Characterization of Genetic Causes of Pediatric Cardiomyopathies – Analysis of PRDM16 as a Major Disease Gene

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Christopher Herbst
2020

This work was carried out in the period of 01.04.2016 to 31.12.2019 under the supervision of PD. Dr. Sabine Klaassen at the Experimental and Clinical Research Center (ECRC) of the MDC and Charité Berlin, Germany.

1st Reviewer: PD Dr. Sabine Klaassen
Genetics of Congenital Heart Disease
Experimental and Clinical Research Center (ECRC) of the MDC and Charité Berlin
Lindenberger Weg 80, 13125 Berlin, Germany
sabine.klaassen@mdc-berlin.de

2nd Reviewer: Prof. Dr. Sigmar Stricker
Biochemistry and Genetics
Institute of Biochemistry
Department of Biology, Chemistry, Pharmacy
Freie Universität Berlin
Thielallee 63, 14295 Berlin, Germany
sigmar.stricker@fu-berlin.de

Date of Defense: 06.07.2020

Table of contents

Summary

Primary cardiomyopathies (CMPs) are genetically heterogeneous disorders with a large number of disease-causing genes. The underlying genetic mechanisms and early pathological events of children with primary CMP are poorly characterized. This work aimed to identify genetic causes of childhood CMP, and to understand the development of early-onset and severe heart muscle disease. Mutation of *PR/SET domain 16 (PRDM16)* was shown to be associated with two types of CMP, dilated CMP (DCM) and left ventricular non-compaction CMP (LVNC). *PRDM16* encodes for a transcriptional regulator with a previously unknown function in the heart.

Eighty unrelated index patients with pediatric primary CMP underwent genetic testing with a panel-based next generation sequencing (NGS) approach of 89 genes. Adverse events such as heart transplantation (HTX) and death were higher for individuals affected by DCM and restrictive CMP than for other types of CMP. Patients with a higher number of variants of interest (VOI) had an increased risk to experience an adverse event. At least one pathogenic or likely pathogenic variant was identified in 30/80 (38%) index patients. In all CMP subgroups, patients carried most frequently VOI in sarcomere genes suggesting them as a major contributor in pediatric primary CMP. Protein and transcript level analysis on heart biopsies from individuals with homozygous mutation of *troponin I3, cardiac type (TNNI3)* revealed that the TNNI3 protein was absent and associated with upregulation of the fetal isoform troponin I1, slow skeletal type (TNNI1). These observations support the clinical importance of sarcomeric mutation in primary CMP. *TNNI3* is the third most important disease gene in pediatric CMP.

The mutational spectrum of *PRDM16* in an extended cohort of 285 patients with different types of heart muscle diseases revealed 16 VOI in total. These alterations occurred in six different cardiac diseases and comprised missense, splice site, frame shift, and stop gain variants. Pathogenic *PRDM16* variants were only found in LVNC patients, confirming the important role of PRDM16 in LVNC pathogenesis. By prediction of posttranslational modification sites, structural features and functional interaction sites of the PRDM16 protein were identified. Functional and biochemical characterization of PRDM16 showed different subcellular localization of CMP variants and the influence of CMP variants on the protein stability of PRDM16. Patient specific CMP variants altered the cellular distribution of PRDM16, with condensed signals differing between the cytoplasm and the nucleus.

In conclusion, the spectrum and number of genetic alteration in pediatric primary CMP determine clinical outcome. *PDRM16* is a frequent CMP gene and translational approaches for potential therapies need to be developed.

Zusammenfassung

Primäre Kardiomyopathien (CMP) sind heterogene, genetisch determinierte Erkrankungen mit einer Vielzahl von krankheitsverursachenden Genen. Die zugrundeliegenden genetischen Mechanismen und früh vorhandenen pathologischen Vorgänge sind bei Kindern mit primärer CMP wenig untersucht. Diese Arbeit beschäftigt sich mit der Identifizierung von genetischen Ursachen der Kardiomyopathie des Kindesalters, insbesondere mit dem Auftreten von früh einsetzenden und schweren Krankheitsverläufen. Genetische Veränderungen von *PR/SET domain 16* (*PRDM16*) führen zu zwei verschiedenen Arten von Kardiomyopathien, der dilatativen CMP (DCM) und der linksventrikulären Noncompaction CMP (LVNC). *PRDM16* kodiert für einen transkriptionellen Regulator mit bis dahin unbekannter Funktion im Herzen.

Achtzig nicht verwandte Indexpatienten mit pädiatrischer primärer CMP wurden mit einem panel-basierten next generation sequencing (NGS) Ansatz einer genetischen Testung auf Veränderungen in 89 Genen untersucht. Unerwünschte Ereignisse wie Herztransplantation und Tod waren bei Individuen mit DCM und restriktiver CMP höher als bei anderen CMP. Patienten mit einer höheren Anzahl von genetischen Varianten von Interesse (VOI) hatten ein höheres Risiko für ein unerwünschtes Ereignis. Bei 30/80 (38%) Indexpatienten wurde mindestens eine wahrscheinlich pathogene oder pathogene Variante identifiziert. Bei allen CMP Subtypen waren VOI in Sarkomergenen am häufigsten und sind somit für die Entstehung der pädiatrischen primären CMP hauptverantwortlich. Auf Protein- und mRNA Ebene von Herzbiopsien von Individuen mit homozygoter Variante konnte gezeigt werden, dass TNNI3 nicht nachweisbar war und stattdessen eine Hochregulation der fetalen Isoform TNNI1 stattfand. Diese Beobachtungen unterstützen die klinische Bedeutung von genetischen Veränderungen in Sarkomergenen bei primärer CMP. TNNI3 ist das dritthäufigste Krankheitsgen bei pädiatrischer CMP.

Das Spektrum der VOI bei *PRDM16* in einer erweiterten Kohorte von 285 Patienten mit verschiedenen Herzmuskelerkrankungen ergab insgesamt 16 verschiedene VOI. Diese Veränderungen traten bei sechs verschiedenen Herzerkrankungen auf und umfasste Punkt-Varianten, Spleiß-Varianten, Frameshift-Varianten und Stopp-Varianten. Pathogene *PRDM16* Varianten wurden nur bei Patienten mit LVNC gefunden, was die Bedeutung von PRDM16 in der Entstehung der LVNC unterstreicht. Durch die Vorhersage von Stellen für posttranslationale Modifikationen wurden strukturelle Eigenschaften und funktionelle Interaktionsstellen des PRDM16 Proteins identifiziert. Die funktionelle und biochemische Charakterisierung von PRDM16 zeigte unterschiedliche subzelluläre Lokalisationen von CMP Varianten und auch die Stabilität vom PRDM16 wurde unterschiedlich durch die jeweiligen CMP Varianten beeinflusst. Patientenspezifische CMP Varianten veränderten die zelluläre

Verteilung von PRDM16 dahingehend, dass kondensierte Signale zwischen Zellkern und Zytoplasma jeweils differierten.

Abschließend fanden wir, dass das Spektrum und die Anzahl der genetischen Veränderungen bei pädiatrischer primärer CMP für den klinischen Ausgang verantwortlich sind. *PRDM16* ist ein häufiges Krankheitsgen für CMP und translationale Ansätze für mögliche Therapien sollten entwickelt werden.

List of Figures

List of Tables

Abbreviations

aa – Amino acid
ACGV – Atlas of cardiac genetic variation
AICD – Automatic implantable cardioverter defibrillator
ARVC – Arrhythmic right ventricular cardiomyopathy
BCA – Bicinchoninic acid assay
BSA – Bovine serum albumin
BVAD – Biventricular assist device
C/EBPß – CCAAT/enhancer binding protein beta
CHX – Cycloheximide
CMP – Cardiomyopathy
CMV – Cytomegalovirus
cp – Cytoplasm
CTBP – C-terminal binding protein
DCM – Dilated cardiomyopathy
DHZB – Deutsches Herzzentrum Berlin
DMEM – Dulbecco's modified eagle medium
DNA – Deoxyribonucleic acid
ds – Double stranded
E. coli – Escherichia coli
ECM – Extracellular matrix
ECMO – Extracorporeal membrane oxygenation
EGFP – Enhanced green fluorescent protein
EHMT1 – Euchromatic histone lysine methyltransferase 1
EVS – Exome variant server
f – Female
FBS – Fetal bovine serum
FLNC – Filamin C
FS – Fractional shortening
gnomAD – Genome aggregation database
GRCh – Genome reference consortium human 37
HCM – Hypertrophic cardiomyopathy
HEK293 – Human embryoid kidney 293
HGMD – Human gene mutation database
HRP – Horse radish peroxidase
HSF – Human splicing finder

HTX – Heart transplantation

IVSd – Interventricular septum thickness at end diastole

LBe – Likely benign

LB – Lysogeny broth

LP – Likely pathogenic

LVEDD – Left ventricular end diastolic diameter

LV-EF – Left ventricular ejection fraction

LVESD – Left ventricular end systolic diameter

LVNC – Left ventricular non-compaction cardiomyopathy

m – Male

MACE – Major adverse cardiovascular event

MAF – Minor allele frequency

MDC – Max Delbrück Center for Molecular Medicine in the Helmholtz Association

MEM – Minimum essential media

Midiprep – Midi preparation

Miniprep – Mini preparation

MIP – Maximum intensity projection

MKL2 – Myocardin-related transcription factor B

n – Nucleus

n.a. – Not applicable

NGS – Next generation sequencing

NP-40 – Nonyl phenoxypolyethoxylethanol

P – Pathogenic

PA – Polyacrylamide

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

PFA – Paraformaldehyde

PGC1 - PPAR-gamma coactivator 1

PPAR – Peroxisome proliferator-activated receptor

PPOLYPHEN-2 – Polymorphism phenotyping v2

PTM – Post-translational modification

PVDF – Polyvinylidene fluoride

qPCR – Quantitative PCR

RCM – Restrictive cardiomyopathy

RIPA – Radioimmunoprecipitation assay buffer

rpm – Rounds per minute

RT – Room temperature

SD – Standard deviation

SDS – Sodium dodecyl sulfate

SIFT – Sorting tolerant from intolerant

SKI – SKI proto-oncogene

SUMO – Small ubiquitin-related modifier

TBS – Tris buffered saline

TBS-T – TBS with Tween20

UCP1 – Mitochondrial brown fat uncoupling protein 1

VAD – Ventricular assist device

VOI – Variant of interest

VUS – Variant of uncertain significance

WHO – World health organization

ZNF516 – Zinc finger protein 516

Abbr. of 89 cardiomyopathy-relevant genes in Table S 1

1 Introduction

In 2016 according to World Health Organization (WHO) data cardiovascular disease represented with 31% the number one cause of death worldwide (https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29). The most frequent diseases in this group are ischemic heart disease, where restricted blood supply leads to adverse events in the heart, strokes, where limited blood flow affects the brain, and hypertensive heart disease caused by high blood pressure. Nevertheless, cardiomyopathy (CMP), myocarditis and endocarditis pooled into one category by the WHO, are included in this group of life-threatening diseases and constitute to 374000 casualties worldwide each year (0.7%, estimates for 2016). This category comprises malformations of the heart due to congenital alterations or inflammatory events. Therefore, these numbers present a necessity to investigate mechanisms leading to these diseases to reduce death rates. Especially congenital forms of cardiomyopathies can be related to genetic alterations and are suited for genetic screening. While mechanisms leading to cardiomyopathy in adults are well-established, the underlying genetic mechanisms, early pathological events and other disease promoting factors are poorly characterized in children. Thus, in this work a pediatric cohort consisting of patients affected by primary, non-syndromic cardiomyopathies was tested for its genetic background. Furthermore the function of many proteins involved in cardiomyopathy relevant processes is well understood, but the impact of small genetic alterations is difficult to predict and requires further investigation. Additionally to well-known cardiomyopathy genes like *myosin heavy chain 7* (*MYH7*) and *myosin binding protein C, cardiac* (*MYBPC3*) genetic testing led to newly discovered disease genes and approximately 100 different genes are associated so far with cardiomyopathy (R. E. Hershberger, Hedges, & Morales, 2013). Many of these genes are poorly characterized in the cardiac setting. Therefore, another purpose of this work was to gain functional knowledge of the involvement of the transcriptional regulator PR/SET domain 16 (PRDM16) for the progression of cardiomyopathies due to small variants.

1.1 Classification of cardiomyopathies and prevalence

The word cardiomyopathy originates from the Greek language and means disease of the heart muscle (καρδίᾱ - heart, mῦs – muscle, πάθος – suffering, disease). It is characterized by complex malformations of the myocardium and manifests itself as abnormalities in the structure and function of the heart (Burke, Cook, Seidman, & Seidman, 2016). Especially resulting in arrhythmia, heart insufficiency and hypoxia can lead to adverse events and death. Cardiomyopathies are distinguished into primary and secondary. Secondary cardiomyopathies are either part of systemic diseases or rather consequences of these for example Fabry's

disease (metabolic) or Diabetes mellitus (endocrine). Primary cardiomyopathies are defined as predominantly affecting the myocardium in the clinically relevant disease processes. Furthermore, primary cardiomyopathies are distributed into the categories genetic/congenital, acquired and mixed. Acquired diseases of the heart muscle can be caused by environmental factors like stress or infection of the myocardium through bacteria and viruses (Takotsubo cardiomyopathy and myocarditis). Further risk factors such as overweight, pregnancy or drug and alcohol abuse may cause acquired cardiomyopathy or at least contribute to the congenital forms. Genetic cardiomyopathies however are caused by alterations of the genome. Some of the congenital forms of cardiomyopathies can also be acquired through environmental factors and therefore represent a mixed form (Maron et al., 2006; Cecchi, Tomberli, & Olivotto, 2012). Such case may occur when a myocarditis develops into a more severe dilated cardiomyopathy. Historically the term idiopathic was used due to unknown mechanisms as to why cardiomyopathies develop and progress. Ongoing investigation relates genetic burdens to the initiation and progression of primary cardiomyopathies. These genetic forms are further distributed into the specific cardiomyopathy phenotypes of dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), arrhythmogenic right ventricular cardiomyopathy (ARVC, due to involvement of both ventricles in new publications referred to as arrhythmogenic ventricular cardiomyopathy - AVC), left ventricular noncompaction cardiomyopathy (LVNC) and ion channelopathies. The subtypes of cardiomyopathy will be further described in the following sections. Because ion channels can also lead to other affected organs and syndromic diseases, they are not further described or discussed in this work.

1.1.1 Dilated cardiomyopathy

Dilated cardiomyopathy is the most common cardiomyopathy form (McKenna, Maron, & Thiene, 2017; Herkert et al., 2018). DCM is characterized by the dilatation of the left ventricle. Additionally global systolic dysfunction has to be observed leading to an ejection fraction smaller than 50%. For the congenital form further factors like coronary artery disease, systemic hypertension or inflammatory heart disease have to be absent. Wall thinning is believed to contribute to the contractile deficiencies. The right ventricle can be affected similarly. These physiological features can manifest in ventricular heart failure (right-, left- and biventricular), arrhythmias, syncope or sudden death (Lipshultz et al., 2019). The prevalence of DCM is estimated at 1:250 to 1:500 in adults and uncommon in children with an accumulation of cases with an age of younger than one year (McKenna, Maron, & Thiene, 2017; Lipshultz et al., 2003).

1.1.2 Hypertrophic cardiomyopathy

Hypertrophic cardiomyopathy is characterized by an enlargement of the septum or ventricular wall, whereas the right ventricles are rarely affected. The hypertrophy can occur in confined segments. Due to wall thickening tachycardia, arrhythmia and sudden cardiac death can occur (often in undiagnosed athletes with HCM through burden of the heart, Cecchi, Tomberli, & Olivotto, 2012). Therefore, implantation of pacemakers are common for advanced HCM additionally to pharmaceutical treatment. Otherwise many HCM patients are asymptomatic or show only mild signs of the disease. The onset of disease is often with advanced age and thus uncommon in children, although often occurring in adults with an estimated frequency of 1:250/500 (Semsarian, Ingles, Maron, & Maron, 2015; McKenna, Maron, & Thiene, 2017). Still, HCM is believed to be rarer than DCM and the second most frequent cardiomyopathy in children (R. E. Hershberger, Hedges, & Morales, 2013; Lee et al., 2017).

1.1.3 Restrictive cardiomyopathy

Restrictive cardiomyopathy is a rare phenotype and characterized by diastolic impairment or abnormal compliance often of both ventricles. The ventricles therefore seem rigid or restricted and as a result the atria are often affected by enlargement, whereas the ventricles are presented normal (not dilated or hypertrophic, normal wall thickness). RCM can result in heart failure, arrhythmias or sudden death and conduction abnormalities and atrial/ventricular arrhythmias as first signs of the disease (M. A. Walsh et al., 2012). As mentioned, RCM is uncommon in adults and children (McKenna, Maron, & Thiene, 2017; Lee et al., 2017).

1.1.4 Arrhythmogenic right ventricular cardiomyopathy

Arrhythmogenic right ventricular cardiomyopathy is also rare and only restricted data of population data is available for this CMP. Nevertheless, a frequency of 1:2000/5000 is estimated and rarer in children (Marcus et al., 2010; Nava et al., 2000; Lee et al., 2017). The disease is characterized by fibrofatty replacement of the myocardium in the ventricles and thus restricted conduction. Ventricular arrhythmias are the consequence and atrial fibrillation can occur (Cecchi, Tomberli, & Olivotto, 2012; Corrado et al., 2015).

1.1.5 Left ventricular noncompaction cardiomyopathy

Left ventricular noncompaction cardiomyopathy is characterized by “spongy” and therefore noncompacted appearance of the left ventricular myocardium due to massive trabeculation and the formation of deep intratrabecular recesses (Oechslin & Jenni, 2011; Lipshultz et al.,

2019). Typical symptoms are palpitations, non-specific chest pain, heart failure symptoms and arrythmias. LVNC is often observed in combination with other cardiomyopathies and without or mild symptoms (Maron et al., 2006). Therefore it is highly debated, if LVNC is a distinct cardiomyopathy leading to adverse events or only prominent with other diseases and thus not an isolated cardiomyopathy. The prevalence of LVNC cases is difficult to estimate, because available large population reports are lacking and diagnosis criteria leading to adverse events are highly discussed. Additionally, it is unclear, if LVNC is a developmental condition or can also be acquired in advanced age (Tian et al., 2017; Kodo et al., 2016).

1.2 Genes causing cardiomyopathies

In contrast to acquired forms of cardiomyopathy, congenital CMPs are caused by genetic alterations, predominantly small changes leading to amino acid changes at one position or modifying the reading frame and therefore resulting in a frameshift. Meanwhile a wide spectrum of genes was associated with this category of disease.

1.2.1 Functional groups of cardiomyopathy-related genes

Only for HCM over 2000 variants have been associated with the onset of this cardiomyopathy detected in over 40 genes (Bondue et al., 2018; Cirino et al., 2017). These findings show, that CMPs are heterogenous diseases considering that DCM is associated with considerably more cardiac genes. Thereby, variation in genes are not entirely specific for one phenotype and mechanisms how different variants in the same gene result in different cardiac phenotypes are hardly understood (R. E. Hershberger et al., 2018; Burke, Cook, Seidman, & Seidman, 2016; Pugh et al., 2014). Genes investigated regarding the onset and progression of CMPs can be classified in functional groups and a set of major genes is presented in Figure 1. Genes included in the functional group of the sarcomere are detected in CMP cases with the highest rate. The sarcomere is the functional unit of the muscle and generates the contractile force for movement due to the interaction of myosin and actin filaments. Therefore, impairment of this system leads to compensative remodeling of the heart and thus to cardiomyopathies (Dadson, Hauck, & Billia, 2017). Alteration of proteins in the Z-disc, the complex scaffold separating two sarcomeres, often leads to structural modifications and again adverse remodeling of the heart (Towbin, 2014). Defects in proteins of desmosomes, cell-cell contacts in cardiomyocytes, can result in decreased capability to withstand mechanical strains in the heart and mitochondrial proteins may influence metabolic features of cardiomyocytes (Lipshultz et al., 2019).
Further classification categorizes CMP-related genes into cytoskeleton (dystrophin-sarcoglycan complex), gene expression leading to impaired transcription of proteins, nuclear

membrane and a few others (Burke, Cook, Seidman, & Seidman, 2016). Due to divergence of genetic burden suggestions have been phrased to change the classification from morphological phenotypes to molecular genetic defect (sarcomeric cardiomyopathy, cell junction cardiomyopathy, ion channel cardiomyopathy, cytoskeletal cardiomyopathy, etc., Thiene, Corrado, & Basso, 2004). Either way, newly discovered genes associated with CMP are discovered repeatedly like *PRDM16* (Arndt et al., 2013) and diagnostical criteria have to be as distinct as possible to choose the clinical treatment accordingly.

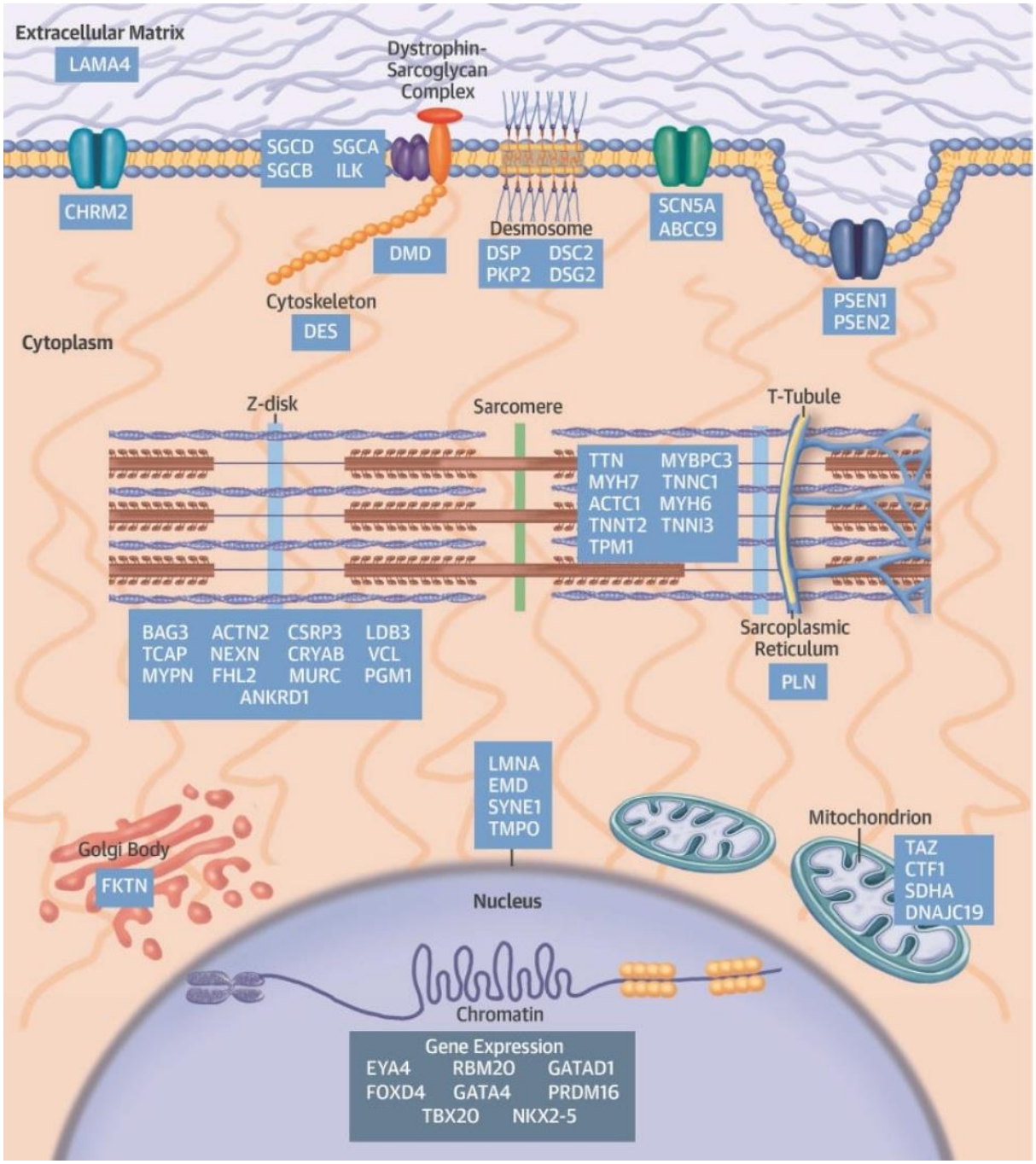


Figure 1: Spectrum of genes associated with CMPs in different cell compartments. Genes are associated with the corresponding compartments in a schematic cell model (**Burke, Cook, Seidman, & Seidman, 2016**).

1.2.2 Abundance of variants in genes causing cardiomyopathy

For the cardiomyopathies DCM, HCM, RCM, ARVC and LVNC distinct gene spectra were detected. A former study suggested, that HCM is a force generating disease of the sarcomere and DCM a force transmitting disease of the mostly the cytoskeleton, whereas ARVC is caused by defects in desmosome and therefore a cell junction disease (McKenna, Maron, & Thiene, 2017). The occurrence of variants in sarcomeric genes in patients with DCM, RCM and LVNC contradicts the afore mentioned suggestion and presents the difficulty to genetically define the different cardiac phenotypes (Bollen & van der Velden, 2017). One precise example is *MYH7*, a major factor for the contraction of sarcomeres. Variants in this gene were detected in more than one type of CMP. Therefore, variants in different functional domains of a protein may contribute differently to the phenotype.

Nevertheless, tendencies can be detected and in HCM cases mainly sarcomere genes are observed (Ingles et al., 2019; Rupp et al., 2019). The majority of variants is observed in the genes *MYBPC3* and *MYH7* (19% and 14.2%), whereas variants in other sarcomere genes are considerably less frequent (R. Walsh et al., 2017). Other non-sarcomeric genes associated with HCM are *galactosidase alpha* (*GLA*), *protein kinase AMP-activated non-catalytic subunit gamma 2* (*PRKAG2*) and *lysosomal associated membrane protein 2* (*LAMP2*) but may present the cardiomyopathy as a secondary symptom (R. E. Hershberger et al., 2018).

DCM is associated with a heterogenous spectrum of genes. Most frequently with 15-25% variants were detected in *titin* (*TTN*, Akinrinade et al., 2019, Schafer et al., 2017). Thereby, only truncating variants were considered for analysis. TTN is the longest known protein with a number of alternatively spliced transcripts. Due to its challenging length it is difficult to predict the effect of single amino acid or short in-frame changes. A cardiac-specific transcript has been analyzed to predict the expression of truncating variants (Roberts et al., 2015). Other less frequent but statistically enriched genes compared to healthy individuals were *MYH7* (5.3%, sarcomere), *lamin A/C* (*LMNA*, 4.4%, nuclear envelope), *troponin T2, cardiac type* (*TNNT2*, 2.9%, sarcomere), *tropomyosin 1* (*TPM1*, 1.9%, sarcomere,) and *desmoplakin* (*DSP*, truncating variants with 2.8%, desmosome, R. Walsh et al., 2017).

For LVNC patients the highest genetic burden was observed with variants in *MYH7*, *MYBPC3* and *TTN* (Oechslin & Klaassen, 2019; Kolokotronis et al., 2019). Therefore again, LVNC cases show similarities to HCM and DCM. Additionally, mitochondrial and desmosomal genes were associated with the disease (Lipshultz et al., 2019).

Patients affected with ARVC present the most distinct and non-overlapping genotype in the group of cardiomyopathies. In 50-60% of ARVC cases variants were detected in desmosomal genes (Hall et al., 2018).

Cases with RCM are mainly affected by alterations in sarcomeric genes (Kaski et al., 2008). The abundance of variants in RCM-specific is only estimated because epidemiological research is lacking so far. Therefore, mainly single cases or individual consanguineous families are reported with the onset of RCM. In less frequent cases RCM patients were described to be affected with variants in *desmin* (*DES)* and *filamin C* (*FLNC*, Lipshultz et al., 2019).

1.3 Next generation sequencing

Due to the need to understand mechanisms on how cardiomyopathies develop, genetic screening of CMP patients becomes more important. Although genetic cause is associated with cardiomyopathies since 1990 (Geisterfer-Lowrance et al., 1990) the genetic basis for this group of diseases remains challenging and new genes are associated repeatedly with the onset of cardiomyopathies partly due to rapid development of new sequencing methods. Since the decoding of the human genome during the Human Genome Project (IHGSC, 2004) the sequencing of a collection of genes (panel-based sequencing), the protein coding exome or the whole genome is improving steadily.

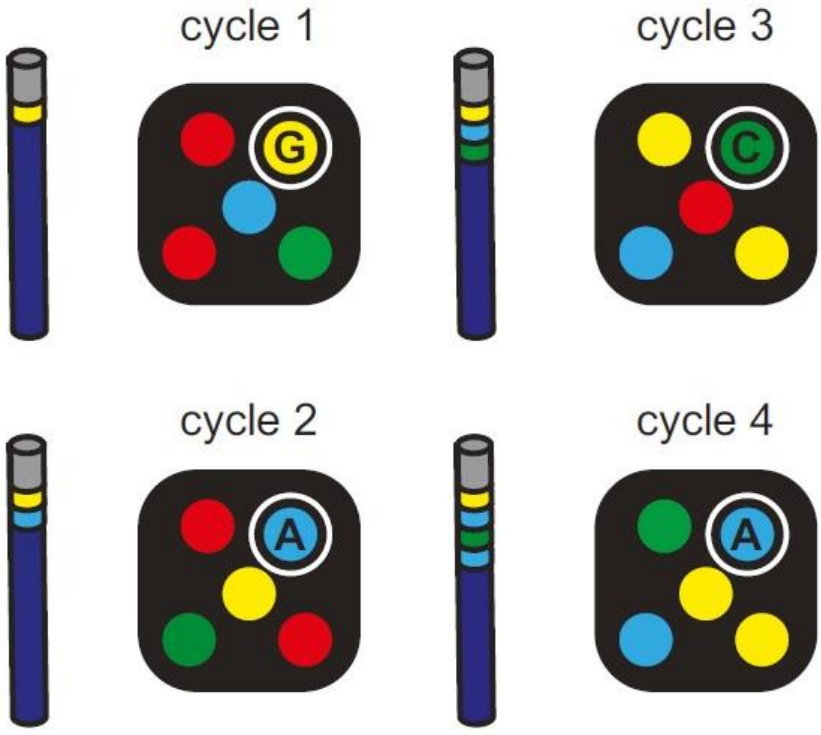


Figure 2: Schematic presentation of the Illumina next generation sequencing (NGS) method. The stepwise sequencing of a DNA fragment in four cycles shows marking of each nucleotide with an additional fluorophor in each cycle (Illumina Inc., 2010).

One major contributor in this field is the company Illumina. The method preferred by Illumina technology is based on sequencing through synthesis of the complementary DNA strand

(Figure 2, Pettersson, Lundeberg, & Ahmadian, 2009; Reuter, Spacek, & Snyder, 2015). The process is using fluorophor conjugated nucleotides and stepwise sequencing of each DNA position in cycles. The fluorophor is used as terminator group and therefore inhibits further DNA polymerization at the 3´ end of the newly synthesized chain in the same cycle (one nucleotide per cycle). After the positional readout the fluorophor is cleaved from the sequence and the next cycle is initiated (cleavage of fluorophor removes terminating feature, Buermans & den Dunnen, 2014; Illumina Inc., 2010). Each nucleotide is according to its base emitting light with specific wavelengths. The complementary DNA strands were captured on a flow cell through previously conjugated adaptor sequences and amplified/clustered to strengthen the signal (Meyer & Kircher, 2010).

1.4 PRDM16 as cause for cardiomyopathies

As mentioned before, genes newly associated with cardiomyopathy are discovered steadily. Currently genetic diagnostic rates are still low. Therefore, a necessity to increase genetic screening in patients is given. Further need to understand the impact of detected variants in regard to the occurrence of more than one genetic alteration (modifying variants) or newly associated genes remains, especially functionally or mechanistically. One such gene is *PRDM16*.

1.4.1 PRDM16 in fat tissue and as important factor for the state of progenitor and stem cells

The function of the transcriptional regulator PRDM16 was described in different tissues and first associated with the development of positive leukemia cells (Mochizuki et al., 2000). However, it is best known for its involvement in the differentiation of fat tissue, especially brown and beige adipocytes. It pushes the cell fate of myoblastic and adipocyte precursors to adipocytes and therefore interacts with different transcription factors (Ohno, Shinoda, Spiegelman, & Kajimura, 2012; Ohno, Shinoda, Ohyama, Sharp, & Kajimura, 2013). Therefore, PRDM16 has been shown to direct brown and beige fat determination and differentiation, acting as a major contributor in a critical complex to control the cell fate switch from myoblastic precursors to brown fat cells. Peroxisome proliferator-activated receptor gamma (PPARγ) agonists induce a white-to-brown fat conversion through stabilization of PRDM16 Protein (Figure 3). On the molecular level, PRDM16 interacts with several transcription factors like CCAAT/enhancer binding protein beta (C/EBPß) regulating adipogenic programs. Moreover, PRDM16 exposes epigenetic regulation of adipogenesis via

interaction with the euchromatic histone lysine methyltransferase 1 (EHMT1). Thus, PRDM16 seems to be involved in many signal pathways or is regulated by such.

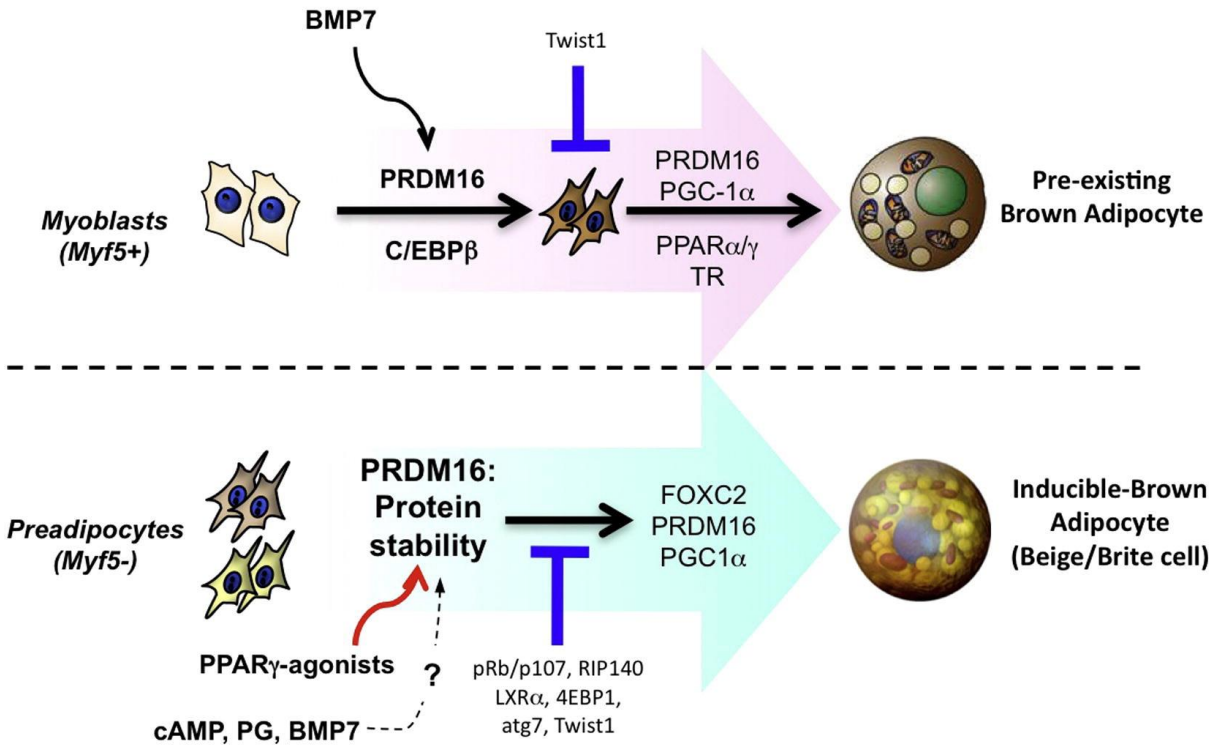


Figure 3: Schematic development of brown and beige adipocytes driven by PRDM16-specific initiation and stabilization. The upper part shows the development of brown adipocytes from myoblasts and trough PRDM16 activated transcription processes and the lower part shows differentiation to beige adipocytes from preadipocytes through stabilization of PRDM16 (**Ohno, Shinoda, Spiegelman, & Kajimura, 2012**).

