAD1
SMAD2	SMAD4	SMAD5	SMAD8	SMARCD3	SMO	SNAI1	SOX17	SOX4	SRF
STIL	SUFU	SUPT5H	SUPT6H	T	TBX6	TCF21	TEAD1	TGFB2	VANGL2
VCAM1	VCL	WNK1	ZFPM1	ZIC2					

3.3.3 Recessive analysis

The recessive analysis was based on a strict frequency filter cut-off of MAF 0.0001 in order to minimize the number of false positive results. Given the fact that the sequences of the parents were present (trio analysis), I checked manually the homozygous or compound heterozygous state of the 915 variants identified.

A total of 60 homozygous non-synonymous variants distributed along 57 genes were identified in 18/78 probands (23%). Among these is worth mentioning the homozygous mutation identified in *CHD7* (G744S), which has already been associated with CHARGE syndrome and is present in HGMD database[202]. *MEGF8* is another potentially interesting gene as it is associated with a congenital recessive disorder, the Carpenter syndrome, in which heart malformations can be present. Interestingly, mutations in *MEGF8* cause a subtype of Carpenter syndrome frequently associated with defective left-right patterning, probably through perturbation of the hedgehog and NODAL signaling[203].

Using the same frequency filter, 206 compound heterozygous variants distributed in 97 genes were detected in the study-population. Among them, I identified 5 genes present in the candidate gene list (Table 19) and specifically *CREBBP*, *EVC2*, *FBN2, UBR1, COL6A1*. All these genes, except *COL6A1,* are reported in OMIM in association with syndromes (mainly recessive) in which CHDs are possible features[193,204–211]. Among these syndromes, the Ellis-Van Creveld Syndrome (Chondroectodermal Dysplasia), is the one with the highest prevalence of CHDs, observed in over 50% of the cases[204,205]. This syndrome is a recessive disorder, due to mutations in *EVC2*[204,205]. Interestingly, the EVC2 protein is found in primary cilia and is thought to regulate the hedgehog signalling pathway[206]. In a single gene carrying compound heterozygous variants, *CREBBP*, one of the two mutations represents a de novo previously described (Table 15).

Finally, among the 54 male cases in the study cohort, I identified a total number of 128 variants on the X chromosome: 3 nonsense, 2 frameshift and 123 missense. In this list, no gene already present in the de novo gene list was identified, but I identified mutations in 4 genes reported as syndromic in the candidate gene list (*ATRX*, *OFD1, NSDHL* and *BCOR*). Diseases associated with *OFD1* include Oral-Facial-Digital Syndrome and Joubert syndrome[212,213]. *OFD1* plays an important role in ciliogenesis regulation and mediates multiple developmental signalling pathways; in particular, it regulates Wnt and hedgehog

signalling and the specification of the left-right axis ciliogenesis regulation[214]. *NSDHL*, a gene associated with CHILD (congenital hemidysplasia with ichthyosiform nevus and limb defects) syndrome, an X-linked condition that is usually lethal for males during gestation. Few males with CHILD syndrome have been reported[215] and focusing on the heart defects, the clinical spectrum includes septal defects[216], unilateral ventricle[217], and a single coronary ostium[218]. *ATRX* encodes for a chromatin remodelling protein related to Alpha Thalassemia/Mental Retardation Syndrome in which cardiac defects have been reported[219]. Finally, *BCOR* is a gene associated with Oculofaciocardiodental (OFCD) syndrome in which several forms of cardiac malformations[220], probably due to uncontrolled Notch activity[221], have been described.

The recessive analysis performed in 49 control trios, identified only one gene present in the candidate gene list, *FANCB* (p=0.012).

In summary, with the recessive analysis, I identified 8 additional cases (10%) carrying mutations in candidate genes. Other two cases (73119 and 69675) were also carriers of de novo variants in genes present in the candidate gene list (*FGFR1* and *CREBBP*).

Table 19. Candidates genes identified with recessive analysis.

Sample ID	Chromosome position	Function	Gene	Protein change	Polyphen2	Sift	CADD	Variant allele
73119	chr1:146763171-146763171	missense	CHD1L	Ser776Arg	benign (0.02)	0.32	12.55	2
73119	chr8:61707678-61707678	missense	CHD7	Gly744Ser	probably damaging (0.99)	0.62	20.2	2
73119	chr8:61765762-61765762	missense		Ala2160Thr	benign (0)	0.82	0.009	2
84496	chr19:42874440-42874440	missense	MEGF8	Ile2312Val	benign (0.003)	1	0.008	2
84496	chrX:13778612-13778612	missense	OFD1	Ala678Gly	benign (0.121)	0.43	12.01	2
84496	chrX:13785309-13785309	missense		Lys888Thr	benign (0,049)	0.16	21.9	2
HLHS0163	chrX:76938830-76938830	missense	ATRX	Val640Ile	benign (0.001)	0.48	2.467	2
HLHS0052	chr15:43262788-43262788	missense	UBR1	Val1463Ile	benign (0)	0.35	15.83	1
HLHS0052	chr15:43299467-43299467	missense		Ser1075Arg	benign (0.25)	0.58	12.24	1
69675	chr16:3786134-3786134	missense	CREBBP	Phe1506Ser	probably damaging (0.99)	0	19.75	1
	chr16:3823772-3823772	missense		Ala777Ser	benign (0.178)	0.96	12.45	1
70270	chr4:5633572-5633572	missense	EVC2	Glu473Gly	possibly damaging (0.71)	0.03	14.15	1
	chr4:5642555-5642555	missense		Ala306Pro	probably damaging (0.94)	0.02	19.44	1
HLHS0049	chr5:127697490-127697490	missense	FBN2	Arg794Gln	probably damaging (0.99)	0.22	33	1
	chr5:127729055-127729055	missense		Tyr380Cys	probably damaging (0.81)	0.04	8.735	1
HLHS0076	chr21:47406975-47406975	missense	COL6A1	Val236Met	possibly damaging (0.65)	0.18	17.01	1
	chr21:47418316-47418316	missense		Tyr527Cys	probably damaging (0.88)	0.18	14.11	1
76357	X 39933948	missense	BCOR	Met217Ile	benign (0.26)	0.01	19.98	1
HLHS0079	X 152036195	missense	NSDHL		benign (0.004)	0.68	9.61	1

Variant allele 1 indicates heterozygous state
Variant allele 2 indicates homozygous state

3.3.4 CNVs analysis

A total of 1499 de novo CNVs were identified among all cases. Selection criteria included deletions or duplications of at least 3 exons and genotype frequency cut-off of 0.0001. I manually investigated with IGV all the 1499 CNVs, in order to reduce the number of false positives. This exercise reduced the number of CNVs to 54, underlying the high number of false positive detected by ExomeDepth (n=1445). The 54 possible CNVs involved 63 genes which were distributed amongst 23 (29.5%) probands (Table 20). These 54 CNVs await validation by qtPCR currently ongoing in Utah. The number of rare CNVs ranged from 1 to 6 per patient, with a median of 2. Among the de novo CNVs identified, 8 (26%) were deletions and 20 (74%) duplications. The average size of the CNVs identified was about 11 kp and only 7 CNVs contained 10 or more exons. 12/54 CNVs were present in DGV (Database of genomic variants) which contains healthy controls populations. Genes present in the candidate gene list were detected in 4 cases: 2 syndromic genes (*BBS5*, *FANCM*), 1 non-syndromic gene (*REL*) and 2 candidates in animal model genes (*USP34*, *NRG1*). *BBS5* encodes a protein that has been directly linked to Bardet-Biedl syndrome, an autosomal recessive ciliopathy disorder[222]. Genes present in the de novo mutation list were detected in 1 case: *MYPN* encodes a sarcomeric protein highly expressed in heart tissue[223]. Among the CNVs detected in candidate genes, only the one in *REL* is present in DGV controls. Among the 29 control trios, I detected 263 de novo CNVs. After the manual visualization step, 9 CNVs within 11 genes remained. Specifically, I detected 8 duplications and 1 deletion in 6 probands encompassing a range of 3-6 exons (average size 4226 bp). None of the CNVs identified were located in regions previously reported in CHDWiki or in genes present in the candidate gene list. Nearly half of them (4/9) were present in DGV (Database of genomic variants) set in at least 2 controls.

In summary, I have identified in 4 cases a de novo CNV in a gene present in the candidate gene list. This result indicates that de novo CNVs can be implicated in the genetic cause of HLHS, however, the validation of these findings is still ongoing, therefore these results should be considered with caution.

Table 20. Candidates genes identified with de novo CNVs analysis.

ID	Chromosome	CNV	size	DGV	Band	CHDWiki	Gene	Depth	exons
HLHS0061	chr1:170927602-170931116	duplication	3515		1q24.3	1q24 loss	MROH9	1.51	3
HLHS0085	chr1:172546679-172547528	deletion	850		1q24.3	1q24 loss	SUCO	0.59	3
69675	chr1:225440003-225454470	duplication	21019	7	1q42.12		DNAH14	1.34	6
73122	chr1:86375633-86437075	duplication	60013		1p22.3		COL24A1	1.44	5
84490	chr1:892274-892652	duplication	379	4	1p36.3	1p36 del	NOC2L	1.35	3
HLHS0151	chr10:46998881-47000256	deletion	1376	8	10q11.22		GPRIN2	0.73	3
HLHS0133	chr10:69955207-69959331	duplication	4125		10q21.3	10q22	**MYPN**	1.46	4
69270	chr11:18111722-18113870	duplication	2149		11p15.1		SAAL1	1.52	3
HLHS0124	chr11:67263649-67264906	deletion	1258	1	11q13.2		PITPNM1	0.61	3
HLHS0145	chr11:9441958-9444686	duplication	2729		11p15.4		IPO7	1.57	3
HLHS0160	chr12:125396260-125398316	duplication	2057		12q24.31		UBC	1.27	4
HLHS0085	chr12:42707492-42711248	deletion	3757		12q12		ZCRB1	0.62	4
84482	chr14:103396261-103396663	duplication	403	2	14q32.32		AMN	1.51	4
HLHS0160	chr14:105173247-105174338	duplication	1092	1	14q32.33		INF2	1.34	3
69270	chr14:45623900-45628482	duplication	4583		14q21.2		FANCM*	1.46	3
HLHS0085	chr14:54947592-54948175	deletion	584	1	14q22.2	14q22-14q23.1 loss	GMFB	0.63	3
HLHS0061	chr14:61509865-61512142	duplication	2278		14q23.1	14q22-14q23.1 loss	SLC38A6	1.58	3
69270	chr14:67769127-67783619	duplication	14493	1	14q23.3		MPP5	1.42	4
69675	chr16:3706639-3727658	deletion	21020		16p13.3	16p13.3 loss	**DNASE1** BC095475 TRAP1	0.52	18
HLHS0139	chr17:45893754-45894060	duplication	307		17q21.32		OSBPL7	1.42	3
HLHS0061	chr19:1011366-1011796	deletion	431	4	19q13.3	19q13.32 loss	TMEM259	0.6	3
HLHS0160	chr19:10671679-10672520	duplication	842		19p13.2	19p13.3pter loss	KRI1	1.33	4
73491	chr19:12501280-12595470	duplication	94190	6	19p13.2		ZNF799 ZNF709 ZNF443	1.29	11
73122	chr19:58904343-58904853	deletion	511	6	19q13.43		RPS5	0.52	3
69270	chr2:100065798-100081436	duplication	15639		2q11.2		REV1	1.36	3
HLHS0097	chr2:128079622-128081816	duplication	2195		2q14.3		MAP3K2	1.41	3

HLHS0133	chr2:15767178-15769820	duplication	2643		2p24.3		DDX1	1.4	5
HLHS0097	chr2:170343579-170344364	duplication	786		2q31.1	2q31 del synd	KLHL41 BBS5*	1.59	3
HLHS0046	chr2:228217230-228220476	duplication	3247		2q36.3		MFF	1.42	4
73491	chr2:61144012-61147243	duplication	3232	3	2p16.1	2p16.1-p15 DELETION SYNDROME	REL~	1.4	4
HLHS0097	chr2:61484008-61484484	duplication	477		2p15	2p15 gain	USP34§	1.64	3
HLHS0061	chr20:58440862-58442901	duplication	2040		20q13.33	20q13.3 loss	SYCP2	1.46	4
69270	chr21:27134623-27137097	duplication	2475		21q21.3	21q21q22 loss	GABPA	1.74	3
HLHS0160	chr22:19964229-19965108	duplication	880	4	22q11.21		**ARVCF**	1.44	3
HLHS0097	chr3:179085227-179085890	duplication	664		3q26.33	3p25.3p26.1 loss	MFN1	1.53	3
73122	chr4:104070275-104098202	duplication	27927		4q24		CENPE	1.34	14
HLHS0049	chr4:146026757-146030400	duplication	3644		4q31.21		ABCE1	1.57	4
HLHS0001	chr4:156276223-156281566	duplication	5344	1	4q32.1		MAP9	1.37	4
73122	chr4:56831810-56865810	duplication	34001		4q12		CEP135	1.42	10
HLHS0085	chr4:68499063-68501275	deletion	2213		4q13.2		UBA6	0.61	4
HLHS0058	chr5:110438044-110441841	deletion	3798	2	5q22.1		WDR36	0.6	6
HLHS0004	chr5:180046253-180047307	duplication	1055		5q35.3		**FLT4**	1.43	3
HLHS0133	chr5:50092816-50137901	duplication	45086		5q11.1		PARP8	1.29	14
73122	chr5:70797389-70840333	duplication	42944	1	5q13.2		BDP1	1.33	21
HLHS0025	chr7:141759272-141762501	duplication	3230	5	7q34		MGAM	1.44	4
66111	chr7:142375063-142482363	deletion	107300	1	7q34		TCRBV11S1A1T PRSS3P2 TCRVB TCRBV5S1A1T AX748042 BV6S4-BJ2S2 TCRBV2S1 BV03S1 J2.2 MTRNR2L6 TCRBV3S1 TCRBV14S1 TCRBV4S1A1T TCRBV19S1P TCRB PRSS1	0.69	18
HLHS0049	chr7:33136105-33136973	duplication	869	2	7p14.3	7p14.3p14.2 loss	RP9	1.53	3
HLHS0085	chr7:81599197-81601179	deletion	1983		7q21.1		CACNA2D1	0.59	3
HLHS0139	chr8:31497501-31498244	deletion	744		8p12	8p12p21	NRG1§	0.65	3
HLHS0091	chr8:48955593-48973387	duplication	17795	1	8q11.21		UBE2V2	1.41	5
HLHS0091	chr8:54882952-54900818	duplication	16633	1	8q11.23		TCEA1	1.49	4
HLHS0049	chr8:87470150-87474074	duplication	3925	1	8q21.3		WWP1	1.42	5

HLHS0133	chr9:74586421-74587757	duplication	1337		9q21.13		C9orf85	1.67	3
HLHS0085	chr9:99324975-99327764	deletion	2790		9q22.33		CDC14B	0.66	4

§: genes present in the Candidate gene list and associated to cardiac defects only in animal models (colour code red)
* : genes present in the Candidate gene list and related to syndromic CHD (colour code green)
~: genes present in the Candidate gene list and related to non-syndromic CHD (colour code blue)
DGV: Database of genomic variants
Depth: values below 1 indicate deletions, values above 1 indicate duplications
Exons: number of exons involved in the detected CNV

3.3.5 Pathway analysis

De novo variants

Performing gene ontology analysis with DAVID in genes identified with de novo analysis, I observed a significant enrichment in pathways named heart development (p=4.4*10-4), embryonic morphogenesis (p=4.9*10-4) and pattern specification process (p=1.4*10-3). Further down in the ranking list, I identified a significant enrichment in hedgehog signalling (p=2.1*10-3) and Notch signalling (p=2.3*10-3). None of these GO terms survived Bonferroni correction, as expected given the small set of mutations considered.