Additionally, PRDM16 was observed with crucial roles in other tissues. Hence, this protein was described to be involved in the processes of palatogenesis, hematopoiesis, neurogenesis and angiogenesis and is defined by a major role in stem cell regulation and homeostasis (Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010; Chuikov, Levi, Smith, & Morrison, 2010; Aguilo et al., 2011; Baizabal et al., 2018; Su et al., 2020).

1.4.2 Heart phenotype of PRDM16 deactivation

Apart from the afore mentioned tissues PRDM16 was associated with the onset of cardiomyopathies. First evidence for cardiac involvement of PRDM16 was generated through a knockout mouse model carrying a splice site alteration resulting in a premature stop and therefore impaired PRDM16 expression (Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010). The variant was lethal in pre- and early postnatal stages in homozygous form. Although the mouse mainly was described for a cleft palate defect, it was shown, that mice were affected by cardiac hypoplasia (Figure 4).

Further evidence was presented in a study, where 1p36 deletion syndrome patients with additional occurring cardiomyopathy were analyzed (Arndt et al., 2013). With this syndromic disease individuals are affected by the partial deletion of the first chromosome. It was suggested that cardiomyopathy is only occurring when certain parts of the chromosome are deleted. The common minimal region of deletion of 18 individuals was positioned in the gene *PRDM16* including the exons 4 to 17. Furthermore, through genetic screening of patients affected by primary non-syndromic cardiomyopathies missense and truncating variants were detected in seven DCM and LVNC cases. One of the truncating variants was then investigated in zebrafish, leading to severe heart phenotypes.

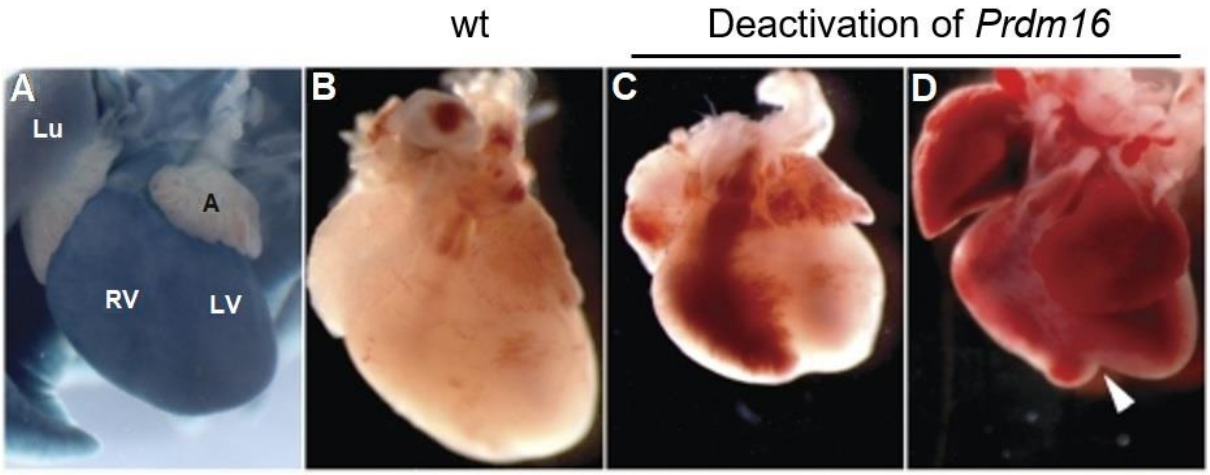


Figure 4: Expression of PRDM16 in mice heart and cardiac hypoplasia due to PRDM16 ko. (A) Expression of PRDM16 in ventricles marked with blue. **(B)** Cardiac proportions of a normal heart (wt) are shown. **(C)/(D)** Through deactivation of PRDM16 the mice heart was affected by hypoplasia (**Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010**).

Since then, further PRDM16-specific publications associated with cardiac research were rare. However, in a study investigating a nonsense variant in the T-box transcription factor 20 (TBX20) PRDM16 was determined as a downstream target of TBX20 and therefore a transforming growth factor beta (TGF-β) signaling modifier. In patient-specific human induced pluripotent stem cell–derived cardiomyocytes and cardiac biopsies the regulation of PRDM16 through TBX20 was observed and its effect on the TGF-β pathway determined

Just recently further proof for the involvement of PRDM16 in cardiac phenotypes was published (Nam, Lim, Ha, Oh, & Kang, 2020). A null variant of PRDM16 was introduced into a cardiac conditional knockout mouse model. With electrocardiographic diagnosis considerable alterations were observed (QRS duration and QTc interval increased). Fibrosis and hypertrophic cardiomyocytes were detected in the affected mice and the expression of cardiac ion channels was dysregulated. Thus, cardiac conduction was impaired.

These findings mainly proof that PRDM16 is involved in the development of the heart and cardiac impairments. Molecular mechanisms mainly remain uncertain and further investigation to understand the effect of genetic alterations in PRDM16 is necessary for prospective therapies.

1.5 Aims of the dissertation

This work aims to identify genetic causes of childhood CMP and to understand the development of early-onset and severe heart muscle disease. Only recently, mutation of *PRDM16* was shown to be associated with two types of CMP, DCM and LVNC. *PRDM16* encodes for a transcriptional regulator with a previously unknown function in the heart. The mutational spectrum of *PRDM16* in extended cohorts of patients with different types of heart muscle diseases will be examined. Furthermore, biochemical exploration of PRDM16 in the context of CMP is another aim of this work.

1. We start with a targeted panel NGS approach covering the established CMP disease genes. The cohort of patients with pediatric CMP will be characterized using clinical parameters. The impact of age of onset of CMP, type of CMP, and genetic vs. sporadic CMP will be evaluated for event-free survival.
2. We ask if the number and type of VOI have an impact on the occurrence of adverse events in patients. The genetic variants will be classified according to ACMG for their pathogenicity. This aim requires to develop and apply specific rules for interpretation in the context of CMP and to establish an in-house evaluation scheme.
3. We hypothesize that we find a gene or a functional subset of genes that are of either high or low risk to develop severe CMP courses. With these patients we will perform further characterization on RNA and protein level.
4. As PRDM16 seems to be important for heart function, we anticipate that genetic alterations in *PRDM16* will be associated with several types of heart muscle disease of different age groups. We will perform mutational analysis of CMP associated genes, specifically in myocarditis, and in a set of adult patients with LVNC. For prediction of posttranslational modification sites and structural elements of PRDM16, protein prediction tools will be used.
5. In addition, this work will cover research on genetic CMP disease mechanisms and biochemical characterization of PRDM16. Specifically, we will explore the subcellular localization of CMP variants and the influence of CMP variants on the protein stability of PRDM16.

This approach is of vital interest to explore CMP pathogenesis, appropriately manage CMP patients, and to explore translational approaches for diagnosis as well as potential therapies.

2 Methods

2.1 Materials

2.1.1 Consumables

Plastic consumables such as Eppendorf tubes, Falcon tubes or cell culture dishes and plates were purchased from companies mentioned in the following list.

Manufacturer	Location
Biozym Scientific GmbH	Hessisch Oldendorf, Germany
Eppendorf AG	Hamburg, Germany
INTEGRA Biosciences GmbH	Biebertal, Germany
Merck KGaA (Sigma-Aldrich, Millipore)	Darmstadt, Germany
Sarstedt	Nümbrecht, Germany
Thermo Fisher Scientific/Applied Biosystems	Darmstadt, Germany
VWR International GmbH/NanoEnTek	Darmstadt, Germany

2.1.2 Sequencing

2.1.2.1 DNA isolation from blood and saliva

DNA was isolated and purified from blood or saliva. The following kits were used.

Chemical/Kit	Manufacturer	Ref
NucleoSpin Blood	MACHEREY-NAGEL	740951
PrepIT-L2P Kit	DNA Genotek	PT-L2P

2.1.2.2 Sanger sequencing

Sanger sequencing was used to verify the results of NGS and investigate the segregation of variants within families. Furthermore, the quality of cloned vectors was validated with this method. The following devices, chemicals and primers were used for Sanger sequencing.

Gene (Transcript)	Primer	Primer sequence	Length
Primer for genomic exon amplification			
LDB3	ghLDB3_ex10_f	AAGTGATGCAACAATGAACACC	22
(NM_007078)	ghLDB3_ex10_r	GTTCCACCACCACTTCAAGC	20
MYBPC3	ghMYBPC3_ex10-11_f	CAACAGTCATCCTCACAGTG	20
(NM_000256)	ghMYBPC3_ex10-11_r	CAGGACCAAGGAGCTGTAG	19
TNNI3	ghTNNI3_ex2_f	AAGTGGGTTTGCGAGTCA	18
(NM_000363.4)	ghTNNI3_ex2_r	CCATCACCACCAAGACCC	18
	ghTNNI3_ex5_f	GGAGCTTGAGAATGGGTGGG	20
	ghTNNI3_ex5_r	GAGCCAAGACTCCACAGACC	20
	ghTNNI3_ex8_f(a)	AGATACTTAGGCATCCAGGGTAG	23
	ghTNNI3_ex8_r(a)	ACAGCCAAGAGTGCTTCACAT	21
	ghTNNI3_ex8_f(b)	GCTACTATTGACCTGAGAATCC	22
	ghTNNI3_ex8_r(b)	ACAGCCAAGAGTGCTTCACAT	21
Primer for amplification of coding sequence embedded in plasmid			
PRDM16	235-256	CCGATCCCAGCAGACTTCGAGC	22
(NM_022114.3)	260-241	CGGAGCTCGAAGTCTGCTGG	20
	621-641	GGTGCACGTGAAGGAAGGCG	20
	872-892	CCAACAAGTACAGCCTGGAGC	21
	1062-1081	GCACATCCGCTCGCAGCACG	20
	1450-1469	GGCTTCAACGAGTACTTTCC	20
	1800-1819	CGGCAGTGACTTTGAGGACG	20
	2009-2028	CCCAGCACTCATTCTTCCCG	20
	2112-2093	CCATTGCCGAGAAGTACTTT	20
	2238-2257	CCACAACTTGCTGGTCAAGG	20
	2553-2572	CCCGCTCCACTACGCCAAGC	20
	2712-2730	GACAGAGAAGCTGGAGAGC	19
	2994-3013	CCGGAACATCCACAACAAGG	20
	3241-3259	GCCAATAGTGAGATGAACC	19
	3498-3518	GGGCTTTGACCACACCCGAAG	21
Primer for cDNA exon amplification for splice site analysis TNNI3 c.24+2T>A			
TNNI3	chTNNI3_ex1_f1	TCACTGACCCTCCAAACG	18
(NM_000363.4)	chTNNI3_ex1_f2	GGGAGTCTCAAGCAGCCC	18
	chTNNI3_ex5_r	TGGCAGCGGGTGCTCAGA	18

Device	Company
DNA Engine Tetrad 2, Peltier Thermal Cycler	Bio-Rad
NanoDrop ND-1000 Spectrophotometer	Thermo Fisher Scientific
3730xl DNA Analyzer	Applied Biosystems

Chemical/Enzyme/Kit	Manufacturer	Ref
BigDye Terminator v3.1	Thermo Fisher Scientific	4337455
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D4540
dNTP set	Rapidozym	GEN-009-250
Ethidium bromide, 1% in H_2O	SIGMA-ALDRICH	46067-50ML-F
Exonuclease I	New England Biolabs	M0293
FIREPol DNA polymerase	SOLIS BIODYNE	04-11-00115
GeneRuler 100bp DNA ladder	Thermo Fisher Scientific	SM0242
GeneRuler 1kb DNA ladder	Thermo Fisher Scientific	SM0313
Hi-Di Formamide	Thermo Fisher Scientific	4311320
Illustra Sephadex G-50 DNA Grade	GE Healthcare Life Sciences	17-0573-02
LE Agarose	Biozyme	840004
Phusion High-Fidelity DNA Polymerase	New England Biolabs Inc.	M0530
rAPid Alkaline Phosphatase	Roche/MERCK	4898141001
Taq DNA polymerase	Qiagen	201205
TERMIPol DNA polymerase	SOLIS BIODYNE	01-03-00500

2.1.2.3 Next-generation sequencing (NGS)

Next-generation sequencing (NGS) was used to identify genetic variants in CMP patients by massive parallel sequencing in a high throughput method. The TruSight Cardio Sequencing panel, including a primer mix to detect 174 cardiac disease genes, was applied. Utilized devices and chemicals are listed below. Corresponding software is listed in the section 2.1.6.

Device	Company
2100 Bioanalyzer	Agilent Technologies
DynaMag-2 Magnet	Applied Biosystems, Thermo Fisher Scientific
DynaMag-PCR Magnet	Invitrogen, Thermo Fisher Scientific
Hybex Microsample Incubator	SciGene
NextSeq 550	Illumina
Qubit 3.0 Fluorometer	Invitrogen, Thermo Fisher Scientific

Chemical/Kit	Manufacturer	Ref
Agilent DNA 1000 Kit	Agilent Technologies	5067-1504
Agilent High Sensitivity DNA Kit	Agilent Technologies	5067-4626
Ethanol ≥99,8 %	Carl Roth	9065.1
NextSeq 500/550 High Output Kit v2	Illumina	FC-404-2005
Qubit dsDNA HS Assay Kit	Invitrogen, Thermo Fisher Scientific	Q32854
TruSight Cardio Sequencing Kit for NextSeq, (48 samples)	Illumina	FC-141-1011

2.1.3 Cell culture

For cell culture approaches mainly the **h**uman **e**mbryoid **k**idney cell line HEK293 was used. Only for initial experiments other cell lines like HeLa (cervical cancer cell line extracted from patient **He**nriette **La**cks) or MCF7 (Michigan Cancer Foundation - 7) cell lines were tested for the expression of specific proteins but were less suited. For the maintenance and transfection of these cell lines the following materials were used.

Chemical/Enzyme/Medium/Kit	Manufacturer	Ref
Cycloheximide	SIGMA-ALDRICH	C7698
FBS	Thermo Fisher Scientific	10270106
HEK293 cells	ATCC	ATCC-CRL-1573
jetPEI kit	Polyplus	101-10N
MEM	Thermo Fisher Scientific	31095-052
OptiMEM	Thermo Fisher Scientific	31985070
PBS pH 7.4	Thermo Fisher Scientific	20012-068
Trypsin	Thermo Fisher Scientific	25200056

2.1.4 Cloning of newly ligated expression vectors

To transfer genes with patient specific variants into expression vectors suited for expression of proteins in human cell lines specific primer, adding distinct restriction sites into the newly designed constructs, were developed and are listed below.

Primer	Primer sequence	Length	Target vector
Primer for ligation of *PRDM16* (NM_022114.3)			
PRDM16_f_XhoI	CTCAGATCTCGAGCTATGCGATCCA AGGCGAGGGCGAGGAAGCTA	45	pEGFP-C1
PRDM16_f_EcoRI	TCAGAATTCCCGATGCGATCCAAGG CGAGGGCGAGGAAGCTA	42	pFLAG-CMV5a
PRDM16_f_HindIII	CTCAAGCTTGCGATGCGATCCAAGG CGAGGGCGAGGAAGCTA	42	pFLAG-CMV6a
PRDM16_r_EcoRI_a	ACTGGAATTCTCAGAGGTGGTTGAT GGGGTGAAATGCTCC	40	pEGFP-C1/ pFLAG-CMV6a
PRDM16_r_BamHI_a	ACTGGGATCCTCAGAGGTGGTTGAT GGGGTGAAATGCTCC	40	pFLAG-CMV5a

After amplification of the gene target (insert) the PCR product was ligated into the new vector and transformed into competent *E. coli*. Additional antibiotic selection and purification completed the cloning process. All necessary components are listed below.

Component	Manufacturer	Ref	Restriction site
Restriction enzymes used for cloning			
EcoRI	New England Biolabs	R3101 S	5‘...G^{v}AATTC...3‘ 3‘...CTTAA^{v}G...5‘
BamHI	New England Biolabs	R0136S	5‘...G^{v}GATCC...3‘ 3‘...CCTAG^{v}G...5‘
HindIII	New England Biolabs	R0104S	5‘...A^{v}AGCTT...3‘ 3‘...TTCGA^{v}A...5‘
XhoI	New England Biolabs	R0146 S	5‘...C^{v}TCGAG...3‘ 3‘...GAGCT^{v}C...5‘
Plasmids used as target vector			
pEGFP-C1	BD Biosciences Clontech	6084-1	MCS*
pFLAG-CMV5a	Sigma-Aldrich	E7523	MCS*
pFLAG-CMV6a	Sigma-Aldrich	E1900	MCS*

* MCS – multiple cloning site

Chemical/Enzyme/Kit	Manufacturer	Ref
Ampicillin	Sigma-Aldrich	A0166-25G
Kanamycin	Sigma-Aldrich	60615-5G
LB Broth with agar (Miller) Powder	Sigma-Aldrich	L3147
LB medium (Luria/Miller)	Carl Roth	X968
NucleoBond Xtra Midi EF	MACHEREY-NAGEL	740420.50
NucleoSpin Plasmid	MACHEREY-NAGEL	740588.250
Phusion DNA Polymerase	New England Biolabs	M0530L
rAPid alkaline phosphatase	Sigma-Aldrich	489813300
T4 ligase	New England Biolabs	M0202 S
XL1 blue competent	Agilent	200249

2.1.5 DNA and protein expression analysis

2.1.5.1 Preparation of human tissue

2.1.5.2 qPCR

For the isolation of RNA, synthesis of cDNA and quantitative PCR the following components were necessary.

Chemical/Enzyme/Kit	Manufacturer	Ref
Ethanol	Thermo Fisher Scientific	20012-068
Isopropanol	Thermo Fisher Scientific	31095-052
PrimeScript RT Reagent Kit	TaKaRa	RR037A
SYBR® Premix Ex Taq™ II (Tli RNase HPlus)	TaKaRa	RR820W
Trizol	Thermo Fisher Scientific	31985070

The Taqman 7500 system from Applied Biosystems, USA, was used for all qPCR measurements. Specific primers were designed to detect and amplify mRNA expressed DNA fragments. The targeted genes and designed primers are listed next.

Gene (Transcript)	Primer	Primer sequence	Length
Primer for cDNA amplification with quantitative PCR			
ACTN2	qhACTN2_ex13-14_f	TTGGAACACCTGGCTGAGA	19
(NM_001278343)	qhACTN2_ex13-14_r	GCCGACTCGTAATCCTTCTG	20
MYBPC3	qhMYBPC3_ex11-12_f	GCATGAGGCGCGATGAGAAGA	21
(NM_000256)	qhMYBPC3_ex11-12_r	CAGCCAGTTCCACGGTCAGC	20
TNNI1	qhTNNI1_ex5-6_f	GGATGAGGAGCGATACGACA	20
(NM_003281)	qhTNNI1_ex5-6_r	GGCGCTTGAACTTCCCAC	18
TNNI3	qhTNNI3_ex3-4_f	CACCAGCCCCAATCAGACG	19
(NM_000363.4)	qhTNNI3_ex3-4_r	CTGCAATTTTCTCGAGGCGG	20
TNNT2	qhTNNT2_ex14-15_f	GAGCTGTGGCAGAGCATCTA	20
(NM_000364)	qhTNNT2_ex14-15_r	ATCCTGTTTCGGAGAACATTG	21

2.1.5.3 Western blotting

For Western blot analysis standard procedures and protocols were used (running and washing buffers for BioRad protocols). Tissue and cells were harvested by adding RIPA buffer. To separate a cytoplasmic fraction from the nucleus of a cell a hypotonic buffer with 20 mM Tris-HCl pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$ was prepared to separate the lysis of each compartment. Further components used for Western blot are listed below.

Chemical/Enzyme/Kit	Manufacturer	Ref
20X Bolt™ MES SDS Running Buffer	Thermo Fisher Scientific	B0002
Ammonium peroxydisulfate	Carl Roth	9592.2
Bolt™ 4-12% Bis-Tris Plus Gels	Thermo Fisher Scientific	NW04125BOX
DNase I, 20.000 U	Thermo Fisher Scientific	18047019
Hypotonic buffer*	In-house	
Methanol	Carl Roth	8388.2
NP-40	Sigma-Aldrich	74385
Powdered milk	Carl Roth	T145.2
Precision Plus Protein™ Dual Xtra Prestained Protein Standards	BioRad	1610377
RIPA buffer 10x	Cell Signaling	9806S
Rothiphorese NF-Acrylamide/Bis-Sol. 30%	Carl Roth	A124.1
SDS	Carl Roth	2326.2
TEMED, ≥99 %	Carl Roth	2367.3
WesternBright ECL	Biozym	541004

Device/Material	Company
Mini-PROTEAN® Tetra electrophoresis system	BioRad
Chemiluminescence imaging system CHEMI only	VWR
PVDF membrane, roll, pore size 0.2 µm	Thermo Fisher Scientific

2.1.5.4 Immunohistochemistry

Samples prepared for immunohistochemistry were first fixated with a formaldehyde solution and then permeabilized with saponin. In further incubation steps antibodies and staining chemicals were applied and supernatants were washed away with PBS. All used chemicals are listed below.

Chemical/Enzyme/Kit	Manufacturer	Ref
Fetal Bovine Serum, heat inactivated, US origin	Thermo Fisher Scientific	16140071
Fluoromount-G Mounting Medium	Science Services	E17984-25
Formaldehyde solution 37%	Sigma-Aldrich	47608-250ML-F
PBS pH 7.4	Thermo Fisher Scientific	20012-068
Saponin	Sigma-Aldrich	47036-50G-F

Cells were seeded on cover slips to prepare samples for immunostained fluorescence microscopy. Therefore, the LSM700 microscope of the Advanced Light microscopy core facility of the MDC Berlin was used.

Device/Material	Company
Cover slips for microscopy (13 mm Ø)	Neolab (Marienfeld)
Microscope slides, VWR superfrost BLUE CUT EDGE	VWR
LSM700	Zeiss

2.1.5.5 Antibodies and cellular imaging chemicals

In the following list there are all antibodies used for this thesis. These antibodies were used in Western blot and immunohistochemistry approaches.

Antibody	Host	Class	Manufacturer	REF
Characterization of TNNI3				
Anti-GAPDH	Mouse	monoclonal	Life Technologies	AM4300
Anti-HSC70/HSP70	Mouse	monoclonal	Enzo Life Science	ADI-SPA-820
Anti-MYBPC3	Sheep	polyclonal	R&D Systems	AF7439
Anti-TNNI1	Rabbit	polyclonal	Sigma-Aldrich	HPA028190
Anti-TNNI3	Rabbit	polyclonal	Thermo Fisher Scientific	PA5-28964
Anti-TNNT2	Mouse	monoclonal	Thermo Fisher Scientific	MA5-12960
Characterization of PRDM16				
Anti-EGFP	Rabbit	polyclonal	Sigma-Aldrich	G1544
Anti-FLAG	Mouse	monoclonal	Sigma-Aldrich	F1804
Anti-LMNA	Mouse	monoclonal	Cell Signaling	4777S
Anti-PRDM16	Rabbit	polyclonal	Abcam	ab106410
Anti-PRDM16	Sheep	polyclonal	R&D systems	AF6295
Anti-PRDM16	Rabbit	polyclonal	Thermo Fisher Scientific	720206
Anti-β-Tubulin	Mouse	monoclonal	UBPBio	Y1060
DAPI			Thermo Fisher Scientific	D1306
Phalloidin Alexa 647			Thermo Fisher Scientific	A22287
Secondary antibodies				
Anti-mouse Alexa 488	Goat		Thermo Fisher Scientific	A-11029
Anti-mouse IgG-HRP	Horse		Cell Signaling	7076
anti-rabbit Alexa 568	Goat		Thermo Fisher Scientific	A-11036
Anti-rabbit IgG-HRP	Goat		Cell Signaling	7074
Anti-sheep Alexa 488	Donkey		Thermo Fisher Scientific	A-11015
Anti-sheep IgG-HRP	Donkey		Thermo Fisher Scientific	A16041

2.1.6 Software

To analyze the data generated in various experiments different software packages were used. Licenses for these packages for these applications were obtained through institution involved for this work (FU Berlin, MDC, Charité, DHZB) or acquired privately.

Software	**Company**
Varfish	Core Unit Bioinformatics (CUBI), Charité Berlin
CorelDRAW Home & Student 2019	Corel Corporation
Microsoft Office 2010/365	Microsoft
SPSS	IBM
GelQuantNet	BiochemLabSolutions
ZEN 3.0 (blue edition)	ZEISS
2100 Bioanalyzer Expert	Agilent Technologies
IGV Version 2.3.97	Broad Institute/Regents of the University of California
Variant Studio Software v3.0	Illumina

2.2 Methods

2.2.1 Genetic characterization of cardiomyopathy patients

2.2.1.1 Clinical assessment of cardiomyopathy patients

Recruitment of CMP patients has been done at the Charité - Universitätsmedizin Berlin and the German Heart Center Berlin (DHZB), Berlin, Germany. Between November 2011 and February 2017 unrelated probands carrying a CMP and equal to or under the age of 18 (≤) were phenotyped according to the design of the study related to this thesis (Figure 5).

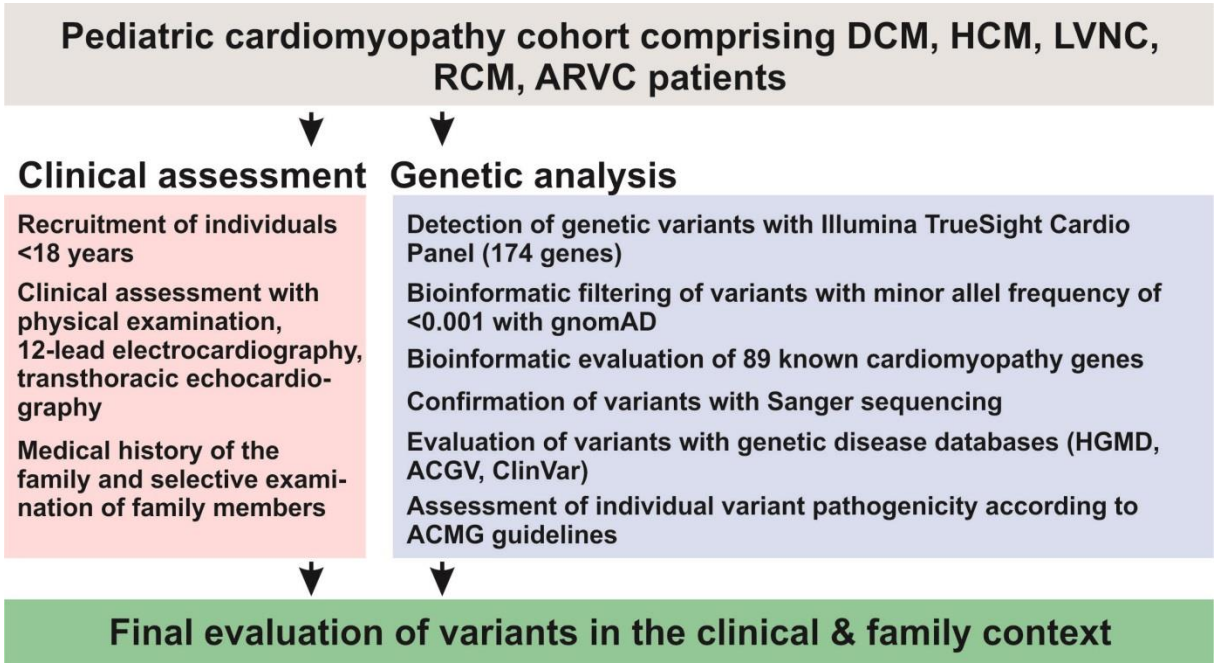


Figure 5: Schematic study design with clinical assessment and genetic analysis for evaluation of genetic variants and their pathogenicity.

Informational consent according to the Declaration of Helsinki was obtained from each patient or their legal guardians (study approval by the local institutional review board: ID EA2/083/13, EA2/131/10, Charité Universitätsmedizin Berlin). For each subject and their available family members evaluation of medical history, physical examination, 12-lead electrocardiography, and transthoracic echocardiography was recorded. A retrospective approach for clinical and echocardiographic data has been realized for patients recruited after heart transplantation (HTX). Probands included in this study had to be diagnosed with a primary CMP, comprising HCM, DCM, RCM, LVNC and ARVC, and occurred non-syndromically. CMPs were classified according to the guidelines of the American Heart Association (Maron et al., 2006). Patients carrying a CMP and an additional structural congenital heart defect were excluded from the

study. Blood samples from the recruited subjects and their available family members were taken for further genetic analysis.

2.2.1.2 Targeted next-generation sequencing

For genetic characterization with NGS DNA was isolated from blood samples of the CMP patients with the NucleoSpin Blood kit (2.1.2.1) or were already available as DNA samples. Amounts of DNA were measured with a Qubit 3 fluorometer using Qubit reagents for dsDNA (2.1.2.3). To investigate alterations in exonic regions of 174 cardiovascular disease genes, a cardio-panel from Illumina was used for NGS (2.1.2.3, Pua et al., 2016). Information of these genes and their implication in structural heart diseases and arrhythmias can be found on https://support.illumina.com/sequencing/sequencing_kits/trusight-cardio-sequencing-kit.html. After diluting the DNA to prescribed amounts library preparation and enrichment has been realized in a multiplex approach. Therefore, DNA fragments of defined length were marked by specific flanking sequences (protocol according to manufacturer). To quantify the DNA libraries a Bioanalyzer 2100 was used, applying the high sensitivity DNA kit or DNA 1000 kit (2.1.2.3). For sequencing the NextSeq 500 system was implemented with mid output cartridge v2 with paired end sequencing (150 cycles) and dual indexing.

2.2.1.3 Alignment and variant calling of NGS data

The multiplexed libraries were decoded using bcl2fastq v2.17.1.14. Alignment of the reads was realized with BWA-MEM v0.7.15 (Li, 2013) aligning to the reference genome GRCh37 (hs37d5.fa). Separate read groups were assigned for all reads from one lane, and duplicates were masked using Samblaster v0.1.24 (Faust & Hall, 2014) was used to mask duplicates and separate read groups were assigned for all reads from one lane. Quality control was realized with FastQC (Andrews, 2010) and due to analyzing the minimal coverage of all sections covered by the Illumina TruSight Cardio Sequencing Kit. To finish the alignment, variants were called with GATK UnifiedGenotyper v3.7 (DePristo et al., 2011).

2.2.1.4 Bioinformatic evaluation and variant classification

For further analysis the called variants were imported into Variant Studio (Illumina, USA) to filter all non-synonymous and splice site variants with a minor allele frequency (MAF, the frequency at which the *second most common* allele occurs in a given population) < 0.001. After additional filtering with a MAF < 0.001 in the Genome Aggregation Database (gnomAD, Lek et al., 2016) and excluding CMP genes only described in single case reports, a selection of 89

CMP genes was further analyzed (Table S 1). The coverage quality control of the variants was again examined with the Integrative Genomics Viewer (2.1.6).

Table 1: Online tools to filter for rare variants and predict the effect and pathogenicity of base exchanges

Category	Name	Website
Population Database	gnomAD	http://gnomad.broadinstitute.org/
	HGMD	http://www.hgmd.cf.ac.uk/ac/index.php
	ACGV	https://cardiodb.org/ACGV/
Missense prediction	SIFT	http://sift.jcvi.org/
	Polyphen-2	http://genetics.bwh.harvard.edu/pph2/
	Mutationtaster*	http://www.mutationtaster.org/
Splice site prediction	MaxEntScan	http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
	NNSplice	http://www.fruitfly.org/seq_tools/splice.html
	HSF3	http://www.umd.be/HSF3/index.html

*Predicts effect of all small base changes (missense, splice site changes, frameshift, indel)

The pathogenicity of the variants was evaluated according to the guidelines of the ACMG (Richards et al., 2015). For the presence of variants in disease databases such as Human Gene Mutation Database (HGMD) and Atlas of Cardiac Genetic Variation (ACGV) were checked. Computational and predictive data like PolyPhen-2 and Sorting Intolerant From Tolerant (SIFT) for missense variants or Human Splice Finder (HSF), Neural Network Splice (NNSplice) and Maximum Entropy Scan (MaxEntScan) for splice site variants was used to predict the impact of the filtered variants. Mutation Taster (MT) served as a multifunctional tool for the impact of all amino acid changing variants. Other criteria comprised functional data, including data from variant and literature databases for information about mutational hotspots, functional domains or well-established functional studies (Ensembl, UniProt, HGMD, ARVD/C Genetic Variants Database and PubMed), segregation data from family members and phenotypical compliance. Corresponding to the guidelines of the ACMG variants were classified as pathogenic (P), likely pathogenic (LP) and variant of uncertain significance (VUS). After application of the filter criteria mentioned in Figure 6 all variants interpreted as VUS, LP and P were summarized with the term variant of interest (VOI). To investigate the influence of different functional origins of the variants, the 89 CMP genes were categorized into functional groups like the sarcomere or the Z-disc (Table S 1, Burke, Cook, Seidman, & Seidman, 2016).