Recessive variants

Gene ontology analysis performed on the list of compound heterozygous and homozygous variants detected in HLHS probands showed a significant enrichment in dynein complex term (p=1.9*10-3), axoneme (p=3.3*10-3), axonemal dynein complex (p=3.3.*10-3), ciliary motility (p=4.1*10-3), axoneme part (p=6.6*10-3) and cilium axoneme (p=1.4*10-2).

Inherited variants

In the 78 probands of the study cohort, I detected 13.788 sequence variations distributed in 8471 genes (missense, nonsense, frameshift, splice, deletion and insertions) using a minor allele frequency cut-off of 0.0001. Using DAVID I identified a significant enrichment in the GO terms extracellular matrix (p=3.6*10-13, p=2.8*10-10 after Bc), plasma membrane (p=4.8*10-12, p=2.9*10-9 after Bc), cytoskeleton (p=4.2*10-7, p=3.4*10-4 after Bc), plasma membrane (p=4..8*10-9, p=3.8*10-6 after Bc) and cilium (p=8*10-6, p=6.3*10-3 after Bc) (Table 21).

Altogether these results indicate a possible role of genes involved in ciliogenesis and early developmental pathways (hedgehog and Notch signalling).

Table 21. GO terms identified with pathway analysis on 13788 variants (i.e. inherited variants).

Category	Term	%	P-Value	Bonferroni	Benjamini
GOTERM_CC_FAT	proteinaceous extracellular matrix	0.3	3.6E-13	2.8E-10	2.8E-10
GOTERM_CC_FAT	extracellular matrix	0.3	3.6E-12	2.9E-9	1.4E-9
GOTERM_CC_FAT	plasma membrane	2.6	4.8E-9	3.8E-6	1.3E-6
GOTERM_CC_FAT	cytoskeleton	1.0	4.2E-7	3.4E-4	8.4E-5
GOTERM_CC_FAT	extracellular matrix part	0.1	6.3E-6	5.0E-3	1.0E-3
GOTERM_CC_FAT	cilium	0.1	8.0E-6	6.3E-3	1.1E-3
GOTERM_CC_FAT	basement membrane	0.1	1.2E-5	9.5E-3	1.4E-3
GOTERM_CC_FAT	cell projection	0.5	4.6E-5	3.6E-2	4.5E-3
GOTERM_CC_FAT	cytoskeletal part	0.7	5.7E-5	4.4E-2	5.0E-3
GOTERM_CC_FAT	myosin complex	0.1	1.0E-4	8.0E-2	8.3E-3
GOTERM_CC_FAT	extracellular region part	0.7	1.1E-4	8.6E-2	8.2E-3
GOTERM_CC_FAT	plasma membrane part	1.5	1.2E-4	9.1E-2	7.9E-3
GOTERM_CC_FAT	microtubule cytoskeleton	0.4	1.6E-4	1.2E-1	9.6E-3
GOTERM_CC_FAT	microtubule	0.2	1.9E-4	1.4E-1	1.1E-2
GOTERM_CC_FAT	axoneme	0.1	2.6E-4	1.9E-1	1.4E-2
GOTERM_CC_FAT	cilium part	0.1	3.2E-4	2.2E-1	1.6E-2
GOTERM_CC_FAT	dynein complex	0.0	3.4E-4	2.4E-1	1.6E-2
GOTERM_CC_FAT	collagen	0.0	5.6E-4	3.6E-1	2.5E-2
GOTERM_CC_FAT	cell projection part	0.2	6.1E-4	3.9E-1	2.5E-2

Category: CC= cellular components; %=genes enriched in the GO term

3.4 Genetic basis of LQTS

WES was performed in a cohort of 26 clinically diagnosed LQTS families negative at the conventional LQTS panel, which contains the five major LQTS genes (*KCNQ1, KCNH2, SCN5A, KCNE1* and *KCNE2*). LQTS represents a leading cause of SCD in the young[29] and the major genes are responsible for approximately 80% of the cases.

3.4.1 LQTS cohort

Among 26 families, in 14 families only the proband was sequenced, in 7 families 2 family members were sequenced, in 4 families 3 family members were sequenced and in 1 family 4 family members were sequenced. When available un-affected family members were also sequenced to improve the analysis.

Using a dominant search, I detected heterozygous mutations (rated pathogenic or likely pathogenic) in known LQTS genes in 11/26 (42%) patients. The genes included *KCNQ1*, *KCNH2* and *CALM3*. In *KCNQ1* (LQT1) I identified 3 missense (2 reported in ClinVar and one novel) and 2 novel CNVs in 5 LQTS families. In *KCNH2* (LQT2) I found 4 missense variants (3 reported in ClinVar and 1 novel) and one novel CNV distributed in 5 LQTS families. Figures 37 and 38 show one LQTS pedigree and the IGV visualization of the CNVs identified in the family, respectively. In 3 affected family members and not in the unaffected one I identified a deletion of five exons (exons 10-15) in *KCNH2* gene. Clinically the cases showed a form of LQTS not exacerbated by exercise activity (typical trigger of the most frequent LQTS form, LQT1). In form of LQTS due to *KCNH2* variants (LQT2) it is usually not present a prolongation of QT interval during and after exercise test. Therefore, the genotype-phenotype correlation further supports the genetic finding. In addition, I identified a non-penetrant LQTS case (rated as unaffected but carrier of the *KCNQ1* disease-causing variant) and I unmasked in a family the segregation of two different mutations, one in *KCNQ1* and the other in *KCNH2*.

In *CALM3* I discovered a novel missense variation, subsequently confirmed as a de novo by Sanger sequencing (Figure 39). Notably, in OMIM there are no *CALM3* cases and in ClinVar only 1 case is submitted by a clinical diagnostic laboratory, while only 3 novel variants have been so far described in literature[224,225]. No additional heterozygous mutations (rated pathogenic of likely pathogenic) in the other known LQTS genes have been identified in this cohort.

In summary, WES showed a high yield (11/26, 40%) in identifying pathogenic variants and in addition, it provided analysis of CNVs.

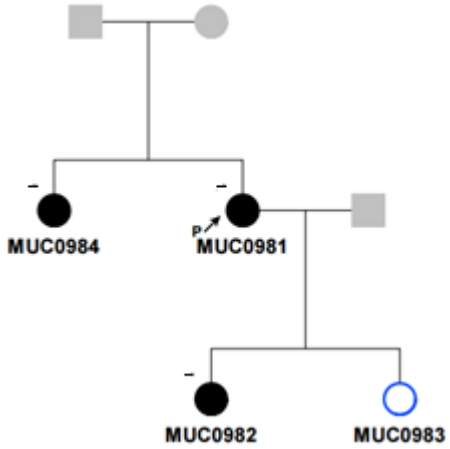

Figure 37. Example of one LQTS family with four affected and one unaffected family members sequenced by WES. In every affected individual a CNV in *KCNH2* was detected (see below).

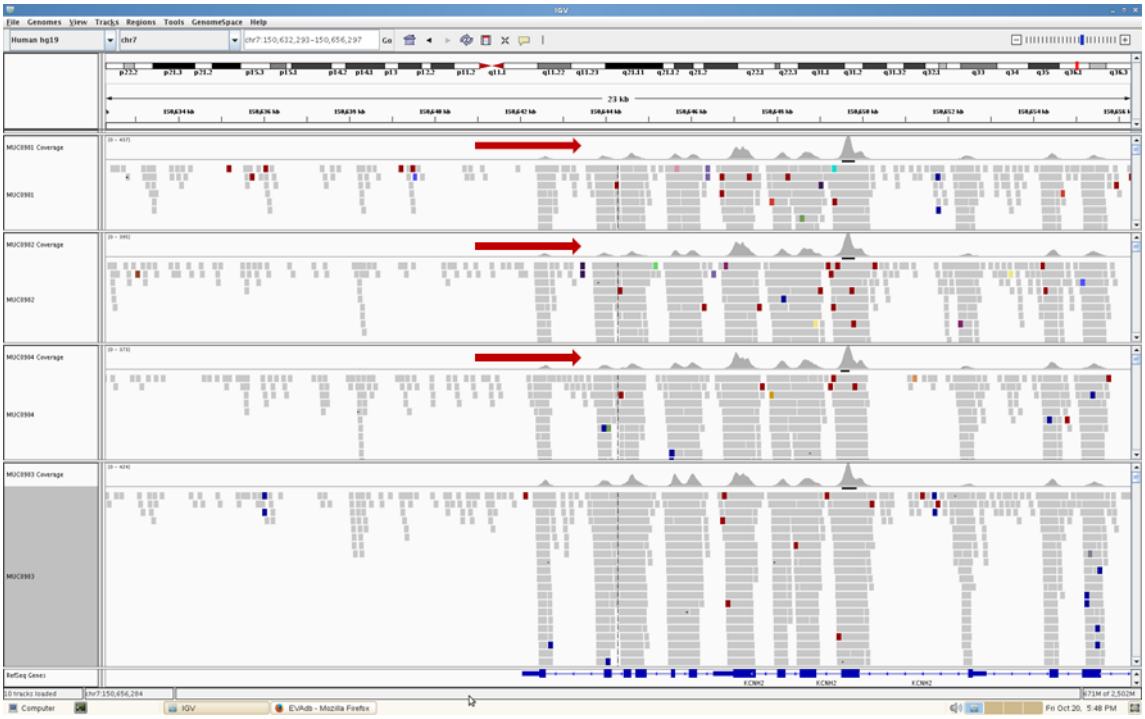

Figure 38. CNVs detected by WES in three affected family members (red arrows) and not present in the unaffected family member and in in-house controls (n=15000 exomes). The CNVs encompassed a deletion from of exones 10-15 in *KCNH2* gene.

Figure 39. Visualization with IGV of a heterozygous variant detected in *CALM3* in one family with LQTS. The variant was subsequently confirmed as a de novo by Sanger sequencing. Only 3 *CALM3* variants are currently reported in the literature in LQTS cases.

3.5 Incidental findings

WES allows the identification not only of mutations responsible for the disease under investigation, but also of variants potentially causing other diseases, namely the "incidental findings" (IFs)[145]. The American College of Medical Genetics and Genomics (ACMG) recommended the return of IFs from a list of 56 actionable genes, which represents conditions with well-known guidelines for prevention and treatment.

After applying standard quality thresholds (SNV quality >30, Mapping quality >50, coverage > 10) and a frequency filter of MAF < 1% to minimize the number of false positive results, among 855 probands, previous patient informed consent, I identified 1484 rare variants in 56 genes present in the ACMG list. These genes are listed in the ACMG recommendations for reporting IFs and 31/56 are associated with cardiac disorders. Among them, 20 are genes related to channelopathies and cardiomyopathies (Table 22). Variants reported as pathogenic or likely pathogenic in ClinVar[133] and variants reported as disease-causing (DM) in HGMD[146] were further investigated. Novel disruptive variants (nonsense, frameshift and canonical splice), were also evaluated. Among the total 1884 variants identified in these genes, 214 were reported as disease-causing in HGMD and 109 variants were annotated in ClinVar as pathogenic or likely pathogenic, with an overlap between the two databases of 94 variants (Figure 40).

For variant classification, I followed stringent criteria for reporting IFs, given the fact that the a priori probability of pathogenicity is lower than in disease-targeted testing[141]. At this phase, I applied an additional disease-specific frequency filter for each gene using the internal database, 1000 Genomes and ExAC[112]. As a further step, I reviewed the primary literature for the remaining variants and only variants putatively producing the phenotype of interest were considered. In all the 56 ACMG genes, only 4 variants met the criteria for the pathogenic rank (two frameshift variants in *BRCA2*, one nonsense variant in *TP53* and one nonsense variant in *KCNH2)* while 6 variants (missense variants in *BRCA2*, *RET*, *DSG2*, *KNCH2*, *SCN5A* and *RYR1*) were classified as likely pathogenic. I further analysed for novel protein truncating mutations (nonsense, frameshift and canonical splice), identifying 3 additional variants located respectively in *MUTYH, TSC2* and *MYBPC3.* Overall, the cumulative prevalence of pathogenic IFs in the cohort of 855 individuals was 1.52%. Focusing on cardiac genes, the variants that met the criteria for pathogenicity were 4 reported (*DSG2*, 2 in *KCNH2* and 1 in *SCN5A*) and 1 novel in *MYBPC3*. As a result, in the diagnostic cohort of 855 probands, 0.6% carries IFs in cardiomyopathy or channelopathy genes. The 4 variants reported in the literature were reported to the patients (Table 23).

Three of the variants were identified in early childhood patients (1-2 years of age) with mitochondrial or neurological phenotype as a primary disorder for which WES was ordered. For these patients, the follow up of the genetic findings is ongoing. In the fourth adult patient (27 years), who was sequenced due to the suspicion of Marfan syndrome, an incidental finding in *SCN5A* has been identified and a clinical follow-up investigation has been initiated based on this report. The results of the clinical cardiological investigation for Brugada syndrome were negative. However, considering the sex and age of the patient, the profile risk of Brugada syndrome (male, age 30-50), its incomplete penetrance and highly variable expressivity, it was decided to screen in other members of the family. The variant was identified also in the father and he is currently monitored by the specialist in arrhythmogenic disease.