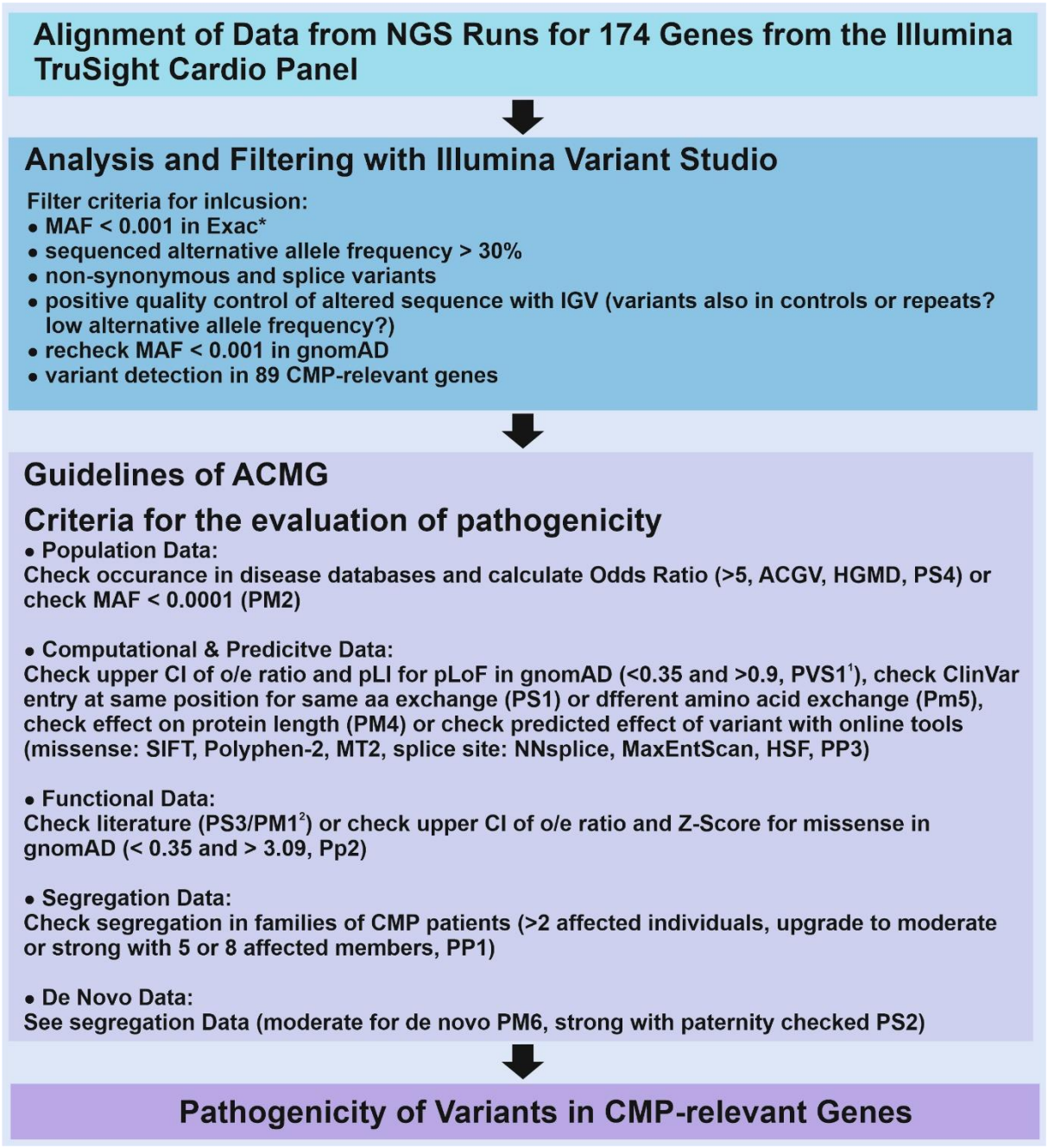


Figure 6: Schematic overview of genetic analysis and evaluation of pathogenicity according to the guidelines of the ACMG. In each category of the guidelines the strongest applicable argument was counted (double counting, Abou Tayoun et al., 2018).

[1]PVS1: If borderline cases occurred, literature and ClinVar entries were taken into consideration.

[2]PM1: Examples for mutational hotspot or well-studied functional domains are MYH7 with its head region (aa 181 to 937, Kelly et al., 2018) and TTN (PSI, Roberts et al., 2015).

***gnomAD was not available for Illumina variant studio**

2.2.1.5 Validation and segregation of variants of interest

To validate the variants detected in CMP patients and to examine the familial segregation (if possible) Sanger sequencing was applied. Standard laboratory protocols were used for this

procedure (Klaassen et al., 2008, 2.1.2.2). Primers for sequencing variants identified by NGS were constructed with Primer3web (Whitehead Institute for Biomedical Research, USA, 2.1.2.2) and verified with the tools Blat and in-silico PCR from the UCSC genome browser (UCSC Genomics Institute, USA). Additional verification was realized with sequence alignment using Ensembl (EMBL-EBI, Great Britain).

2.2.1.6 Statistical analysis

Frequencies and percentages were used for variables defined as categorical. Continuous values were presented as means and interquartile ranges (IQR). Kaplan-Meier curves were generated to analyze the survival or event free progression of disease in the CMP cohort. The probability of survival was therefore contextualized with the phenotype (DCM, HCM, etc.), number of VOI and the sporadic or genetic appearance of the CMP. For calculations the time point zero was defined as the age at diagnosis. The log-rank test was used to compare the differences in estimated survival curves of ≥ 2 groups. Statistical significance was considered with p value < 0.05 (probability value). SPSS v24.0 was used for data analysis (2.1.6).

2.2.1.7 Genetic variation in an extended cohort

To further analyze the distribution of CMP related variants the initial pediatric CMP cohort was extended with CMP and myocarditis patients from all age groups. The samples were registered for an in-house cohort. The recruitment, which started in 1999, and the genetic handling were similar to the samples from the pediatric cohort. DNA was either extracted from blood samples or was already prepared. NGS, alignment, variant calling, bioinformatic evaluation and variant classification was done as described before (see 2.2.1.2 to 2.2.1.4). Family members were mostly not available.

2.2.2 Functional analysis of cardiomyopathy genes

2.2.2.1 Cloning of *PRDM16* wildtype and variants

Plasmids carrying the wt and variants containing *PRDM16* gene were already available in the laboratory. The backbone of these plasmids was the pcDNA-Dest53 vector. Variants chosen for analysis comprised the amino acid exchanges p.E271K, p.P291L, p.R525Pfs*79, p.N816S and p.L887P. Therefore, four missense and one truncating (frameshift, which leads to a premature stop) variant were included. All of these variants were formerly described to cause DCM or LVNC (Arndt et al., 2013).

For overexpression of PRDM16 in the human cell line HEK293 *PRDM16* constructs were reamplified without the backbone and containing specific restriction sites using PCR (2.1.4). After cutting the plasmids pEGFP-C1, pFLAG-CMV5a and pFLAG-CMV6a with specific restriction enzymes and dephosphorylating their open DNA ends, the ligation into the vectors was realized with the T4 DNA ligase from NEB (2.1.4). The restriction enzymes used in this approach were EcoRI (pFLAG-CMV5a, pFLAG-CMV6a and pEGFP-C1), BamHI (pFLAG-CMV5a), HindIII (pFLAG-CMV6a) and XhoI (pEGFP-C1).
Then competent *E. coli* XL1-Blue cells were transformed with these vectors for positive selection and amplification of the plasmids. Therefore LB-plates containing the corresponding antibiotics kanamycin or ampicillin were inoculated with the transformed bacteria. After incubation overnight and at 37°C the grown colonies were used to inoculate LB-master plates and 1.5 ml of liquid LB medium with the corresponding antibiotic for doing a miniprep with the NucleoSpin Plasmid kit the following day (2.1.4, protocol according to manufacturer's instruction). For higher yields of DNA and after quality control liquid LB medium of 160 to 200 ml was inoculated and the midiprep NucleoBond Xtra Midi EF kit was used for plasmid isolation (2.1.4, protocol according to manufacturer's instruction).

2.2.2.2 Quality control of cloned vectors

To check the quality and sequence of the *PRDM16* constructs two general methods were used. The first step was to estimate the length of the newly ligated plasmids in an agarose gel run with a standard DNA ladder and to compare it to the calculated length available in online databases. Therefore a 1% agarose gel and the GeneRuler 1 kb DNA ladder were used (2.1.2.2). The gels were loaded with plasmids in supercoiled, linearized or double digested form (no restriction enzyme, one restriction enzyme or two restriction enzymes corresponding to the enzymes used in 2.2.2.2).
The second step was to sequence the inserted *PRDM16* constructs and its up and downstream transitions into the vector to check the right insertion into the vector and if false variants were inserted into the DNA during the cloning procedure. For the sequencing reaction the BigDye Terminator v3.1 cycle sequencing kit was used (2.1.2.2). The concentration of the DNA was spanning from 100 to 200 ng/µl. The protocol was adjusted to smaller amounts of BigDye Terminator solution and additional TERMIPol DNA polymerase to generate more stable and longer sequencing strands (2.1.2.2). For sequencing a 3730 DNA Analyzer was used. The sequences were base called with ABI sequencing analysis and analyzed with the software LaserGene, Seqman (2.1.2.2).

2.2.2.3 Maintenance of cell culture

HEK293 cells were used for overexpression of PRDM16. They were thawed from stocks stored in liquid nitrogen with a low passage number and splitted twice a week using trypsin (0,25% Trypsin-EDTA, 0) to detach the surface attached cells (HEK293 cells splitted in the ratio from 1:8 to 1:12).

2.2.2.4 Cell count and seeding

For experiments the cells were seeded in specific cell numbers and therefore counted with the automated cell counter system EVE (0). Accordingly, the cells were trypsinated, centrifuged (500 rpm, 5 min, RT) and resuspended in a defined volume of cell culture medium (DMEM low glucose/MEM plus 5% FBS). For counting the cells 10 µl of cell culture was mixed with 10 µl of trypan blue and pipetted into a counting slide available from the same manufacturer (10 µl into the slide). The cell counter system is giving the unit cells per ml. With this the number of cells can be calculated with the following formula.

$$v(wanted\ cell\ number)\ \mu l = \frac{wanted\ cell\ number}{measured\ cell\ number\ ml^{-1}} \cdot 1000$$

For microscopy cells were seeded on glass cover slips. Therefore, up to three cover slips were put in a well of a six well plate. To increase the attachment of the cells to the cover slips and wells, the plates were incubated with 1 ml DMEM F-12 medium per well, including a 1:100 dilution of Geltrex. Seeding without Geltrex was sufficient for samples used in Western blotting.

2.2.2.5 Transfection of cells

For transfection of HEK293 cells jetPEI was used (0). Therefore, cells were seeded in six-well or 12-well plates with a respective cell number for the relevant experiment and planed time for transfection (250k to 400k cells per well for six-well plates and one to two days of incubation and 150k to 200k for 12-well plates one to two days of incubation). The transfection protocol was started with a confluency of the cells at 70 to 80%. Dilution of 3 µg of DNA with water resulted in 100 µl of DNA solution. After 6 µl of jetPEI were diluted with 94 NaCl solution it was mixed with the DNA solution with following vortexing immediately and spinning down to incubate for 15 to 30 minutes. In the meantime, cell culture medium was changed to 1 ml of OptiMEM for better transfection results. After incubation the DNA/jetPEI mix was added dropwise to the wells. To distribute the mix equally in the wells the plates were agitated gently (not circular to avoid concentration of cells in the middle). To prepare cells to investigate under the microscope glass cover slips.

2.2.2.6 Determination of protein concentration (BCA assay)

To compare different cell or tissue lysates and quantify amounts of detected proteins it is important to use the same concentration of proteins. To determine the concentration a bicinchoninic acid (BCA) assay was applied. Within this assay the reduction of Cu^{2+} ions to Cu^{+} in highly basic medium due to proteins is used to chelate bicinchoninic acid with Cu^{+} , which produces a purple colored complex with strong light absorption at 562 nm and proportional to the protein concentration. Therefore, lysates were treated according to the manufacturer's protocol and the respective absorption was measured to calculate the concentration (Thermo Fisher Scientific, USA, REF 23250).

2.2.2.7 Stabilization assay of proteins

The stability or turnover of proteins was tested to investigate the effect of variants compared to the wild type form. To gain this information HEK293 cells were transfected as described in 2.2.2.5. After 24 h of transfection incubation, cycloheximide (CHX) was added to the wells to inhibit the expression of new proteins. Therefore, protein levels decreased, and the different amounts could be detected after certain time points of 0h, 8h, 16 h and 24 h. The cells were then harvested with RIPA buffer, treated with DNase I to digest DNA and centrifuged (14800 x g, 5 min, at RT) to separate debris from the lysate. The samples were frozen at -20°C for later use or directly used for further experimentation like BCA assay and Western blotting. To calculate the half time of the proteins an exponential fit was applied, and the corresponding formula was used to calculate the time at 50% of maximal protein detection.

$$y = e^{-ax} \; with \; normalized \; values \; (starting \; at \; 1 \; at \; intersection \; with \; y - axis)$$

With y being the signal from protein detection follows for 50% protein (0.5 for normalized protein signals):

$$0.5 = e^{-ax} \quad \rightarrow \quad x = \frac{\ln 0.5}{-a} = T_{1/2}$$

2.2.2.8 Nuclear separation

To separate a cytosolic from a nuclear fraction to observe the distribution of specific proteins in these compartments, cells were first harvested by scraping to loosen them from the surface. After collecting the cells in a new tube, they were washed with PBS (1 ml per well in a six-well plate, centrifuged for 5 min at 1000 rpm at 4°C). In the next step the cells were treated with a hypotonic buffer to increase the inner osmotic pressure (2.1.5.3). Therefore 50 µl of the buffer

was added to approx. 5 x 10^5 cells. After resuspension and incubation for 15 minutes on ice the detergent NP40 was added in a final concentration of 0.5%. To mix the samples they were vortexed for ten seconds on the highest setting and centrifuged with 3000 rpm for 30 minutes at 4°C to separate the cytoplasmic fraction from the nuclei. Therefore, the supernatant was collected in a new tube and stored at -20°C for further use. The pellet containing the nuclei was then resuspended in 50 µl RIPA buffer with protease inhibitor for 30 minutes on ice, vortexing the mix every ten minutes. To separate the nuclear fraction from debris and precipitated DNA the samples were centrifuged for 30 minutes at 14000 x g and 4°C and the supernatant was transferred to a new tube for storage at -20°C or further experimental use.

2.2.2.9 Immunofluorescence staining

To stain different proteins in HEK293 cultures on cover slips cells were fixated with 4% PFA and incubated for 30 min at 4°C (2.1.5.4). From this point on every step was carried out at RT. After the fixation the cells were washed three times with PBS for five to ten minutes and the membrane of the cells was permeabilized for ten minutes with a solution of 0.1% saponin and 3% BSA solved in PBS buffer. After three further washing steps in PBS for five to ten minutes the samples were blocked with 3% BSA in TBS for one hour. Again, a washing step was applied with PBS for five to ten minutes and the primary antibody was added to the samples in the corresponding dilution released by the manufacturer. For incubation the samples were left in the antibody dilution for one hour with four additional washing steps afterwards. The secondary antibodies, conjugated with a fluorophor, were added for another 20 minutes. During the first of the last four washing steps DAPI was added to PBS in a 1:300 dilution to mark the nuclei. After washing the cover slips in water to remove salt remnants, they were embedded onto object slides with fluoromount-G (2.1.5.4). After drying the samples were ready for observation.

2.2.2.10 Preparation of human heart tissue

Cardiac biopsies from patients with severe CMP phenotypes were obtained during cardiac surgery. To analyze the protein and mRNA expression levels of these samples they were frozen with liquid nitrogen and mechanically crushed. For measuring mRNA levels, total RNA isolation was applied and therefore the crushed samples were collected with a cooled spatula, weighed and solved in 1 ml of TRizol reagent (2.1.5.2). Further steps are described in the section RNA isolation (2.2.2.9). For the investigation of proteins in these samples 100 µl of RIPA buffer mixed with protease inhibitor (Pierce, Thermo Fisher Scientific, Waltham, USA) was used instead of TRizol. Afterwards the samples were homogenized with an ultra turrax

disperser. A centrifugation step pelleted the debris and the protein concentration was measured with a BCA assay (2.2.2.6).

2.2.2.11 RNA isolation

To isolate RNA cells seeded in a six-well plate or prepared cardiac tissues were harvested with adding 1 ml of TRizol (2.1.5.2) to the samples. Then the cells/tissues of interest were centrifuged for five minutes at full speed and RT (14800 x rpm) before starting with the RNA-isolation and to purify the sample from cell debris (cell pellet). To separate RNA from other cell components 0.2 ml chloroform was added to the mix and centrifugation (10000 x g for 15 min at 4°C), after vigorously shaking the samples for at least 15 seconds, led the separation of three different phases (upper aqueous phase, interphase, lower chloroform phase). The upper phase, which contained the RNA, was then pipetted into a new RNase-free tube. The lower organic and the interphase, which contained DNA and proteins, were discarded. For precipitation of RNA 0.5 ml isopropanol was added to the samples and incubated for ten minutes. After a further centrifugation step for ten minutes with 10000 x g at 4°C a white RNA pellet should be visible if the amounts were sufficient. In the next step the samples were washed with 1 ml of 75% Ethanol (diluted in DEPC water) and centrifuged for five minutes at RT with 21100 x g (full speed) followed by discarding the supernatant and letting the samples dry at 80°C. The pellets were then solved in 100 to 200 µl of DEPC water and stored at -20°C or directly used for experimentation.

2.2.2.12 Real time quantitative PCR

Expression levels of mRNA were measured using the standard method real time qPCR. Therefore, tissue of interest was prepared as described in 2.2.2.8 and 2.2.2.9. Transcription of RNA to cDNA was done according to the standard protocol of the PrimeScript RT reagent kit (Takara, Kusatsu Japan). For the detection of cDNA the TB Green Premix Ex Taq II from Takara was used (Kusatsu, Japan). After mixing reaction reagents with the cDNA samples and primers DNA levels were measured with a Taqman 7500 (2.1.5.2). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an endogenous control and the detected signals were normalized with the ΔΔCT method (Kolanczyk et al., 2011; Seifert et al., 2009). Different primers to amplify die cDNA during the PCR are listed in 2.1.5.2.

2.2.2.13 Western Blotting

The presence or absence and the distribution of proteins can indicate mechanisms for pathological development and functions of specific proteins. Therefore, Western blots were

performed. First protein samples were loaded onto SDS-PA gels. After running the gels the proteins were loaded onto PVDF membranes using a wet blot system from Bio-Rad (2.1.5.3). Blocking of membranes was done with 5% milk powder and 0.2% NP-40 in TBS-T buffer (blocking buffer). After incubation with the primary antibody for 3 h or overnight in blocking buffer (used primary antibodies in 2.1.5.3) the membranes were washed three to four times with TBS-T for 10 min. In the next step the membranes were incubated with the secondary antibody (conjugated with HRP) for 1 h in blocking buffer and again washed three to four times with TBS-T for 10 min. For detection of the marked proteins a mix of luminol/ECL enhancer and peroxide was added to the membrane in a 1:1 ratio as reaction mix for the HRP (2.1.5.3). After incubation for 1 min the membranes were drained from excess reagent and the emitted light was detected with a Chemiluminescence imaging system CHEMI only system.

2.2.2.14 Prediction of post-translational modification sites and structural elements

To predict structural features and functional interaction sites, online prediction tools were used. Therefore, the amino acid chains from proteins of interest were loaded into the programs. For post-translational modification (PTM) tools common tools were used to predict sites for acetylation, methylation, phosphorylation, etc. (Audagnotto & Dal Peraro, 2017, http://gps.biocuckoo.cn). Additional information for structural and conserved features was collected from programs like PSIPRED, RaptorX, Spider2 or ConSurf. All applied online tools with positive results are listed in the supplement (Table 2). PTM prediction was always applied with the highest stringency. With more than one tool per PTM category only hits at the same amino acid position with ≥ half of the applied prediction programs were considered as true hit. Exceptions are marked in figures. For the secondary structure the most frequent predicted structure (α-helix, β-sheet, coil) was determined for each amino acid position. The average of normalized values for disorder were used to predict ordered regions in proteins. Only ConSurf was used for conservational features. The same procedure was applied for accessible surface are (ASA).

Table 2: Online Prediction tools used to determine possible functional sites or domains from proteins of interest (SUMOylation coming from small ubiquitin-related modifier – SUMO)

Category	Name	Website
General PTM prediction	ProSite	https://prosite.expasy.org/
	PhosphoSitePlus	https://www.phosphosite.org/psrSearchAction
	dbPTM	http://dbptm.mbc.nctu.edu.tw/
Acetylation	NetAcet	http://www.cbs.dtu.dk/services/NetAcet/
	PAIL	http://bdmpail.biocuckoo.org/prediction.php
	GPS-PAIL	http://pail.biocuckoo.org/
	PSKAcePred	http://bioinfo.ncu.edu.cn/inquiries_PSKAcePred.aspx
	ASEB	http://bioinfo.bjmu.edu.cn/huac/predict_p/
Glycosylation	GLYCOPP1.0	http://crdd.osdd.net/raghava/glycopp/submit.html
	NetNGlyc1.0	http://www.cbs.dtu.dk/services/NetNGlyc/
	NetOGlyc4.0	http://www.cbs.dtu.dk/services/NetOGlyc/
	YinOYang1.2	http://www.cbs.dtu.dk/services/YinOYang/
Methylation	PMeS	http://bioinfo.ncu.edu.cn/inquiries_PMeS.aspx
	GPS-MSP	http://msp.biocuckoo.org/
	PSSME	http://bioinfo.ncu.edu.cn/PSSMe.aspx
	iMethyl-PseACC	http://www.jci-bioinfo.cn/iMethyl-PseAAC
S-Nitrosylation	iSNO-PseAAC	http://app.aporc.org/iSNO-PseAAC/
	GPS-SNO	http://sno.biocuckoo.org/
Phosphorylation	ScanSite	https://scansite4.mit.edu/4.0/#scanProtein
	KinasePhos2.0	http://www.jci-bioinfo.cn/pSumo-CD
	ProSite	https://prosite.expasy.org/
	GPS	http://gps.biocuckoo.cn/online.php
SUMOylation	GPS-SUMO	http://sumosp.biocuckoo.org/online.php
	JASSA	http://www.jassa.fr/
	pSUMO-CD	http://www.jci-bioinfo.cn/pSumo-CD
	SUMOAMVR	http://bioinfo.ncu.edu.cn/SUMOAMVR_Prediction.aspx
Ubiquitination	iUbiq-Lys	http://www.jci-bioinfo.cn/iUbiq-Lys
	UbiProber	http://bioinfo.ncu.edu.cn/ubiprober.aspx
	UbPred	http://www.ubpred.org/
Conservation	ConSurf	http://consurf.tau.ac.il/
Accessible Surface Area	NetSurf2.0	http://www.cbs.dtu.dk/services/NetSurfP/
	Spider2	http://sparks-lab.org/server/SPIDER2/
Disorder	DisoPred	http://bioinf.cs.ucl.ac.uk/psipred/
	IUPred	https://iupred2a.elte.hu/plot
Secondary Structure	PsiPred	http://bioinf.cs.ucl.ac.uk/psipred/
	NetSurf2.0	http://www.cbs.dtu.dk/services/NetSurfP/
	SPIDER2	http://sparks-lab.org/server/SPIDER2/
	JPRED	http://www.compbio.dundee.ac.uk/jpred/index.html
	RaptorX	http://raptorx.uchicago.edu/

3 Results

3.1 Phenotypic and genetic characterization of pediatric patients affected by primary cardiomyopathy

To investigate the genetic mechanisms of primary CMPs in pediatric cases not affected by systemic or syndromic disorders a prospective study at the DHZB and Charité Berlin was established. During enrollment of this study 83 pediatric patients (age ≤ 18 years) were recruited and examined for analysis of their cardiac phenotype and the underlying genetic variation. 80 subjects of this cohort (CMP-80) were genotyped with NGS (3.1.1/3.1.2), whereas a sub cohort of 60 patients in this study (CMP-60) underwent additional follow-up examinations (3.1.3, overlap of 57 patients). For all index patients the pathogenicity of the genetic information originated from NGS was determined according to the guidelines of the ACMG (Figure 6) and available family members were checked for segregation of the detected CMP-relevant variants.

3.1.1 Characteristics of pediatric patients with primary cardiomyopathy

During the study period 80 patients (CMP-80) were recruited for testing genetic variants in 89 CMP-relevant genes included in the Illumina TruSight Cardio Panel. Therefore, cases affected by DCM (34), HCM (23), LVNC (14), RCM (7) and ARVC (2) were analyzed. Only two individuals had ARVC and were excluded in further analysis. The overall gender distribution was almost equal with 38 females (f) and 42 males (m), but differentiated for the single CMP´s, most notably for DCM with 23 females and 11 males and HCM with 5 females to 18 males. The mean age of the pediatric cohort was 4.8 years. In general DCM, LVNC, RCM and ARVC patients were slightly younger than the average age, whereas HCM patients were older and had by far the widest range in the distribution of age (IQR = 0.6 – 12.8). Within the cohort CMP-80, 63 patients were younger than 10 years at the date of diagnosis (79%) including 34 patients younger than 1 year (43%). Considering these age groups the distribution of the different phenotypes did not differ drastically (Figure 7 A - C) although the rate of DCM cases decreased slightly with ageing in comparison to the total cohort, whereas the rate of HCM patients increased. With the shortest range in the distribution of age (IQR = 1.5 – 4.2 years) the RCM subgroup had the most defined age and therefore appeared in the age group younger than 10 years. Family screening was available in 61 families with positive CMP results in 27 (available family members in 76.3 % of all patients), where at least one additional family member was affected by CMP. The highest number of positive family screenings was observed in the subgroup of HCM with 12 (52.2%) out of 19 families (82.6%), whereas the lowest number of

familial CMP was observed in the subgroup of DCM with 8 (23.5%) out of 23 families included (67.7%). For a short clinical characterization IVSd, LVEDD, LV-EF and FS are listed in Table 3. The IVSd, presenting the thickness of the septum, was increased for the HCM subgroup, displaying hypertrophied myocardium (Z-score of 10.6).

Table 3: Characterization of CMP-phenotypes during enrollment and diagnosis of pediatric cohort CMP-80

Phenotype (n)	All (80)	DCM (34)	HCM (23)	LVNC (14)	RCM (7)	ARVC (2)
Sex						
Females, n (%)	38 (47.5)	23 (67.7)	5 (21.7)	7 (50)	3 (42.9)	0 (0)
Males, n (%)	42 (52.5)	11 (32.3)	18 (78.3)	7 (50)	4 (57.1)	2 (100)
Age						
in years, mean	4.8	3.2	7.2	3.9	4.4	n.a.
IQR	0.2-9.0	0.1-4.8	0.6-12.8	0.2-8.2	1.5-4.2	n.a.
Family screening						
Included, n (%)	61 (76.3)	23 (67.7)	19 (82.6)	12 (85.7)	5 (71.4)	n.a.
Positive result, n (%)	27 (33.8)	8 (23.5)	12 (52.2)	5 (35.7)	1 (14.3)	n.a.
IVSd						
mean mm (±SD)	9.1 (7.6)	5.5 (1.8)	16.7 (9.6)	5.5 (1.3)	5.0 (1.3)	n.a.
mean z-score (±SD)	3.6 (6.8)	0.4 (1.6)	10.6 (8.4)	0.4 (0.6)	0.2 (0.8)	n.a.
LVEDD						
mean mm (±SD)	39.4 (13.0)	46.2 (13.7)	33.8 (11.1)	38.6 (7.1)	26.0 (4.4)	n.a.
mean z-score (±SD)	2.8 (5.1)	7.0 (4.2)	-1.4 (2.8)	2.5 (3.0)	-2.0 (0.7)	n.a.
LV-EF, mean % (±SD)	46.5 (22.4)	29.7 (15.4)	66.8 (14.7)	44.0 (19.5)	62.1 (12.6)	n.a.
FS, mean % (±SD)	27.6 (16.8)	15.1 (6.7)	44.7 (16.5)	26.2 (11.5)	31.5 (11.6)	n.a.
MACE, n (%)	39 (48.8)	22 (64.7)	8 (34.8)	3 (21.4)	6 (85.7)	n.a.
AICD, n (%)	8 (10)	1 (2.9)	7 (30.4)	0 (0)	0 (0)	n.a.
VAD/BVAD, n (%)	22 (27.5)	16 (47.1)	1 (4.4)	2 (14.3)	3 (42.9)	n.a.
ECMO, n (%)	6 (7.5)	5 (14.7)	0 (0)	0 (0)	1 (14.3)	n.a.
HTX						
with enrollment, n (%)	17 (21.3)	13 (38.2)	0 (0)	3 (21.4)	1 (14.3)	n.a.
during study, n (%)	9 (11.3)	4 (11.8)	1 (4.4)	0 (0)	4 (57.1)	n.a.
Deceased, n (%)	5 (6.3)	2 (5.9)	1 (4.4)	0 (0)	2 (28.6)	n.a.

Abbreviations: IVSd - interventricular septum thickness at end diastole, LVEDD - left ventricular end diastolic diameter (Z score, normal reference range -2 < +2, m-mode), LV-EF - left ventricular ejection fraction (auto 4CH monoplan) , FS - fractional shortening, MACE - Major adverse cardiovascular events, AICD - automatic implantable cardioverter defibrillator, VAD - ventricular assist device, BVAD - biVAD, ECMO - extracorporeal membrane oxygenation, HTX - heart transplantation; n.a. - not applicable, SD - standard deviation

The LVEDD and LV-EF are measurements for the function of the left ventricle and pose the diameter at the end diastole and the fraction of blood pumped out. The LVEDD was increased

and the LV-EF was decreased for DCM and LVNC (LVEDD Z-Score of 7 and 2.5, LV-EF of 29.7% and 44%, respectively). The fractional shortening is a term for the ratio between LVEDD and LVESD (normal range: 25% - 45%, LVESD – left ventricular end systolic diameter). Therefore, changes to the norm are another characteristic for DCM and was down to 15.1% for the DCM subgroup. The FS for LVNC was slightly reduced (26.2%). Other criteria not mentioned in Table 3 but used for the distinction of the CMPs are trabeculation and non-compacted myocardium for LVNC and remodeling of the atria and restricted movement of the heart for RCM.

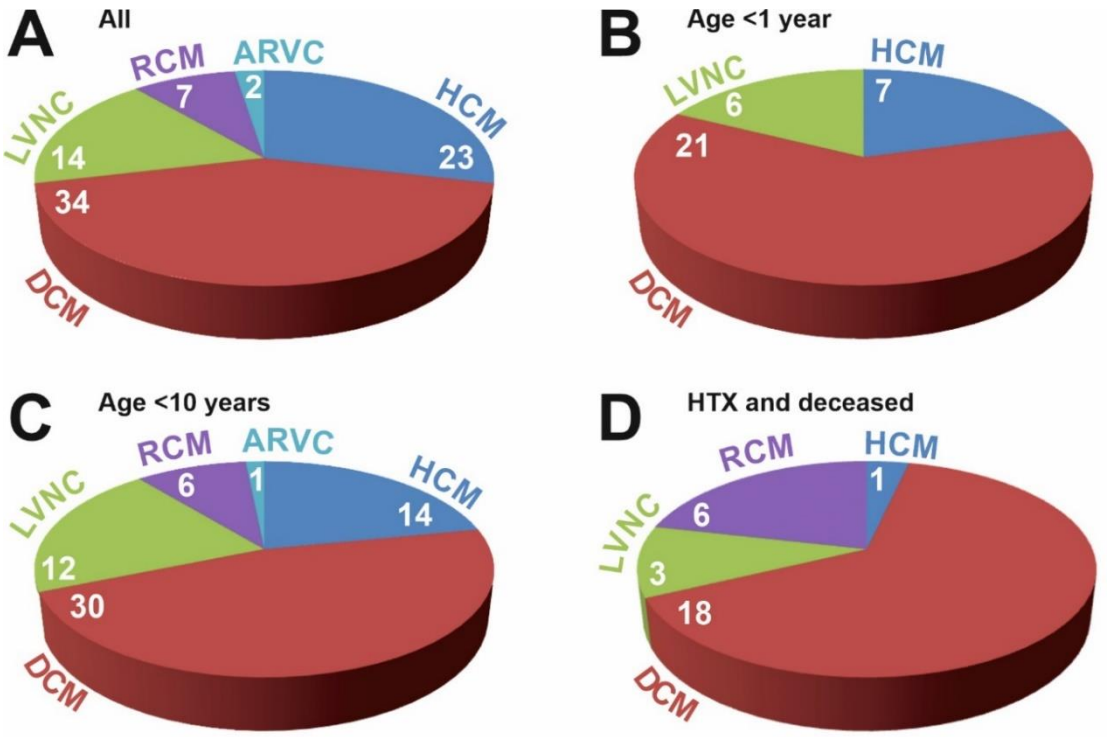


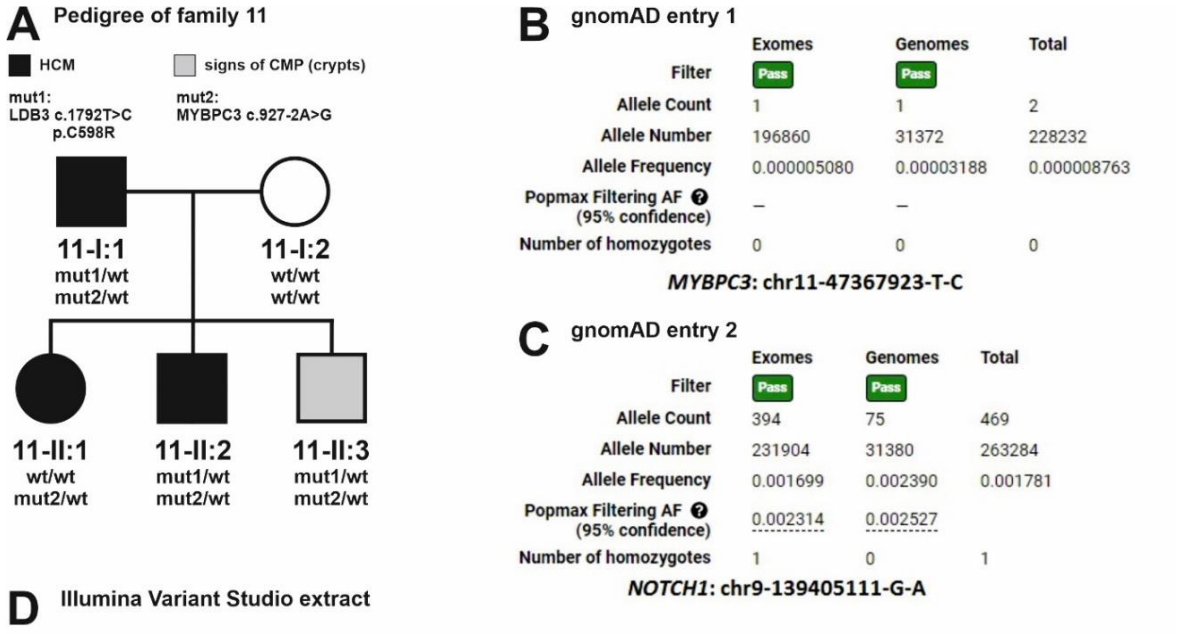
Figure 7: Distribution of CMP phenotypes in cohort with 80 primary and pediatric CMP patients (CMP-80). **(A)** All patients with indicated CMP subgroup. **(B)** Patients with an age <1 yearsindicated CMP subgroup. **(C)** Patients with an age <10 years with indicated CMP subgroup. **(D)** Patients who underwent heart transplantation (HTX) or deceased patients with indicated CMP subgroup.

To assess the severity of the phenotypes the events implantation of an automatic implantable cardioverter defibrillator (AICD), a ventricular assist device (VAD) or a device for extracorporeal membrane oxygenation (ECMO), heart transplantation and death were summarized as major adverse cardiovascular events (MACE). In total 39 patients (48.8%) were affected by MACE. With 85.7% and 64.7% RCM and DCM patients had the highest rates and therefore a more severe disease progress. In contrast, the percentage of MACE in individuals with HCM and LVNC was considerably lower with 34.8% and 21.4%, respectively. Especially the occurrence of HTX and death was of notice. DCM and RCM patients had rates

of 53% and 85.7% in this category whereas the average was 36.4% (Figure 7 D, distribution of CMP phenotypes in the subgroup of HTX and deceased patients).