Table 22. ACMG gene list recommendations for reporting IFs. Among the 56 actionable genes, 31 are cardiac genes, of which 20 are associated with channelopathies (such as LQTS) or cardiomyopathies (such as ARVC). [145]

	Phenotype	MIM-disorder	PMID-Gene Reviews entry	Typical age of onset	Gene	MIM-gene	Inheritance[a]	Variants to report[b]
	Familial hypercholesterolemia	143890 603776	No GeneReviews entry	Child/adult	*LDLR* *APOB* *PCSK9*	606945 107730 607786	SD SD AD	KP and EP KP
4 channelopathy genes	Catecholaminergic polymorphic ventricular tachycardia	604772			*RYR2*	180902	AD	KP
4 channelopathy genes	Romano–Ward long QT syndrome types 1, 2, and 3, Brugada syndrome	192500 613688 603830 601144	20301308	Child/adult	*KCNQ1* *KCNH2* *SCN5A*	607542 152427 600163	AD	KP and EP
16 cardiomyopathy genes	Hypertrophic cardiomyopathy, dilated cardiomyopathy	115197 192600 601494 613690 115196 608751 612098 600858 301500 608758 115200	20301725	Child/adult	*MYBPC3* *MYH7* *TNNT2* *TNNI3* *TPM1* *MYL3* *ACTC1* *PRKAG2* *GLA* *MYL2* *LMNA*	600958 160760 191045 191044 191010 160790 102540 602743 300644 160781 150330	AD XL AD 	KP and EP KP KP and EP KP KP and EP (hemi, het, hom) KP KP and EP
16 cardiomyopathy genes	Arrhythmogenic right-ventricular cardiomyopathy	609040 604400 610476 607450 610193	20301310	Child/adult	*PKP2* *DSP* *DSC2* *TMEM43* *DSG2*	602861 125647 125645 612048 125671	AD 	KP and EP KP KP and EP
	Marfan syndrome, Loeys–Dietz syndromes, and familial thoracic aortic aneurysms and dissections	154700 609192 608967 610168 610380 613795 611788	20301510 20301312 20301299	Child/adult	*FBN1* *TGFBR1* *TGFBR2* *SMAD3* *ACTA2* *MYLK* *MYH11*	134797 190181 190182 603109 102620 600922 160745	AD 	KP and EP
	Ehlers–Danlos syndrome, vascular type	130050	20301667	Child/adult	*COL3A1*	120180	AD	KP and EP

Tot: 20 genes

Green et al, ACMG Recommendations 2013

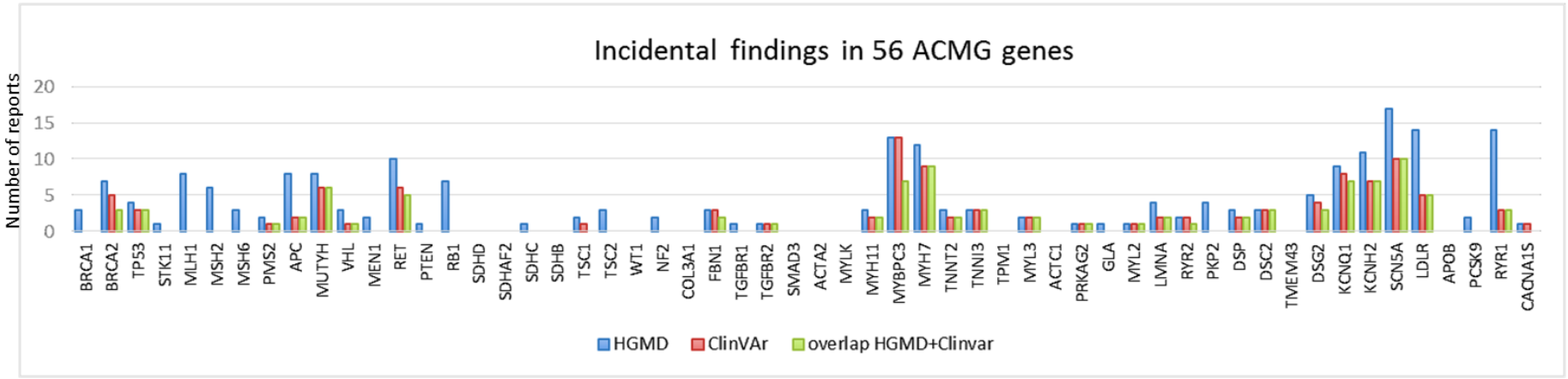

Figure 40. Number of variants identified in the 56 actionable ACMG genes and reported in HGMD and/or ClinVar.

Table 23. List of IFs in cardiac genes identified and reported to the patients

Age	Sex	Classification	Primary diseases	Gene	Variant	Allele	HGMD	ClinVar	ExAC EUR	ExAC AFR
1 y	F	LP	Mitochondrial disease	DSG2	missense	1	DM	Pathogenic	33354--2--0	4899--1--0
1 y	F	P	Developmental delay	KCNH2	nonsense	1	DM	Pathogenic	33209--0--0	5149--1--0
2 y	M	LP	Mental retardation	KCNH2	missense	1	DM	Pathogenic,Pathogenic, Likely Pathogenic	33312--1--0	5192--0--0
27 y	F	LP	Marfan syndrome	SCN5A	missense	1	DM	Pathogenic,Pathogenic	33361--1--0	5190--0--0

LP= likely pathogenic; P= pathogenic; Allele 1= heterozygous; DM= disease-causing mutation; Allele frequency of each variant in the different exome databases (IHG Helmholtz, 1000 Genomes, ExAC) is reported. EUR denotes the European non-Finnish and AFR denotes the African sub-populations of ExAC..

IV. Discussion

Cardiac disorders are a genetically very heterogeneous group of diseases, with hundreds of genes and molecular pathways implicated in their pathogenesis. Clinical presentation can be diverse, with blurred boundaries and overlapping phenotypes.

This broad clinical and genetic spectrum challenges their molecular diagnosis. To face such genetic complexity, an un-targeted molecular genetic approach can be particularly suited. During the last decade, WES has become the primary discovery tool in human genetics and it represents an extraordinary potential for most of the clinical fields. Indeed, WES allows the identification of genetic defects in both novel and known disease-associated genes. WES is mainly used in gene discovery of Mendelian disorders (e.g. long QT syndrome (LQTS), arrhythmogenic right ventricular cardiomyopathy (ARVC), but has also been successfully applied to elucidate the complex genetic basis of congenital heart defects (CHDs).

During my PhD, I extensively investigated the role of WES and its application in cardiological disorders. The identification of the underlying mutation is crucial for these disorders as it does not only provide the patient with a molecular diagnosis, but also allows directs family screening and can potentially guide therapy. Genetic funding can have a high impact on life-threatening conditions, potentially improving prevention and intervention. Moreover, the discovery of the underlying molecular basis improves patients' condition and, in many cases, can facilitate novel treatment trials and support the development of new therapeutic strategies. In this scenario, WES discoveries can effectively be seen as a first step towards "precision medicine" interventions in cardiac disorders.

However, the yield of genetic testing with WES technologies is variable across cardiac diseases, ranging from 80% in long QT syndrome to 50% in ARVC cases. In other words, one patient out of four for LQTS and one patient out of two for ARVC cannot receive any possible explanation of their disease, which carries the risk of sudden cardiac death in more than 20% of the cases.

In this study, I applied WES to discover novel genes in Mendelian dominant and recessive forms of cardiomyopathy (e.g. ARVC and mitochondrial cardiomyopathy) as well as novel variants in known disease-associated genes (e.g. long QT syndrome). The identification of novel genes (or pathways) offers the opportunity to better understand the pathomechanism of the underlying disorder. To this end, the possibility of linking genetic discoveries here described with functional assessment studies (gene variant functional analysis and iPCS-derived cardiomyocytes platforms) is highly relevant and will be covered in this section.

Two examples of novel genes identified in this study are discussed in 4.1. In addition, by applying the de novo analysis in index-unaffected parent trios I also found a new pathway in an extremely severe form of congenital heart defects, hypoplastic left heart syndrome (HLHS), for which the pathological mechanism was largely unknown.

In case of identification of variants in known genes, I have explored the impact of WES findings in the management of patients with ARVC and mitochondrial cardiomyopathy in 4.2. and also the perspective of personalised treatment options in *ACAD9* patients in 4.5.

During my PhD studies, I have also applied different type of analyses in order to identify causative variants in a cohort of LQTS individuals already screened with conventional LQTS diagnostic panel. Finally, I have evaluated the presence of incidental findings (IFs), an emerging demanding issue in clinical genomics, in the internal diagnostic cohort.

4.1 Detection of Novel Candidate Disease Genes

In this study, I have applied WES using filters for both dominant and recessive pattern of inheritance to discover novel genes in cardiac disorders. I could successfully identify variants in novel candidate disease genes related to ARVC, namely *CDH2* and *MYH10*.

4.1.1 ARVC and cell-adhesion genes

Cardiomyocytes are connected by the intercalated discs, specialized structures essential for cell-cell adhesion plus electrical and mechanical interaction between cardiomyocytes[151,226].

In cardiac desmosomes, two transmembrane cadherins, desmoglein-2 (DSG2) and desmocollin-2 (DSC2), maintain cell-cell adhesion. Moreover, they linked proteins such as plakoglobin (JUP) and plakophilin-2 (PKP2), which connect with desmoplakin (DSP) and desmin filaments[226]. The structure of myocardium is preserved by these specialized cell-adhesion junctions that link the cells to the network of intermediate filaments[227,228]. Therefore, defects of these adhesion junctions can cause decreased connection between cardiomyocytes and consequent fibro-fatty tissue replacement. These molecular changes lead to the generation of life-threatening arrhythmias, which are the predominant clinical manifestation of ARVC[147,226] (Figure 1).

ARVC is mainly associated with variants in genes encoding for desmosome proteins, but recently mutations in CTNNA3-encoded αT-catenin have also been implicated in the mechanism of the disease[147]. The αT-catenin protein acts as a cytoskeletal adhesion protein which links desmosomal proteins and adherens junctions[151]. By applying WES for variant

discovery on ARVC families, I have identified two novel candidate genes that belong to this cell-cell adhesion network.

4.1.1.1 Additional ARVC cases carrying *CDH2* variants

In a three-generation South African family of European ancestry with dominant ARVC inheritance, I used WES selecting 2 cousins, the most distant-related affected family members. The number of variants shared by the two affected family members was still high (Figure 24), but applying quality and frequency filters, the final number of variants was greatly reduced to 13[147]. Among them, the *CDH2* variant p.Gln229Pro resulted as the only variant predicted to be damaging by all prediction tools considered and absent in exome databases of controls[147]. In addition, the biological evaluation added a strong support to consider the *CDH2* variant as the causative one[147]. Moreover, a parametric linkage analysis further strengthened this finding. Indeed, among all the genomic regions of positive linkage using a LOD score>2, the only variant present in an area of linkage was the *CDH2* variant[147].

In order to support the association between *CDH2* variants and ARVC phenotype, a genotype-negative cohort of 73 unrelated ARVC cases was screened by our collaborators (Prof. Mayosi). Notably, an additional *CDH2* variant (p.Asp407Asn, Figure 41) was identified in one ARVC case of white South African ancestry[147]. He was a young male who presented at the age of 15 years similarly to patient III:2, with palpitations and pre-syncopal event linked with ventricular tachycardia[147]. The diagnosis of ARVC was based on the electrocardiographical findings (such as T-wave inversion from V1 through V6, late potentials, sustained ventricular tachycardia with left bundle branch block morphology) and structural changes (marked RV dilatation and multiple aneurysms of the RV free wall) on echocardiography, angiography, and cardiac magnetic resonance imaging[147]. Moreover, histological analysis of the RV endomyocardial biopsy underlined fibrofatty replacement of cardiomyocytes[147]. The missense *CDH2* heterozygous variant identified in this case affected a highly conserved amino-acid residue, it was predicted to be damaging by all in silico tools. In addition, the variant was not present in his unaffected mother and in internal exome database[147].

CDH2 pathogenic variants were detected in 2.7% (2/74) South African ARVC patients[147]. This prevalence is much higher than the reported one in the ExAC database, where among 60706 exomes 244 nonsynonymous variants with MAF<0.0001 are present, but only 61, present in 123 subjects (0,2%), are predicted to be damaging[125].[147] There is a statistically

significant difference (p=0.0005) in the prevalence of rare variants with damaging predictions between the ARVC cohort (2.7%) and the general population (0.2%)[147].

Regarding the biological aspect, Cadherin 2 or N-cadherin belongs to the cadherin superfamily of Ca2+-dependent cell surface adhesion proteins[147,229]. In cardiac tissue, this protein is present at the intercalated disc, where the cell-cell adhesion between cardiomyocytes depend on the gap junctions, the fascia adherens junctions, and the desmosome[147,229]. Notably, the junctions of the intercalated disc are formed by a mixture of desmosomal and fascia adherens proteins, called area composite, a hybrid functional area[151,230]. Accordingly, the pathomechanism underlying ARVC may involve not only the desmosomal area but also the area composita[147.] The function of CDH2 in the intercalated disc is also suggested in a Cdh2 mouse model, where the cardiac-specific deletion of cadherin 2 in the adult mouse heart was link to the loss of the intercalated disc structure[147,150]. Interestingly, the mouse model also showed mild dilated cardiomyopathy with spontaneous ventricular arrhythmias causing sudden cardiac death[147].

In conclusion, these data offer novel insight into the pathogenesis of ARVC, extending the mechanism to the area composite through the identification of disease causing variants in *CDH2*[147].

Some limitations of this study include the absence of functional data and number of affected patients is still limited, underling that these findings need to be validated in large cohort of ARVC cases[147].

However, to reinforce these findings, it is important to mention that recently Ackerman et al. identified with WES the same p.Asp407Asn-*CDH2* variant in an additional ARVC family [226]. The variant was present among 3 affected family members and not present in the unaffected family member[226].