3.1.2 Distribution and pathogenicity of rare genetic variants

The analysis process is exemplarily shown for family 11 (Figure 8). In this family with five members three were affected by HCM (11-I:1/11-II:1/11-II:2, Figure 8 A) whereas one showed first signs of a CMP (11-II:3). Therefore, only the mother (11-I:2) was unaffected. After the signals from the NGS runs were aligned to the corresponding genes, the data was further processed with Variant Studio to eliminate synonymous variants, technical artefacts and variants occurring in variable regions and repeats from rare missense, in frame indel or truncating variants (nonsense, frameshift or splice site).



D Illumina Variant Studio extract

Gene	Variant	Chr	Coordinate	Genotype	Allelic Depths	Consequence	HGVSc	HGVSp
ABCC9	CA>CA/C	12	22063251	het	118,115	splice_region_variant	NM_020297.2:c.1165-6delT	
ABCC9	G>G/T	12	22066003	het	172,77	splice_region_variant	NM_020297.2:c.817-3C>A	
ALMS1	TGGAGGA>TGGAGGAGGA/T	2	73613031	het	21,33,110	inframe_deletion	NM_015120.4:c.69_74delGGAGGA	NP_055935.4:p.Glu27_Glu28del
ALMS1	TGGAGGA>T/TGGAGGAGGA	2	73613031	het	21,33,110	inframe_insertion	NM_015120.4:c.72_74dupGGA	NP_055935.4:p.Glu28dup
ALMS1	G>G/T	2	73678130	het	173,76	missense_variant	NM_015120.4:c.4473G>T	NP_055935.4:p.Glu1491Asp
ALMS1	A>A/T	2	73678393	het	172,76	missense_variant	NM_015120.4:c.4736A>T	NP_055935.4:p.Asn1579Ile
GAA	A>AGCAGCGG/AGCAGCGG	17	78081526	hom	2,240	splice_region_variant	NM_000152.3:c.858+7_858+8insAGCGGGC	
KCNH2	T>T/G	7	150644513	het	115,132	missense_variant	NM_000238.3:c.3055A>C	NP_000229.1:p.Thr1019Pro
LDB3	T>T/C	10	88477836	het	134,112	missense_variant	NM_001171610.1:c.1807T>C	NP_001165081.1:p.Cys603Arg
MYBPC3	T>T/C	11	47367923	het	118,125	splice_acceptor_variant	NM_000256.3:c.927-2A>G	
MYH6	AG>AG/A	14	23858271	het	99,139	splice_region_variant	NM_002471.3:c.3979-8delC	
MYLK	T>T/G	3	123419215	het	163,83	missense_variant	NM_053025.3:c.3100A>C	NP_444253.3:p.Thr1034Pro
NOTCH1	T>T/G	9	139399311	het	142,108	missense_variant	NM_017617.3:c.4832A>C	NP_060087.3:p.His1611Pro
TRDN	G>G/GA	6	123581789	het	75,92	splice_region_variant	NM_006073.3:c.1805-7dupT	

Figure 8: Example for analysis with data generated from NGS. (A) A pedigree of family 11 affected by HCM is presented. **(B)/(C)** Two gnomAD entries of variants detected in family 11 are shown. **(D)** NGS data was analyzed with Variant Studio from Illumina. Shown are all detected variants for family 11.

In Figure 8 D there are 14 variants listed for family 11. All genetic alterations with a MAF of higher than 0.001 in ExAC, an alternative variant frequency of lower than 30% (all sequences covering the same region presented the alternative allele with less than 30%) and detected as synonymous variants were previously excluded. After elimination of false positives only the variants in *MYBPC3* and *LIM domain binding 3* (*LDB3*) were considered as VOI. Examples for false positives are the deletion of an adenine (A) at the end of an A repeat in *ATP binding cassette subfamily C member 9* (*ABCC9*, Figure S 4) and the exchange of thymine to guanine (G) between two G repeats in *potassium voltage-gated channel subfamily H member 2* (*KCNH2*, Figure S 5). Strong arguments for both variants not to be considered for further analysis are the appearance in controls and the relatively low alternative variant frequency (near 30%). In comparison the variants in *LDB3* and *MYBPC3* are absent in controls and both alleles are distributed almost evenly with 3% and 1% difference (Figure S 6, Figure S 7). Because in Variant Studio only data from ExAC was available the MAF of all relevant variants were checked again in gnomAD (Figure 8 B/C). In the case of *MYBPC3* the MAF was considerably smaller than the threshold of 0.001 whereas the MAF of the *notch receptor 1* (*NOTCH1*) variant was above the threshold and was therefore excluded from further analysis. Families were then checked for segregation. In the case of family 11 *MYBPC3* NM_000256.3:c.927-2A>G segregated with all affected family members.11-II:1 is not carrying

Table 4: Characterization of genetic VOI detected in pediatric CMP-cohort CMP-80

	Pathogenic[1]	**Likely pathogenic[1]**	**Uncertain significance[1]**	**Total**
Variants according to phenotype, *n*	**17 (14%)**	**15 (12%)**	**94 (74%)**	**126 (100%)**
DCM	4	4	44	52
HCM	11	2	26	39
LVNC	1	4	13	18
RCM	1	2	10	13
ARVC	0	3	1	4
Category of variants, *n*				
De novo	5	2	0	7
Not *de novo* (inherited)	4	9	55	68
Novel[2]	2	9	42	53
Not novel (known)	15	6	52	73
Missense	11	9	83	103
Indel/frameshift	2	4	3	9
Stop gain	1	1	1	3
Splice site	3	1	7	11

[1]Classification according to Richards et al., 2015

[2]Novel indicates variants that have not been annotated in genetic disease reference databases ClinVar, ACGV, HGMD

LDB3 NM_001171610.1:c.1807T>C indicating that the variant does not segregate with the disease phenotype in all family members. With a complete and checked dataset for each patient and family the pathogenicity of the variants was determined according to the ACMG guidelines (Figure 6). All 80 patients and family members were analyzed in the same way.

In all patients of the CMP-80 cohort 126 VOI were detected (Table 4). 17 and 15 variants were classified as pathogenic (P) and likely pathogenic (LP, 14% and 12%). Variants of uncertain significance occurred 94 times (74%). Most P and LP variants were detected in HCM patients, whereas individuals affected by DCM were carrying the most VUS variants. Seven of all VOI were detected as *de novo*. Five of these variants were interpreted as P and two as LP. With 58% over half of all VOI were already detected in disease databases like ClinVar. The frequency of novel variants considered as P/LP or VUS did not differ considerably with 34% and 45%. Not surprisingly most VOI were caused by single missense exchanges (103). Stronger alterations such as frameshift or nonsense variants were much rarer, but also shifted to be P/LP (frameshift: six out of nine P/LP; Stop gain: two out of three P/LP; splice site: four out of 11 P/LP).

VOI were then considered per number of patients. 1.58 variants were detected per patient. In the subgroups these numbers differed slightly (DCM: 1.53, HCM: 1.7, LVNC: 1.29, RCM: 1.86). In 16 patients no VOI were observed and 36 patients were carrying more than one VOI (1 VOI, n = 28; 2 VOI, n = 20; 3 VOI, n = 12; >3 VOI, n = 4). P and LP variants occurred in 30 of 80 index patients (38%). Complex genotypes were only detected in two individuals carrying two LP variants each (Table S 5). In 21 patients with and 34 patients without a P/LP variant (26%, 43%) sequencing resulted in an additional VUS variant. P and LP variants were detected with a higher rate in HCM patients with over half of all patients carrying at least one variant (0.57 variants per patient; values for other CMP: DCM 0.24 variants per patient, LVNC: 0.357 variants per patient, RCM: 0.429 variants per patient).

3.1.3 Spectrum of rare variants

A more detailed analysis of the VOI detected in the pediatric CMP-80 cohort showed that mainly changes in major CMP genes were observed. Thus, VOI were most frequently identified in *MYH7* (16%), *MYBPC3* (9%), *troponin I3, cardiac type* (*TNNI3*, 7.5%), *DSP* (7.5%), *LDB3* (6%), and myopalladin (*MYPN*, 6%) (Figure 9). The order was slightly changed in the different subgroups, but *MYH7* remained the most frequent gene except for RCM patients (Figure 9 B - E). In individuals diagnosed with RCM truncating *TTN* variants were the most frequent with three splice site variations. In general truncating *TTN* DNA changes were very rare with one additional single base deletion in a DCM case (4). The distribution of rare variants in the HCM group was very distinct and similar to the general cohort with *MYH7*, *MYBPC3* and *LDB3* the

most abundant genes with observed variation. With VOI in 31 different genes the subdivision of DCM patients was the most heterogeneous one.

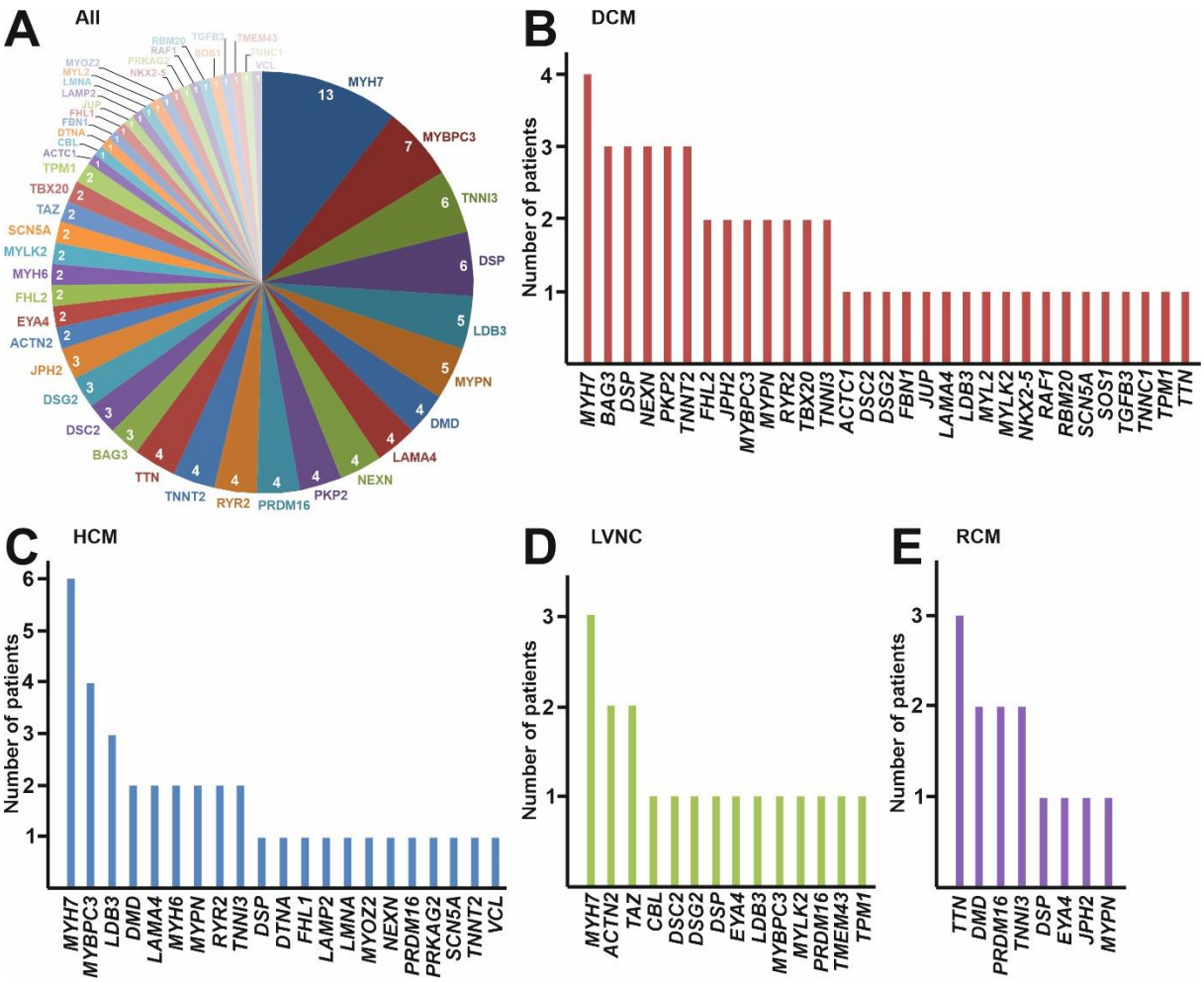


Figure 9: Distribution of genes with VOI in CMP patients (CMP-80). (A) The number of CMP patients with a VOI in the indicated CMP gene is shown in a pie chart. **(B)** The number of patients affected by DCM **(B)**, HCM **(C)**, LVNC **(D)** and RCM **(E)** with a VOI in the indicated CMP gene is shown (full name of genes in **Table S 1**).

To get an overview of the general mechanisms for the manifestation of CMP in this cohort all 89 CMP genes were pooled into functional groups (Table S 1). As expected, sarcomere genes were most frequent in pediatric CMP patients in general and likewise in the subgroups (Figure 10 A - E). Surprisingly VOI in desmosome genes were second most frequent in the complete cohort and mainly detected in DCM and LVNC patients (in total cohort: 20%, in DCM subgroup: 26%, in LVNC subgroup: 21%). Nearly as frequent as VOI in the desmosome group were VOI detected in Z-disc genes. Sarcomere and Z-disc genes mainly contribute to the build-up of the sarcomere, the structural element of muscle cells, and were most abundantly detected with VOI in the HCM group (sarcomere genes: 57%, Z-disc: 26%). Additionally of notice is the fact, that some functional categories were specific to the CMP groups. Patients with DCM were the

only ones with VOI in the category protein quality control, whereas VOI in the group of dystrophin complex were very specific to individuals affected by HCM and RCM. VOI in genes affecting the mitochondria were only observed in the subgroup of LVNC. The group transcription/splicing was most specific for RCM and DNA changes for extracellular matrix (ECM) genes were only observed in DCM and HCM patients. The distribution of rare variants in the different functional groups from patients younger than one year was similar to the complete cohort with only minor differences (Figure 10 F). The order from highest to lowest frequency of the groups was nearly the same starting with sarcomere (49% to 35%), desmosomes (20% to 24%) and Z-disc (19% to 18%) and only considerably different in cellular signaling (6% to 15%). Of note are possible overrepresentations of groups comprising a large number of genes or exons in Figure 10.

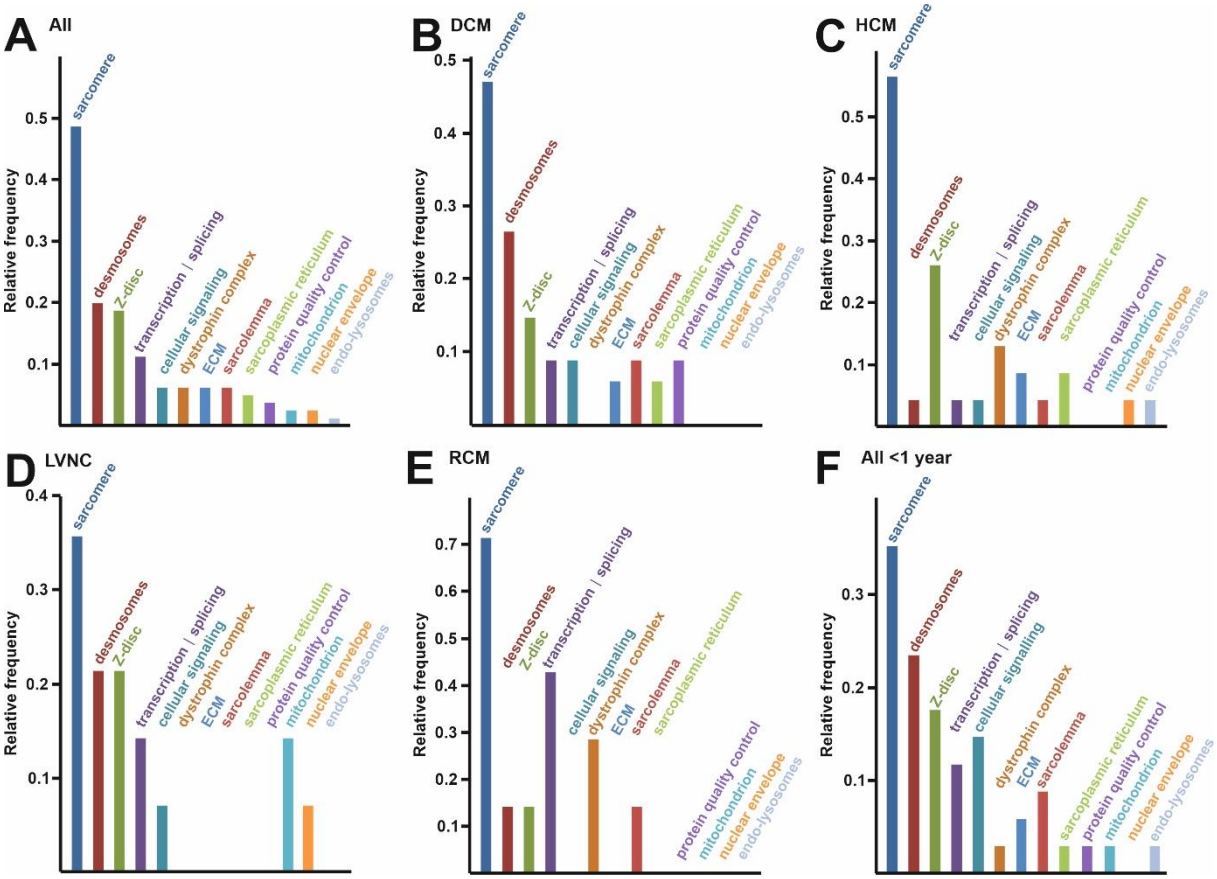


Figure 10: Relative frequency of VOI in CMP genes detected in samples from CMP patients (CMP-80) divided into functional groups. The relative frequency of VOI in functional groups is shown for all **(A)**, DCM **(B)**, HCM **(C)**, LVNC **(D)** and RCM **(E)** patients and patients with age < 1 year **(F)**. The classification of CMP genes into functional groups can be seen in **Table S 1.**

Most VOI were detected in the heterozygous state. Exceptions were observed with two homozygous (*nexilin* abbr. *NEXN*, *TNNI3*) and two compound heterozygous variants (*desmocollin 2* abbr. *DSC2*, *MYBPC3*). The case involving the *MYBPC3* variants was

considered notable, because one allele was affected by the exchange p.S858R and the other allele was completely deleted (Table S 5). Hemizygous variants occurred in *dystrophin* (*DMD*, 1), *four and a half LIM domains 1* (*FHL1*, 1) and *tafazzin* (*TAZ*, 2).

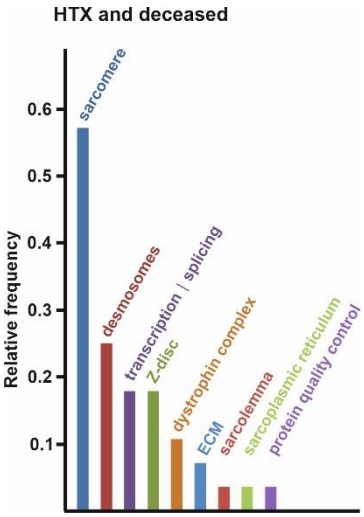


Figure 11: Relative frequency of VOI in CMP genes detected in samples from HTX and deceased patients from the CMP cohort (CMP-80) divided into functional groups. In addition to **Figure 10** the distribution of VOI categorized into functional groups is shown for severe CMP cases who underwent heart transplantation or were deceased.

HTX and deceased patients (n = 28) represented a faction of this cohort with a very severe form of CMP and are therefore separately mentioned (Figure 7 D). Most frequently VOI were detected in *DSP* (4 VUS), *TTN* (2 LP, 2 VUS), *DMD* (3 VUS), *NEXN* (3 VUS), *PRDM16* (3 VUS) and *TNNI3* (1 LP, 2 P). Thus, all four truncating *TTN* variants in the CMP-80 cohort were also present in the HTX and deceased patients more precisely in 3 cases of RCM and in one case of DCM (Table S 5). Compared to all patients the frequency of VOI in the functional groups was similar with most detected VOI in the groups of sarcomere, desmosomes, transcription/splicing and Z-disc, although the values for transcription/splicing and desmosomes were slightly increased and frequencies for cellular signaling, mitochondria, nuclear envelope and endo-lysosomes were absent. Two or more VOI occurred in 15 HTX and deceased patients, seven were carrying one VOI and six individuals were observed to carry no VOI.

3.1.4 Clinical course of early onset cardiomyopathy in a selected pediatric cohort

To further investigate the process and severity of CMPs in children a selected group of the pediatric cohort CMP-80 was examined in follow up investigations (CMP-60). To calculate a value for the development of the disease the occurrence of major adverse cardiovascular events (MACE) like implantation or application of defibrillators, ventricular assist devices, heart transplantation or death was observed over time. 60 CMP patients were repeatedly examined. This sub cohort included 21 cases of DCM, 17 cases of HCM, 15 cases of LVNC, five cases

of RCM and two cases of ARVC. The phenotypic characteristics were similar to the general cohort. Therefore, the gender was slightly shifted to males with exceptions for DCM (slightly shifted to females: 12 f, 9 m) and HCM (drastically shifted to males: 4 f, 13 m). The mean age was 5.4 years and again the group of HCM was an outlier with 7.5 years (IQR = 0.3 – 9.1 years).

Table 5: Characterization of selected CMP patients (CMP-60) for correlation of major adverse cardiovascular events (MACE) over time

Phenotype (n)	All (60)	DCM (21)	HCM (17)	LVNC (15)	RCM (5)	ARVC (2)
Sex						
Females, n (%)	25 (41.7)	12 (57.1)	4 (23.5)	7 (50)	2 (42.9)	0 (0)
Males, n (%)	35 (58.3)	9 (42.9)	13 (76.5)	8 (50)	3 (57.1)	2 (100)
Age						
in years, mean	5.4	4.3	7.5	3.5	5.5	n.a.
IQR	0.3-9.1	0.2-6.2	1.5-12.1	0.2-6.0	1.8-4.6	n.a.
MACE, n (%)	**24 (40)**	**11 (52.4)**	**5 (29.4)**	**3 (20)**	**5 (100)**	**n.a.**
AICD, n (%)	6 (8.3)	1 (4.8)	5 (23.5)	0 (0)	0 (0)	n.a.
VAD, n (%)	13 (21.7)	8 (38.1)	0 (0)	2 (13.3)	3 (60.0)	n.a.
ECMO, n (%)	3 (5.0)	2 (9.5)	0 (0)	0 (0)	1 (20.0)	n.a.
HTX						
with enrollment, n (%)	8 (13.3)	5 (23.8)	0 (0)	3 (20.0)	0 (0)	n.a.
during study, n (%)	8 (13.3)	4 (19.0)	0 (0)	0 (0)	4 (80.0)	n.a.
Deceased, n (%)	2 (3.3)	0 (0)	0 (0)	0 (0)	2 (40.0)	n.a.

Regarding MACE's most events occurred in DCM patients although RCM cases showed the highest frequency (11 from 21 DCM and five from five RCM patients with MACE's, Figure 12 B). Due to arrhythmias often occurring in HCM patients, defibrillators were mostly implanted in this group (HCM: five, DCM: one). For preventive reasons and due to no recorded shocks from these devices AICD's were not included for further calculations. Otherwise other events were absent with HCM. The requirement to introduce supporting circulatory devices like ventricular assist device or extracorporeal membrane oxygenation was often leading to further procedures and therefore ending in HTX or death. Deceased patients were only present in the RCM group. ARVC patients were not analyzed due to low sample size.

With these results Kaplan-Meier plots were generated and analyzed (Figure 12). Hence, event-free survival in percent was correlated with age in years. Adverse events were observed in 32% of index patients (Table 5, Figure 12, be aware of defibrillator implantations not taken into consideration, because of preventive character). In Figure 12 B the probability of event-free

survival was investigated regarding the different CMP subgroups. As described RCM had the lowest rate followed by DCM, LVNC and HCM. With a p-value of p<0.001 the groups had a significantly different outcome for adverse events. Also of interest was the disease course with different numbers of VOI detected in the index patients (Figure 12 C). Individuals with >1 VOI observed, had a significant higher rate for adverse cardiac events (p-value = 0.021). Furthermore the difference in outcome for sporadic and genetic CMP was tested and resulted in worse prognosis for genetic CMP patients (not significantly: p-value = 0.305, Figure 12 D).

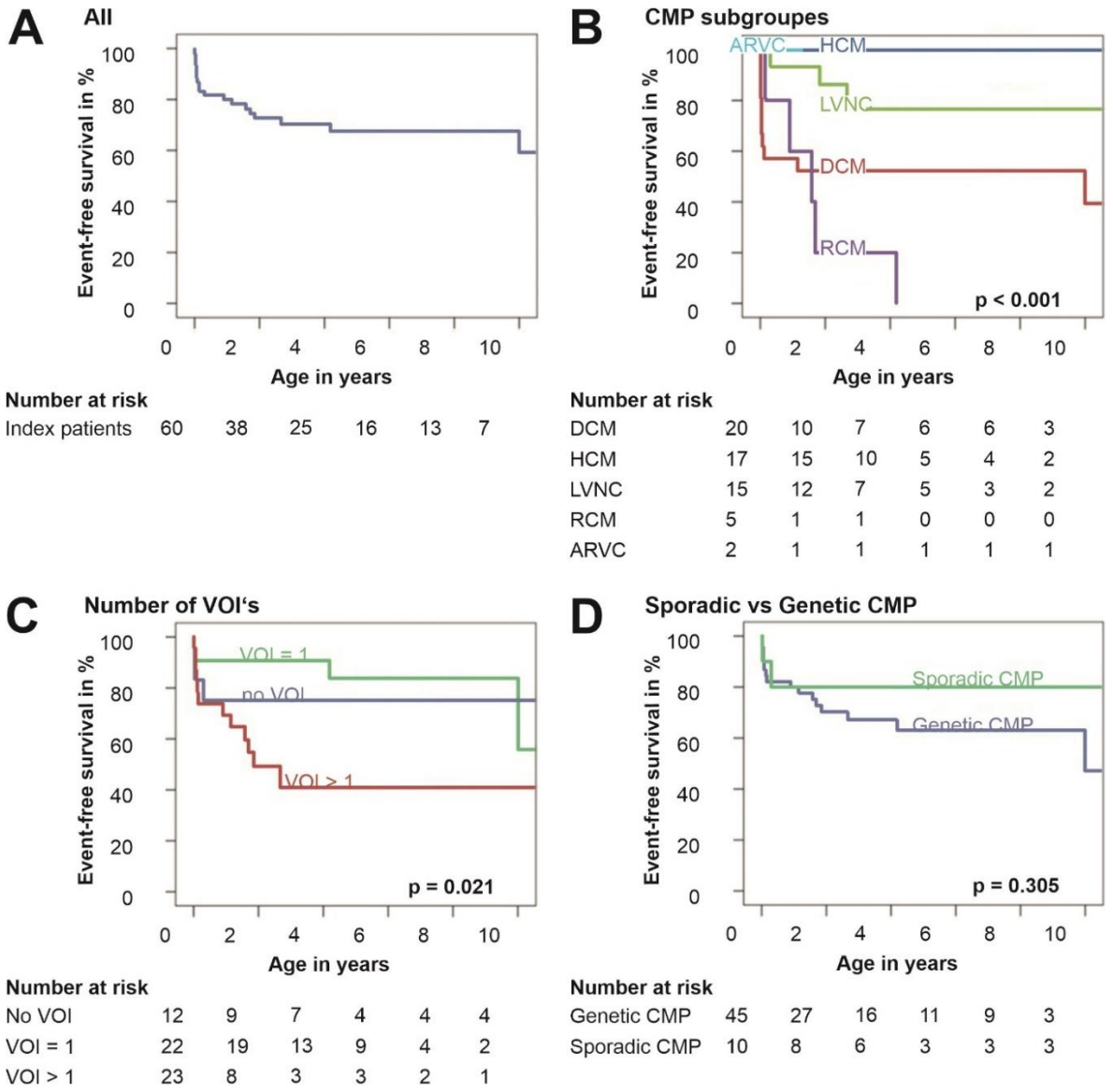


Figure 12: Kaplan-Meier curves of event-based survival in pediatric cohort of 60 CMP patients (CMP-60). The event-free survival percentage of CMP patients is presented over the years. **(A)** The correlation was observed in all patients, **(B)** between CMP subgroups, **(C)** between subjects carrying no, one or more than one VOI and **(D)** between the groups diagnosed with sporadic and genetic CMP. Events are defined as death, heart transplantation, and mechanical circulatory support. No VOI but positive family history of CMP was only seen in one index patient who was therefore excluded from the Kaplan-Meier analysis.

3.1.5 Early onset of cardiomyopathy in cases carrying variants in TNNI3

As mentioned in 3.1.3, variants of sarcomere genes were most common in the pediatric CMP cohort and within this group *MYH7*, *MYBPC3* and *TNNI3* exhibited the most genetic alterations (number variants (P or LP): *MYH7* – 13 (7), *MYBPC3* – 8 (6, two variants in one patient), *TNNI3* – 6 (5), Figure 9 A, Figure 10 A). In this selection *TNNI3* was of notice because of two reasons. Firstly, a set of variants agglomerated in the c-terminus of the protein and secondly, in one patient a homozygous truncating variant was detected. Therefore, available tissue and blood samples from these patients and inherent family members were further analyzed (Figure 13, Table S 5, Table S 6). An additional case of DCM from an in-house cohort was included in these analyses due to a detected homozygous truncating variant in *TNNI3*. In conclusion, seven genetic alterations of *TNNI3* were detected in seven patients (DCM n = 2, HCM n = 2, RCM n = 2, LVNC n = 1). Six of these variants were interpreted as P or LP (VUS variant was not considered in further analysis). Two variants were inherited as homozygous and truncating. One variant presented a splice site modification directly after the second exon. The other variant resulted in a frameshift leading to a premature stop codon in the IT arm region shortly behind a troponin C1, slow skeletal and cardiac type (TNNC1) binding region of TNNI3 (Figure 13). The other four variants clustered at the c-terminus, which is expected to be important in actin and tropomyosin binding. All four variants occurred in a heterozygous state and three led to an amino acid exchange (missense), whereas one was a truncating variant at the very end of the amino acid chain at position 209 from 210 amino acids (Figure 13).

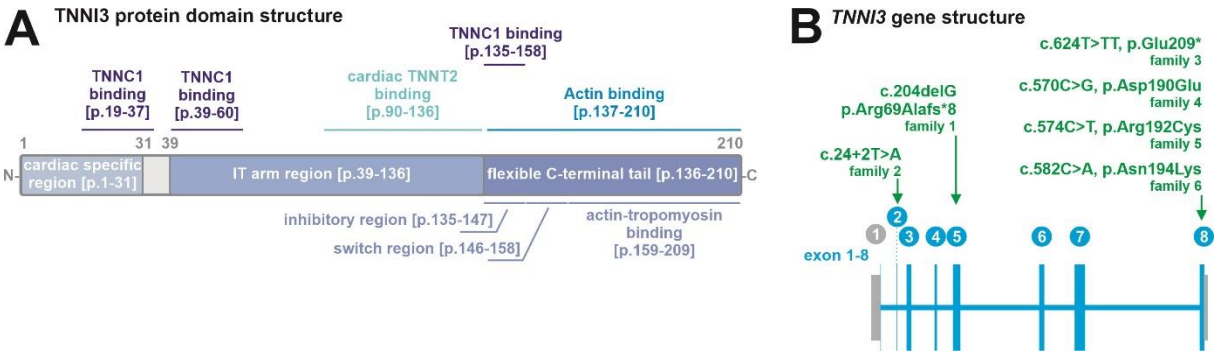


Figure 13: TNNI3 functional protein domain and gene structure with six indicated TNNI3 variants. (A) The protein domain structure of TNNI3 includes binding sites to TNNC1 (purple), TNNT2 (cyan), ACTC2 (blue) and TPM1 (light purple). Therefore, truncating variants near the n-terminus lose the ability to bind to most of these proteins whereas binding capabilities of C-terminal variants to actin and tropomyosin are influenced. **(B)** Two truncating variants in exons two and five are predicted to lead to early truncation of TNNI3 and four missense variants in exon eight indicate a mutational hotspot for variants to cause CMPs.

In both homozygous truncating cases, affected by DCM/LVNC and an early disease progression (1.1 and 0.8 years old at diagnosis and with HTX or deceased), the parents were

consanguineous, which led to the homozygous inheritance of the genetic variants. In family 1 blood samples from both parents were available and therefore checked for the segregation of the variant (Figure 14 A, Figure S 3). In family 2 only heart tissue from the index patient 2-II:3 was available, and segregation could not be checked, but an extended pedigree was available (Figure 14 B, Figure S 2). To analyze the expression of TNNI3 qPCR and Western blotting approaches were used. In Figure 14 C *TNNI3* mRNA levels were drastically reduced in patient 1-II:1 whereas patient 2-II:3 showed only little differences. In both samples a distinct increase in signal was detected for the gene *troponin I1, slow skeletal type* (*TNNI1*), the fetal form of *TNNI*. Smaller changes were also detected in *TNNT2* (increase) and *MYBPC3* (decrease).