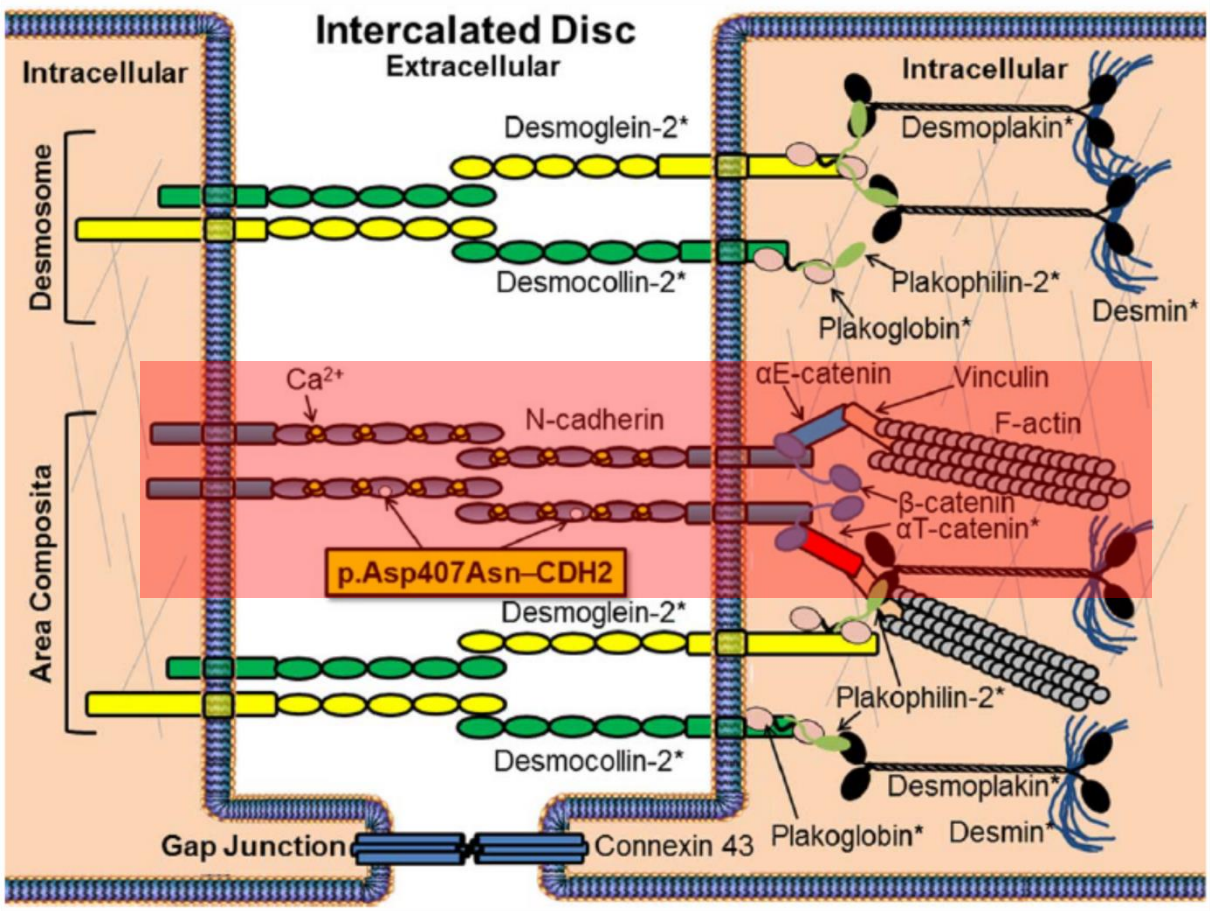

Figure 41. Intercalated disc (ID) divided in gap junctions, desmosomes, and adherens junctions. Gap junctions include connexins, mainly connexin-43. Desmosomes are responsible of cell-cell adhesion thought desmosomal cadherins that connect to plakoglobin-2 and desmoplakin, which then link desmin intermediate filaments. The adherens junction are formed by N-cadherin, which is heart-specific and has a calcium-dependent ectodomain. The adherens junctions are located near the desmosomes, creating a hybrid zone called the area composita. Ncadherin is extracellular and connects to the cell via β-catenin (or plakoglobin) and α-catenin, which then binds to vinculin and then F-actin. αT-catenin can bind plakophillin-2 (desmosome), as well as β-catenin (plakoglobin), which has been shown to be an essential component for the stability of the area composita. [Taken from [226]].

4.1.1.2 RhoA/MRTF-A signalling and cardiomyocytic identity

To verify the importance of cytoskeleton remodelling in driving pathological myocyte-to-adipocyte switch, I analysed WES data from ARVC patients lacking mutations in classical disease-causing genes. Indeed, if this mechano-signaling is relevant in ARVC, then genes belonging to the cytoskeletal pathways may be targets of pathogenic mutations[151]. Interestingly, in one of four cases, I identified a heterozygous nonsense mutation (c.1729C>T; p.R577*) in *MYH10*. This gene encodes the non-muscle myosin IIB (NMIIB) of the actomyosin cytoskeleton, which has been reported to regulate both actin dynamics as well as accumulation of active RhoA-GTP at the adherens junctions[151,231]. Functional studies conducted by Prof. Moretti and Prof. Laugwitz group on iPSC with *MYH10* defect (c.1729C>T; p.R577*) have shown how genetic perturbation of RhoA-mediated cytoskeleton remodelling can induce myocyte-to-adipocyte switch[151]. However, important limitation

remains the necessity of identifying additional patients carrying variants in these pathways. Further screening of these genes in ARVC cohorts is currently ongoing.

4.2 Identification of novel variants in known disease-associated genes

Among the many applications of WES, the identification of novel variants in known disease-associated genes plays a key role and was performed in this study, in ARVC (*DSP*) and in mitochondrial cardiomyopathy (*ACAD9*). WES allows to analyse more than 20000 genes at the same time with a reasonable price, thus avoiding time-consuming lists of candidate gene list and an odyssey of multiple target gene testing. This is particularly important for multisystemic disorders (such as mitochondrial disorders) or for isolated cardiac idiopathic form (such as idiopathic ventricular fibrillation, IVF). The routinely use of WES in molecular diagnostics offers a continuous possibility of expansion of the phenotypes spectrum associated with a particular gene, leading to a faster collection of genotyped cohorts (such as *ACAD9* patients). This is extremely useful in terms of genotype-phenotype correlations studies and epidemiological evaluation of genetic findings identified by WES.

4.2.1 Left-dominant arrhythmogenic cardiomyopathy and *DSP* variants

Mutations of the cell-to-cell adhesion protein desmoplakin (*DSP*) are known to be the cause of arrhythmogenic right ventricular cardiomyopathy (ARVC), but can lead to very heterogeneous phenotypes. Accordingly, the consensus paper of HRS/EHRA suggested the use of the term ARVC (arrhythmogenic ventricular cardiomyopathy) for the desmosomal diseases[232]. These are cardiomyopathies with an increased arrhythmogenic risk, most of which are autosomal dominant. However, after the revision of the diagnostic criteria by the Task Force Diagnostic Criteria 2010, the diagnosis of ARVC has been made suitable only for classical ARVC phenotype (with or without LV involvement)[5]. The left-dominant arrhythmogenic cardiomyopathy (LDAC) eludes these diagnostic criteria[233].

Left ventricular cardiac involvement is more common in *DSP* mutation carriers than a pure ARVC phenotype. A subepicardial late gadolinium enhancement at cardiac MRI is the only early sign of this disease, which otherwise often manifests itself as sudden cardiac death[233].

In a German family, I detected in all affected members with DNA available a novel frameshift variant located in the desmoplakin gene (*DSP*) associated with ARVC with LV involvement (left-dominant arrhythmogenic cardiomyopathy, LDAC). This finding allows an early identification of family members at risk and preventive interventions.

Such findings highlight the power of WES in elucidating the genetic causes of apparently idiopathic ventricular fibrillation cases. Moreover, WES could recognize an unusual presentation of the disease, allowing the expansion of the phenotypic spectrum associated with *DSP* variants. Notably, only recently LOF mutations in *DSP* have been specifically associated with left-dominant forms of ARVC[234]. Therefore, the presence of *DSP* LOF mutations should alert to the likely development of LV dysfunction.

4.2.2 *ACAD9* deficiency: clinical and genetic spectrum

Complex I deficiency is the most common biochemical defect in mitochondrial disorders [235] and *ACAD9* probably represents one of the most common causes of mitochondrial respiratory chain deficiency, with a conservative estimate of 27 new patients born every year in Europe (Table 10)[250].

The variants of the 67 patients from 47 families with *ACAD9* deficiency were distributed across the entire gene, with no genotype-phenotype correlation in specific regions of the gene or functional domains of the protein[250]. Moreover, no individual carried biallelic loss of function alleles, suggesting that the complete loss of *ACAD9* protein is probably not compatible with life[250].

Although cardiomyopathy is not always present, the vast majority of patients presented with hypertrophic cardiomyopathy, lactic acidosis, muscle weakness, and exercise intolerance[250].
The cohort is clearly divided in two main subgroups of *ACAD9* deficient patients. Those with an early presentation (i.e. in the first year of life) showed an extremely severe and poorer outcome compared to those who presented a later in life[250].
This observation can also influence the decision of heart transplantation. Four patients of the studied cohort underwent heart transplantation. Unfortunately, in the two patients with onset in the first year of life, the heart transplantation was not successful. On the other hand, the two patients with onset after the first year of age developed normally. Longitudinal studies are needed to identify which patients with *ACAD9* deficiency should be considered for heart transplantation[250].
This study offers new insight in the clinical and genetic spectrum of *ACAD9* deficiency and underlines the efficacy of characterizing larger genotyped cohorts for the novel genes identified by WES.

4.2.3 Isolated mitochondrial cardiomyopathy in *ELAC2* cases

Cardiomyopathy is estimated to occur in 20–40% of children with mitochondrial diseases but, given the lack of routinely screening for cardiac involvement, this figure is likely underestimated. The most common form of cardiomyopathy is hypertrophic, however, dilated cardiomyopathy and left ventricular (LV) non-compaction also occur relatively frequently. Conduction system and bradyarrhythmias, in addition to ventricular pre-excitation and tachyarrhythmias, are the most commonly encountered arrhythmias in mitochondrial disorders[49,50].
On the other hand, when a mitochondrial condition affects at the initial stage selectively the heart, the mitochondrial cardiomyopathy may be clinically indistinguishable from other genetic determined cardiomyopathies[236]. Hence, a mitochondrial disease has also to be considered in idiopathic isolated forms of cardiomyopathies. Conventional cardiomyopathy gene panel often do not include all the mitochondrial cardiomyopathy-associated genes and only the increasing use of WES unmasked such mitochondrial cases.
Mutations in *ELAC2* have been described in association with a multiple-organ system disorder, encompassing early-onset hypertrophic cardiomyopathy, developmental delay and complex I deficiency[237,238]. Using WES bi-allelic *ELAC2* variants in two unrelated patients with severe early-onset cardiomyopathy were identified. In both patients, mitochondrial disease was initially not considered and clinical features such as failure to thrive, occasionally

elevated lactate levels and neurological impairment were considered as consequences of heart failure. These findings underline the utility of WES in atypical disease presentation and stress the importance of including mitochondrial cardiomyopathy, especially *ELAC2* deficiency. Moreover, *ELAC2* cases presented in this study perfectly picture an important issue related to the treatment of mitochondrial cardiomyopathy: the option of cardiac transplantation. Indeed, for mitochondrial disorders, there are no clear guidelines that can help clinicians treating these patients and in most of the case decisions rely on individual opinion. One of the two *ELAC2* case presented here was successfully transplanted at the age of 5 years and currently does not show any neurological or multisystem manifestation. In the *ACAD9* cohort, the survival patients showed a good outcome even after many years of follow up. Although preliminary, this evidence supports the inclusion of mitochondrial patients in heart transplantation list, especially for those patients presenting after the age of 1 year.

4.3 WES and new insight into the molecular basis of CHDs

4.3.1 Primary cilium gene defects in HLHS cases

In the HLHS cohort, I identified de novo variants influencing the occurrence of the congenital heart defect (CHDs) under study in 11/78 cases (14%). Specifically, 4 cases (5.1%) carried mutations related to monogenic syndromes in which CHDs are a feature (*NOTCH2, CREBBP, CHD7, FGFR2*), whereas in 7 additional cases (9.0%) I have identified de novo mutations in genes known to be implicated in early cardiac development (*FGFR1, GLI1*, *GLI2, MACF1, ETS2, PTCH1*, *PTCH2*). The pathway enrichment analysis focused on the de novo mutations, highlighted an overrepresentation of mutations in genes and pathways important for cardiac development, indicating a partial role of the de novo mutations in the genesis of HLHS and underscoring the genetic heterogeneity associated with this disease.

Focusing on the recessive analysis, I have identified 8 additional cases (10%) carrying mutations in candidate genes, specifically *OFD1*, *ATRX, EVC2, FBN2, BCOR, COL6A1 UBR1, NSDHL*. These genes are reported in OMIM in association with syndromes (mainly recessive) in which CHDs are possible features[193,204–211]. Among these syndromes, the Ellis-Van Creveld Syndrome is the one with the highest prevalence of CHDs, observed in over 50% of the cases[204,205]. This syndrome is a recessive disorder, due to mutations in *EVC2*[204,205]. Interestingly, the EVC2 protein is found in primary cilia and is thought to regulate the hedgehog signalling pathway. A connection between the sonic hedgehog (HH) signalling pathway and the cilia is well established[206]. Specifically, EVC and EVC2 co-localize at

primary cilia in a mutually dependent manner and the EVC/EVC2 complex mediates hedgehog signal transduction from Smoothened to the GLI transcription factors[206].

Moreover, pathway analysis performed on the list of recessive and inherited genes have underlined a significant enrichment of GO terms related with ciliogenesis.

Through a rigorous interrogation of known and suspected human CHDs genes using available bioinformatics data resources, I have provided new insights into the genetic landscape of HLHS. The functional evidence of these genetic findings is currently under investigation by the collaborators in Munich (Prof. Laugwitz, Prof. Moretti and Dr. Krane) and USA (Prof. Gruber).

4.3.2 Defective primary cilium/autophagy/cell-cycle axis in HLHS

As part of the PhD work of Svenja Laue (group of Prof. Alessandra Moretti and Karl-Ludwig Laugwitz) the hiPSCs from 3 individuals of HLHS cohort (70315, 69270, 73119) were generated and differentiated into cardiovascular cell lineages for molecular characterization. HLHS-affected hiPSCs formed cardiomyocytes did not show any structural abnormalities but exhibited transcriptional reduction of key cardiac markers that can be assigned to left cardiac structures. Moreover, while apoptosis rate was not affected in HLHS-cardiovascular progenitors cells (CVPCs), their specification was altered due to a cell cycle progression delay. This correlated to impaired cilia function and reduced response to autophagy induction. Inhibition of autophagy in healthy CVPCs was sufficient to recapitulate the observed HLHS disease phenotype. This confirms the results of my analysis with WES of three HLHS patients and their parents, which revealed de novo mutations in genes involved in cilia and autophagy processes (*MACF1*, *FGFR1* and *DENND5B*) (Table 15). Further analysis of the potential role of *DENND5B* in HLHS demonstrated that the identified de novo mutation impairs binding of DENND5B to RAB39, a key player in autophagosome formation. Moreover, morpholino knockdown of DENND5B in Xenopus and zebrafish revealed a crucial and novel role of this gene in cardiogenesis and cilia formation and function. Together, this work uncovered an unprecedented role of autophagy in human cardiovascular cell lineage specification and demonstrated defective primary cilium/ autophagy/cell-cycle axis in HLHS.

4.4 WES in clinical genomics

4.4.1 LQTS cohort

To date, sixteen genes have been identified in association with long QT syndrome (LQTS)[31]. The three main genes, *KCNQ1* (LQT1), *KCNH2* (LQT2), and *SCN5A* (LQT3), account for approximately 75% clinically definite LQTS, reaching almost entirely the 80% of the diagnostic yield of genetic testing in LQTS[31]. The use of WES and particularly the performance of trio analysis in severe isolated cases of LQTS led in 2013 to the discovery of missense de novo variants in *CALM1* and *CALM2* genes responsible for a very malignant neonatal form of LQTS[36,37]. Notably, the *CALM1* patient was analysed through the Helmholtz internal pipeline. Since then, *CALM* genes have been implicated in the pathogenesis of LQTS and many efforts are ongoing to elucidate the physiological mechanism underlying these extremely severe forms of the disease and in vitro CRISPR therapeutic strategies have been performed. These results represent an important step toward precision medicine[239]. Very recently, Ackerman group performed WES in a cohort of 39 LQTS individuals with no mutations for all known LQTS-susceptibility genes: *AKAP9, ANKB, CACNA1C, CAV3, KCNE1, KCNE2, KCNH2, KCNJ2, KCNJ5, KCNQ1, SCN4B, SCN5A, SNTA1*, and *TRDN*. Among these 39 cases, 4 carried a missense variant in *CALM1* and *CALM2* genes[240]. However, they did not report the discovery of variants in known LQTS genes as the focus of their work was only on the genetic findings in *CALM* genes.

In the cohort assessed in this study (26 LQTS families), a variant in one of the three *CALM* genes (*CALM3*) was identified in one case. The phenotype of this index patient corresponds with the early and extremely severe presentation of the other *CALM* patients. Ackerman group in 2015 had reported the first LQTS case due to a *CALM3* variant[38]. So far, only three LQTS/*CALM3* have been described in literature[38,241]. My analysis confirms the presence of *CALM* variants in genotype negative LQTS patients (1/26=3%) and support the concept of a LQTS "calmodulinophaty" phenotype[241]. In addition, these findings underline the possibility of identification of variants in known LQTS gene, suggesting limits in the conventional diagnostics panels. Especially for cases screened years ago, a reanalysis with more advanced technology is needed. Another crucial aspect that emerged with WES analysis of LQTS cases is the presence of CNVs in known LQTS genes. In 2011, an array-based study conducted by Barc and colleagues showed that CNVs in *KNCQ1* and *KCNH2* are responsible for around 3% of the cases with no point mutations identified with conventional LQTS panel[242]. The presence of CNVs in 3 out 26 LQTS cases in the cohort studied here highlights this important

aspect. An important limitation of the study is the lack of availability of DNA to prove the CNVs identified. However, the cosegregation among affected family members and the genotype-phenotype correlation offered a convincing evidence for the plausibility of these findings. Still, CNVs analysis needs to be improved and results should be interpreted carefully. With the improvement and systematic use of this type of analysis, the use of WES to diagnose LQTS cases will be even more comprehensive and justified. Taken all together, WES has identified in 11 out of 26 LQTS cases (40%) a genetic finding in known LQTS disease-associated genes, compared with the negative results of a conventional gene panel.

4.4.2 Incidental findings

This is an exciting time to work in clinical genomics, since the genomic sequencing market has experienced and exponential growth. However, this enormous amount of genomic data produced brought also a new demanding issue in clinical practice.

WES is a routine molecular diagnostic test for patients with genetic disorders[159,243–245]. However, this technique allows the identification not only of mutations responsible for the disease under investigation, but also of variants potentially causing other diseases, the so-called "incidental findings" (IFs)[145].

The American College of Medical Genetics and Genomics (ACMG) stated that IFs should be reported based on clinical validity and utility and indicated a list of 56 actionable genes[145]. Among the genes listed, nearly half (20/56) are major genes associated with channelopathies and cardiomyopathies (Table 22).

The problem of unexpected positive findings by WES is relatively common. Even the testing of healthy individuals could result in an unexpected detection of pathogenic variants responsible for late-onset disease. The ACMG recommendation on reporting IFs has raised intense and controversial debates in the clinical genomic community. Arguments against are represented by the unknown evidence of benefit (following the medical principal precept "primum non nocere"), risks, the lack of evidence-based data that could suggest recommendations ad hoc and, most important, patient's consensus. On the other hand, a strong argument in favour (and also the main principle behind the ACMG recommendations) is that the recognition of any known disease-associated gene could be clinically useful, especially in terms of prevention of life-threatening disorders such as cancer and cardiovascular disease contained in the ACMG list[145]. In the first large exome cohorts in whom the incidental findings were investigated, the percentages of patients carrying IFs in ACMG genes were quite high (5-10%)[246,247]. With the introduction of database of exome

controls like ExAC, the criteria for selection of the variant became more stringent. It became then clear that the frequency of the variant is the first mandatory step to select IFs, independently from the clinical report in database of known variants[141]. Moreover, the introduction of database like ClinVar helped the genetists in the odyssey of variant interpretation. Indeed, ClinVar clearly presents the classification in benign, VUS, probably pathogenic and pathogenic variants and the review of the variants is constantly updated. These distinctions are the basis to try to automatize the process of identification of IFs and reduce the time of analysis. However, the current status of databases of reported variants is far to be perfect and this lack of completeness affects dramatically the interpretation of IFs. For IFs the a priori probability of pathogenicity is much lower than in a standard genetic testing process, since the phenotypic manifestations are missing. In this context, we are currently facing a reversion of the standard process in patient care: from genotype to phenotype and not vice versa. It is important to bear in mind that standard criteria of pathogenicity applied in conventional molecular genetic testing are too loose if transferred in the context of IFs interpretation. For disorders with incomplete penetrance and variable expressivity (e.g. cardiomyopathy and channelopathies) this issue becomes even more crucial. Cardiologists are also currently not prepared to deal with these genetic findings. In order to improve the status of reporting IFs, in 2015 Amendola and colleagues proposed a revision of ACMG classification criteria adapted to IFs issue, thus providing a more strict and specific classification[141]. They also tested this new classification system among different clinical and research laboratories and showed that 1-2% of the European population (4300 exomes) carry a likely pathogenic or pathogenic variant in one of the 56 ACMG genes[141]. However, they used as a cohort a publicly available database, with no information about the age, sex and clinical record of the patient and no possibility of follow up of the genetic findings[141].

Following the criteria revised by Amendola and colleagues [141], I analysed the internal cohort of 855 individuals and found a pathogenic or likely pathogenic variant in one of the 56 ACMG genes in 1.52% of the individuals. In 4 patients of the studied cohort (0.6%) I have identified an IF in channelopathy or cardiomyopathy genes. The IFs identified were located in an ARVC gene (*DSG2*), LQTS genes (*KCNH2*) and *SCN5A* (Brugada syndrome). As already mentioned, these disorders are a frequent cause of sudden cardiac death (SCD) that in many cases can occur as first clinical manifestation. Since there are well-established criteria for diagnosis and treatment of such disorders, the report of probably pathogenic variants in these genes seems particularly useful[145]. Moreover, for the majority of the genes, like *KCNH2* in LQTS, a genotype-phenotype correlation is known and the trigger events of the cardiac

condition are also quite specific. Important to mention, in channelopathy disorders certain medicaments (such as anestethics) should be avoided because they can trigger a cardiac event. Analysing a family carrying a *SCN5A* variant, it was also possible to point out that the IFs are not "single patient" issue, but requires the investigation of the families with selection of carriers and not carriers. It is important to stress that the genetic finding is an indication, "a warrant" for the cardiologists who are responsible for the patient, but it cannot substitute the clinical evaluation and standard clinical diagnostic criteria for assessing the risk profile. Cardiologists should be aware of this issue since an interdisciplinary approach is fundamental.

4.5 Personalized treatment options

Personalized treatment option represents the future of clinical genomics as suggested in mitochondrial cardiomyopathy patients and severe cases with long QT syndrome.

A step forward has been made precisely in the *ACAD9* patients treated with riboflavin (Figure 33 C, D). The supplementation with riboflavin showed improvement in complex I activity in the majority of patient fibroblasts, and most patients were reported to have clinical improvements after treatment[250]. Patients presenting within the first year of life show a significantly better survival when treated with riboflavin[250]. One limitation of this observation is that most of the deaths occurred at the end of the first year of life and that might possibly induce a selection bias. Before drove any conclusion, detailed data about the riboflavin treatment in more patients are needed. Still, these findings are supported by single case reports[250]. In *ACAD9* cohort, families 1 and 33 are particularly informative. In both families the first child (Ind. 1, Ind. 45) died within the first two years of life without riboflavin treatment, while the younger affected siblings (Ind. 2, Ind.45 and Ind. 46), in whom supplementation was begun immediately upon diagnosis, are currently still alive (aged 10, 1.5 and 11 years, respectively)[250]. Paired data on fibroblasts and patient riboflavin treatment were available for eight patients, six of which showed parallel beneficial effects. Further cellular studies are necessary to define the mode of action of riboflavin in *ACAD9* deficiency[250].

In conclusion, these data add new arguments for the development of future clinical trials of riboflavin. The improvement of *ACAD9* deficient patients after riboflavin supports a trial of riboflavin and should be administrated in all phenotypically-consistent patients meanwhile the genetic investigations are underway. For the patients with an early presentation such easy therapeutic intervention can be really represent the difference between life and death[250].

The identification of *CALM* variants in 1-3% of the LQTS cases and the recent findings showing a CRISPR targeted approach and functional studies on the arrhythmogenic

mechanism involved[248] have opened also important scenarios for personalized treatment of this extremely malignant form of LQTS. It is crucial to consider that for these patients the conventional treatment for LQTS (β-blockers) is not sufficient to prevent the cardiac events and, due to multiple episodes of cardiac arrests, the ICD implantation is required. A gene therapy approach would have an enormous impact on the clinical outcome of these infants.

V. Outlook

5.1 WES in diagnostics of cardiac disorders

Currently, in the molecular diagnostics of cardiac disease, three NGS approaches are mainly used: gene panels, WES and sporadically whole genome sequencing (WGS). Gene panels can provide a high coverage of selected genes, but they only consider a predefined set of genes. This affects the possibility of identifying novel disease-associated genes and given that WES is constantly enlarging the number of genes of interest, new panels need to be often re-designed. The design and production of such panels are economically reasonable only if large numbers of samples are analyzed. As mentioned several times in this work, exome sequencing overcomes these problems by attempting to capture all coding genes. However, the coverage of targeted regions may be insufficient in exome sequencing and some region of interest might not be targeted at all. Moreover, WES analysis can also detect incidental findings, pathogenic variants in actionable genes associated with disorders not related to the primary disorder under investigation. Therefore, trained clinicians and genetic counselors are required to interpret and communicate the results to the patients.

5.2 Future perspective: complementary strategies (WGS and RNA-seq)

WGS provides complete coverage of the entire genome. Compared to exome, WGS does not include the exon capture step; therefore, the limitations due to the PCR steps are solved. In the near future, WGS will probably overcome the use of WES, following also a further drop in sequencing costs. However, with the enormous amount of genomic data produced by WGS for each sample (3-5 million between SNVs and small indels, but also 2,500 structural variants) WGS also arises the big issue of prioritization and interpretation of genomic findings.

Recently, it has become popular a complementary approach obtained by WES or WGS and RNA-seq data[157]. Indeed, RNA-seq allows direct probing of detected variants for their effect on RNA abundance and RNA sequence and not only pure genomic information (Figure 42)[98].

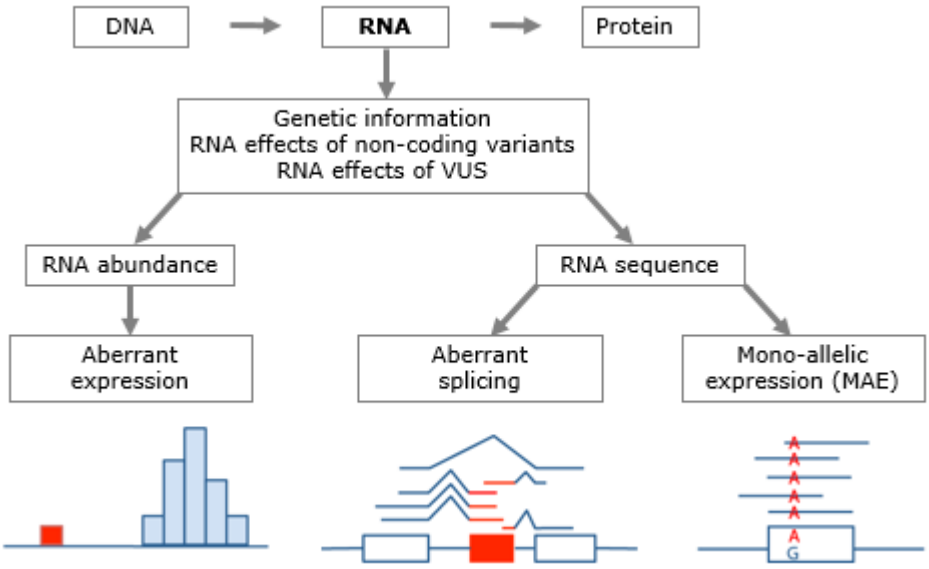

Figure 42. Different approaches to use RNA-seq in clinical diagnostics. [Taken from[98,157]].

The effect of a variant on RNA abundance can be measured by studying expression outliers using statistical analysis on entire gene sets[98]. The aberrant expression can be due to variants in non-coding regulatory regions (such as promoters, enhancers, and suppressors), but also variants located in the coding region[98]. In addition, RNA-seq can show differences in RNA sequence in cases of mono-allelic expression (MAE). MAE is found upon transcriptional silencing or post-transcriptional degradation of one allele. Altered RNA sequence can be related to aberrant splicing, a known cause of Mendelian disorders.[98]

Recently, two different studies used aberrant splicing and allele specific as well as differential expression to diagnose cases previously sequenced by WES but still elusive a genetic diagnosis. Kremer et al., with the analyse RNA-seq data could solve the diagnosis in 5/47 mitochondrial disease patients previously sequenced by WES [157]; while Cummings et al. found a genetic explanation in 17/50 cases with rare neuromuscular disorder [249]. These data pointed out the power of this complementary approach in improving the diagnostic yield of genetic disorders and it should be considered also for cardiac disease. However, in order to establish the RNA-seq application in clinical diagnostics of rare disease, further efforts concerning general standards and best practices are required.[120]

5.3 Clinical trials (large genotyped cohorts)

The identification of the genetic defect and providing definite molecular diagnoses are central aspects of precision medicine[3]. Individualized patient treatment is affected by the current knowledge of disease mechanism and the tendency of using phenotypic-based diagnostic approach. The broad phenotypic spectrum of a disease and the underlying genetic heterogeneity further challenge the clinical practice[3]. With the advent and use of NGS technology in clinical practice, phenotypic expansions of the disease spectrum are becoming more and more common[3].

The study on the *ACAD9* cohort represents an emblematic example in this sense. *ACAD9* is a mitochondrial gene identified by WES only few years ago (2010)[177] and it is currently considered one of the most frequent genetic cause of mitochondrial disorders. The data presented here on a cohort of 67 *ACAD9* patients enlarged the clinical and genetic spectrum of *ACAD9* deficiency and offered a successful example of rapid recruitment and characterization of large genotype cohorts through the use of well-established international networks on mitochondrial disorders[250]. These findings also provide valuable insight for the development of future clinical trials with riboflavin in these patients. Indeed, despite the scientific progress discussed thus far, it is important to underline that most of these disorders do not have any treatment. Among mitochondrial cardiomyopathy cases, there are still more than 70 diseases with less than 10 patients described. With the increasing number of patients, genotype-phenotype correlations will improve and that will support families counseling: the better the understanding of the pathomechanism, the greater the chance of identifying new treatment options.