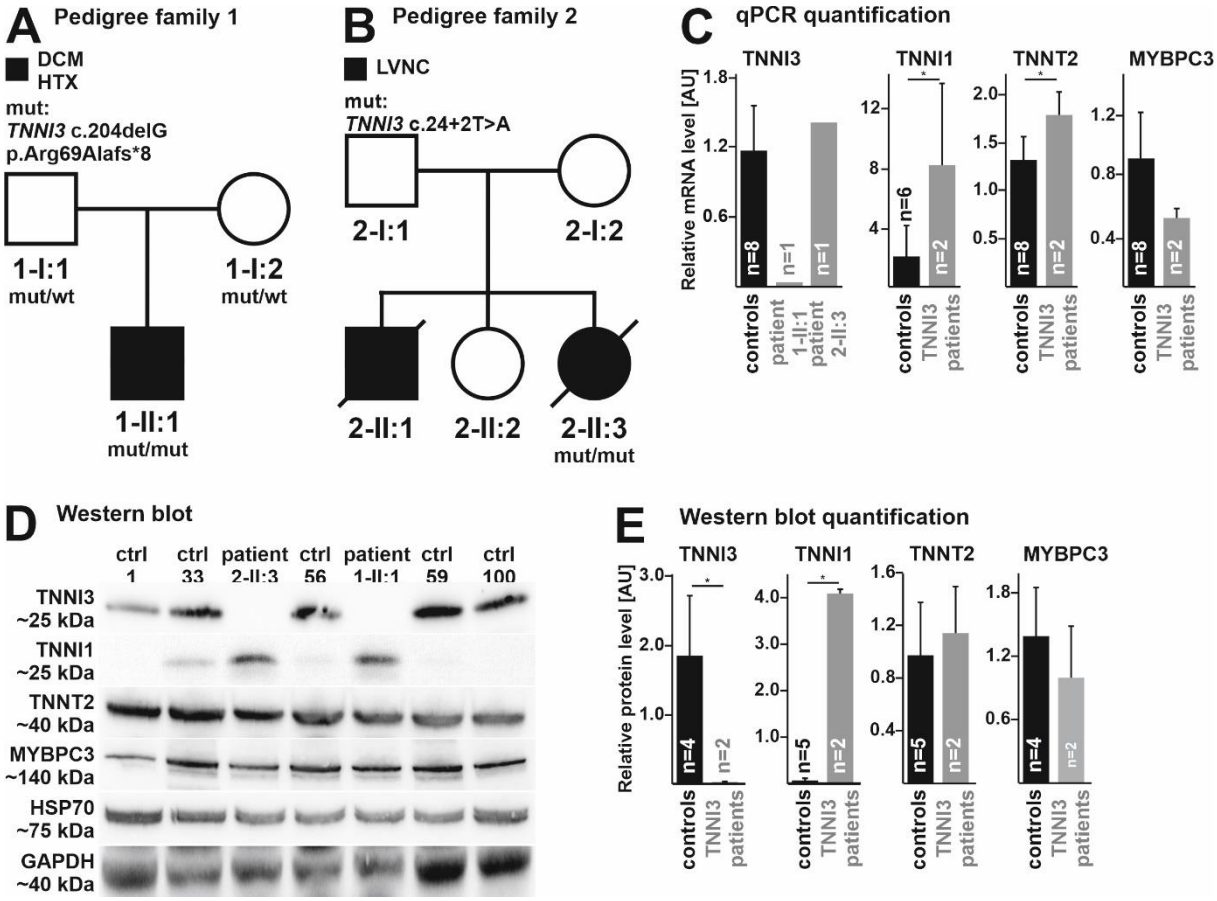


Figure 14: Pedigrees, mRNA analysis and protein expression analysis of two patients with a homozygous and truncating variant in TNNI3. (A)/(B) Pedigrees of two families carrying truncating variants in TNNI3 are presented. Genotypes are shown as wt or mut, representing the wildtype or altered allele. For individuals with no genotype no samples for sequencing were available. Phenotypes and the variants on coding sequence and protein level are indicated. Males are represented with squares and females with circles. Affected individuals have a filled form and deceased family members are marked with a diagonal line. **(C)** Analysis of mRNA levels show transcription of different targets (TNNI3, TNNI1, TNNT2 and MYBPC3). Expression of mRNA was normalized to GAPDH with a following Student's *t*-test to determine statistical significance (* for $p<0.05$). Heat shock 70 kDa protein 1A (HSP70) was used as loading control (normalization) and Student's *t*-test was performed to determine statistical significance (* for $p<0.05$). **(D)** Protein levels from heart tissues were analyzed to compare the expression of different targets (TNNI3, TNNI1, TNNT2 and MYBPC3). **(E)** Expression levels from signals in (D) were quantified.

Similar events were examined on protein level. Changes in protein concentrations are shown in Figure 14 D. Whereas TNNI3 was absent in both index patients, TNNI1 was drastically increased. Levels of TNNT2 and MYBPC3 showed small changes. Quantification of these signals showed significant changes for TNNI3 and TNNI1 (Figure 14 E).

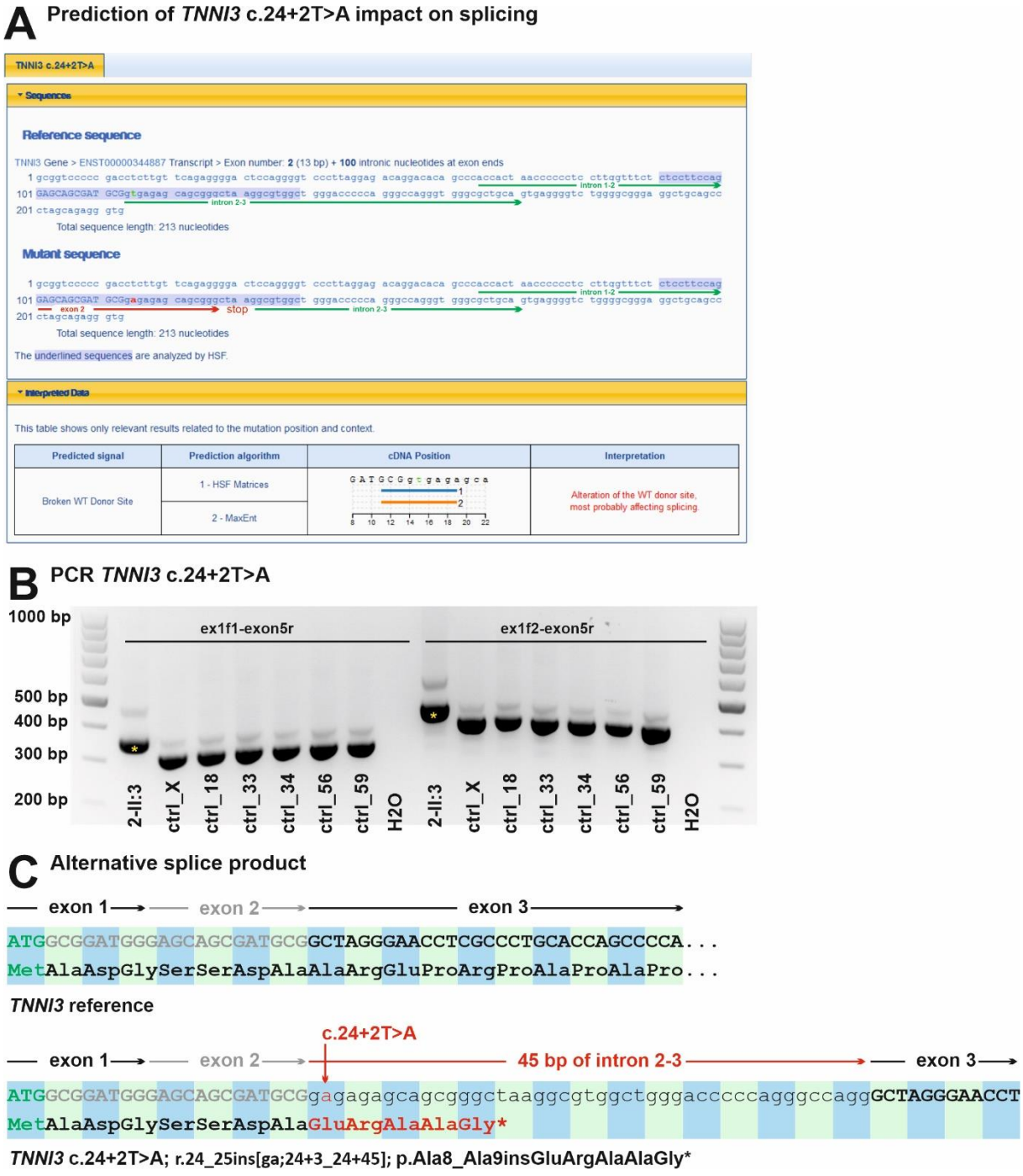


Figure 15: Prediction, mRNA expression and sequencing of the TNNI3 splice site variant c.24+2T>A. (A) *In silico* analysis of the base exchange TNNI3 c.24+2T>A with HSF (**http://www.umd.be/HSF3/index.html**) predicted a negative effect and an altered splice site. **(B)** A PCR product, amplified from cDNA originating from heart tissue samples of patient 2-II:3, shows altered length in fragments (yellow star). **(C)** Sequencing of the elongated cDNA fragment in (B) showed intronic extension of exon 2 leading to a premature stop.

To further investigate the effect of the splice site variation in *TNNI3* c.24+2T>A *in silico* prediction, amplification from *TNNI3* of affected parts in the mRNA and sequencing were used. The prediction of the online tool HSF was positive for changes of the wt donor splice site (Figure 15 A). Thus cDNA was generated from isolated RNA originating from the patient 2-II:3. Primer designed to amplify parts of *TNNI3* between exon 1 and 5 were used to generate the altered splice products seen in Figure 15 B. In comparison to control PCR products mRNA affected by *TNNI3* c.24+2A>A was elongated and alternated by alternative or mis-splicing. In the following sequencing of these products five additional amino acids were identified and a premature stop codon was implemented as illustrated in Figure 15 C.

In the same manner a case of RCM diagnosed at the age of 3.8 years and with HTX at 4.5 years was analyzed. The phenotype of this patient was associated with a heterozygous variant in *TNNI3* occurring as truncating at the end of the protein. The base insertion c.624t>TT led to a premature stop at the amino acid position 209 and therefore concluded in the truncation of the last two amino acids in TNNI3. The variant occurred *de novo* and was not detected in the parents (Figure 16 A). This LP variant is likely to have an effect because the parents without the variant are not affected by a cardiac phenotype. On mRNA level tissue samples of patient 3-II:1 showed increases in signals for *TNNI3* and *TNNI1*. The gene *MYBPC3* was slightly downregulated, whereas mRNA levels of *actinin alpha 2* (*ACTN2*) and *TNNT2* were similar to the control signals. In Western blotting experiments TNNI3 was still present in same amounts of protein as in controls. Only the signal for TNNI3 in the sample 2-II:3 was absent (Figure 14 D). Regarding TNNI1 there was no increase in signal observed for patient 3-II:1, which distinguished again from the homozygous truncating variant.

As mentioned before, three cases of CMP with a detected heterozygous missense variant in *TNNI3* were observed in the pediatric CMP cohort. Like p.Glu209* these base exchanges clustered near the c-terminus in a region associated with actin and tropomyosin binding (p.Asp190Glu, p.Arg192Cys and p.Asn194Lys, Figure 17). The three index patients were affected by HCM, RCM and DCM, whereas the RCM patient had the most adverse disease development diagnosed with CMP at the age of 1.8 years and undergoing HTX at age 3.4 years. The individuals affected by HCM and DCM were diagnosed at the age of 11.2 and 9.3 years. In family 4 the father was already deceased, and a history of sudden cardiac death was documented in the family. Thus, the segregation was not fully checked, but the probability is high that the LP variant p.Asp190Glu in *TNNI3* segregated in the affected members of this family. For the other two families parental blood samples were available. In both cases the variants occurred *de novo* and were not detected in the parents. Therefore, the observed *TNNI3* alterations, similar to the heterozygous truncating variant, are likely to influence the heart in these patients, because the individuals without p.Arg192Cys and p.Asn194Lys are not affected by any cardiac phenotype. For preventive reasons patient 4-II:1 underwent ICD

implantation and patient 6-II:1 was at risk for potential heart failure, because similar events occurred in a deceased twin brother. Due to the lack of heart tissue no further experiments were performed.

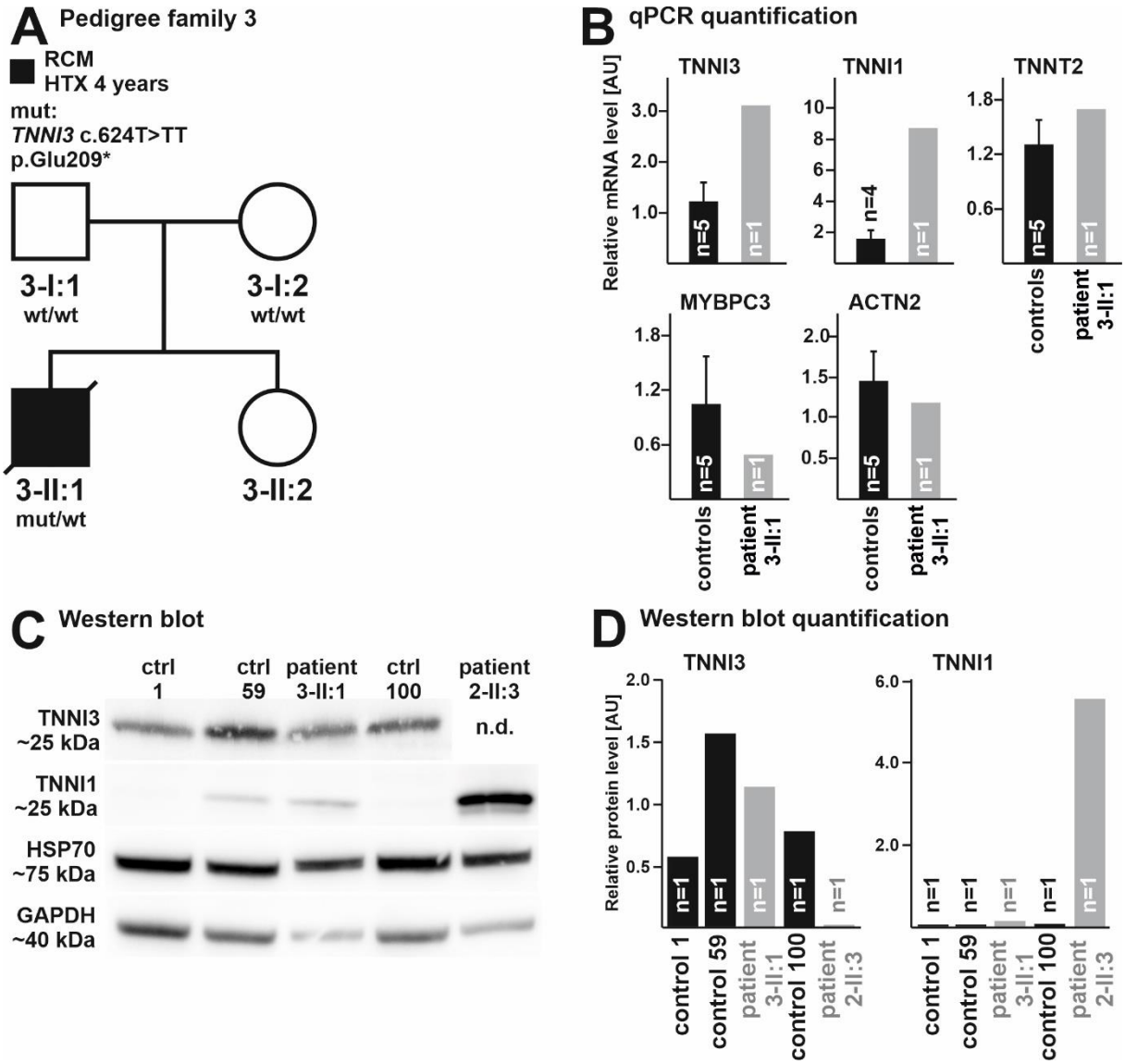


Figure 16: Pedigree, mRNA analysis and protein expression analysis of a patient with a truncating and C-terminal variant in TNNI3. (A) A pedigree of a family carrying a truncating and C-terminal variant in TNNI3 is presented. Genotypes are shown as wt or mut, representing the wildtype or altered (mutated) allele. For individuals with no genotype no samples for sequencing were available. Phenotypes and the variants on coding sequence and protein level are indicated. Males are represented with squares and females with circles. Affected individuals have a filled form and deceased family members are marked with a diagonal line. **(B)** Analysis of mRNA levels show transcription of different targets (TNNI3, TNNI1, TNNT2, MYBPC3 and ACTN2). Expression of mRNA was normalized to GAPDH. **(C)** Protein levels from heart tissues were analyzed to compare the expression of different targets (TNNI3 and TNNI1). **(D)** Expression levels from signals in (C) were quantified. HSP70 was used as loading control (normalization). The results for patient 2-II:3 were used from **Figure 14**.

In conclusion, two homozygous variants in *TNNI3* leading to truncations near the n-terminus showed absent signals for TNNI3 on protein level probably due to non-sense mediated decay and a drastic increase in expression of TNNI1. Both cases showed a very severe and early disease development. Experiments with a heterozygous truncating variant at the c-terminus did not show any significant alteration in protein expression. In this case the CMP progression was still severe, but the onset of the phenotype was with higher age. Similar statements can be made for the three mentioned missense variants. Of note is the heterogeneity of the disease development with *TNNI3* variants. Four CMP subtypes were associated with *TNNI3* variants in the CMP-80 cohort and the onset varied from 0.8 to 11.2 years.

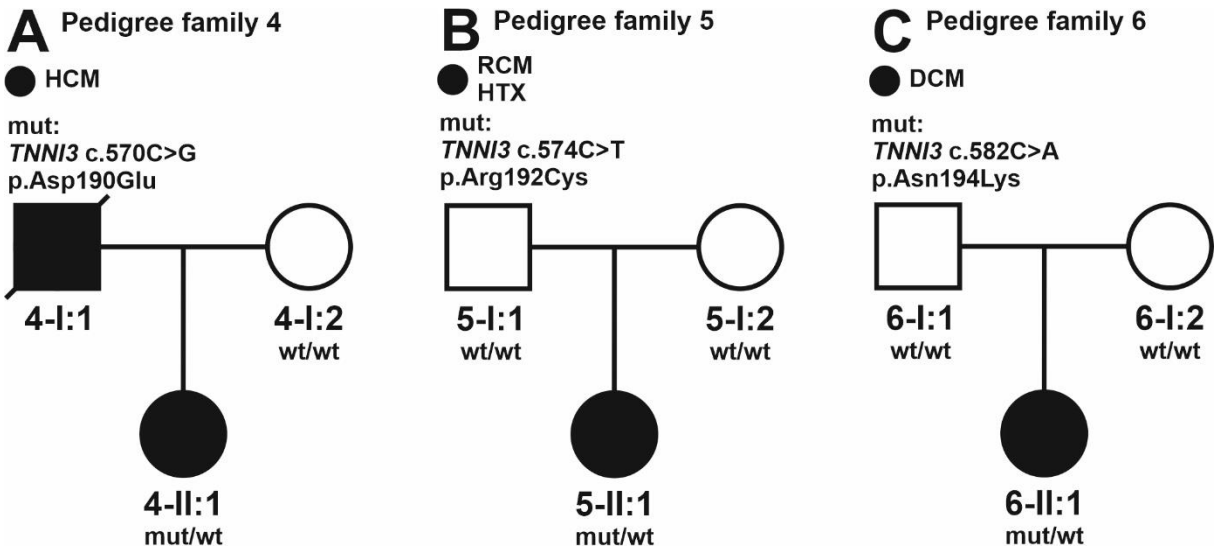


Figure 17: Pedigrees of patients with C-terminal missense variants in TNNI3. (A) – (C) Pedigrees of three families carrying C-terminal variants in TNNI3 are presented. Genotypes are shown as wt or mut, representing the wildtype or altered (mutated) allele. For individuals with no genotype no samples for sequencing were available. Phenotypes and the variants on coding sequence and protein level are indicated. Males are represented with squares and females with circles. Affected individuals have a filled form and deceased family members are marked with a diagonal line.

3.2 Detection of *PRDM16* variants in patients with cardiac phenotypes – Genetic and functional analysis

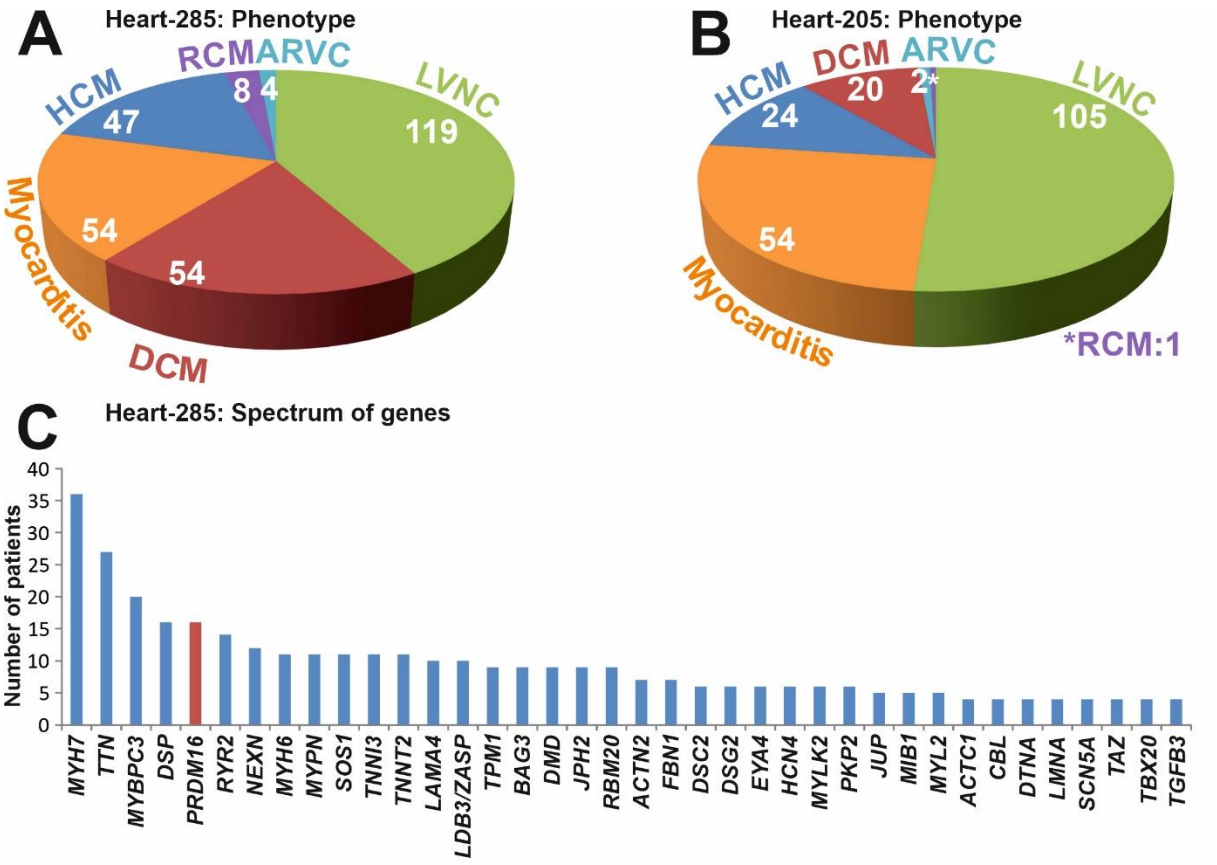
VOI were most frequently detected in the genetic groups of sarcomere, Z-disc and desmosomes genes. With four detected variants *PRDM16* was as frequent as *plakophilin 2* (*PKP2*), *TNNT2* or *TTN* (truncating) and an exception outside of the afore mentioned groups. PRDM16 is a transcriptional regulator and therefore listed in the group of transcription and splicing. Further genetic and functional analysis of PRDM16 can help to better understand cardiac mechanisms and signal pathways in the heart. Hence, this chapter investigates the genetic and functional impact of *PRDM16* variants for CMP in an extended genetic cohort.

3.2.1 High number of *PRDM16* variants in an extended cohort of cardiac disease phenotypes

To further investigate the frequency of *PRDM16* variants in patients with the phenotype of CMP and myocarditis an in-house cohort was added to the existing pediatric CMP-80 cohort. CMP specific genes of an additional 205 patients with CMP or myocarditis (Heart-205) were tested for genetic alterations. In total an assembly of 285 individuals was investigated (Heart-285). In this cohort individuals of all age groups were included (181 pediatric cases with age ≤ 18 years). The gender was shifted to males (112 f, 174 m). VOI were detected in 226 cases with 320 missense, 58 splice site, 33 frame shift and 15 stop gain variants. On average a patient was carrying 1.49 VOI. The in-house cohort comprised mainly LVNC cases and therefore the distribution of phenotypes shifted in comparison to the pediatric CMP cohort (Figure 18 B). Most patients were diagnosed with LVNC (n = 119) followed by similar amounts of myocarditis, DCM and HCM patients with around 50 cases each (Figure 18 A). RCM and ARVC were again underrepresented with eight and four cases. The five genes with most detected VOI were *MYH7* (n = 36), *TTN* (n = 27, only possible truncating variants), *MYBPC3* (n = 20), *DSP* (n = 16) and *PRDM16* (n = 16, Figure 18 C). Whereas *MYH7*, *MYBPC3* and *TTN* were expected with high numbers of variations in this cohort, cases with cardiac phenotypes caused by *PRDM16* variants are rarely known. Therefore, the detected frequency was surprisingly high. The frequency of VOI in major CMP genes like *TNNI3*, *TNNT2*, *LDB3* or *TPM1* was considerably lower (n = 11/11/10/10).

To gain deeper understanding about the effect of *PRDM16* variants, further investigations were performed. An overview of *PRDM16* variants showed altered sequences in 16 different patients, whereas two variants occurred in one patient and one variant was detected twice (Table 6). Most of these patients were diagnosed with LVNC (n = 7) followed by two cases of ARVC, DCM, HCM and RCM each and one case of myocarditis. Therefore, VOI in *PRDM16*

were detected in all investigated phenotypes. The gender of the 16 patients prevailed slightly to males (f = 6, m = 10) and with a range from 0.6 to 55 years all age groups were present.



Variants detected in three patients: BRAF, FHL1, NKX2-5, PRKAG2, RAF1, TMEM43, VCL
Variants detected in two patients: ABCC9, ANKRD1, CALR3, DES, FHL2, HRAS, LAMP2, MYL3, MYOY2, PDLIM3, TNNC1
Variants detected in one patients: CRYAB, GLA, HSPB8, KRAS, MAP2K1, MAP2K2, NRAS, SGCD, SHOC2, TCAP, TTR

Figure 18: Distribution of cardiac phenotypes from additional in-house cohort and frequency of VOI in CMP relevant genes in patients. (A) The distribution of phenotypes from 285 CMP and myocarditis patients shows the composition of the initial pediatric cohort and an additional in-house cohort. **(B)** The distribution of phenotypes from 205 CMP and myocarditis patients shows the composition of an additional in-house cohort. **(C)** The frequency of different genes with a variant detected in CMP patients is shown (full name of genes in **Table S 1**).

The spectrum of the detected *PRDM16* variants included 11 missense, three truncating and two splice site affecting variants. Two stop gain and one frame shift variant of the truncating variants were considered as P, whereas the rest was interpreted as VUS (11 missense and two splice site variants). Due to a MAF in the reference database gnomAD of smaller than 0.1% all VOI were considered as rare, but variants with an amino acid position of 702 and smaller were at least detected with a MAF of 0.01% and lower. The only exception was variant p.G791D (gnomAD MAF = 0). Eight of these variants were not previously reported with cardiac phenotypes and three variants were already published.

Table 6: Heterozygous genetic Variants detected in *PRDM16* (transcript: ENST00000270722) in the cohort Heart-285

Diagnosis	Sex	Age in years	Genomic location	cDNA position	Protein position	gnomAD MAF	ACMG pathogenicity	Novel	ClinVar	PubMed ID
Myocarditis	m	13,7	1:3102766	c.115G>A	p.E39K	0.0000789	VUS	yes	no	-
HCM	m	0,6	1:3313149	c.668G>T	p.G223V	0.0000318	VUS	yes	no	-
ARVC[1]	m	38,6	1:3319454	c.776C>T	p.A259V	0.0000572	VUS	yes	no	-
LVNC	m	4	1:3322136	c.1110C>A	p.D370E	0.0000041	VUS	yes	no	-
LVNC	m	44,6	1:3328329	c.1573dup	p.R525Pfs*79	0	P	no	P	23768516
LVNC	f	10,5	1:3328388	c.1627C>T	p.Q543*	0	P	yes	no	-
LVNC	f	28,9	1:3328646	c.1885G>C	p.V629L	0	VUS	yes	no	-
ARVC[1]	m	38,6	1:3328817	c.2056A>G	p.T686A	0.0000678	VUS	yes	no	-
LVNC	f	21,5	1:3328865	c.2104A>T	p.K702*	0	P	no	P	23768516
DCM	f	2,5	1:3328948	c.2187C>G	p.F729L	0.00039	VUS	no	LBe	-
HCM	m	15	1:3329057	c.2296G>A	p.G766S	0.00029	VUS	no	LBe	-
RCM	m	17	1:3329133	c.2372G>A	p.G791D	0	VUS	yes	no	-
LVNC[2]	m	19,9	1:3329208	c.2447A>G	p.N816S	0.000232	VUS	no	P	23768516
RCM[2]	m	5	1:3329208	c.2447A>G	p.N816S	0.000232	VUS	no	P	23768516
DCM	f	9,9	1:3329370	c.2603+6C>T	-	0.000414	VUS	no	LBe	-
ARVC	f	53,5	1:3331186	c.2666C>T	p.P889L	0.000177	VUS	no	VUS	-
LVNC	m	55,0	1:3331216	c.2691+5G>A	-	0.000201	VUS	no	LBe	-

Abbreviations: MAF – minor allele frequency, ACMG – American college of Medical Genetics and Genomics, VUS – variant with uncertain significance, P – pathogenic, LBe – likely benign

[1]two variants in same patient

[2]same variant in two patients

To analyze the distribution of known and novel *PRDM16* variants, their amino acid position was highlighted in a schematic protein domain structure (Figure 19). Homology alignments and similarity tools predicted an n-terminal SET domain at amino acid position 82 to 211 and two separated zinc finger domains (aa position 230 to 445 and 951 to 1032). The first zinc finger domain contained seven zinc fingers and was located near and behind the SET domain. The second zinc finger was located near the c-terminus and consisted of three zinc fingers. At position p.459-557 a proline rich region was located and already published data provided information on isolated SKI proto-oncogene (SKI), c-terminal binding protein 1 (CTBP1), c-terminal binding protein 2 (CTBP2) and zinc finger protein 516 (ZNF516) interaction sites in the second half of the sequence (PFDLT and PLDLS motifs at aa position 774 to 778 and 804 to 808 mediate binding of CTBP, Kajimura et al., 2008, Takahata et al., 2009). The detected variants in this work were distributed evenly throughout the protein and no clear cluster emerged. Regions without any alteration so far were the SET domain, the second zinc finger domain and the following n-terminal region. The pathogenic nonsense and frameshift variants located between the two zinc finger domains were predicted to express approximately half of the protein and therefore leading to loss-of function of PRDM16. Functional regions containing missense variants were observed between the amino acids 223 to 370, potentially changing properties of the first zinc finger domain and 629 to 889, maybe interfering with published protein binding sites.

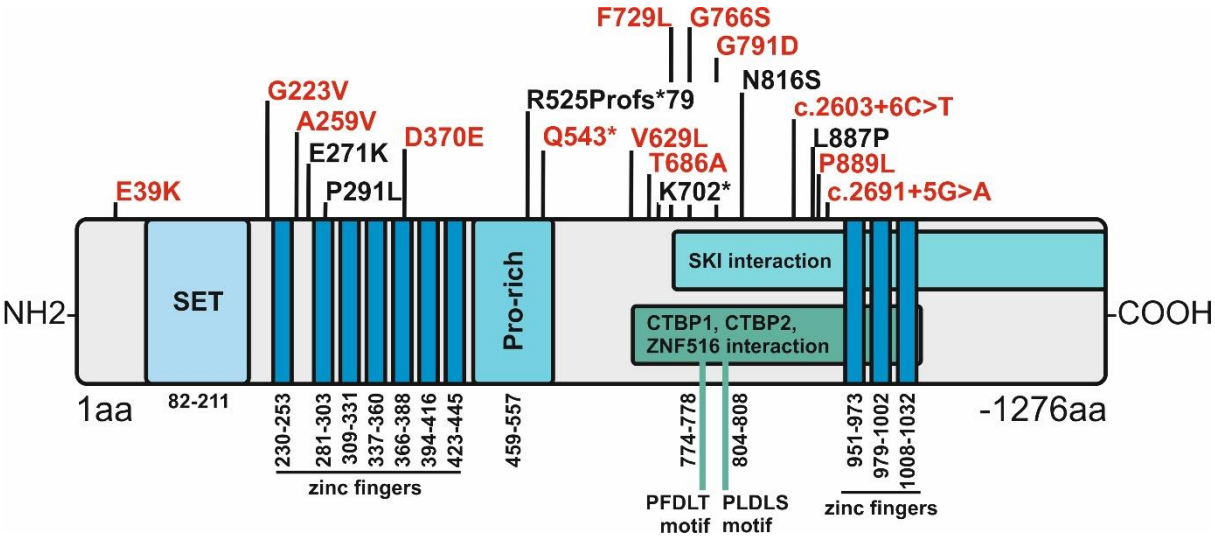


Figure 19: Schematic protein domain structure of PRDM16. Shown is the sequence of predicted functional domains in PRDM16. It consists of an n-terminal SET domain followed by seven zinc fingers of the C2H2-type. A second zinc finger domain near the c-terminus consists of three zinc fingers (C2H2 type). Shortly after the first zinc finger domain a proline rich region is following. At the c-terminus the interaction sites with SKI (TGF-β signaling), CTBP1/2 (PLDLS motif) and ZNF515 are shown. Above the scheme are variants colored in black (described in literature) and red (newly detected in cardiac phenotype cohort of 285 patients).

3.2.2 Bioinformatic characterization of PRDM16

Additional features of PRDM16 were identified using protein prediction tools. PRDM16 was examined for PTM, secondary structure, conserved regions, disordered sections through the sequence and parts with high surface accessibility (Figure 20). Positive results for PTM were obtained for acetylation, methylation, phosphorylation, glycosylation, SUMOylation and S-nitrosylation. Data was collected from different prediction tools and hits, exceeding certain thresholds, were considered as true results. Putative PTM sites were distributed over the entire PRDM16 protein. On note, in the region p.540- 608 six PTM clustered. This suggests an exposed surface accessibility and a potentially critical functional region. At position p.K915 an acetylation and SUMOylation site were predicted. This lysine residue was described to be involved in stabilization of PRDM16 through SUMOylation (Chen, Huang, Pan, Zhu, & Wang, 2018). Stabilization of PRDM16 protects the protein from degradation and enables binding of downstream interaction partners for further stabilization and signaling. No PTM corelated with VOI detected throughout genetic analysis. Protein conservation tools, taking into consideration homology and similarity alignments, identified functional regions e.g. the zinc finger domains. Thereby nine grades of conservation were possible spanning from 1 to 9 and from low to high conservation. Positions exceeding the threshold of 5 for being conserved were marked by a red dot above the graph. Other sections were characterized by changing conservation values and only peaked for small sequences like the PFDLT and PLDLS motifs. Surprisingly the SET domain showed a relatively low conservation, although secondary structural elements were especially predicted in this region and the zinc finger domains and should be affected by steric properties of amino acid residues changes. Further structural elements, marked as red for α-helices and as green for β-sheets, were located at the position from amino acid 680 to 762 and at the c-terminus. These findings coincided with results from surface accessibility calculations. Values for surface accessibility were normalized to 1 (exposed = maximal value 1, buried = minimal value 0). Again values exceeding the threshold were marked as red dot above graph (threshold = 0.25). In general, predicted structural elements seemed more buried than flexible regions. Therefore, low surface accessibility was observed for the SET and zinc finger domains, the region between amino acid 680 to 762 and the c-terminus. The presence of structural elements positively coincided with ordered protein regions. For these predictions again normalized values were used and 1 presented maximal disorder, whereas 0 was the minimal amount and 0.5 was used as threshold to characterize a position as disordered. The SET and zinc finger domains and the afore mentioned region from position 680 to 762 rarely passed the threshold for disorder and strengthened the prediction of secondary structure. Most disordered region located at the termini, between position 521 to 654 and between 784 and 866 and contained possible protein binding sites marked in red on the graph.