Bibliography

1. Cirino, A. L. *et al.* Role of Genetic Testing in Inherited Cardiovascular Disease. *JAMA Cardiol.* **2,** 1153 (2017).
2. Sturm, A. C. & Hershberger, R. E. Genetic testing in cardiovascular medicine. *Curr. Opin. Cardiol.* **28,** 1 (2013).
3. Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17,** 507–522 (2016).
4. Parikh, V. N. & Ashley, E. A. Next-Generation Sequencing in Cardiovascular Disease. *Circulation* **135,** 406–409 (2017).
5. Marcus, F. I. *et al.* Diagnosis of Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia: Proposed Modification of the Task Force Criteria. *Circulation* **121,** 1533–1541 (2010).
6. Pinamonti, B., Brun, F., Mestroni, L. & Sinagra, G. Arrhythmogenic right ventricular cardiomyopathy: From genetics to diagnostic and therapeutic challenges. *World J. Cardiol.* **6,** 1234–44 (2014).
7. Corrado, D. & Thiene, G. Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia: Clinical Impact of Molecular Genetic Studies. *Circulation* **113,** 1634–1637 (2006).
8. Delmar, M. & McKenna, W. J. The Cardiac Desmosome and Arrhythmogenic Cardiomyopathies: From Gene to Disease. *Circ. Res.* **107,** 700–714 (2010).
9. Gerull, B. *et al.* Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat. Genet.* **36,** 1162–1164 (2004).
10. Rampazzo, A. *et al.* Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. *Am. J. Hum. Genet.* **71,** 1200–6 (2002).
11. Syrris, P. *et al.* Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. *Am. J. Hum. Genet.* **79,** 978–84 (2006).
12. Pilichou, K. *et al.* Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy. *Circulation* **113,** 1171–1179 (2006).
13. McKoy, G. *et al.* Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet (London, England)* **355,** 2119–24 (2000).
14. Campuzano, O. *et al.* Genetics of arrhythmogenic right ventricular cardiomyopathy. *J. Med. Genet.* **50,** 280–289 (2013).
15. Fressart, V. *et al.* Desmosomal gene analysis in arrhythmogenic right ventricular dysplasia/cardiomyopathy: spectrum of mutations and clinical impact in practice. *EP Eur.* **12,** 861–868 (2010).
16. Watkins, H., Ashrafian, H. & Redwood, C. Inherited Cardiomyopathies. *N. Engl. J. Med.* **364,** 1643–1656 (2011).
17. Merner, N. D. *et al.* Arrhythmogenic Right Ventricular Cardiomyopathy Type 5 Is a Fully Penetrant, Lethal Arrhythmic Disorder Caused by a Missense Mutation in the TMEM43 Gene. *Am. J. Hum. Genet.* **82,** 809–821 (2008).
18. Tiso, N. *et al.* Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum. Mol. Genet.* **10,** 189–94 (2001).
19. Roux-Buisson, N. *et al.* Prevalence and significance of rare RYR2 variants in arrhythmogenic right ventricular cardiomyopathy/dysplasia: Results of a systematic screening. *Hear. Rhythm* **11,** 1999–2009 (2014).

20. BEFFAGNA, G. *et al.* Regulatory mutations in transforming growth factor-?3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc. Res.* **65,** 366–373 (2005).
21. Erkapic, D. *et al.* Electrical storm in a patient with arrhythmogenic right ventricular cardiomyopathy and SCN5A mutation. *Europace* **10,** 884–887 (2008).
22. Klauke, B. *et al.* De novo desmin-mutation N116S is associated with arrhythmogenic right ventricular cardiomyopathy. *Hum. Mol. Genet.* **19,** 4595–4607 (2010).
23. Taylor, M. *et al.* Genetic Variation in Titin in Arrhythmogenic Right Ventricular Cardiomyopathy-Overlap Syndromes. *Circulation* **124,** 876–885 (2011).
24. van der Zwaag, P. A. *et al.* Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. *Eur. J. Heart Fail.* **14,** 1199–1207 (2012).
25. Quarta, G. *et al.* Mutations in the Lamin A/C gene mimic arrhythmogenic right ventricular cardiomyopathy. *Eur. Heart J.* **33,** 1128–1136 (2012).
26. van Hengel, J. *et al.* Mutations in the area composita protein αT-catenin are associated with arrhythmogenic right ventricular cardiomyopathy. *Eur. Heart J.* **34,** 201–210 (2013).
27. Lazzarini, E. *et al.* The ARVD/C Genetic Variants Database: 2014 Update. *Hum. Mutat.* **36,** 403–410 (2015).
28. Link, M. S. *et al.* Ventricular Arrhythmias in the North American Multidisciplinary Study of ARVC. *J. Am. Coll. Cardiol.* **64,** 119–125 (2014).
29. Tester, D. J. & Ackerman, M. J. Postmortem long QT syndrome genetic testing for sudden unexplained death in the young. *J. Am. Coll. Cardiol.* **49,** 240–6 (2007).
30. Schwartz, P. J. *et al.* Prevalence of the Congenital Long-QT Syndrome. *Circulation* **120,** 1761–1767 (2009).
31. Schwartz, P. J., Ackerman, M. J., George, A. L., Wilde, A. A. M. & Wilde, A. A. M. Impact of genetics on the clinical management of channelopathies. *J. Am. Coll. Cardiol.* **62,** 169–180 (2013).
32. Heijman, J. *et al.* Dominant-negative control of cAMP-dependent IKs upregulation in human long-QT syndrome type 1. *Circ. Res.* **110,** 211–9 (2012).
33. Anderson, C. L. *et al.* Most LQT2 mutations reduce Kv11.1 (hERG) current by a class 2 (trafficking-deficient) mechanism. *Circulation* **113,** 365–73 (2006).
34. Vatta, M. *et al.* Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. *Circulation* **114,** 2104–12 (2006).
35. Medeiros-Domingo, A. *et al.* SCN4B-encoded sodium channel beta4 subunit in congenital long-QT syndrome. *Circulation* **116,** 134–42 (2007).
36. Crotti, L. *et al.* Calmodulin Mutations Associated With Recurrent Cardiac Arrest in Infants. *Circulation* **127,** 1009–1017 (2013).
37. Makita, N. *et al.* Novel Calmodulin Mutations Associated With Congenital Arrhythmia Susceptibility. *Circ. Cardiovasc. Genet.* **7,** 466–474 (2014).
38. Reed, G. J., Boczek, N. J., Etheridge, S. P. & Ackerman, M. J. CALM3 mutation associated with long QT syndrome. *Hear. Rhythm* **12,** 419–22 (2015).
39. Altmann, H. M. *et al.* Homozygous/Compound Heterozygous Triadin Mutations Associated With Autosomal-Recessive Long-QT Syndrome and Pediatric Sudden Cardiac Arrest: Elucidation of the Triadin Knockout Syndrome. *Circulation* **131,** 2051–2060 (2015).
40. Priori, S. & Napolitano, C. in *Hurt's the heart* (ed. Valentin Fuster, Robert A. Harrington, Jagat Narula, Z. J. E.) 2208 (McGraw-Hill Professional, 2017).
41. Fermini, B. & Fossa, A. A. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nat. Rev. Drug Discov.* **2,** 439–447 (2003).

42. Mayr, J. A. *et al.* Spectrum of combined respiratory chain defects. *J. Inherit. Metab. Dis.* **38,** 629–40 (2015).
43. Reinecke, F., Smeitink, J. A. M. & van der Westhuizen, F. H. OXPHOS gene expression and control in mitochondrial disorders. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1792,** 1113–1121 (2009).
44. Gorman, G. S. *et al.* Mitochondrial diseases. **2,** 16080 (2016).
45. Holmgren, D. *et al.* Cardiomyopathy in children with mitochondrial disease; clinical course and cardiological findings. *Eur. Heart J.* **24,** 280–8 (2003).
46. Scaglia, F. *et al.* Clinical Spectrum, Morbidity, and Mortality in 113 Pediatric Patients With Mitochondrial Disease. *Pediatrics* **114,** (2004).
47. Limongelli, G. *et al.* Prevalence and natural history of heart disease in adults with primary mitochondrial respiratory chain disease. *Eur. J. Heart Fail.* **12,** 114–21 (2010).
48. Tang, S., Batra, A., Zhang, Y., Ebenroth, E. S. & Huang, T. Left ventricular noncompaction is associated with mutations in the mitochondrial genome. *Mitochondrion* **10,** 350–7 (2010).
49. Finsterer, J. & Kothari, S. Cardiac manifestations of primary mitochondrial disorders. *Int. J. Cardiol.* **177,** 754–63 (2014).
50. Bates, M. G. D. *et al.* Cardiac involvement in mitochondrial DNA disease: clinical spectrum, diagnosis, and management. *Eur. Heart J.* **33,** 3023–33 (2012).
51. Horvath, R. & Chinnery, P. F. The Effect of Neurological Genomics and Personalized Mitochondrial Medicine. *JAMA Neurol.* **74,** 11 (2017).
52. Wortmann, S., Mayr, J., Nuoffer, J., Prokisch, H. & Sperl, W. A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era. *Neuropediatrics* **48,** 309–314 (2017).
53. Alston, C. L., Rocha, M. C., Lax, N. Z., Turnbull, D. M. & Taylor, R. W. The genetics and pathology of mitochondrial disease. *J. Pathol.* **241,** 236–250 (2017).
54. Goldstein, A. C., Bhatia, P. & Vento, J. M. Mitochondrial Disease in Childhood: Nuclear Encoded. *Neurotherapeutics* **10,** 212–226 (2013).
55. Li, A. H. *et al.* Whole exome sequencing in 342 congenital cardiac left sided lesion cases reveals extensive genetic heterogeneity and complex inheritance patterns. *Genome Med.* **9,** 95 (2017).
56. van der Linde, D. *et al.* Birth Prevalence of Congenital Heart Disease Worldwide. *J. Am. Coll. Cardiol.* **58,** 2241–2247 (2011).
57. Jefferies, J. L. *et al.* Cardiovascular findings in duplication 17p11.2 syndrome. *Genet. Med.* **14,** 90–94 (2012).
58. Ware, S. M. & Jefferies, J. L. New Genetic Insights into Congenital Heart Disease. *J. Clin. Exp. Cardiolog.* **S8,** (2012).
59. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424,** 443–447 (2003).
60. Lalani, S. R. & Belmont, J. W. Genetic basis of congenital cardiovascular malformations. *Eur. J. Med. Genet.* **57,** 402–413 (2014).
61. Kerstjens-Frederikse, W. S. *et al.* Left ventricular outflow tract obstruction: should cardiac screening be offered to first-degree relatives? *Heart* **97,** 1228–32 (2011).
62. McBride, K. L. *et al.* Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. *Am. J. Med. Genet. A* **134A,** 180–6 (2005).
63. Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N. Engl. J. Med.* **376,** 21–31 (2017).
64. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498,** 220–223 (2013).
65. COX, H. & WILSON, D. I. GENETICS OF HYPOPLASTIC LEFT HEART

SYNDROME. *Fetal Matern. Med. Rev.* **18,** 103 (2007).
66. Brownell, L. G. & Shokeir, M. H. Inheritance of hypoplastic left heart syndrome (HLHS): further observations. *Clin. Genet.* **9,** 245–9 (1976).
67. Morris, C. D., Outcalt, J. & Menashe, V. D. Hypoplastic left heart syndrome: natural history in a geographically defined population. *Pediatrics* **85,** 977–83 (1990).
68. Hoffman, J. I. E. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39,** 1890–900 (2002).
69. Tikkanen, J. & Heinonen, O. P. Risk factors for hypoplastic left heart syndrome. *Teratology* **50,** 112–7 (1994).
70. Brenner, J. I., Berg, K. A., Schneider, D. S., Clark, E. B. & Boughman, J. A. Cardiac malformations in relatives of infants with hypoplastic left-heart syndrome. *Am. J. Dis. Child.* **143,** 1492–4 (1989).
71. Burn, J. *et al.* Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet (London, England)* **351,** 311–6 (1998).
72. Anderson, R. C. Congenital cardiac malformations in 109 sets of twins and triplets. *Am. J. Cardiol.* **39,** 1045–50 (1977).
73. Shokeir, M. H. Hypoplastic left heart syndrome: an autosomal recessive disorder. *Clin. Genet.* **2,** 7–14 (1971).
74. Wessels, M. W. *et al.* Autosomal dominant inheritance of left ventricular outflow tract obstruction. *Am. J. Med. Genet. Part A* **134A,** 171–179 (2005).
75. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461,** 272–276 (2009).
76. Gilbert, W. Why genes in pieces? *Nature* **271,** 501 (1978).
77. Kuhlenbäumer, G., Hullmann, J. & Appenzeller, S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum. Mutat.* **32,** 144–151 (2011).
78. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33 Suppl,** 228–37 (2003).
79. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12,** 745–755 (2011).
80. Petersen, B.-S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D. & Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet.* **18,** 14 (2017).
81. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci.* **106,** 19096–19101 (2009).
82. 1000 Genomes Project Consortium, G. A. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
83. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
84. Narasimhan, V. M., Xue, Y. & Tyler-Smith, C. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? *Trends Mol. Med.* **22,** 341–351 (2016).
85. Zhang, J., Chiodini, R., Badr, A. & Zhang, G. The impact of next-generation sequencing on genomics. *J. Genet. Genomics* **38,** 95–109 (2011).
86. Wong, L.-J. C. *Next generation sequencing : translation to clinical diagnostics*. (Springer, 2013).
87. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–7 (1977).
88. Moorthie, S., Mattocks, C. J. & Wright, C. F. Review of massively parallel DNA sequencing technologies. *Hugo J.* **5,** 1–12 (2011).
89. national human genome research institute. The Cost of Sequencing a Human Genome -