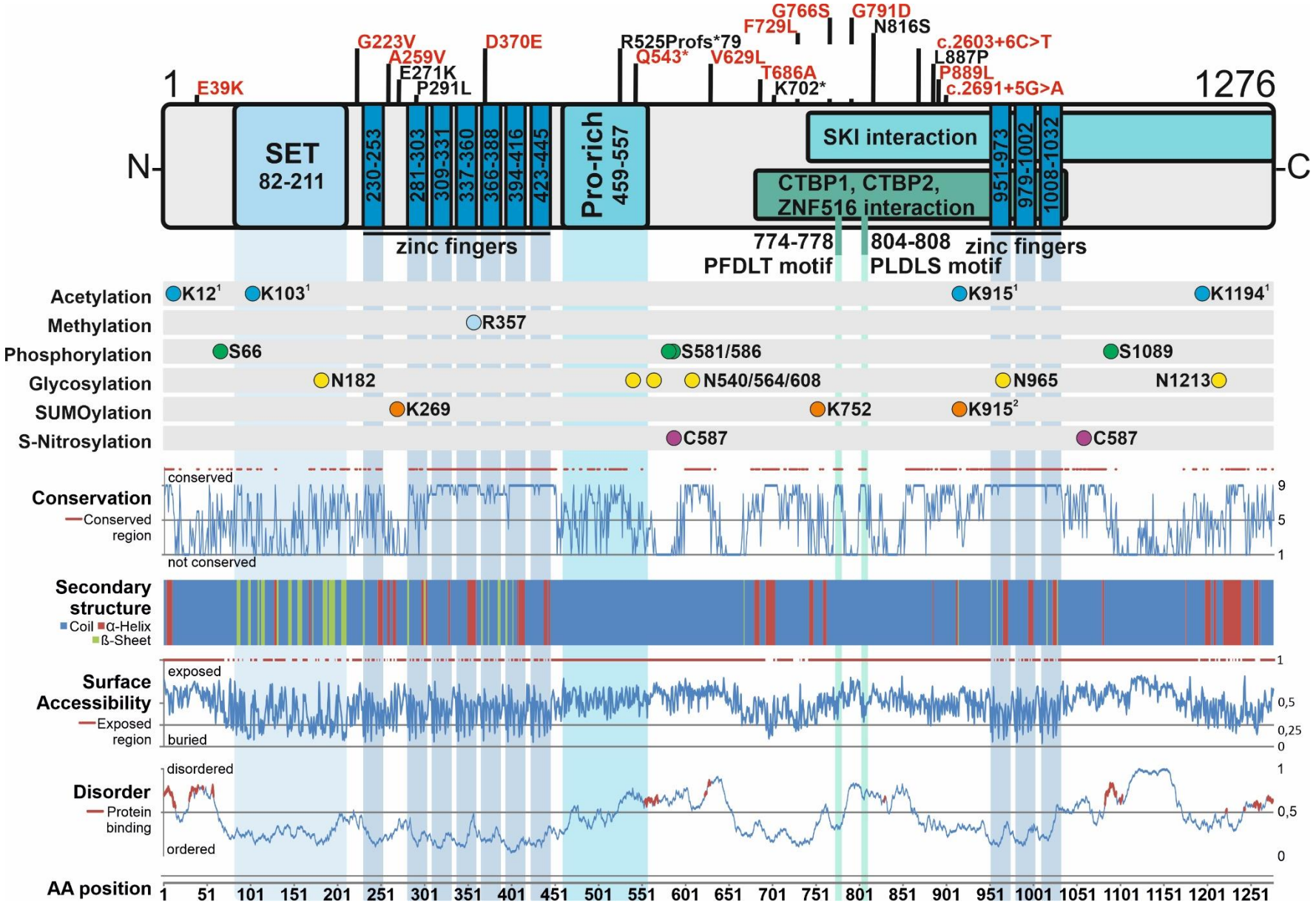

G766S
G791D
F729L
N816S
G223V
D370E
R525Profs*79
c.2603+6C>T
A259V
Q543*
V629L
L887P
E271K
T686A
P889L
P291L
K702*
c.2691+5G>A
1
E39K
1276
N-
SET
82-211
230-253
281-303
309-331
337-360
366-388
394-416
423-445
Pro-rich
459-557
SKI interaction
CTBP1, CTBP2,
ZNF516 interaction
951-973
979-1002
1008-1032
-C
zinc fingers
774-778
PFDLT motif
804-808
PLDLS motif
zinc fingers
Acetylation
K12[1]
K103[1]
K915[1]
K1194[1]
Methylation
R357
Phosphorylation
S66
S581/586
S1089
Glycosylation
N182
N540/564/608
N965
N1213
SUMOylation
K269
K752
K915[2]
S-Nitrosylation
C587
C587
Conservation
Conserved region
conserved
not conserved
9
5
1
Secondary structure
Coil
α-Helix
ß-Sheet
Surface Accessibility
Exposed region
exposed
buried
1
0,5
0,25
0
Disorder
Protein binding
disordered
ordered
1
0,5
0
AA position
1 51 101 151 201 251 301 351 401 451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101 1151 1201 1251

Figure 20: Analysis of predicted functional features from PRDM16. Shown is the schematic functional structure of PRDM16 at the bottom of the figure with novel (red) and published (black) genetic variants above the model. Predicted PMT are shown for acetylation, methylation, phosphorylation, glycosylation, SUMOylation and S-nitrosylation. Predictions for other modification categories like palmitoylation were also applied . The following graphs show predictions for PRDM16 protein conservation, secondary structure (coil, α-helix, β-sheet), accessible surface area and disorder. The numbers of the corresponding amino acids are shown at the bottom.

[1]Acetylation sites were predicted with three out of six online tools and not with at least four and were therefore selected with less stringency than the other predictions of PTM.

[2]The SUMOylation at position K915 did not fulfill the criteria to be listed but was described experimentally in the literature and is therefore considered as SUMOylation site (**Chen, Huang, Pan, Zhu, & Wang, 2018**).

3.2.3 Cloning of *PRDM16* and quality control

For expression already available *PRDM16* in wt (PRDM16_wt), p.E271K (PRDM16_v1), p.P291L (PRDM16_v2), p.R525Pfs*79 (PRDM16_v3fs), p.N816S (PRDM16_v4) and p.L887P (PRDM16_v5) form embedded in the pcDNA-Dest53 vector was used to amplify the gene, ligate it into the new vector and transform the constructs in *E. coli* (XL1-blue). Vectors used for expression in human cell lines with and without protein tags were pFLAG-CMV5/6a and pEGFP-C1. Selection of positively transformed *E. coli* was expressed through antibiotic resistance (pFLAG-CMV5/6a - ampicillin, pEGFP-C1 - kanamycin). After inoculation of LB medium with the transformed bacteria a midiprep was applied to purify the plasmid DNA from the cultures. This process with additional quality control steps is described in the following part.

Table 7: DNA-constructs for the expression of PRDM16 in wt form and containing different variants

Name		**Protein tag**	**Expression**	
Backbone	**Insert**		**Vector**	**Variant**
pFLAG_CMV5a	PRDM16_wt	-	PRDM16_no_tag	wt
pFLAG_CMV5a	PRDM16_E271K	-	PRDM16_no_tag	v1
pFLAG_CMV5a	PRDM16_P291L	-	PRDM16_no_tag	v2
pFLAG_CMV5a	PRDM16_R525Pfs*79	-	PRDM16_no_tag	v3fs
pFLAG_CMV5a	PRDM16_N816S	-	PRDM16_no_tag	v4
pFLAG_CMV5a	PRDM16_L887P	-	PRDM16_no_tag	v5
pFLAG_CMV6a	PRDM16_wt	n-term. FLAG	FLAG+PRDM16	wt
pFLAG_CMV6a	PRDM16_E271K	n-term. FLAG	FLAG+PRDM16	v1
pFLAG_CMV6a	PRDM16_P291L	n-term. FLAG	FLAG+PRDM16	v2
pFLAG_CMV6a	PRDM16_R525Pfs*79	n-term. FLAG	FLAG+PRDM16	v3fs
pFLAG_CMV6a	PRDM16_N816S	n-term. FLAG	FLAG+PRDM16	v4
pFLAG_CMV6a	PRDM16_L887P	n-term. FLAG	FLAG+PRDM16	v5
pEGFP-C1	PRDM16_wt	n-term. EGFP	EGFP+PRDM16	wt
pEGFP-C1	PRDM16_E271K	n-term. EGFP	EGFP+PRDM16	v1
pEGFP-C1	PRDM16_P291L	n-term. EGFP	EGFP+PRDM16	v2
pEGFP-C1	PRDM16_R525Pfs*79	n-term. EGFP	EGFP+PRDM16	v3fs
pEGFP-C1	PRDM16_N816S	n-term. EGFP	EGFP+PRDM16	v4
pEGFP-C1	PRDM16_L887P	n-term. EGFP	EGFP+PRDM16	v5

All used constructs are listed in Table 7. These plasmids were created to express PRDM16 with its mentioned variants in three forms, without any tag, with a FLAG-tag and with an EGFP-tag. The protein tags were expressed n-terminally to ensure the expression of the tagged variant PRDM16_v3fs. For untagged protein, a plasmid with c-terminal FLAG-tag was chosen. Ligation primers in this case were designed to leave the stop codon of *PRDM16* intact to exclude the protein tag. Upscaling to midi cultures and purification of plasmid DNA through midiprep resulted in plasmid concentrations of 0.8 to 3.4 µg/ml.

The plasmids were then checked for the DNA length on agarose gels (Figure 21).The expected length of the backbone vectors with *PRDM16* was 8.5 kb. For a quick check the plasmids were not linearized. Therefore, gel bands were visible for a lower and stronger supercoiled and a higher nicked form of plasmids. Supercoiled vectors move faster through agarose gels and thus appeared lower than the expected 8.5 kb (around 6 kb). In the example of Figure 21 all constructs had at least one clone with the corresponding height except PRDM16_wt. To get the full set, new clones had been picked and checked for the right height as seen in the right part of the figure. For further quality checks a sequencing step followed afterwards to detect possible base exchanges and small deletions or insertions.

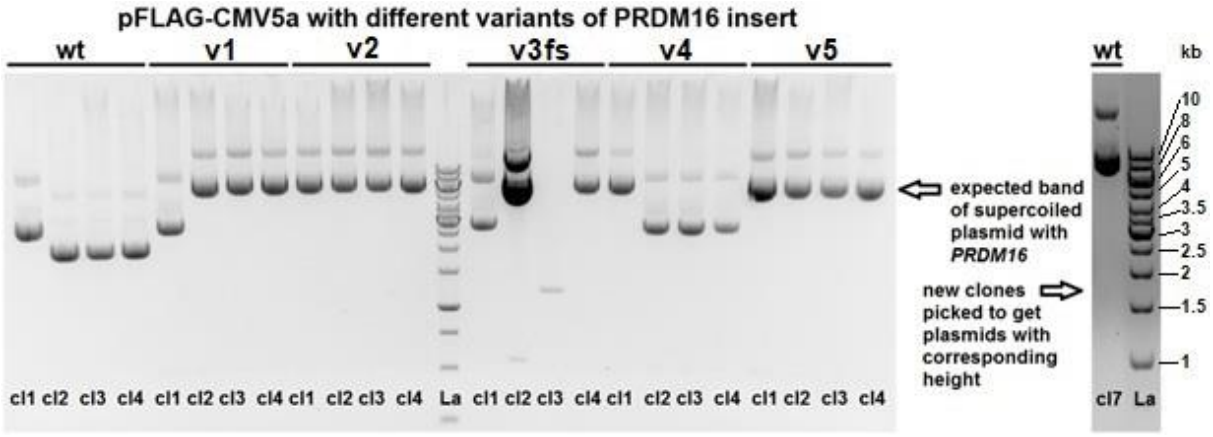


Figure 21: Ligation approach of *PRDM16* inserts into the pFLAG-CMV5a vector. A minipreps preceded by PCR, ligation and transformation into XL1 blue *E. coli* cells with the pFLAG-CMV5a vector and *PRDM16* inserts were loaded onto an agarose gel (1%) to check the corresponding length in supercoiled from. If one approach did not lead to a clone with the right height (wt cl1-4), new clones had to be picked (wt cl7). Clones containing the corresponding plasmids with the right height were picked to inoculate a miniprep culture for higher yields of DNA and deeper purification.

In a final quality check, all cloned plasmids were validated with Sanger sequencing. In Figure 22 two positive and two negative examples are shown. Most alterations of the DNA appeared at the termini. Base exchanges of altered start and stop codons would lead to missing initiation of PRDM16 expression or its elongation. Clone one and two in Figure 22 were affected by this and therefore not considered for further processing. All clones presented the right base

exchange for the variant PRDM16_v2 normally coded by the codon CCC and changed to CTC. The entire gene was sequenced in this matter.

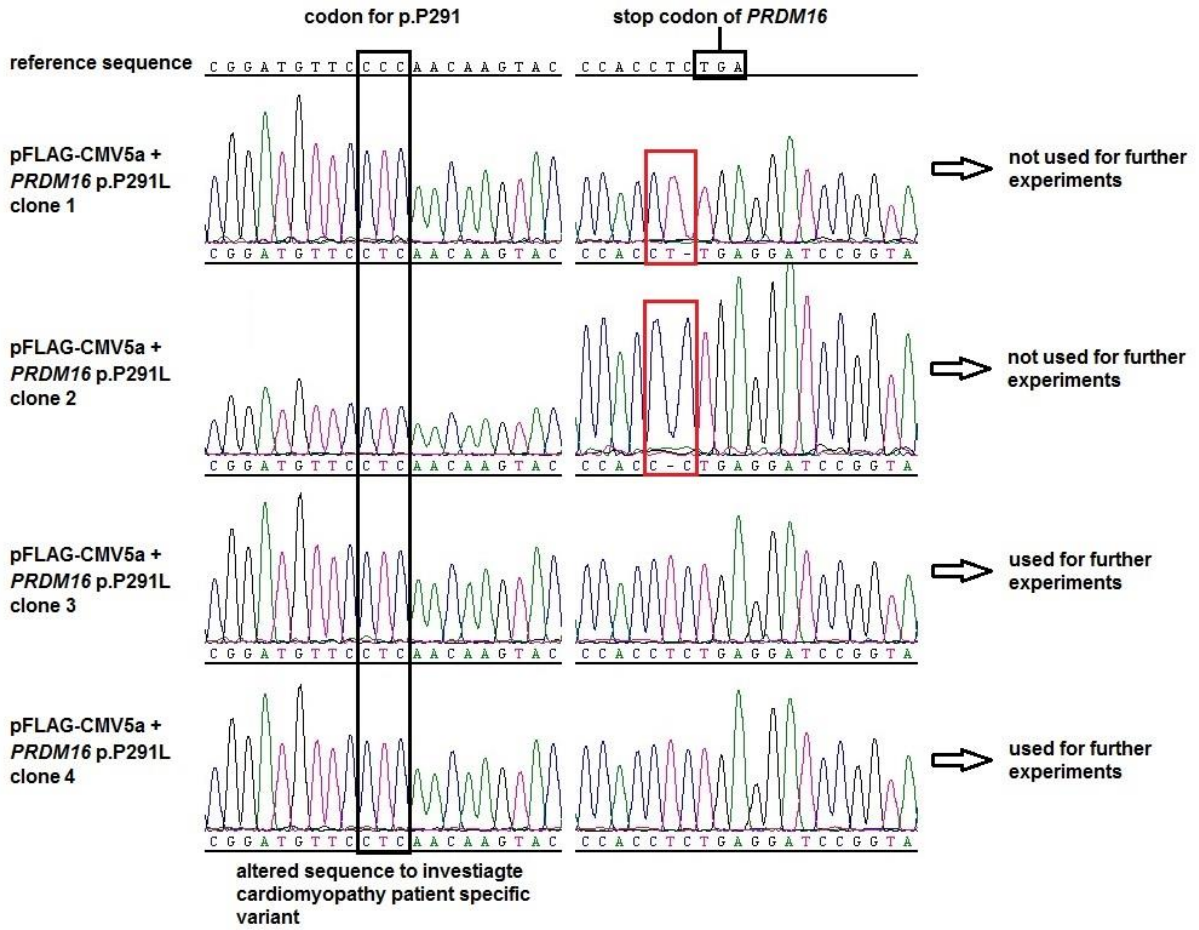


Figure 22: Sequencing of pFLAG-CMV5a plasmids containing PRDM16_v2. After checking the size of newly cloned vector/*PRDM16* constructs on an agarose gel (**Figure 21**), the DNA base sequence had to analyzed for small deletions/inserts and base exchanges. Therefore, the BigDye Terminator v3.1 protocol was applied, and samples were analyzed with the 3730 DNA Analyzer.

A last step to ensure the correctly designed expression vectors were restriction enzyme digests (Figure 23). Due to restriction sites in the multiple cloning sites of vectors restriction enzymes could be used to confirm the length of insert and vector in linear form. Two approaches were followed. A single digest confirmed the length of the entire expression plasmid, expected to be 8.5 kb and a double digest separated the backbone vector and the insert (backbone vector with 4.7 kb and *PRDM16* insert with 3.8 kb). For the single digest the restriction enzyme EcoRI was used. For the double digest BamHI was added to the reaction. The linearization with EcoRI showed bands with heights at approximately 8.5 kb. No additional bands were observed and the expected difference in length to the empty vector was visible. The double digest clearly showed the separation of insert and backbone with the later exactly on the same height as the

empty vector. Again no additional unexpected bands were visible and therefore the samples were clear of contaminations.

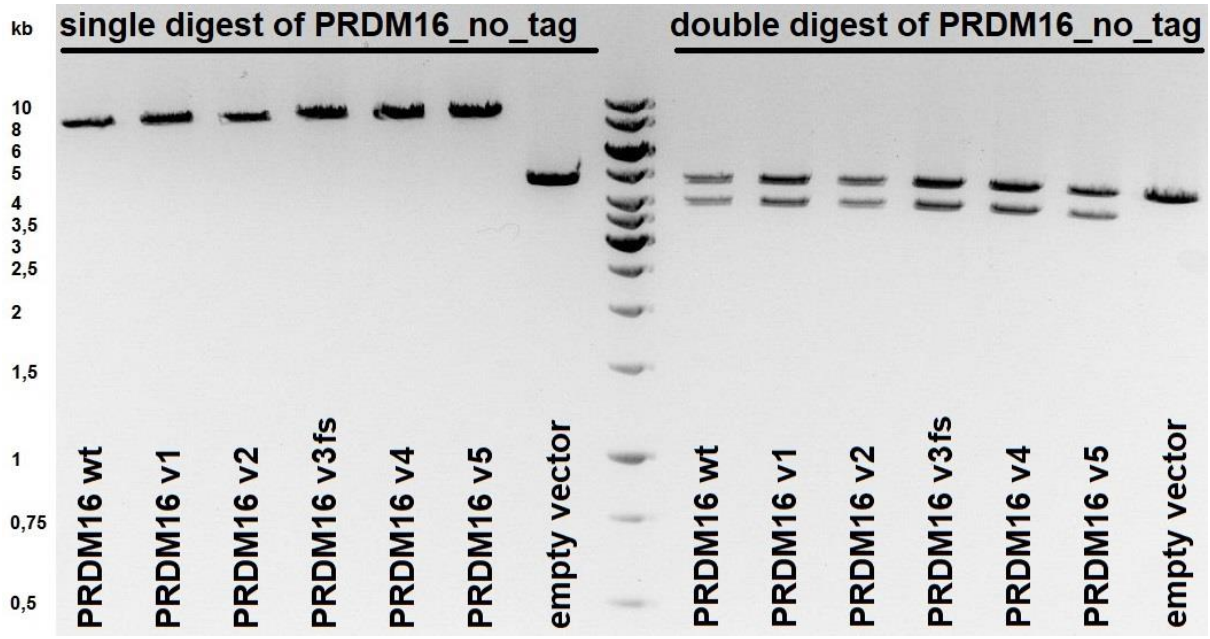


Figure 23: Restriction enzyme digest of pFLAG-CMV5a with *PRDM16* inserts. To prove the correct length of the *PRDM16* insert into the vector pFLAG-CMV6a, a restriction enzyme digest with EcoRI in a single digest approach and EcoRI + BamHI in a double digest approach were used and loaded onto an agarose gel (1%).

3.2.4 Overexpression of PRDM16 wt and variants

All investigated *PRDM16* variants were detected in patients with CMP. Zebrafish studies showed that PRDM16 truncation induces CMP. However, the impact of missense variants on *PRDM16* function is unknown. Thus, further experiments were conducted to explore an impact of *PRDM16* missense variants on subcellular localization and protein stability.

3.2.4.1 Subcellular localization analysis of cardiomyopathy-associated PRDM16 mutants with immunostaining

To investigate the subcellular location of PRDM16, HEK293 cells were transfected with different *PRDM16* constructs and subsequently analyzed with immunostaining and four-channel confocal microscopy. Several cell types were tested for the overexpression of PRDM16. Cell lines like Hela and A549 are characterized by no intrinsic PRDM16 expression and therefore more sensitive to PRDM16 transfection and expression (unstable and early cell death after transfection). MCF7 cells exhibit endogenous PRDM16, but expression of this protein was more heterogenous than in HEK293 cells. Therefore, HEK293 cells were chosen for further investigation. Initially, the impact of the EGFP and FLAG protein tags on

PRDM16_wt subcellular localization was tested (Figure 24). Hence, with PRDM16_wt transfected HEK293 cells were fixated and analyzed with antibodies against PRDM16 and the FLAG-tag. EGFP emitted fluorescence without marking. The nuclei were detected with DAPI (grey) and phalloidin (magenta) was used to mark actin filaments. In each row of Figure 24 the expression of one construct is shown in three pictures.

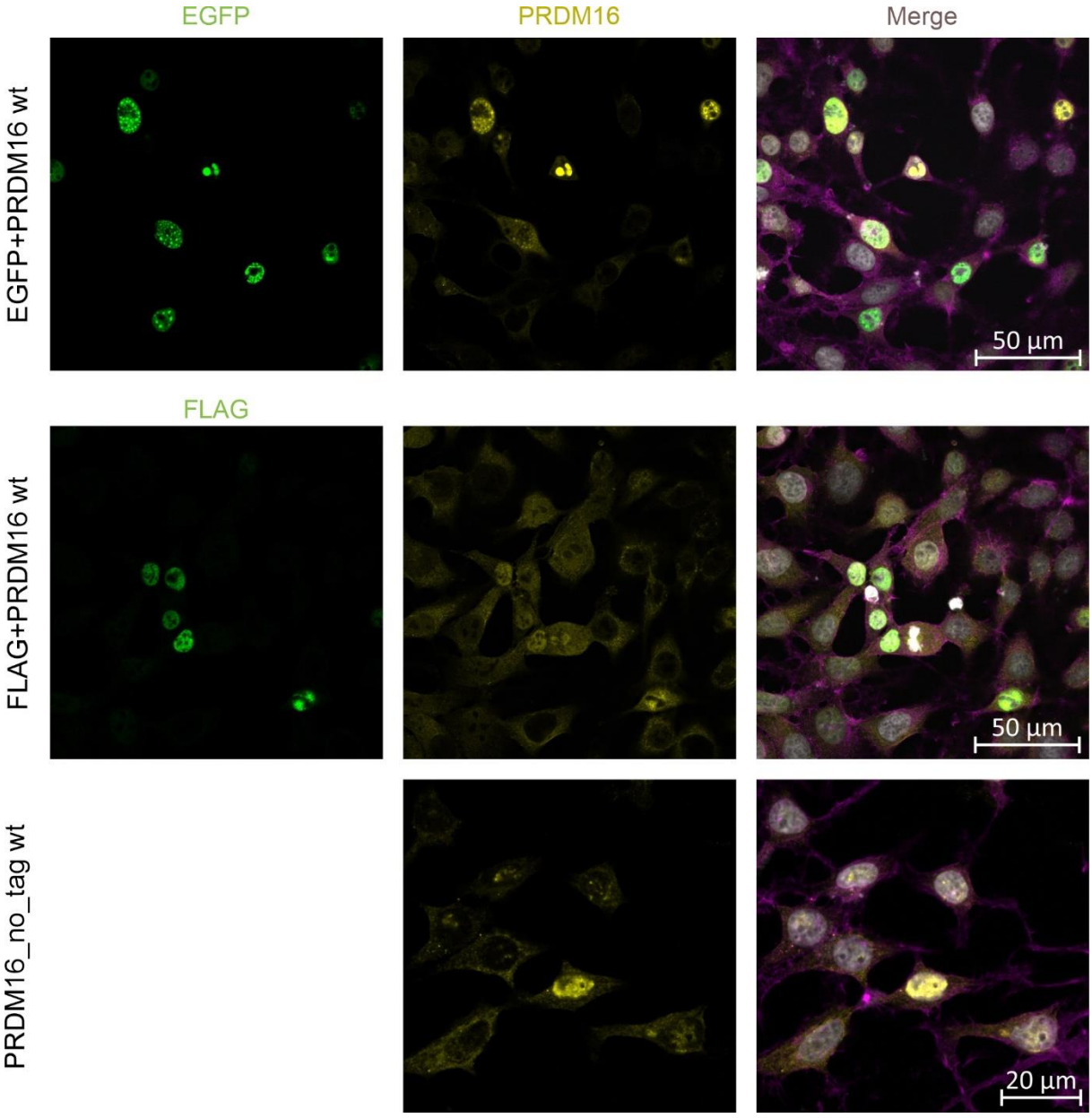


Figure 24: PRDM16 expression with protein tag or in untagged form in immunostained HEK293 cells. To investigate the location and distribution of overexpressed RPDM16 HEK293 cells were transfected with EGFP (top) and FLAG (middle) tagged PRDM16 and compared to cells with untagged overexpressed PRDM16 (bottom). Fluorescence signals were either directly detected with EGFP being a fluorophor or due to immunostaining, where fluorophor conjugated antibodies were used. Nuclei were stained with DAPI (white). Green represented detection of the EGFP- and FLAG-tag, yellow was used for PRDM16 and magenta for phalloidin staining.

HEK293 cells transfected with EGFP+PRDM16_wt expressed strong EGFP signals in the nuclei (green) and PRDM16 signals in nuclei as well as the cytoplasm (yellow) (Figure 24). Unexpectedly, both signals did not merge completely as the EGFP staining was not present in the cytoplasm. A similar pattern was observed for HEK293 cells transfected with FLAG+PRDM16_wt. With this approach the PRDM16 signal was more abundant in the cytoplasm. The samples transfected with PRDM16_no_tag_wt expressed signals in the cytoplasm and showed the localization of PRDM16 in the nucleus in different intensities, when compared from cell to cell. Therefore, the protein seemed to localize in both fractions, the nucleus and the cytoplasm. The subcellular analysis revealed discrepancies between untagged and tagged PRDM16 protein with pronounced nuclear accumulation of tagged PRDM16. These observations suggest an impact of n- or c-terminal tagging for correct PRDM16 dynamics between the cytoplasm and nucleus. For further analysis only PRDM16_no_tag proteins were used. For these initial approaches the antibody anti-PRDM16 from Thermo Fisher Scientific was used. For further approaches (Figure 25, Figure 26) the antibody anti-PRDM16 from R&D systems was used except for the PRDM16_v3fs variant.

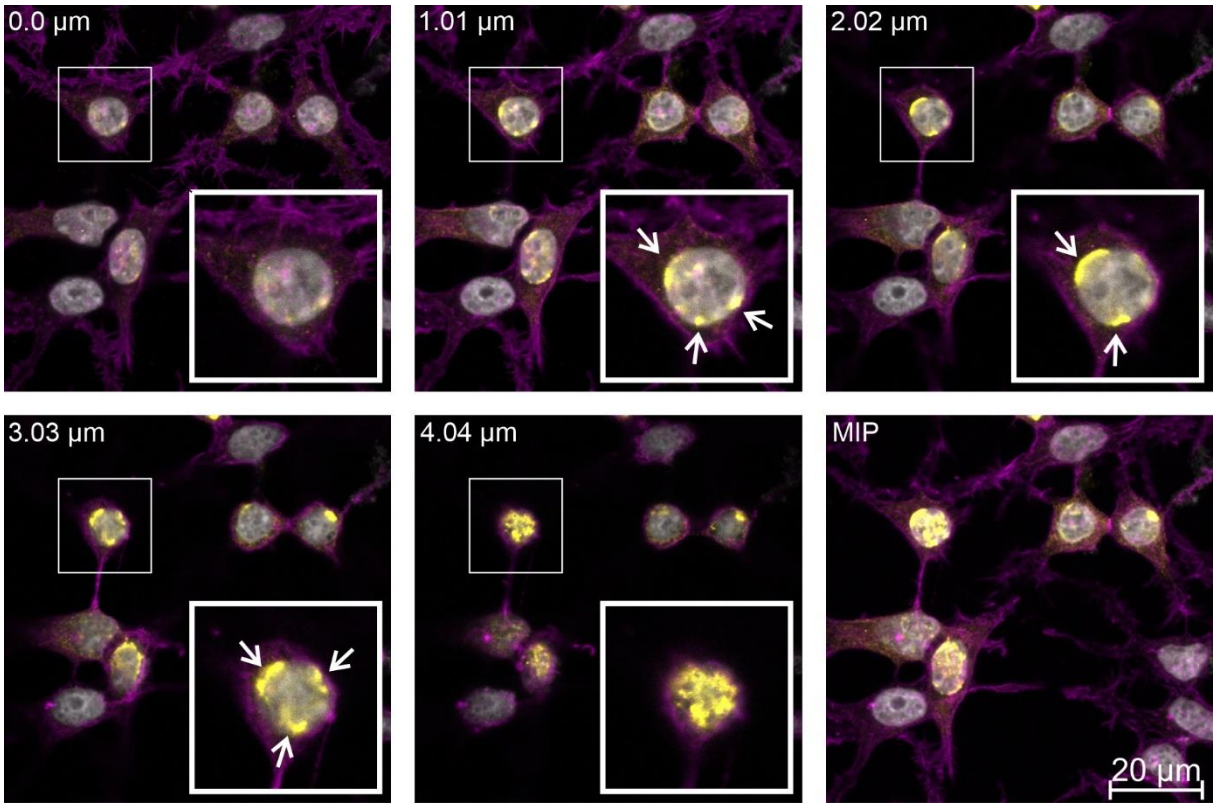


Figure 25: Distribution of untagged overexpressed PRDM16 in three-dimensional microscopy images of HEK293 cells. HEK293 cells were transfected with untagged PRDM16 and stained with DAPI (white), PRDM16 (yellow) and phalloidin (magenta). A series of images with different depths were taken for the same section (z-stack). Therefore, the location of PRDM16 could be determined in three different dimensions. White boxes were used to enlarge and highlight specific sections. The five images in different planes (0 – 4.04 µm) were merged into a maximum intensity projection (MIP).

To further investigate the subcellular distribution of PRDM16 in a cell three-dimensional records (z-stacks) were generated with confocal microscopy. The signals of PRDM16_wt were detected in the cytoplasm and the nucleus (Figure 25). At the lowest level the yellow signal was visible in small spots. In further planes the intensity of the signals increased in every compartment and detected PRDM16 concentrated at the edge of the DAPI marked nucleus (highlighted in the white box with white arrows). On the highest level the yellow signal accumulated and overlapped with DAPI detection in the nucleus. In conclusion PRDM16 was located in both the cytoplasm and nucleus and accumulated at the nuclear membrane. The maximum intensity projection (MIP) showed the accumulation at and in the nucleus.

To test the impact of *PRDM16* variants detected in CMP patients HEK293 cells were transfected with six *PRDM16* mutant constructs (Table 7) and subsequently analyzed with immunostaining and confocal microscopy. These constructs comprised PRDM16_wt, PRDM16_v1, PRDM16_v2, PRDM16_v3fs, PRDM16_v4 and PRDM16_v5. In Figure 26 signals detected for the PRDM16_wt protein were compared with immunostained samples from cells transfected with variants containing *PRDM16*. Three different stainings were used to analyze the samples. PRDM16 was detected in yellow whereas the nuclei and actin filaments were again stained with DAPI (grey) and phalloidin (magenta). The images for PRDM16_wt transfected cells showed similar results as observed before with PRDM16 localization in the cytoplasm and nucleus. Thereby some cells showed a shift in localization to the nucleus or the cytoplasm (highlighted with an arrow for a shift to the nucleus or marked with a star for a shift into the cytoplasm). Overall the signal of PRDM16_wt was evenly distributed between nucleus and cytoplasm as observed before in Figure 24 and Figure 25. For PRDM16_v1 and PRDM16_v2 a change of localization to the cytoplasm was observed and signals in the nuclei were much less distinct as for the wt (missense variants in first zinc finger domain). The opposite was examined for PRDM16_v3fs (frame shift variant after the first zinc finger domain). Signals for PRDM16 were strongly accumulated in the nuclei (only an arrow to highlight cells with PRDM16 localized in the nuclei). Of note was also the availability of the signal for PRDM16_v3fs because a premature stop codon may lead to regulatory degradation of the partly expressed protein, which was not observed in this case. The sample PRDM16_v4 expressed similar results to the wt whereas the signals of PRDM16_v5 again shifted to the nuclei, but in a more attenuated form compared to PRDM16_v3fs. In conclusion, an even distribution for PRDM16_wt in cytoplasm and nucleus was confirmed, but the introduction of variants to this protein seemed to change the conditions for localization in the cell. Signals of PRDM16 shifted to the nucleus drastically for PRDM16_v3fs and with weaker effects in PRDM16_v5. For PRDM16_v1 and PRDM16_v2 PRDM16 was detected with weaker intensity in the nuclei and stronger signals in the cytoplasm. Therefore, further analyzing the distribution of PRDM16 in the two fractions nucleus and cytoplasm seemed promising.

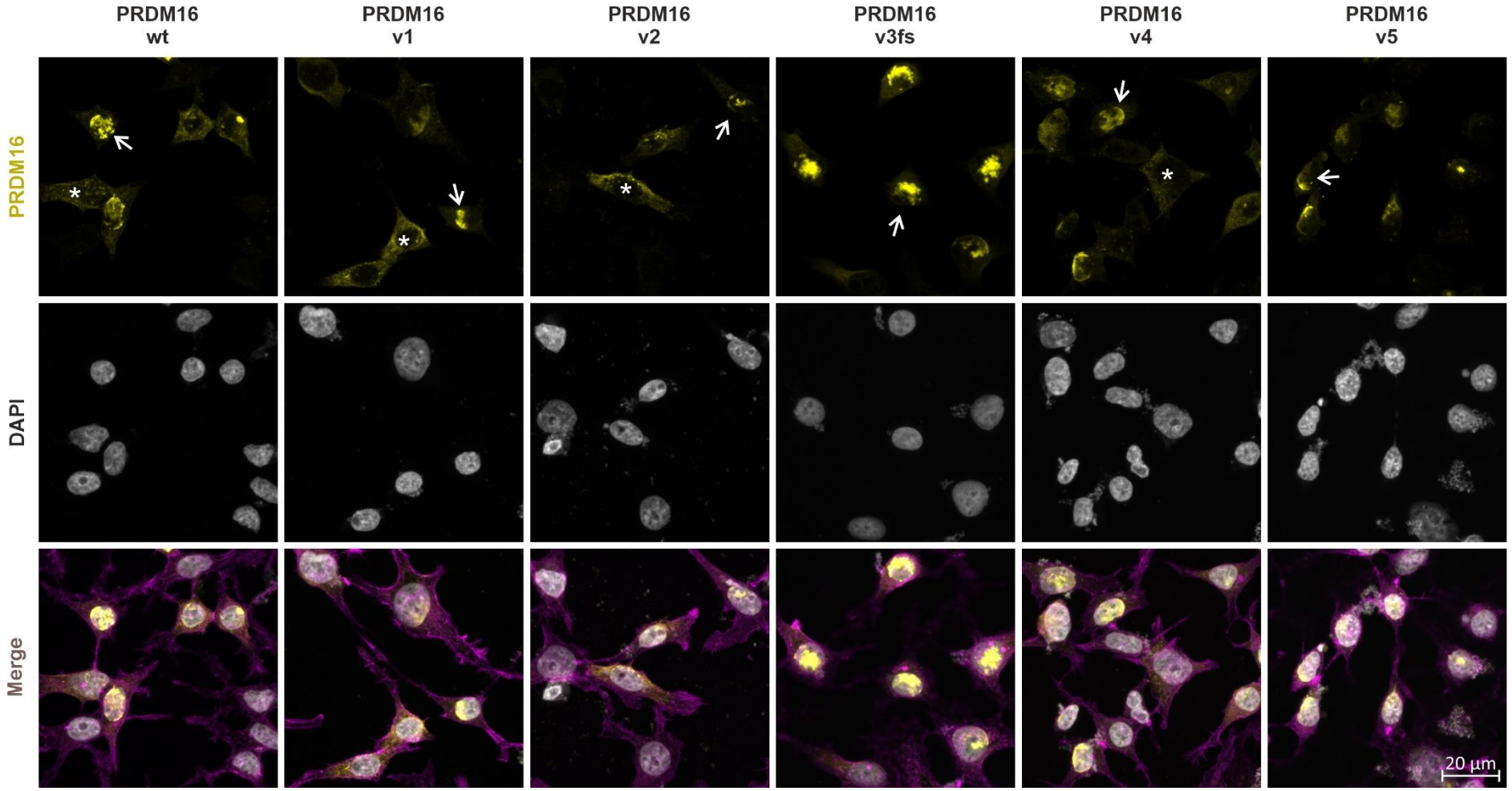


Figure 26: Immunostaining of PRDM16_wt form and CMP variants in HEK293 cells. HEK293 cells were transfected with untagged PRDM16 and stained with DAPI (white), PRDM16 (yellow) and phalloidin (magenta). Variants detected in CMP patients were introduced into the wt protein to investigate functional effects. These five variants contained four missense variants (v1 – E271K, v2 – P291L, v4 - N816S and v5 – L887P) and a frameshift variant with a premature stop codon (v3fs – R525Pfs*79).

3.2.4.2 Localization and stability analysis of PRDM16 through Western blotting

With immunofluorescence staining differences between tagged and untagged PRDM16 were observed. Thus, for more validity these findings were reevaluated with Western blotting (Figure 27). HEK293 cells were transfected with the EGFP+PRDM16, FLAG+PRDM16 and PRDM16_no_tag constructs encoding the PRDM16_wt and PRDM16_v1 – 5 forms. With every approach PRDM16 was expressed and could be detected. In general, all variants and the wt protein could be detected in the three approaches (tagged expression with EGFP or FLAG and untagged expression). Interestingly for the frame shift variant PRDM16_v3fs, containing a predicted premature stop codon, stable signals were observed, and regulatory degradation did not lead to absence of PRDM16. For the proteins expressed with tags specific antibodies against their tags were used, which confirmed the expression of artificially introduced PRDM16 into the cells and the corresponding weights of the different constructs due to the detection with a variety of antibodies (anti-EGFP, anti-FLAG and anti-PRDM16). The theoretical molecular weight of PRDM16 is 140.3 kDa. PRDM16_no_tag_wt and PRDM16_FLAG_wt were detected with approx. 170 kDa PRDM16_EGFP_wt was detected at 200 kDa due to the EGFP tag contributing 27.7 kDa to the overall protein mass. The detected signals for the tagged PRDM16 were sharp and no additional protein bands were visible. The shift in height compared to the untagged protein was accordingly due to conjugated tag. The signals for PRDM16_v3fs were appropriately reduced by 74 kDa due to the afore mentioned truncation (expected molecular weight without tag: 66.3 kDa). The results for untagged PRDM16 were similar but differed in the occurrence of a second band clearly marked by a red box. A second band also appeared for PRDM16_v4 and PRDM16_v5 with EGFP-tagged PRDM16 but will not be considered for further discussion due to the absence in other experiments. The tagged PRDM16 wt constructs were used as expression controls as seen in the detection of PRDM16_no_tag (in the right image of Figure 27, c1 – EGFP+PRDM16, c2 – FLAG+PRDM16). A second band was also visible in PRDM16 detection for c2. The expression of the different variants differed in intensity and was also observed in following experiments not shown, especially for the wt expression. Furthermore, a smear appeared with PRDM16 specific detection and therefore different antibodies against this protein were tested (Figure 28). In conclusion the expression of PRDM16 with protein tags again distinguished from untagged protein but confirmed the expression of the introduced plasmids used during transfection. Therefore, these constructs were further used as expression controls. The detected signals for PRDM16 were all considerably higher in molecular weight for unclear reasons. The truncating variant was stably expressed and showed expected reduction in height for the expected truncation.

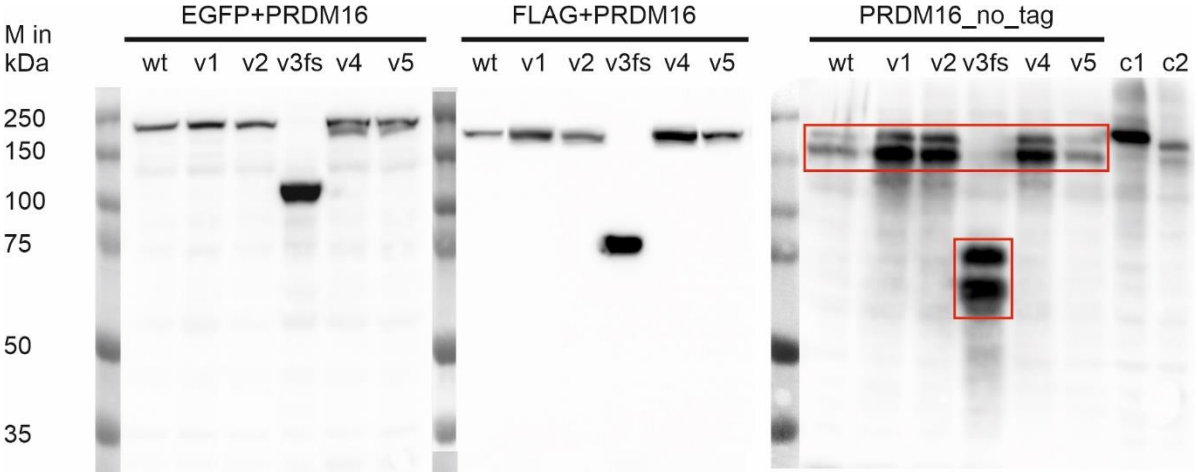


Figure 27: Western blot detection of HEK293 lysates with overexpressed PRDM16 carrying a protein tag or in untagged form. HEK293 cells were transfected with EGFP (left) and FLAG (middle) tagged PRDM16 and compared to cells with untagged overexpressed PRDM16 (right). For constructs comprising the EGFP or FLAG-tags, specific antibodies were used to only mark the protein tags. In the right part of the figure an antibody against PRDM16 was used to detect the untagged protein. EGFP- and FLAG-tag containing PRDM16 samples were used as controls (c1 – PRDM16 + EGFP, c2 – PRDM16 + FLAG) to verify the corresponding height and specificity. For every detection the wt and variants of PRDM16 were applied (v1 – E271K, v2 – P291L, v3fs – R525Pfs*79, v4 – N816S and v5 – L887P).

As mentioned before three antibodies to detect PRDM16 were tested for the expression of this protein (Figure 28). Although for the first two antibodies operating instructions suggested a predicted molecular weight of 140 kDa, specific PRDM16 bands could be observed at 170 kDa. Double bands were also common for the three different detections. Thus, the occurrence of the second PRDM16 specific band was confirmed. For the first antibody additional bands occurred with increased and decreased heights (above 250 kDa and below 150 kDa). Due to the immunization against a c-terminal peptide of PRDM16 during the creation of the Abcam antibody the frame shift variant PRDM16_v3fs was not detected. Therefore, the PRDM16_v3fs sample could be used as a c-terminal PRDM16 specific control and as seen in the left image of Figure 28 for detection of unspecific protein bands. Similar events took place for the R&D Systems antibody. The immunogen of the Thermo Fisher Scientific antibody was an n-terminal peptide and therefore PRDM16_v3fs could be detected. There were only minor additional bands with lower molecular weight detected for this antibody, but a severe smear was visible. Bands with lower weight were probably degraded elements of PRDM16 and therefore not evitable. The sharpest bands were detected with the R&D systems antibody. No smear and only minor additional bands with lower molecular weight were observed. As mentioned before the variant PRDM16_v3fs could not be detected due to a c-terminal immunogen used for antibody production. Nevertheless, further investigations were performed with the R&D Systems antibody. If detection of the truncating variant PRDM16_v3fs was necessary, the Thermo Fisher Scientific antibody was applied.

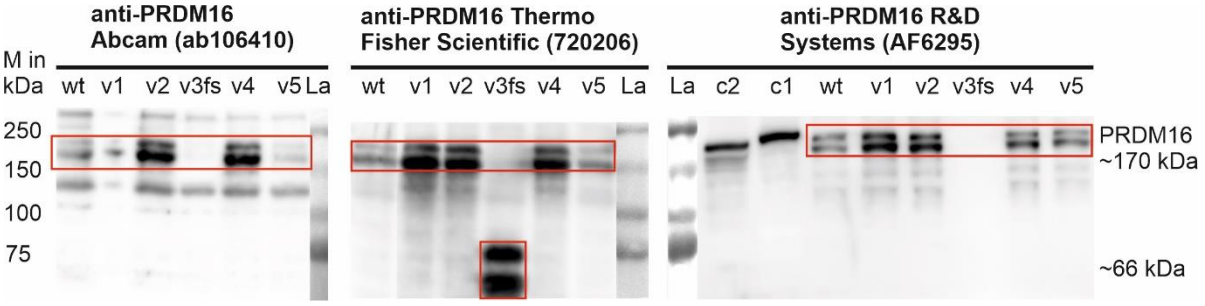


Figure 28: Antibody test specific for PRDM16 and detection of PRDM16 variants with Western blotting. Western blots were generated to compare the expression of PRDM16 wt and variants (v1 – E271K, v2 – P291L, v3fs – R525Pfs*79, v4 – N816S and v5 – L887P) with three different antibodies detecting PRDM16. PRDM16+EGFP_wt (c1) and PRDM16+FLAG_wt (c2) were used as a loading controls. PRDM16 specific bands were marked with a red box.

Subcellular analysis of overexpressed PRDM16_no_tag in PRDM16_wt and PRDM16_v1 to v5 form revealed diverse localizations. With the following approach mainly changes between the concentration of PRDM16 in a cytoplasmic and nuclear fraction were examined. To quantify this shift a procedure to separate the cytoplasmic fraction from the nuclear fraction has been applied to HEK293 cells transfected with PRDM16_no_tag constructs. Following detection with Western blot was used to characterize the distribution of PRDM16 (Figure 29). First the nuclear separation method was tested with the PRDM16_wt construct. Similar to the general signal distribution in immunostained samples for PRDM16_wt no differences were visible for PRDM16 specific bands detected in the cytoplasmic and nuclear fraction (cytoplasm - cp, nucleus - n, overexpressed, Figure 29 A). The separation of both fractions could be confirmed with the control detection of β-Tubulin at 50 kDa. β-Tubulin appeared enriched in the cytoplasm due to its role for the microtubule network. For detection of the cytoplasmic or nuclear fraction β-Tubulin and LMNA, an inner nuclear membrane marker, emerged as suited control proteins. In comparison the detection of endogenous expressed PRDM16 showed very weak or no signals although the protein concentration was amplified considerably (adherent cells harvested from two 75 cm^2 cell culture flasks for endogenous detection to two wells of a standard six-well plate with 9.5 cm^2). Probably due to the high protein concentration in the endogenous samples there were not a distinct separation of the fraction as seen for the β-Tubulin specific bands. To compare the expression of PRDM16 carrying different variants the same procedure was applied to the PRDM16_no_tag constructs PRDM16_wt and PRDM16_v1 to v5 (Figure 29 B/C). After blotting the protein samples from an SDS polyacrylamide gel onto a PVDF membrane the gel was stained with Coomassie staining solution to check for the protein transfer and homogenous protein loading (Figure 29 B). As seen for the ladder (La) the transfer worked although plenty of protein remained in the gel in the case of the samples. Of note, nuclear and cytoplasmic fraction showed diverse band

pattern confirmed successful separation. Moreover, the samples were evenly loaded as seen in the image. Specific expression bands for PRDM16, typically examined in overexpression approaches, were not observed. More evidence for a successful separation of the two fractions could be confirmed due to the Lamin C and β-Tubulin signals (Figure 29 C). Lamin C was absent in the cp fraction and β-Tubulin was absent in the nuclear fractions. The signals detected for PRDM16 however differed slightly. In the wt sample the strongest signal was examined for the lower band in the cytoplasmic fraction and with the intensity of the other PRDM16 bands evenly distributed the localization of PRDM16 was slightly shifted to the cp. Differences for the samples PRDM16_v1 to v5 were observed for the upper band in the cp fraction resulting in weaker signals. Weaker signals in the lower band of the cp fraction also led to the impression, that PRDM16 was present with higher concentrations in n fraction for the variants PRDM16_v2, PRDM16_v4 and PRDM16_v5.

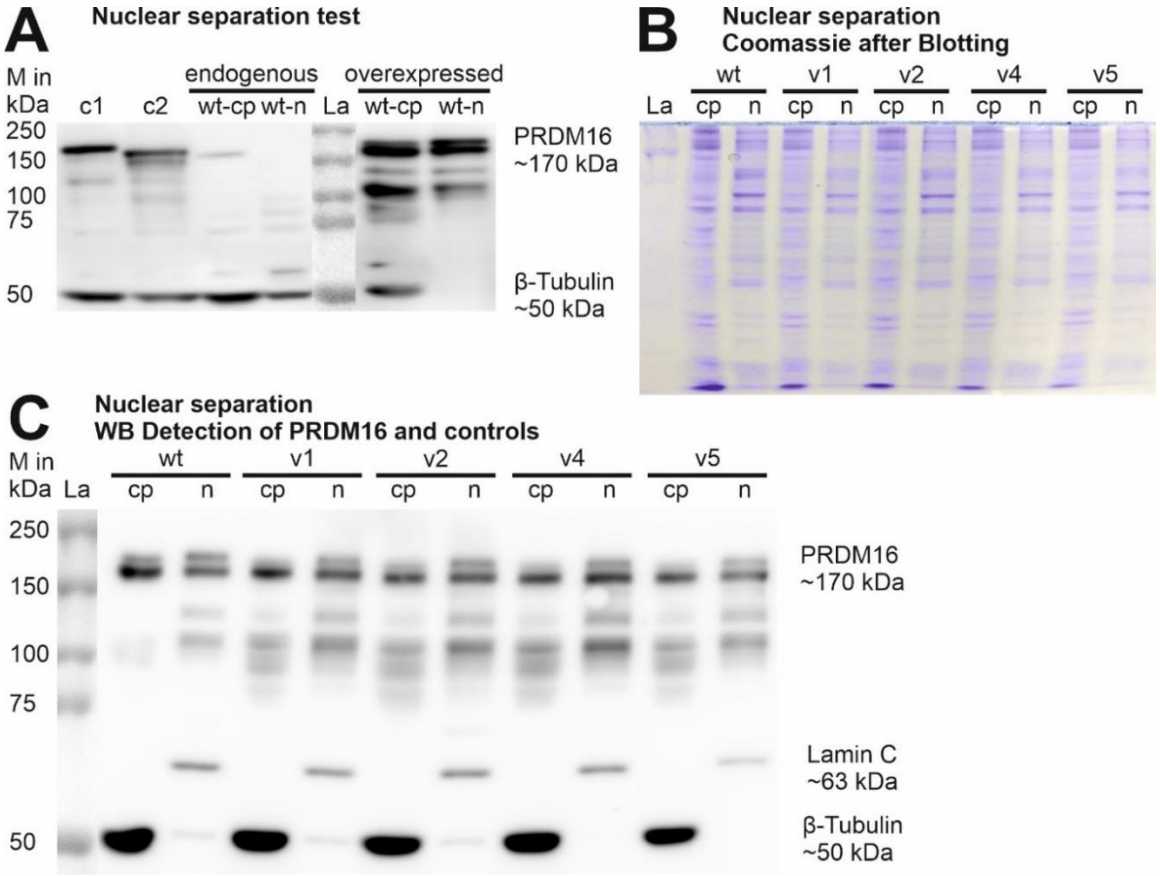


Figure 29: Detection of PRDM16 wt and variants in separated cell fractions (nucleus and cytoplasm) with Western blotting. Western blots were generated to compare the expression of PRDM16_wt and variants (PRDM16_v1 – v5) in the cytoplasm (cp) and the nucleus (n). These two sections were separated and examined separately. **(A)** The two fractions for endogenous PRDM16 expression and overexpressed PRDM16_wt are shown and compared to PRDM16+EGFP_wt (c1) and PRDM16+FLAG_wt (c2) control expression. **(B/C)** Overexpression of PRDM16_no_tag constructs was then compared. The Coomassie staining of the SDS polyacrylamide gel further processed to the Western blot in C was used to show successful transfer to the blotting membrane (see ladder - La) and equal loading of the samples.

For more validity the same experiment was performed at least four times under same conditions, but quantified signals differed too much and produced very heterogeneous results. The main problem here were considerably weaker signals for wt samples although same amounts of protein were loaded (confirmed by analysis of Coomassie stained gels). In conclusion differences in PRDM16 signals were observed, but the results were too heterogenous for further analysis. A major concern was the stability of the PRDM16_wt expression. Without distinct detection for this construct, results from the other variants cannot be compared.

To further investigate the stability of PRDM16_wt and variants, HEK293 cells were transfected with *PRDM16* and harvested after CHX incubation for 0h, 8h, 16h and 24 h). The samples were analyzed via Western blot. In Figure 30 a set of PRDM16_wt and the five different variants PRDM16_v1 to v5 from the stabilization approach are presented. For each transfection the four time points at 0, 8, 16 and 24 h of incubation after addition of CHX to inhibit protein expression are shown.

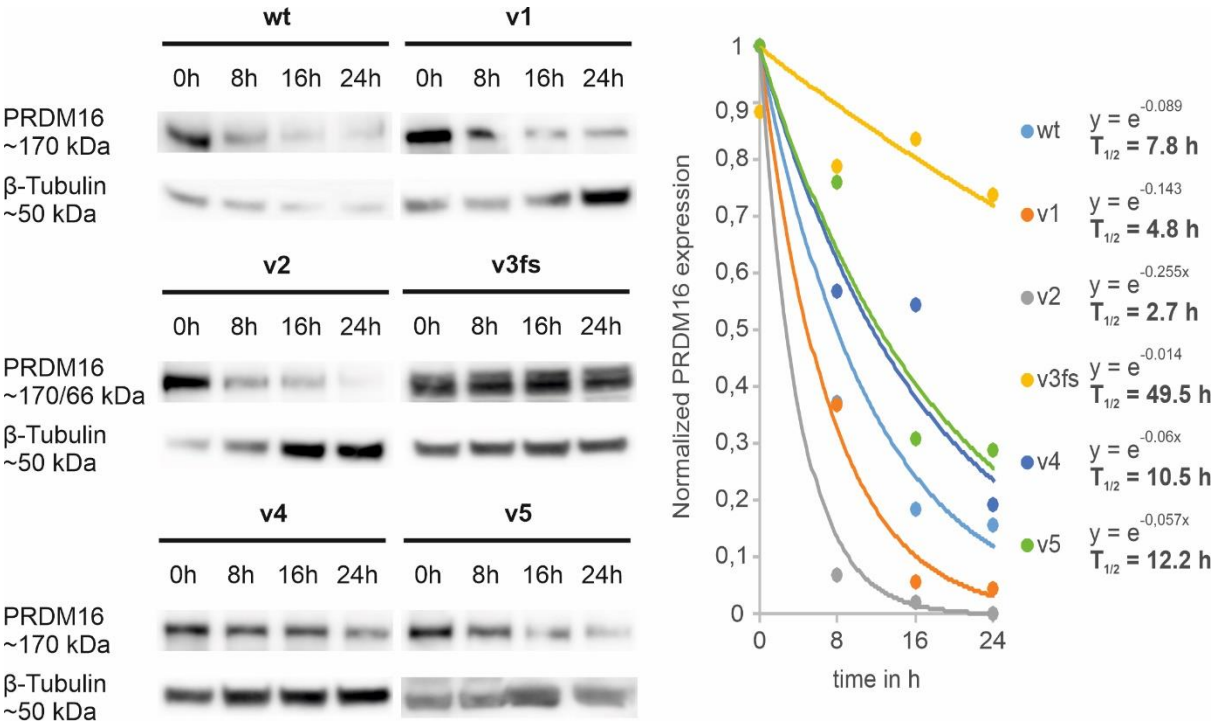


Figure 30: Time series of PRDM16 expression after CHX incubation for stability determination. HEK293 cells were transfected with *PRDM16*_wt and PRDM16_v1 to v5 constructs. After 24 h of incubation with CHX (protein expression inhibitor) differences of stability in *PRDM16* variants were determined. Lysates were generated after a series of time points (0 h, 8 h, 16 h, 24 h) and detected with Western blotting (left). The expression levels at different time points were correlated with the intensity of bands detected in WB to calculate the half time of the corresponding protein (right).

PRDM16 expression detected with a molecular weight of approximately 170 kDa or 66 kDa for the frameshift variant was compared to the expression control of β-tubulin at around 50 kDa. With the quantified intensities of these samples an exponential fit was applied and with the resulted formula the half-life of the protein was determined. Visually the PRDM16_wt, PRDM16_v1 and PRDM16_v2 were observed with strong decline of signal, whereas decline in signal for PRDM16_v4 and PRDM16_v5 was elongated and PRDM16_v3fs showed no decrease in intensity and seemed therefore stable through the entire incubation time. Noticeable here were the observed quantities of the β-tubulin detection. Homogenous protein loading was achieved with BCA analysis and subsequent adjusted protein loading. With consistent protein concentrations unstable proteins were degraded, whereas more stable proteins relatively increased. This seemed to be the case for β-tubulin most visible for the sample PRDM16_v2 (Figure 30). Quantities of β-tubulin again were used as loading controls. The calculated half-lives $T_{1/2}$ in Figure 30 corresponded with the visual estimations. Whereas the wt expressed a $T_{1/2}$ of 7.8 h, the variants PRDM16_v1 and PRDM16_v2 were determined with shortened half-lives of 4.8 h and 2.7 h. The $T_{1/2}$ values for PRDM16_v4 and PRDM16_v5 were elongated with 10.5 h and 12.2 h. The result with the biggest difference to the wt was observed in cells transfected with the frameshift variant PRDM16_v3fs. The signal intensity did not decrease considerably and therefore resulted in a calculated $T_{1/2}$ with 49.5 h. These results suggest difference in protein stability of *PRDM16* variants, but due to the heterogenous results should be analyzed with caution and thus further confirmation is needed.

4 Discussion

In this thesis, the genetic basis of a cohort comprising 80 pediatric cases of primary non-syndromic CMP were analyzed. The frequency of VOI in CMP-relevant genes and genotype-phenotype correlations were evaluated. DCM followed by HCM patients represented the majority of the CMP cases with decreased numbers for LVNC and RCM and nearly no cases of ARVC. Overall, the gender was distributed equally, but differed for DCM and HCM (more affected females for DCM and more affected males for HCM). Considering the age, the HCM group showed later disease onset and a wider range. Adverse events were considerably higher for individuals affected by DCM and RCM, including considerably more frequent heart transplantations and deaths. Adverse events were correlated with the number detected VOI and expressed a higher risk to experience such events with increasing number of VOI.

Most frequently VOI were detected in sarcomere, Z-disc or desmosome genes. These groups contained the disease genes with most frequently detected VOI in *MYH7*, *MYBPC3*, *TNNI3* and *DSP*. Especially VOI in *MYH7* and *MYBPC3* accumulated in HCM patients. The other groups showed more heterogeneity and RCM cases represented an exception with truncating *TTN* VOI most frequently detected.

In the group of most frequently detected VOI *TNNI3* was the third most important gene. Of note were *TNNI3* variants due to a homozygous splice site variant leading to a complete loss of TNNI3 in the myocardium and resulting in the compensatory increase of TNNI1. Confirmed was this case with a second homozygous frame shift variant examined with similar results detected in an additional pediatric CMP case.

PRDM16 is a transcriptional regulator and several variants were detected in pediatric CMP. Further availability of patients with adverse cardiac phenotypes identified 16 variants detected in *PRDM16*. These alterations occurred in six different cardiac diseases and comprised missense, splice site frame shift and stop gain variants. Through bioinformatic analysis structural and accessible regions of the protein were revealed and associated with the detected variants. Immunohistochemistry experiments showed different properties for the expression of PRDM16 when tagged compared to untagged form. Furthermore, *PRDM16* containing CMP variants mediated altered localization in the cell with condensed signals in the cytoplasm or the nucleus. Therefore, the cytoplasmic and nucleic fractions were analyzed separately in Western blot approaches. Stabilization test of PRDM16 wt and variants showed differences in the protein half-life time correlating with localization in the cells suggesting more stability of PRDM16 when located in the nucleus.

4.1 Clinical characterization of pediatric cohort

In the recruitment period of pediatric patients with primary CMP mainly DCM and HCM cases were diagnosed. These results are consistent with current findings (Lee et al., 2017). Less frequently were cases with LVNC and RCM. Children affected by ARVC were rarely recruited. The sex of the patients was equally distributed overall but differed for HCM and DCM cases. The gender in the other subgroups was equally distributed (LVNC) or the group was too small to comment (RCM, ARVC). Males were affected more frequently by HCM, whereas more females were affected by DCM. For HCM cases both in adults and children a higher incidence for males is known (Lu et al., 2020; van Driel, Nijenkamp, Huurman, Michels, & van der Velden, 2019). As for DCM patients, findings are contradictory. In adults, all-cause mortality was described as an independent predictor for females (Doesch et al., 2014). An Australian childhood cohort presented a gender distribution with 56% females affected by DCM (103 out of 184, Nugent et al., 2003), which coincides with characteristics of the cohort described in this thesis. Contradictory to that, in a cohort containing 1426 pediatric DCM cases 769 males were affected by the CMP (54%, Towbin et al., 2006). In these publications differences between ethnic groups were also described as an influential factor with indigenous and African American patients diagnosed with a higher incidence. These factors may alter the gender specific distribution. Additionally, DCM can be caused by different environmental factors and is therefore more susceptible to fluctuations in their patient's characteristics (Lipshultz et al., 2019).

Age specific effects compared to the other phenotype groups were only observed for HCM. The onset was at later age and the age range was wider. In a study of childhood patients affected by HCM, the poorest prognosis has been examined with an age younger than one year. In cases with an age higher than one year no age specific associations were observed (Colan et al., 2007). Partly these findings were confirmed with results in this thesis. For HCM cases survival seems independent from the age of diagnosis. Only one patient in this sub cohort underwent heart transplantation and died half a year after the operation with an age at diagnosis of 0.3. Therefore, in this case the early onset of the disease increased the severity, but six different cases with patients younger than one year affected by HCM presented different results and did not show any occurrences of adverse events. Reasons for this observation could be a milder and elongated progression of the disease or higher thresholds to develop adverse events due to hypertrophy of the heart compared to other disease symptoms such as dilatation (see athlete's heart, Baggish & Wood, 2011).

In our cohort, adverse events were mainly observed in DCM and RCM patients, whereas HCM patients were nearly event free. In other studies, an early onset is often described as risk factor to develop a more severe phenotype for HCM and DCM (Lipshultz et al., 2013; Alexander et

al., 2018; Colan et al., 2007). Regarding age, 11 DCM patients observed with adverse events were younger than one year (seven were older). LVNC cases affected by MACEs seemed also age specific and were diagnosed with an age younger than one year, although the occurrence with three cases was rare. RCM patients were older than one year and up to 15.9 years old at diagnosis. Therefore, in this subgroup the development of adverse events seemed age independent, although these cases presented the worst prognosis together with DCM patients.

4.2 Interpretation of genetic variants detected in 80 pediatric cardiomyopathy patients and efficacy of genetic screening

According to the guidelines of the American College of Medical Genetics and Genomics variants were categorized as VUS, LP or P in this thesis (VOI, Richards et al., 2015). LP and P variants are considered to be disease-causing according to current ACMG guidelines for CMP. Therefore, in 30 of 80 patients at least one P or LP variant was detected concluding in 38% of cases the underlying genetic alteration has been discovered.

4.2.1 Cardiomyopathy specific observations

The highest rate of VOI was observed in HCM patients with 57%. Most frequently VOI were detected in *MYH7* and *MYBPC3*. In all HCM patients 37% of VOI were observed in these two genes (10 out of 37). In P and LP variants this term went up to 69% (9 out of 13). Former studies showed similar results with slightly higher frequencies over 75% (Morita et al., 2008). Categorized by functionality VOI detected in HCM patients were mainly observed in sarcomere and Z-disc genes (57% and 26%). Thus, HCM cases presented the most distinct genotype and as described in the literature HCM is a disease of the sarcomere (Thierfelder et al., 1994). Interestingly, the proportion of VOI in *MYH7* and *MYBPC3* decreases with earlier onset (three *MYH7* and one *MYBPC3* VOI in 14 patients < 10 years of age, one *MYH7* VOI in seven patients < one year of age). Therefore, VOI in *MYH7* and *MYBPC3* seem to manifest in adult CMP patients with a later onset of the disease.

DCM cases represented the biggest group in this work. VOI in these cases were detected in a heterogenous spectrum of genes. *MYH7* remains the gene with most frequently observed VOI in this sub cohort with detection in four patients, but not as clearly as within HCM cases. Altered genes not considerably less frequent were *BCL2 associated athanogene 3* (*BAG3*), *DSP*, *NEXN*, *PKP2* and *TNNT2* (detected in three patients). In P and LP variants only one variant was detected in *MYH7* and *MYBPC3* each. Interestingly, P and LP variants comprised only sarcomere genes. Including VUS, VOI were most commonly detected in sarcomere, desmosome and Z-disc genes. DCM is known to be heterogenous with different spectra of

disease genes for adult, pediatric and infant cases (Pugh et al., 2014). In adults truncating *TTN* and *LMNA* variants are predominating. In infants the occurrence of variants in these genes is described to be drastically decreased. Here, this statement is confirmed for infants. In contrast to the published data this observation is not limited to infants with an age younger than two years and in this work observed in all pediatric age groups. In infants it is also described that alterations in *MYH7* are predominantly, which cannot be entirely confirmed in this work. Compared to HCM patients the frequency of *MYH7* decreases partly with age and mainly with severity (occurrence of MACE's or death as term for severity, 11 from 18 HTX or deceased patients ≤ 1 year old).

RCM patients were primarily affected by VOI in sarcomere and transcription/splicing genes. Pathogenic and likely pathogenic variants were only detected twice in TNNI3 and once in *TTN*. In a former study variants in sarcomere genes were associated with the development of RCM in children, which coincides with the findings in this thesis (Kaski et al., 2008). In adults, genetic analysis is restricted to consanguineous families and case reports (Peled et al., 2014). The interpretation of these results is limited due to the small numbers of diagnosed RCM cases (Kaski et al., 2008).

Patients affected by LVNC showed heterogenous genotypes. *MYH7* again was the most frequent gene with detected VOI followed by *ACTN2* and *TAZ*. Similar genotypes were detected in other studies, again with the absence of *TTN* variants in pediatric cases (van Waning et al., 2018; Wang et al., 2017). Complex *MYBPC3* genotypes specific for pediatric cases were not examined in this work. Sarcomere, desmosome and Z-disc genes were most prominent in the spectrum of LVNC VOI. Here, the proportion of *MYH7* and *MYBPC3* VOI did not change considerably regarding age and severity of the disease. The P and LP variants contained mainly sarcomere genes.

In conclusion, genes categorized into the functional groups of sarcomere, desmosome and Z-disc genes were most frequently detected with VOI. Sarcomere and Z-disc genes are influencing the generation of the force and rhythm used for normal cardiac function and were expected to lead to adverse remodeling of the heart. Interestingly desmosome genes are not commonly associated with CMPs except of ARVC (58% of ARVC patients carried variants in desmosomal genes and 93.5% of all ARVC related variants were detected in desmosome genes , Lazzarini et al., 2015). In this work only two patients were diagnosed with ARVC. In these two patients three VOI in desmosome genes were detected accordingly (DSC2:p.I345T, desmoglein 2 (DSG2):p.340V*, PKP2:p.R573Efs*4). Desmosomes mediate strong cell-cell contacts and localize in tissues with intense mechanical stress like the heart. Thus, mutations in genes of the desmosomes may influence mechanical cardiac properties.

4.2.2 Specific cardiomyopathy genes

In general, *MYH7* and *MYBPC3* showed most frequently VOI in the pediatric CMP-80 cohort. VOI were frequently observed on HCM, DCM and LVNC patients. This observation changed with earlier onset and increasing severity. In conclusion, variants in *MYH7* and *MYBPC3* are mostly not as adverse as variants in other CMP relevant genes and therefore manifest in adults with a later disease development. In more detail, disease onset in 13 patients with *MYH7* VOI ranges from infancy to adolescence. Five out of eight VOI were previously reported and detected in both adults and children (Table S 6). This again suggests an age independency of *MYH7* variants causing the development of CMP. Seven *MYH7* missense variants were interpreted as P or LP. All seven P or LP missense variants localize within the HCM cluster region ranging from amino acid position 167 to 931 supporting their pathogenic character (Kelly et al., 2018). Moreover, a reliable MYH7 disease and risk identification is necessary due to the high detection rate of already known missense variants (only four novel MYH7 variants out of 13, Richards et al., 2015; Kelly et al., 2018). Nevertheless, the assessment and rate of pathogenic *MYH7* variants is highly specific to the availability of comprehensive genetic and functional data. Similar events are applicable for *MYBPC3*. Three out of six VOI were detected in pediatric and adult cohorts and six out of eight variants determined as P or LP (Table S 7). Like for *MYH7* disease specific effects of *MYBPC3* are well known and often described for CMP patients especially cases with HCM (Toepfer et al., 2019).

As discussed before in section 4.2.1 observation of *TTN* VOI was of note. In this study, *TTN* missense variants were excluded, due to their currently uncertain interpretation. Nevertheless, four heterozygous VOI in *TTN* were detected. These genetic alterations included a frameshift variant, inducing a premature stop of translation and three changes near splice sites potentially affecting splicing. One DCM and two RCM cases, observed with VOI in *TTN*, underwent HTX. This may result from a link of *TTN* truncation with pediatric RCM, eventually influenced by a combination of a more complex genetic alteration, and a critical importance of both *TTN* alleles for postnatal development (Peled et al., 2014). In conclusion, VOI in *TTN* were rarely detected in all CMP groups with exceptions in RCM cases. In line with these results, former publications showed a much lower frequency of truncating *TTN* variants in pediatric cases with DCM in comparison to adults (Pugh et al., 2014; Fatkin et al., 2016).