National Human Genome Research Institute (NHGRI). (2016). Available at: https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/.
90. Goodwin, S., McPherson, J. D. & Richard McCombie, W. Coming of age: ten years of next-generation sequencing technologies. (2016). doi:10.1038/nrg.2016.49
91. Demkow, U. & Płoski, R. *Clinical applications for next-generation sequencing*.
92. Illumina | Sequencing and array-based solutions for genetic research. (2015). Available at: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf,2015.
93. Clark, M. J. *et al.* Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **29,** 908–914 (2011).
94. Agilent | Chemical Analysis, Life Sciences, and Diagnostics. Available at: https://www.agilent.com/home.
95. Wieland, T., Burkhard Rost, U.-P., Fabian Theis, U.-P. & Tim Strom, P.-D. M. Next-Generation Sequencing Data Analysis. (2015).
96. Illumina | Sequencing and array-based solutions for genetic research. (2015).
97. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
98. Kremer, L. S. Discovery and validation of coding and non-coding pathogenic variants in mitochondrial disorders.
99. Coregenomics. (almost) everything you wanted to know about @illumina HiSeq 4000...and some stuff you didn't. Available at: http://core-genomics.blogspot.de/2016/01/almost-everything-you-wanted-to-know.html.
100. Church, D. M. *et al.* Extending reference assembly models. *Genome Biol.* **16,** 13 (2015).
101. Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* **23,** 5866–5878 (2014).
102. Platzer, M. The human genome and its upcoming dynamics. *Genome Dyn.* **2,** 1–16 (2006).
103. Gymrek, M. *et al.* Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* **48,** 22–9 (2016).
104. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–85 (1998).
105. Lam, H. Y. K. *et al.* Performance comparison of whole-genome sequencing platforms. *Nat. Biotechnol.* **30,** 78–82 (2011).
106. Goldfeder, R. L. *et al.* Medical implications of technical accuracy in genome sequencing. *Genome Med.* **8,** 24 (2016).
107. Zook, J. M. *et al.* Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **32,** 246–51 (2014).
108. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).
109. Dewey, F. E. *et al.* Clinical Interpretation and Implications of Whole-Genome Sequencing. *JAMA* **311,** 1035 (2014).
110. Li, J. B. & Church, G. M. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nat. Neurosci.* **16,** 1518–22 (2013).
111. Li, J. B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324,** 1210–3 (2009).
112. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
113. gnomAD browser. Available at: http://gnomad.broadinstitute.org/about. (Accessed: 19th November 2017)
114. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of

human genetic variants. *Nat. Genet.* **46,** 310–5 (2014).
115. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7,** 575–6 (2010).
116. Schwarz JM, Cooper DN, Schuelke M & Seelow D. MutationTaster. Available at: http://www.mutationtaster.org/.
117. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. in *Current Protocols in Human Genetics* **Chapter 7,** 7.20.1-7.20.41 (John Wiley & Sons, Inc., 2013).
118. Hu, J. & Ng, P. C. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS One* **8,** e77940 (2013).
119. Eck, S. H. Identification of genetic variation using Next-Generation Sequencing. (2014).
120. Schwarzmayr, T. ¨ F UR Assembly and Analysis of Next-Generation Sequencing Data.
121. Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* **14,** 681–691 (2013).
122. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet. Test. Mol. Biomarkers* **14,** 533–7 (2010).
123. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–5 (2014).
124. Min, L. *et al.* Computational Analysis of Missense Variants of G Protein-Coupled Receptors Involved in the Neuroendocrine Regulation of Reproduction. *Neuroendocrinology* **103,** 230–9 (2016).
125. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
126. Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* **19,** 1151–1158 (2017).
127. Li, M. M. *et al.* Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *J. Mol. Diagnostics* **19,** 4–23 (2017).
128. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508,** 469–476 (2014).
129. Andres, A. M. *et al.* Targets of Balancing Selection in the Human Genome. *Mol. Biol. Evol.* **26,** 2755–2764 (2009).
130. Gershon, E. S. & Grennan, K. S. Genetic and genomic analyses as a basis for new diagnostic nosologies.
131. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84,** 524–533 (2009).
132. Stenson, P. D. *et al.* in *Current Protocols in Bioinformatics* **Chapter 1,** Unit1.13 (John Wiley & Sons, Inc., 2012).
133. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44,** D862-8 (2016).
134. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43,** D789–D798 (2015).
135. Blake, J. A. *et al.* Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Res.* **45,** D723–D729 (2017).
136. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).
137. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* **38,** 1767–71 (2010).
138. Plagnol, V. *et al.* A robust model for read count data in exome sequencing experiments

and implications for copy number variant calling. *Bioinformatics* **28,** 2747–54 (2012).
139. Tan, R. *et al.* An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data. *Hum. Mutat.* **35,** 899–907 (2014).
140. Amendola, L. M. *et al.* Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am. J. Hum. Genet.* **98,** 1067–1076 (2016).
141. Amendola, L. M. *et al.* Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res.* **25,** 305–315 (2015).
142. Smith, L. A., Douglas, J., Braxton, A. A. & Kramer, K. Reporting Incidental Findings in Clinical Whole Exome Sequencing: Incorporation of the 2013 ACMG Recommendations into Current Practices of Genetic Counseling. *J. Genet. Couns.* **24,** 654–662 (2015).
143. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17,** 405–423 (2015).
144. Ashley, E. A. *et al.* Clinical assessment incorporating a personal genome. *Lancet* **375,** 1525–1535 (2010).
145. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15,** 565–74 (2013).
146. Xue, Y. *et al.* Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.* **91,** 1022–32 (2012).
147. Mayosi, B. M. *et al.* Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ. Cardiovasc. Genet.* **10,** e001605 (2017).
148. Matolweni, L. O. *et al.* Arrhythmogenic right ventricular cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes. *BMC Med. Genet.* **7,** 29 (2006).
149. Munclinger, M. J., Patel, J. J. & Mitha, A. S. Follow-up of patients with arrhythmogenic right ventricular cardiomyopathy dysplasia. *S. Afr. Med. J.* **90,** 61–8 (2000).
150. Kostetskii, I. *et al.* Induced Deletion of the N-Cadherin Gene in the Heart Leads to Dissolution of the Intercalated Disc Structure. *Circ. Res.* **96,** 346–354 (2005).
151. Dorn, T. *et al.* Interplay of cell–cell contacts and RhoA/MRTF-A signaling regulates cardiomyocyte identity. *EMBO J.* e98133 (2018). doi:10.15252/embj.201798133
152. Priya, R. *et al.* Feedback regulation through myosin II confers robustness on RhoA signalling at E-cadherin junctions. *Nat. Cell Biol.* **17,** 1282–93 (2015).
153. Corrado, D. *et al.* Spectrum of clinicopathologic manifestations of arrhythmogenic right ventricular cardiomyopathy/dysplasia: a multicenter study. *J. Am. Coll. Cardiol.* **30,** 1512–20 (1997).
154. Distelmaier, F., Haack, T. B., Wortmann, S. B., Mayr, J. A. & Prokisch, H. Treatable mitochondrial diseases: cofactor metabolism and beyond. *Brain* **140,** e11 (2017).
155. Enns, G. M. Pediatric mitochondrial diseases and the heart. *Curr. Opin. Pediatr.* 1 (2017). doi:10.1097/MOP.0000000000000535
156. Magner, M. *et al.* TMEM70 deficiency: long-term outcome of 48 patients. *J. Inherit. Metab. Dis.* **38,** 417–426 (2015).
157. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8,** 15824 (2017).
158. Fassone, E. & Rahman, S. Complex I deficiency: clinical features, biochemistry and molecular genetics. *J. Med. Genet.* **49,** 578–590 (2012).

159. Taylor, R. W. *et al.* Use of Whole-Exome Sequencing to Determine the Genetic Basis of Multiple Mitochondrial Respiratory Chain Complex Deficiencies. *JAMA* **312,** 68 (2014).
160. Götz, A. *et al.* Exome Sequencing Identifies Mitochondrial Alanyl-tRNA Synthetase Mutations in Infantile Mitochondrial Cardiomyopathy. *Am. J. Hum. Genet.* **88,** 635–642 (2011).
161. Ghezzi, D. *et al.* Mutations of the Mitochondrial-tRNA Modifier MTO1 Cause Hypertrophic Cardiomyopathy and Lactic Acidosis. *Am. J. Hum. Genet.* **90,** 1079–1087 (2012).
162. Kopajtich, R. *et al.* Mutations in GTPBP3 Cause a Mitochondrial Translation Defect Associated with Hypertrophic Cardiomyopathy, Lactic Acidosis, and Encephalopathy. *Am. J. Hum. Genet.* **95,** 708–720 (2014).
163. Powell, C. A. *et al.* TRMT5 Mutations Cause a Defect in Post-transcriptional Modification of Mitochondrial tRNA Associated with Multiple Respiratory-Chain Deficiencies. *Am. J. Hum. Genet.* **97,** 319–328 (2015).
164. Bhoj, E. J. *et al.* Pathologic Variants of the Mitochondrial Phosphate Carrier SLC25A3: Two New Patients and Expansion of the Cardiomyopathy/Skeletal Myopathy Phenotype With and Without Lactic Acidosis. *JIMD Rep.* **19,** 59–66 (2015).
165. Lahrouchi, N. *et al.* Exome sequencing identifies primary carnitine deficiency in a family with cardiomyopathy and sudden death. *Eur. J. Hum. Genet.* **25,** 783–787 (2017).
166. Kennedy, H. *et al.* Sudden Cardiac Death Due to Deficiency of the Mitochondrial Inorganic Pyrophosphatase PPA2. *Am. J. Hum. Genet.* **99,** 674–682 (2016).
167. Desbats, M. A., Lunardi, G., Doimo, M., Trevisson, E. & Salviati, L. Genetic bases and clinical manifestations of coenzyme Q10 (CoQ10) deficiency. *J. Inherit. Metab. Dis.* **38,** 145–156 (2015).
168. Brea-Calvo, G. *et al.* COQ4 mutations cause a broad spectrum of mitochondrial disorders associated with CoQ10 deficiency. *Am. J. Hum. Genet.* **96,** 309–17 (2015).
169. Papadopoulou, L. C. *et al.* Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene. *Nat. Genet.* **23,** 333–7 (1999).
170. Freisinger, P., Horvath, R., Macmillan, C., Peters, J. & Jaksch, M. Reversion of hypertrophic cardiomyopathy in a patient with deficiency of the mitochondrial copper binding protein Sco2: Is there a potential effect of copper? *J. Inherit. Metab. Dis.* **27,** 67–79 (2004).
171. Vernon, H. J., Sandlers, Y., McClellan, R. & Kelley, R. I. Clinical laboratory studies in Barth Syndrome. *Mol. Genet. Metab.* **112,** 143–147 (2014).
172. Jefferies, J. L. Barth syndrome. *Am. J. Med. Genet. C. Semin. Med. Genet.* **163C,** 198–205 (2013).
173. Dudek, J. *et al.* Cardiac-specific succinate dehydrogenase deficiency in Barth syndrome. *EMBO Mol. Med.* **8,** 139–54 (2016).
174. Mayr, J. A. *et al.* Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. *Am. J. Hum. Genet.* **90,** 314–20 (2012).
175. Ojala, T. *et al.* New mutation of mitochondrial DNAJC19 causing dilated and noncompaction cardiomyopathy, anemia, ataxia, and male genital anomalies. *Pediatr. Res.* **72,** 432–437 (2012).
176. Haack, T. B. *et al.* Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement. *Ann. Clin. Transl. Neurol.* **2,** 492–509 (2015).
177. Haack, T. B. *et al.* Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat. Genet.* **42,** 1131–1134 (2010).
178. Gerards, M. *et al.* Riboflavin-responsive oxidative phosphorylation complex I

deficiency caused by defective ACAD9: new function for an old gene. *Brain* **134,** 210–9 (2011).
179. Nouws, J. *et al.* Acyl-CoA Dehydrogenase 9 Is Required for the Biogenesis of Oxidative Phosphorylation Complex I. *Cell Metab.* **12,** 283–294 (2010).
180. Haack, T. B. *et al.* Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J. Med. Genet.* **49,** 277–83 (2012).
181. Calvo, S. E. *et al.* Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.* **4,** 118ra10 (2012).
182. Garone, C. *et al.* Mitochondrial encephalomyopathy due to a novel mutation in ACAD9. *JAMA Neurol.* **70,** 1177–9 (2013).
183. Nouws, J. *et al.* A Patient with Complex I Deficiency Caused by a Novel ACAD9 Mutation Not Responding to Riboflavin Treatment. doi:10.1007/8904_2013_242
184. Lagoutte-Renosi, J. *et al.* Lethal Neonatal Progression of Fetal Cardiomegaly Associated to ACAD9 Deficiency. *JIMD Rep.* **28,** 1 (2015).
185. Collet, M. *et al.* High incidence and variable clinical outcome of cardiac hypertrophy due to ACAD9 mutations in childhood. *Eur. J. Hum. Genet.* **24,** 1112–1116 (2016).
186. Kohda, M. *et al.* A Comprehensive Genomic Analysis Reveals the Genetic Landscape of Mitochondrial Respiratory Chain Complex Deficiencies. *PLOS Genet.* **12,** e1005679 (2016).
187. Leslie, N. *et al.* Neonatal multiorgan failure due to ACAD9 mutation and complex I deficiency with mitochondrial hyperplasia in liver, cardiac myocytes, skeletal muscle, and renal tubules. *Hum. Pathol.* **49,** 27–32 (2016).
188. Pronicka, E. *et al.* New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J. Transl. Med.* **14,** 174 (2016).
189. Aintablian, H. K., Narayanan, V., Belnap, N., Ramsey, K. & Grebe, T. A. An atypical presentation of ACAD9 deficiency: Diagnosis by whole exome sequencing broadens the phenotypic spectrum and alters treatment approach. *Mol. Genet. Metab. reports* **10,** 38–44 (2017).
190. Fragaki, K. *et al.* Severe defect in mitochondrial complex I assembly with mitochondrial DNA deletions in *ACAD9* -deficient mild myopathy. *Muscle Nerve* **55,** 919–922 (2017).
191. Canalis, E. & Zanotti, S. Hajdu-Cheney syndrome: a review. *Orphanet J. Rare Dis.* **9,** 200 (2014).
192. Jongmans, M. C. J. *et al.* CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J. Med. Genet.* **43,** 306–314 (2005).
193. Hennekam, R. C. M. Rubinstein–Taybi syndrome. *Eur. J. Hum. Genet.* **14,** 981–985 (2006).
194. Bottero, L., Cinalli, G., Labrune, P., Lajeunie, E. & Renier, D. Antley-Bixler syndrome. Description of two new cases and a review of the literature. *Childs. Nerv. Syst.* **13,** 275–80; discussion 281 (1997).
195. Chen, H.-J. *et al.* The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev.* **20,** 1933–1945 (2006).
196. Varjosalo, M. & Taipale, J. Hedgehog: functions and mechanisms. *Genes Dev.* **22,** 2454–2472 (2008).
197. Zhang, J. *et al.* Frs2alpha-deficiency in cardiac progenitors disrupts a subset of FGF signals required for outflow tract morphogenesis. *Development* **135,** 3611–22 (2008).
198. Park, E. J. *et al.* An FGF autocrine loop initiated in second heart field mesoderm regulates morphogenesis at the arterial pole of the heart. *Development* **135,** 3599–610 (2008).
199. Nakamura, T., Sano, M., Songyang, Z. & Schneider, M. D. A Wnt- and -catenin-

dependent pathway for mammalian cardiac myogenesis. *Proc. Natl. Acad. Sci.* **100,** 5834–5839 (2003).
200. Islas, J. F. *et al.* Transcription factors ETS2 and MESP1 transdifferentiate human dermal fibroblasts into cardiac progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13016–21 (2012).
201. Bujalka, H. *et al.* MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. *PLoS Biol.* **11,** e1001625 (2013).
202. Jain, S. *et al.* Unique phenotype in a patient with CHARGE syndrome. *Int. J. Pediatr. Endocrinol.* **2011,** 11 (2011).
203. Twigg, S. R. F. *et al.* Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. *Am. J. Hum. Genet.* **91,** 897–905 (2012).
204. Tompson, S. W. J. *et al.* Sequencing EVC and EVC2 identifies mutations in two-thirds of Ellis-van Creveld syndrome patients. *Hum. Genet.* **120,** 663–70 (2007).
205. Hills, C. B., Kochilas, L., Schimmenti, L. A. & Moller, J. H. Ellis–van Creveld Syndrome and Congenital Heart Defects: Presentation of an Additional 32 Cases. *Pediatr. Cardiol.* **32,** 977–982 (2011).
206. Yang, C., Chen, W., Chen, Y. & Jiang, J. Smoothened transduces Hedgehog signal by forming a complex with Evc/Evc2. *Cell Res.* **22,** 1593–1604 (2012).
207. Kantarci, S. *et al.* Mutations in LRP2, which encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. *Nat. Genet.* **39,** 957–9 (2007).
208. Pober, B. R., Longoni, M. & Noonan, K. M. A review of Donnai-Barrow and facio-oculo-acoustico-renal (DB/FOAR) syndrome: Clinical features and differential diagnosis. *Birth Defects Res. Part A Clin. Mol. Teratol.* **85,** 76–81 (2009).
209. Christ, A. *et al.* LRP2 Is an Auxiliary SHH Receptor Required to Condition the Forebrain Ventral Midline for Inductive Signals. *Dev. Cell* **22,** 268–278 (2012).
210. Tunçbilek, E. & Alanay, Y. Congenital contractural arachnodactyly (Beals syndrome). *Orphanet J. Rare Dis.* **1,** 20 (2006).
211. Alpay, F., Gül, D., Lenk, M. K. & Oğur, G. Severe intrauterine growth retardation, aged facial appearance, and congenital heart disease in a newborn with Johanson-Blizzard syndrome. *Pediatr. Cardiol.* **21,** 389–90 (2000).
212. Toriello, H. V, Franco, B., Bruel, A.-L. & Thauvin-Robinet, C. *Oral-Facial-Digital Syndrome Type I. GeneReviews(®)* (1993).
213. Parisi, M. & Glass, I. *Joubert Syndrome. GeneReviews(®)* (1993).
214. Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nat. Rev. Genet.* **11,** 331–44 (2010).
215. Ackerman, C. *et al.* An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am. J. Hum. Genet.* **91,** 646–59 (2012).
216. Happle, R., Effendy, I., Megahed, M., Orlow, S. J. & Küster, W. CHILD syndrome in a boy. *Am. J. Med. Genet.* **62,** 192–194 (1996).
217. König, A. *et al.* A novel missense mutation of NSDHL in an unusual case of CHILD syndrome showing bilateral, almost symmetric involvement. *J. Am. Acad. Dermatol.* **46,** 594–6 (2002).
218. Falek, A., Heath, C. W., Ebbin, A. J. & McLean, W. R. Unilateral limb and skin deformities with congenital heart disease in two siblings: a lethal syndrome. *J. Pediatr.* **73,** 910–3 (1968).
219. Tang, T. T. & McCreadie, S. R. Q--congenital hemidysplasia with ichthyosis. *Birth Defects Orig. Artic. Ser.* **10,** 257–61 (1974).

220. Hunkapiller, J., Singla, V., Seol, A. & Reiter, J. F. The ciliogenic protein Oral-Facial-Digital 1 regulates the neuronal differentiation of embryonic stem cells. *Stem Cells Dev.* **20,** 831–41 (2011).
221. Tanaka, K., Kato, A., Angelocci, C., Watanabe, M. & Kato, Y. A potential molecular pathogenesis of cardiac/laterality defects in Oculo-Facio-Cardio-Dental syndrome. *Dev. Biol.* **387,** 28–36 (2014).
222. Seo, S. *et al.* A Novel Protein LZTFL1 Regulates Ciliary Trafficking of the BBSome and Smoothened. *PLoS Genet.* **7,** e1002358 (2011).
223. Bang, M. L. *et al.* Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. *J. Cell Biol.* **153,** 413–27 (2001).
224. Reed, G. J., Boczek, N. J., Etheridge, S. P. & Ackerman, M. J. CALM3 mutation associated with long QT syndrome. *Hear. Rhythm* **12,** 419–422 (2015).
225. Chaix, M.-A. *et al.* Novel CALM3 mutations in pediatric long QT syndrome patients support a CALM3-specific calmodulinopathy. *Hear. case reports* **2,** 250–254 (2016).
226. Turkowski, K. L., Tester, D. J., Bos, J. M., Haugaa, K. H. & Ackerman, M. J. Whole exome sequencing with genomic triangulation implicates CDH2-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy. *Congenit. Heart Dis.* **12,** 226–235 (2017).
227. Rampazzo, A., Calore, M., van Hengel, J. & van Roy, F. Intercalated Discs and Arrhythmogenic Cardiomyopathy. *Circ. Cardiovasc. Genet.* **7,** 930–940 (2014).
228. Sheikh, F., Ross, R. S. & Chen, J. Cell-cell connection to cardiac disease. *Trends Cardiovasc. Med.* **19,** 182–90 (2009).
229. Knudsen, K. A., Soler, A. P., Johnson, K. R. & Wheelock, M. J. Interaction of alpha-actinin with the cadherin/catenin cell-cell adhesion complex via alpha-catenin. *J. Cell Biol.* **130,** 67–77 (1995).
230. BORRMANN, C. *et al.* The area composita of adhering junctions connecting heart muscle cells of vertebrates. II. Colocalizations of desmosomal and fascia adhaerens molecules in the intercalated disk. *Eur. J. Cell Biol.* **85,** 469–485 (2006).
231. Rex, C. S. *et al.* Myosin IIb Regulates Actin Dynamics during Synaptic Plasticity and Memory Formation. *Neuron* **67,** 603–617 (2010).
232. McKenna, W. J. *et al.* Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. *Br. Heart J.* **71,** 215–8 (1994).
233. Sen-Chowdhry, S. *et al.* Left-Dominant Arrhythmogenic Cardiomyopathy. *J. Am. Coll. Cardiol.* **52,** 2175–2187 (2008).
234. López-Ayala, J. M. *et al.* Desmoplakin truncations and arrhythmogenic left ventricular cardiomyopathy: characterizing a phenotype. *EP Eur.* **16,** 1838–1846 (2014).
235. Brunel-Guitton, C., Levtova, A. & Sasarman, F. Mitochondrial Diseases and Cardiomyopathies. *Can. J. Cardiol.* **31,** 1360–1376 (2015).
236. El-Hattab, A. W. & Scaglia, F. Mitochondrial Cardiomyopathies. *Front. Cardiovasc. Med.* **3,** 25 (2016).
237. Haack, T. B. *et al.* ELAC2 Mutations Cause a Mitochondrial RNA Processing Defect Associated with Hypertrophic Cardiomyopathy. *Am. J. Hum. Genet.* **93,** 211–223 (2013).
238. Akawi, N. A. *et al.* A homozygous splicing mutation in ELAC2 suggests phenotypic variability including intellectual disability with minimal cardiac involvement. *Orphanet J. Rare Dis.* **11,** 139 (2016).
239. Limpitikul, W. B. *et al.* A Precision Medicine Approach to the Rescue of Function on

Malignant Calmodulinopathic Long-QT Syndrome. *Circ. Res.* **120,** 39–48 (2017).
240. Boczek, N. J. *et al.* Spectrum and Prevalence of CALM1-, CALM2-, and CALM3-Encoded Calmodulin Variants in Long QT Syndrome and Functional Characterization of a Novel Long QT Syndrome-Associated Calmodulin Missense Variant, E141G. *Circ. Cardiovasc. Genet.* **9,** 136–146 (2016).
241. Chaix, M.-A. *et al.* Novel CALM3 mutations in pediatric long QT syndrome patients support a CALM3-specific calmodulinopathy. *Hear. case reports* **2,** 250–254 (2016).
242. Barc, J. *et al.* Screening for copy number variation in genes associated with the long QT syndrome: clinical relevance. *J. Am. Coll. Cardiol.* **57,** 40–7 (2011).
243. Seidelmann, S. B. *et al.* Application of Whole Exome Sequencing in the Clinical Diagnosis and Management of Inherited Cardiovascular Diseases in AdultsCLINICAL PERSPECTIVE. *Circ. Cardiovasc. Genet.* **10,** (2017).
244. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci.* **106,** 19096–19101 (2009).
245. Wortmann, S. B., Koolen, D. A., Smeitink, J. A., van den Heuvel, L. & Rodenburg, R. J. Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J. Inherit. Metab. Dis.* **38,** 437–43 (2015).
246. Lawrence, L. *et al.* The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience. *Genet. Med.* **16,** 741–750 (2014).
247. Lee, H. *et al.* Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA* **312,** 1880 (2014).
248. Rocchetti, M. *et al.* Elucidating arrhythmogenic mechanisms of long-QT syndrome CALM1-F142L mutation in patient-specific induced pluripotent stem cell-derived cardiomyocytes. *Cardiovasc. Res.* **113,** 531–541 (2017).
249. Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9,** eaal5209 (2017).
250. Repp BR*, Mastantuono E* *et al.* Clinical, biochemical and genetic spectrum of 70 patients with ACAD9 deficiency: is riboflavin supplementation effective? *Orphanet Journal of Rare Disease*s; in Press (OJRD-D-17-00384). (2018).

Appendix

6.1 List of figures and tables

6.2 List of publications

1. **Repp BR*, Mastantuono E***, Alston CL, Schiff M, Haack T, Rötig A, Ardissone A, Lombes A, Catarino C, Diodato D, Schottmann G, Poulton J, Burlina A, Jonckheere A, Munnich A, Rolinski B, Ghezzi D, Rokicki D, Wellesley D, Martinelli D, Lamantea E, Ostergaard E, Ewa P, Pierre G, Smeets HJM, Ika W, Scurr I, de Coo IFM, Moroni I, Smet J, Mayr JA, de Meirleir L, Shuelke M, Zeviani M, Morscher RJ, McFarland R, Seneca S, Klopstock T, Meitinger T, Wieland T, Strom T, Herberg U, Ahting U, Sperl W, Nassogne MC, Freisinger P, Van Coster R, Strecker V, Taylor RW, Häberle J, Jerry V, Prokisch H, Wortmann S. **(*equal contribution)** *Clinical, biochemical and genetic spectrum of 67 patients with ACAD9 deficiency: is riboflavin supplementation effective?* Orphanet Journal of Rare Diseases; in revision (OJRD-D-17-00384)

2. Dorn T, Haas J, Parrotta EI, Zawada D, Ayetey H, Santamaria G, Iop L, **Mastantuono E**, Sinnecker D, Goedel1 A, Dirschinger RJ, My I, Laue1 S, Baarlink C, Graf E, Hinke R, Cuda G, Kääb S, Grace AA, Grosse R, Kupatt C, Meitinger T, Smith AG, Moretti A, Karl-Ludwig Laugwitz *Interplay of cell-cell contacts and RhoA/MRTF-A signalling regulates cardiomyocytic identity.* The EMBO Journal; in revision.

3. **Mastantuono E,** Wolf CM, Prokisch H. *Genetics on mitochondrial cardiomyopathy.* In Springer - "Genetic Causes of Cardiovascular Disease". in revision (book chapter)

4. Feichtinger RG, Oláhová M, Kishita Y, Garone C, Kremer LS, Yagi M, Uchiumi T, Jourdain AA, Thompson K, D'Souza AR, Kopajtich R, Alston CL, Koch J, Sperl W, **Mastantuono E**, Strom TM, Wortmann SB, Meitinger T, Pierre G, Chinnery PF, Chrzanowska-Lightowlers ZM, Lightowlers RN, DiMauro S, Calvo SE, Mootha VK, Moggio M, Sciacco M, Comi GP, Ronchi D, Murayama K, Ohtake A, Rebelo-Guiomar P, Kohda M, Kang D, Mayr JA, Taylor RW, Okazaki Y, Minczuk M, Prokisch H. *Biallelic C1QBP Mutations Cause Severe Neonatal-, Childhood-, or Later-Onset Cardiomyopathy Associated with Combined Respiratory-Chain Deficiencies.* Am J Hum Genet. 2017 Oct 5;101(4):525-538. doi: 10.1016/j.ajhg.2017.08.015. Epub 2017 Sep 21.

5. Mayosi BM, Fish M, Shaboodien G, **Mastantuono E**, Kraus S, Wieland T, Kotta MC, Chin A, Laing N, Ntusi NB, Chong M, Horsfall C, Pimstone SN, Gentilini D, Parati G, Strom TM, Meitinger T, Pare G, Schwartz PJ, Crotti L. *Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy.* Circ Cardiovasc Genet. 2017 Apr;10(2). pii: e001605. doi: 10.1161/CIRCGENETICS.116.001605.

6. Crotti L, Lahtinen AM, Spazzolini C, **Mastantuono E**, Cristina Monti M, Morassutto C, Parati G, Heradien M, Goosen A, Lichtner P, Meitinger T, Brink PA, Kontula K, Swan H, Schwartz PJ. *Response by Crotti et al to Letter Regarding Article, "Genetic Modifiers for the Long-QT Syndrome: How Important Is the Role of Variants in the 3' Untranslated Region of KCNQ1?"* Circ Cardiovasc Genet. 2016 Dec;9(6):581-582

7. Crotti L, Lahtinen AM, Spazzolini C, **Mastantuono E**, Monti MC, Morassutto C, Parati G, Heradien M, Goosen A, Lichtner P, Meitinger T, Brink PA, Kontula K, Swan H, Schwartz PJ. *Genetic Modifiers for the Long-QT Syndrome: How Important Is the Role of Variants in the 3' Untranslated Region of KCNQ1?* Circ Cardiovasc Genet. 2016 Aug;9(4):330-9. doi: 10.1161/CIRCGENETICS.116.001419.

8. Itoh H, Crotti L, Aiba T, Spazzolini C, Denjoy I, Fressart V, Hayashi K, Nakajima T, Ohno S, Makiyama T, Wu J, Hasegawa K, **Mastantuono E**, Dagradi F, Pedrazzini M, Yamagishi M, Berthet M, Murakami Y, Shimizu W, Guicheney P, Schwartz PJ, Horie M. *The genetics underlying acquired long QT syndrome: impact for genetic screening.* Eur Heart J. 2016 May 7;37(18):1456-64. doi: 10.1093/eurheartj/ehv695. Epub 2015 Dec 28

9. Makita N, Yagihara N, Crotti L, Johnson CN, Beckmann BM, Roh MS, Shigemizu D, Lichtner P, Ishikawa T, Aiba T, Homfray T, Behr ER, Klug D, Denjoy I, **Mastantuono E**, Theisen D, Tsunoda T, Satake W, Toda T, Nakagawa H, Tsuji Y, Tsuchiya T, Yamamoto H, Miyamoto Y, Endo N, Kimura A, Ozaki K, Motomura H, Suda K, Tanaka T, Schwartz PJ, Meitinger T, Kääb S, Guicheney P, Shimizu W, Bhuiyan ZA, Watanabe H, Chazin WJ, George AL Jr. *Novel calmodulin mutations associated with congenital arrhythmia susceptibility*. Circ Cardiovasc Genet. 2014 Aug;7(4):466-74. doi:10.1161/CIRCGENETICS.113.000459. Epub 2014 Jun 10