4.2.3 Analysis of segregation in family members

In 77% of families, the inheritance of genetic alterations and clinical status in first degree members was tested. With this approach seven *de novo* variants (10%) were detected. In another study a higher ratio of *de novo* variants was previously described (Vasilescu et al., 2018). In the mentioned study 46% *de novo* variants in a cohort of infant childhood-onset CMP

were reported (median diagnosis age of 0.33 years) and 33% of cases were affected by a systemic disorder. Due to numerous *de novo* variants associated with cardiac disease observed as standing variation in ExAC, the association of this type of variant with disease onset is questioned. Thus, these alterations are less involved in monogenic disease models or present less risk as contributors for cardiac disease (Paludan-Muller et al., 2017).

4.2.4 Interpretation on genetic variants according to the guidelines of the American College of Medical Genetics and Genomics

In this work the guidelines of the American College of Medical Genetics and Genomics were used to interpret the pathogenicity of detected variants (Figure 6, Richards et al., 2015). Different aspects influential on the outcome of pathogenicity, like population data, computational and predictive data, functional data, segregation data and *de novo* data, were used to evaluate variants. The categories allele data, other databases or other data were not included into the evaluation, because these aspects could not be properly applied for this cohort. The principle of this method was to activate fields or arguments in the afore mentioned sections with graded impacts of the pathogenic character (supporting, moderate, strong and very strong). The sum of all arguments then determined the overall pathogenicity of the variant, in this work categorized as uncertain (VUS), likely pathogenic (LP) or pathogenic (P). This method was developed in 2015 and underwent refinement to the current implication described in this chapter. Thus, only the highest argument activated for one section is now counted for the interpretation of pathogenicity (Abou Tayoun et al., 2018).

For population data two points were important. The absence of variants in population databases, including rare variants with a minor allele frequency of 0.01% and less, and the statistically enriched occurrence of these variants in disease population databases. For the first point the MAF was chosen to be lower than 0.1% to exclude variants acting as a modifying factor for further analysis. Recent studies showed modifying aspects of variants distinguishing a mild phenotype from a severe one (Gifford et al., 2019). Therefore, the threshold for exclusion should not be too low. The threshold to activate this field for a moderate pathogenic argument was chosen to be 0.01% to consider the prevalence of CMP in the population (Lek et al., 2016: rare variant defined with MAF < 0.1%; Kelly et al., 2018: MAF < 0.004% recommend for MYH7 variants). For the second point disease databases like the Atlas of Cardiac Genetic Variation are necessary to calculate statistical values for the enrichment of variants in CMP patients (Odds Ratio > 5 is usually used). Data from cohort studies with CMP phenotypes can be used additionally to increase statistical relevance. Limitations here are the lack of disease databases and missing updates on the existing ones. A Database collecting the sequencing results of published cohorts for any specific disease would be of help. With the variant database ClinVar

first approaches were implemented but further refinement is needed (variants listed without case numbers or frequency of the detected variant, Landrum & Kattman, 2018).

The second category, computational and predictive data, is mainly determined by predictive online tools and databases. For the supporting argument PP3 the predicted effect of a variant was considered for the pathogenicity. Therefore, the procedure was optimized in the way, that at least two out of three tools had to predict an altering effect. Thus, different outcome of predictive tools were taken into consideration. The moderate arguments PM4 and PM5 as well as the strong argument PS1 were conditional on their availability on disease databases. To determine the validity for these arguments the occurrence of variants was checked on ClinVar. The very strong argument PVS1 depended on available data on loss of function effects of genes, when absent or degraded. The knowledge from knockout mice makes it difficult to apply if the heterozygous state was not analyzed and limits the possibilities to interpret PVS1. The population database gnomAD implemented values on their site to estimate the effect of truncating variants in a gene for such circumstances. Thus, the value Probability of being loss-of-function intolerant (pLI) and the ratio observed single nucleotide variation to expected single nucleotide variation (o/e) were introduced. Important here is the occurrence and frequency of truncating variants in the population database. Therefore, the observed truncating variants are compared to the amount of calculated expected variants resulting in an o/e value or ratio. This value is further used to calculate the pLI, which represents the probability of belonging to genes affected by haploinsufficiency (Lek et al., 2016). Thresholds for both terms should be chosen stringent and it is important to consider protein length for the confidential interval (CI) of o/e . GnomAD suggests a threshold for the upper CI<0.35 of the 90% confidential interval and an pLI≥0.9. Thus, it is expected, that truncating variants in genes affected by haploinsufficiency are less frequent than calculated and the o/e ratio is reduced. As described before published data for null variants and haploinsufficiency as well as frequently occurring P and LP truncating variants on ClinVar were considered for the activation of the argument PVS1.

For the section functional data mainly described effects of variants or domains of the affected proteins were used to activate the corresponding fields. One exception is the supporting field PP2. Similar to PVS1 the calculated value Z-score and o/e of the population database gnomAD were used to estimate the effect of missense variants. The suggested threshold for the Z-score was ≥3.09 and for the upper 90% CI again <0.35 (Samocha et al., 2014; Lek et al., 2016). For functional relevance frequency of pathogenic and benign variants were additionally checked on ClinVar when needed. The fields PM1 and PS3 are mainly reliable on literature or databases with gathered information taken from published facts regarding the function and impact of variants or functional domains. Especially these categories are vaguely defined and can be interpreted differently. It can be argued that only distinct results restricted by examined in humans and *in vivo* can lead to the activation of the PS3 argument and therefore any other

experiment would be excluded as false negative. On the other site well characterized functional domains can be very long with regions in which variants can be tolerated may resulting in variants determined as false positive. Thus, the interpretation of functional data has to become more distinct and uniformly assessed.

In contrast, the sections segregation and *de novo* data are well defined. For segregation data the occurrence of variants in affected family members is determined in meioses. A threshold of at least three meioses is considered as describing the supporting character of this argument and therefore activated (Kelly et al., 2018). With increasing evidence and higher numbers of affected family members carrying the same variant, this argument is gaining more value and can be upgraded (≥5 meioses for moderate, ≥8 meioses for strong). Fields regarding the *de novo* character of a variant are similarly dependent on the detection of variants in family members. The moderate character of a variant detected only in an affected index patient and absent in unaffected parents can be upgraded to strong when paternity is determined.

In conclusion, Richards et al. introduced a tool to better estimate the pathogenic character of genetic alterations. The approach was complex and well thought, but not clear enough in some arguments. The availability of new predictive tools and ongoing functional assessment of genetic variation is necessary for refinement of this method, but especially functional effects of variants are rarely proven and need further investigation before linking more genes to specific phenotypes.

4.2.5 Implications for genetic diagnosis in pediatric cardiomyopathy cohorts

Whether whole genome sequencing (WGS) or whole exome sequencing (WES) in CMP will be used as a diagnostical tool to provide extensive insight into the underlying genotype is currently highly debated (Cirino et al., 2017). One major concern in this field is the paucity of genetic and functional evidence for variants detected by WGS (Minoche et al., 2019). In another study, WES yielded in a diagnosis rate of more than 50% in DCM patients with an onset <18 years (Herkert et al., 2018). Other recent publications stated rates of 26% to 39% for the identification of pathogenic variants in pediatric CMP (Ouellette et al., 2018; Vasilescu et al., 2018). The results of this work suggest genotype determination in pediatric CMP patients as soon as possible to induce appropriate therapeutic treatment. Here, a major focus should be on sarcomere genes as most frequently group for detected VOI. Therefore, genetic testing by NGS improves risk stratification for pediatric CMP patients and can be of much use in familial cases to diagnose potential CMP cases exposed to a genetic burden (Wilcox & Hershberger, 2018). With this approach the majority of genetic causes for CMP probably will be detected, but severe cases with unclear genotypes need further analysis. With increasing complexity detected in genetically inherited diseases the necessity to apply methods like WES

and WGS becomes demanding (Gifford et al., 2019). Additionally functional analysis of genetic alterations have to be implemented more to better understand disease mechanisms and improve the interpretation of VOI.

4.3 Increased numbers of genetic variants in a cardiomyopathy patient may present a higher risk factor

As mentioned before, in 38% of all cases the probability is high, that the genetic basis for the development of a CMP was detected with at least one P or LP variant. Concluding this result, 62% are still unresolved. One explanation could be the occurrence of CMP initiated by environmental factors. A virus infection of the myocardium could lead to more severe symptoms resulting in a DCM. The involvement of a variant in a gene unknown to have an effect on the development of the heart would be another explanation. A third possibility is the existence of complex phenotypes. In theory in this scenario not only one variant is the genetic basis of the CMP. Thus the monogenic approach is shifting to an oligogenic one. The contribution to the cardiac remodeling in these cases is open for discussion and experimentation, but Gifford et al. showed in 2019 that at least three variants in cardiac genes contributed to the development of a severe LVNC. Patient specific heterozygous missense variants in the genes *MYH7*, *myocardin-related transcription factor B* (*MKL2)* and *NK2 homeobox 5* (*NKX2-5*) led to an early onset of the CMP. Severe events were reported for three children, all carrying variants in the afore mentioned genes. The father was asymptotic but showed subtle signs of LVNC. He carried the two variants in MYH7 and MKL2, whereas the mother, without any signs of CMP, carried a variant in *NKX2-5*. During whole-exome sequencing only variants detected in all three children were included. Further filtering for cardiac expression left only three variants, two overlapping with the father in *MYH7*, *MKL2* and one overlapping with mother in *NKX2-5*. The authors augmented with evidence in genetically modified mice and human induced pluripotent stem cell–derived cardiomyocytes, that *NKX2-5* operated as a genetic modifier in this case and is therefore lowering the threshold to develop the disease and its severity. Of note, the MAF of the variant in *NKX2-5* was only 0.0012 and would have been sorted out in our cohort. Hence the definition of rare variants should be further discussed.

In the cohort described in this study the number of VOI, including variants determined as VUS, P and LP, was correlated with MACE's in section 3.1.4. Aside from the results that DCM and RCM had the most severe disease progression and patients with genetic CMP had worse prognosis than sporadic cases, the number of VOI affected the disease progression significantly and the proportion of event free cases was smaller for patients carrying more than

one VOI. This strengthens the hypothesis of oligogenic disease onset and a modifying character of some detected genetic alterations.

4.4 TNNI switch in cases with homozygous TNNI3 variants

In this study *TNNI3* variants were of note, because they were detected quite frequently (six VOI, third most frequent gene) and heart tissue of two patients with either DCM (1-II:1, p.Arg69Alafs*) or LVNC (2-II:3, c.24+2T>A) carrying homozygous truncating *TNNI3* variants were available (Table S 8). Heterozygous *TNNI3* variants described as disease causing are already published in numerous cases with DCM, HCM, and RCM. Nevertheless, homozygous alteration of *TNNI3* is a rarely described (Burke, Cook, Seidman, & Seidman, 2016). So far, three homozygous *TNNI3* cases in DCM patients have been reported. These three cases comprised the missense variant p.Ala2Val (DCM), the synonymous and exonic splice site variant c.G150A (p.Lys50Lys) and a genomic deletion containing the exon 8 of *TNNI3* and the entire *TNNT2* gene (Belkaya et al., 2017; Murphy et al., 2004; Streff et al., 2019). The absence of the TNNI3 protein in cardiac tissue was shown in this work for patients carrying homozygous and truncating *TNNI3* variants. Furthermore, TNNI1 was upregulated probably to compensate the lack of TNNI3. The TNNI1 protein presents the fetal isoform of troponin I and therefore these results suggest an impairment of TNNI isoform switch form fetal to postnatal cardiac development. Thus, the absence of TNNI3 actually resulted in restrained signaling to downregulate TNNI1, normally observed in the troponin I switch during postnatal cardiac development and human induced pluripotent stem cell–derived cardiomyocytes maturation (Bedada et al., 2014; Sasse et al., 1993). Actually, this switch is already described in mice and was now proven in human tissue (Huang et al., 1999). It is believed that TNNI1 partly compensates the absence of TNNI3 but is also depleted after a certain amount of time. Thus, the postnatal heart is functional but is increasingly constrained in its properties over time. Both patients carrying homozygous truncating *TNNI3* variants were diagnosed with young age (1.1 and 0.8 years old) and suffered from adverse cardiac events (HTX at 1.8 years of age and death after 1.3 years), supporting the afore mentioned model. On molecular level TNNI switching regulates Ca^{2+} sensitivity, the ability to respond to adrenergic stimulation of the contractile system and resistance to hypoxia and acidosis (Schiaffino, Gorza, & Ausoni, 1993). To conclude, a restrained TNNI1 to TNNI3 switch in CMP patients carrying a truncating and homozygous TNNI3 mutation may prevent adequate adaptation to postnatal heart physiology. Although the cardiac specific n-terminus mediates interaction with the regulatory n-terminus of the cardiac TNNC1 (Hwang, Cai, Pineda-Sanabria, Corson, & Sykes, 2014), its specific role remains unclear but proves to be essential for the mechanical strains of the heart.

In the remaining cases of CMP carrying a heterozygous variant a depletion of TNNI3 and an increase of TNNI1 compared to controls was not observed. Thus, different disease mechanisms probably determine the progression of the disease in these patients. The onset and development of the CMP in these individuals coincides with this statement (onset ≥ 1.8 years, occurrence of adverse events ≥ 3.4 years, Table S 8). The clustering of VOI near the c-terminus suggest impaired interaction with other sarcomere genes like TPM1 or cardiac actin (ACTC1). At least for patient 3-II:1 mRNA and protein levels were determined but presented no clear result. The available heart tissue was limited and therefore protein levels of all sarcomere protein remained to be tested. Especially levels of TPM1 and ACTC1 could have revealed new functional insights because of reported interaction sites of TNNI3 with these proteins (van den Wijngaard et al., 2011; Akhter, Bueltmann, Huang, & Jin, 2014; Zhang, Akhter, Mottl, & Jin, 2011).

4.5 PRDM16 in patients with cardiomyopathy

As mentioned before, the rate of pediatric CMP cases with known genotypes does not exceed far above 50% for all described disease genes so far. To exceed these limitations, the search for such genes is ongoing and the function of many newly described genes is unknown or rarely described. Such gene is *PRDM16*. First described in 2013 in association with CMP the cardiac function of this gene remains unknown due to the absence of further publications (Arndt et al., 2013). In an approach to determine the minimal overlapping genomic sequence of 1p36 deletion syndrome patients with CMP involvement, a sequence including *PRDM16* exons four to 17 was determined. Truncating and missense variants were then observed in patients diagnosed with primary non-syndromic DCM and LVNC. In further genetically modified zebra fish experiments the effect of a truncating variant on adverse heart development was confirmed. In this work *PRDM16* was detected as the fourth most frequent gene with 17 VOI after *MYH7*, *TTN* and *MYBPC3* in an extended cardiac phenotype cohort of 285 patients. These VOI were not age specific but tended to be detected more often in males. Interestingly variants in *PRDM16* were detected in all six different phenotypes included in this cohort. Most of these patients were affected by LVNC with 8 patients possibly due to the high number of LVNC cases recruited in this cohort (6.7% detection rate). For HCM and DCM cases VOI were detected in two individuals each. Thus, the rate of *PRDM16* variants were lower with 4.2% and 3.7%. Although two patients carried *PRDM16* variants for RCM and ARVC cases the population has to be increased to determine statistically relevant terms (RCM: n = 8, ARVC: n = 4). The sub cohort of myocarditis only expressed a rate of 1.9%. Nevertheless, these observations suggest, that PRDM16 may influence general cardiac disease mechanisms in early developmental stages as suggested by former studies and especially in LVNC cases

(Kodo et al., 2016; Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010). Another study predicts an increased variant burden for truncating but nor for non-truncating variants due to the number of observed rare variants in the population database gnomAD (MAF < 0.0001, Mazzarotto et al., 2020). Seven truncating and 1900 non-truncating variants in 120147 alleles in gnomAD controls were compared to six truncating and non-truncating variants each detected in 444 affected LVNC patients using the Fisher's Exact test. A significant enrichment was determined for truncating variants. Similar values can be calculated within the extended cohort of this work. Three pathogenic and truncating variants, detected in LVNC patients, are significantly enriched, when processed the same way as described in the publication mentioned before ($p < 0.00001$). Therefore, PRDM16 can be described as specific for LVNC cases. The difference between truncating and non-truncating variants coincides with the o/e values, Z-scores and the pLI for *PRDM16* in gnomAD described in section 4.2.4 (Z-score (missense) = 1.35, o/e (missense) = 0.87, pLI (truncating) = 1, o/e (truncating) = 0.08). Nevertheless, the impact of missense variants in this gene remains unknown and requires more functional insight. So far, no homozygous *PRDM16* variant was reported. The detection and analysis of such a variant could increase insight into the disease mechanism of this gene. At least in mice homozygous truncating variants were shown to be lethal in early fetal stages (Bjork, Turbe-Doan, Prysak, Herron, & Beier, 2010).

4.6 Functional data of *PRDM16* variants

The function of PRDM16 was mainly assessed in brown and beige fat tissue and its myoblastic precursors (Ohno, Shinoda, Spiegelman, & Kajimura, 2012). Other tissue specific publications linked the function of PRDM16 with the processes of hematopoiesis and neurogenesis (Aguilo et al., 2011; Baizabal et al., 2018; Chuikov, Levi, Smith, & Morrison, 2010). Until this point functional knowledge for PRDM16 in cardiac context is rare and only available from knockout experiments in mice, zebra fishes and human induced pluripotent stem cells. Mechanistic insights of truncating and non-truncating variants of *PRDM16* are unknown. Here, an approach was started to investigate the effects of such variants.

4.6.1 Mutation of PRDM16 affects subcellular localization

To investigate the subcellular location of PRDM16, HEK293 cells were transfected with *PRDM16* constructs and immunostained with PRDM16 specific antibodies. In former studies PRDM16 was detected in the nucleus supporting a role of PRDM16 as transcriptional regulator (Baizabal et al., 2018; Dempersmier et al., 2015). In this work, it is shown, that PRDM16 is not entirely localized in the nucleus and therefore confirms similar observations in other

publications (Pinheiro et al., 2012). Overexpression of PRDM16 in HEK293 cells showed fluorescence signals in the nucleus and the cytoplasm. Thereby, cells were observed with the signal entirely in the nucleus, whereas other cells mainly showed PRDM16 expression in the cytoplasm. This suggests a cytosolic function of PRDM16 for instance in initial histone modification (Pinheiro et al., 2012). Additionally, protein tagged expression of PRDM16 seemed to shift the signal into the nucleus (EGFP- and FLAG-tag). Therefore, different observations were determined with tagged and untagged PRDM16. Results gathered with tagged PRDM16 constructs may have to be reevaluated or viewed with caution (e.g. Chen, Huang, Pan, Zhu, & Wang, 2018; Takahata et al., 2009). Most work published on PRDM16 so far describes knockout approaches or experimentation with overexpressed and tagged PRDM16. Endogenous PRDM16 is partly cell type specific and lowly concentrated in cells and detection is thus challenging. In this work, to reduce artificial influences the untagged protein was used to analyze the expression of PRDM16 in cells. In further experiments, it was shown that overexpressed PRDM16 accumulated at the nuclear membrane. In three dimensional images the signal of PRDM16 is mainly distributed around the nucleus and not directly located within the nucleus. Due to the absence of a nuclear localization sequence, PRDM16 has to interact with proteins located to the nucleus. So far interaction with such nuclear localized protein was not proven but seems highly regulated with the strong accumulation of PRDM16 at the nuclear membrane (Figure 25). Otherwise functions outside the nucleus were already described (Pinheiro et al., 2012).

Overexpression of CMP *PRDM16* variants revealed different subcellular localization pattern. The *PRDM16* frameshift variant p.R525Pfs*79 accumulated in the nuclei. Therefore, the truncating *PRDM16* variant was not eliminated by regulatory mechanisms such as nonsense-meditated decay (as described before for TNNI3 in this work or MYBPC3, Carrier, Schlossarek, Willis, & Eschenhagen, 2010; Seeger et al., 2019). The afore mentioned accumulation may induce dysregulated nuclear function and transcription e.g. impaired chromatin or nuclear envelope modeling (Pinheiro et al., 2012). The *PRDM16* missense variants p.E271K and p.P291L were more abundant in the cytoplasm while the variants p.N816S and p.L887P accumulated in the nucleus.

The introduction of truncating and missense variants seems to change properties of the protein in such way, that PRDM16 is differently transported through the cell. Therefore, mechanisms including methyltransferase activity and methylation of histone 3 outside of the nucleus (Pinheiro et al., 2012) or activation of gene expression in the nucleus through transcription factors/regulators like mitochondrial brown fat uncoupling protein 1 (UCP1) and PPAR-gamma coactivator 1-alpha (PGC1α, Ohno, Shinoda, Spiegelman, & Kajimura, 2012; Dempersmier et al., 2015) can be impaired with changed concentrations of PRDM16. In conclusion, these results suggest a dose-dependent disease mechanism in an epigenetic setup.

4.6.2 N-terminal protein tags and patient-specific cardiomyopathy variants influence subcellular localization of PRDMD16

Using Western blot the expression of the untagged protein was again compared with expression of PRDM16 conjugated with n-terminal protein tags (EGFP/FLAG). Of note here were the detection of a PRDM16-specific band higher than expected and the occurrence of a double band mainly observed in the untagged protein. Due to uniformly loading PRDM16 with the detergent SDS during sample preparation, intrinsic features such as charge were eliminated (Reynolds & Tanford, 1970; Smith, 1984). Therefore, the observed higher molecular weight and the occurrence of double bands were probably associated with post-translational modifications. In a former study, SUMOylation of PRDM16 was described to stabilize the protein (Chen, Huang, Pan, Zhu, & Wang, 2018). Here, three SUMOylation sites were predicted (Figure 20) and one was already published at position p.K915. With 11 kDa the difference between expected and appearing bands can be caused by triple SUMOylation (~30 to 35 kDa), whereas one SUMOylation site would be positioned before the amino acid 525, because the frameshift variant was observed with ~10 kDa difference. For the untagged protein an additional band was observed with 10 to 20 kDa difference below the corresponding PRDM16 bands determined with the tagged protein. The same difference was determined before (Chen, Huang, Pan, Zhu, & Wang, 2018). It was suggested that this difference represents the shift between a stable and unstable form of PRDM16 due to SUMOylation. An antibody for SUMO 2/3 was applied to HEK293 samples with overexpressed PRDM16 but no specific band was observed due to complex SUMOylation patterns in the cell (appearance of many bands). Co-immunoprecipitation would be an alternative approach to further investigate the covalent binding of SUMO. With side-directed mutagenesis predicted SUMOylated amino acid residues can be exchanged or deactivated and the distinct SUMOylation sites would be exposed. Furthermore, treatment of *PRDM16* transfected cells with SUMOylation inhibitors can additionally confirm the published data (Chen, Huang, Pan, Zhu, & Wang, 2018). The inhibition would result in reduced molecular weight of PRDM16. *In vitro* experiments in which PRDM16 is SUMOylated or deSUMOylated can also end in promising results. Other post-translational modifications are also possible as predicted in Figure 20, but the change in molecular weight is too drastic to be caused by these alterations. With Western blot similar results were obtained for the frameshift variant (PRDM16_v3fs). This variant was not degraded and therefore not affected by regulatory mRNA or protein degradation (discussed in 4.6.1).

To investigate differences in expression in the nucleus and cytoplasm the corresponding fractions were separated with specific lysis buffers and centrifugation steps for each fraction. Because this procedure allows concentration of the lysates, an endogenous PRDM16 detection was applied and only very weak or no bands were detected. Thus,

endogenous PRDM16 seems to be expressed in very low concentrations, although HEK293 cells are believed to have one of the highest PRDM16 expression levels in cultured cell types (Thul et al., 2017, https://www.proteinatlas.org/). In overexpressed form there are at least small visible effects, when expressed with missense variants of *PRDM16*. The expression pattern between nucleus and cytoplasm exhibits different amounts of PRDM16 for each fraction and also for the upper and lower band of PRDM16, suggesting the variants influence the switch between localization in the nucleus or cytoplasm and the stable or unstable form of PRDM16. Thus, hyperstabilization of PRDM16 mutants in the nucleus is a possible disease mechanism. The experiment was performed at least four times and although the separation worked fine (lamin C and β-tubulin bands) the results were too heterogenous to generate a clear result. In three of the four experiments the wt was expressed weakly, suggesting more limited and unstable expression or more sensitive dose-dependency.

In conclusion, the introduction of patient-specific variants of *PRDM16* in HEK293 cells result in changed localization and stabilization properties, but this cell culture approach seems limited for the purpose of stable PRDM16 expression for further functional insight. There is evidence, that the stability of PRDM16 is controlled by its state of SUMOylation, but further experimentation is necessary.

4.6.3 Patient-specific cardiomyopathy variants influence protein stability of PRDM16

With another approach the protein turnover of PRDM16 was analyzed. The expression of PRDM16 transfected cells was inhibited with cycloheximide for different time intervals from 0 to 24 h. The harvested cells were then analyzed with Western blot and the half-life $T_{1/2}$ of the corresponding protein was determined. The $T_{1/2}$ for the PRDM16 wt was 7.8 h and is consistent with published data (Chen, Huang, Pan, Zhu, & Wang, 2018). Surprisingly, the stability of the frameshift variant PRMD16_v3fs was drastically increased with 49.5 h. With such elongation the exponential fit is not very accurate, because the curve seems almost linear. To increase accuracy the incubation with CHX has to be elongated. These observations suggest aggregation of the truncated PRDM16 protein or hyperstabilization after nuclear translocation (aggregation in myopathy, Meister-Broekema et al., 2018). These findings are consistent with published data, that PRDM16 is regulated by stabilizing processes (Ohno, Shinoda, Spiegelman, & Kajimura, 2012; Chen, Huang, Pan, Zhu, & Wang, 2018). As for the other constructs an elongation to 10.5 and 12.2 h was observed for variants PRDM16_v4 and PRDM16_v5, whereas the half-life of variants PRDM16_v1 and PRDM16_v2 was shortened to 4.8 and 2.7 h. When compared with the results presented in Figure 26, similarities can be observed. For PRDM16_v1 an PRDM16_v2 the PRDM16 signal is shifted to the cytoplasm

and decreased in stability, whereas for PRDM16_v4 and PRDM16_v5 increased detection was observed in the nucleus and PRDM16 carrying these variants seemed more stable. Therefore, the stability of PRDM16 probably is associated with localization in the cell. But whether due to higher stability more PRDM16 can be transported into the nucleus or due to localization in the nucleus the protein is protected from degradation remains unclear. However, the link between stability and localization raises the need to understand the implementation of SUMO to stabilize the protein. To gain validity, here again, a set of four attempts have been applied, but heterogenous results especially for wt did not confirm the determined results. As seen in Figure 30 the signal detected for PRDM16 seems weak for wt compared to the other variants and although the same amounts of protein were applied. Further results for PRDM16_wt were even weaker and more difficult to detect, suggesting cytotoxic effects of high PRDM16 concentrations in the cell.

To conclude, different variants of *PRDM16* resulted in decrease or increase of stability and thus an effect of these variants was observed. This suggests similar effects may can be observed in patients carrying *PRDM16* variants and favor cardiac remodeling. Opposing results were observed for the variants PRDM16_v1 and PRDM16_v2 compared to PRDM16_v3fs to PRDM16_v5 similarly observed as in localization approaches (Figure 26). Therefore, stability can be associated with localization. Nonetheless, one limitation occurred repeatedly, affecting the expression of the wt in all experiments, seemingly weakened in almost all experiments suggesting some cytotoxic effect of PRDM16 in the cell.

4.7 Limitations

With 80 pediatric cases affected by primary CMP more patients than originally planned were recruited. Nevertheless, patients with RCM and ARVC were underrepresented in this cohort. Furthermore, extension of the families with deep phenotyping if available may increase insight into the genetic mechanism and segregation of VOI detected in CMP genes. This would also increase the detection of carriers with pathogenic CMP variants and therefore potential CMP patients in familial cases. With more segregation data, evidence for the pathogenicity of VOI would increase. Although 126 VOI were detected, only 38% of patients were observed with at least one P or LP variant. Thus, the genetic basis in 62% of pediatric CMP cases remains unclear. To increase this rate whole exome and genome sequencing could be applied to identify genetic alteration outside of the selection of disease genes investigated in this work. Although the number of variants in possible disease genes would increase, but the functional characterization to estimate its pathogenicity and the complexity of the genotype would increase too. There are also economic advantages for the panel-based procedure used in this work, although WES and WGS are becoming more available through further development of

the methods. Further insight or elucidation of CMP cases can be gained through investigation of complex phenotypes as discussed in 4.3. To better understand disease mechanisms can also increase the rate of clarified genotypes.

Further limitations were presented during pathogenicity assessment using the ACMG guidelines (Figure 6, Richards et al., 2015). When released in 2015, many arguments of this guideline were phrased with a large range of interpretation. Meanwhile, several studies suggested restrictions on the guidelines generating more distinct results (Abou Tayoun et al., 2018; Kelly et al., 2018; Brnich et al., 2019; Riggs et al., 2020). Prediction tools for this purpose are constantly improved and became more precise, but functional data remains essential for pathogenicity interpretation of genetic variants. Furthermore, the ACMG needs to address distinct regulations on how and when arguments in the category “Functional data” are activated and in which degree. Until disease mechanisms are not fully resolved these guidelines have to be adapted to the current knowledge. One example for these circumstances are presented through the high number of P and LP variants for *MYH7* and *MYBPC3*, functionally better characterized than most other CMP genes, and low numbers of P and LP variants in genes not well characterized, such as *LDB3* or *DSP*. Another example is the impact of more than one modifying variant forming a complex phenotype and initiating a severe phenotype (Gifford et al., 2019).

The main issue for investigating the effects of homo- and heterozygous *TNNI3* variants was missing heart biopsies. The available tissue was used to determine the amounts of some sarcomere and Z-disk genes. With more samples investigation of TNNI3 interaction partners could have been extended. Especially for the heterozygous missense variants accumulated at the c-terminus of TNNI3 no heart tissue was available and possible changes in interaction with described interaction partners like TPM1 and ACTC1 (van den Wijngaard et al., 2011; Akhter, Bueltmann, Huang, & Jin, 2014; Zhang, Akhter, Mottl, & Jin, 2011). Further functional analysis is required to better understand the adverse influence of these genetic variants. To mimic the structural setup of the heart engineered heart tissue (EHT) composed of human induced stem cell- derived cardiomyocytes could be used for such purpose. Biopsies of other cell types like fibroblasts or peripheral blood can be taken from patients carrying *TNNI3* variants and differentiated to cardiomyocytes (Takahashi et al., 2007; Yamanaka, 2010; Mannhardt et al., 2016). In EHT’s force production of the contractile sarcomere system can be measured. Additional immunostaining experiments of 2D cardiomyocyte culture may reveal disrupted interactions between the sarcomere proteins and give further insights into to the mechanisms of the disease onset caused by *TNNI3* variants.

The overexpression of PRDM16 in human cell lines works in principal and promising results can be achieved in such settings. Furthermore, PRDM16 expression seems highly regulated. Due to increased cell death with high PRDM16 concentrations and high lethality in mice when

knocked out PRDM16 is believed to be dose-dependent to function normally and cytotoxic when overdosed or reduced. This suggest, heterozygous knockout and knockin models to introduce CMP-specific VOI in animals and cells are better suited to investigate the effects of PRDM16 modification. Again, induced pluripotent stem cell-derived cardiomyocytes can be used to overcome restrictions to functionally examine *PRDM16* variants in similar ways as described for *TNNI3*. Further implementing additional readouts such as increasing reactive oxygen species levels would strengthen evidence for the effect of *PRDM16* variants (Chuikov, Levi, Smith, & Morrison, 2010). Additionally, PRDM16 production in *E. coli* did not yield significant amounts of protein. Therefore, *in vitro* approaches could not be implemented. For *in vitro* experiments other expression systems like yeast, human cell lines or cell-free expression within cell lysates rather than in living cells to decrease cytotoxic effects of PRDM16 should be tested to gain high yields for PRDM16 (Gregorio, Levine, & Oza, 2019).

4.8 Outlook

This study describes the identification of genetic defects in known CMP disease genes and performed functional analysis of CMP associated VOI identified in *PRDM16*. Thus, this work covers research on genetic CMP disease mechanisms and biochemical characterization of PRDM16. Both areas are of vital interest to explore CMP pathogenesis, appropriately manage CMP patients, and to explore translational approaches for diagnosis as well as potential therapies.

This work focused on a relatively large pediatric CMP cohort. Further strengthening of the main findings gained in this study can be achieved by expanding the number of DCM, HCM, and LVNC patients. Together with advanced clinical characterization this may identify CMP specific subgroups that are of either high or low risk to develop severe CMP courses. Such information would be critical for appropriated patient management. The relative high number of CMP patients without detected VOI can be due to multiple aspects for instance insufficient interpretation of VOI, unknown genetic disease mechanisms, genomic alterations affecting known disease genes, or oligomeric disease mechanisms. In this regard is the analysis of CMP patients without VOI for other genomic alterations a promising approach. WES or WGS will facilitate identification of genetic defects in novel CMP disease genes or chromosomal alterations affecting non-coding or genomic regions. Improved interpretation of VOI will be facilitated by appropriated gene/protein specific assays allowing efficient functional analysis of missense variants. This would help to identify critical alterations *in vitro* or *in vivo*.

An interesting aspect of this work is the observation that oligomeric VOI induce more severe phenotypes. Oligomeric disease mechanisms or the identification of genetic modifier VOI are currently underexplored in CMP (Gifford et al., 2019). Here it would be of interest to investigate

genetic modifier and their impact on well-defined pathogenic or likely pathogenic CMP VOI. Such analysis must be performed in larger cohorts to ensure sufficient power to identify or validate candidate VOI. Complementary analysis *in vivo* or *in vitro* will be essential to model the tissue specific interplay of different VOI.

Here, *PRDM16* was identified as frequently mutated in pediatric and adult CMP. This unexpected finding should motivate further biochemical exploration of PRDM16 in the context of CMP. The molecular understanding of PRDM16 was mainly achieved from studies in adipocytes and adipose tissue. These studies identified PRDM16 as transcriptional integrator of critical transcription factors such as C/EBPβ, CTBP1/2, PGC1b, PPARα, or PPARγ (Fog, Galli, & Lund, 2012). Moreover, PRDM16 exposes histone H3K4 and H3K9 methyltransferase activity regulating chromatin organization and transcriptional programs (Pinheiro et al., 2012; Zhou et al., 2016). The activity of transcriptional programs is highly specific in postmitotic cardiomyocytes. Thus, identification of chromatin regions controlled by PRDM16 and epigenetic marks in heart tissue is of critical value for further understanding of PRDM16 in the heart. Another aspect of transcriptional mechanisms inducing CMP is their activation during development or in adulthood (Kodo et al., 2016; de Soysa et al., 2019).

5 References

Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., . . . ClinGen Sequence Variant Interpretation Working, G. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat, 39*(11), 1517-1524. doi:10.1002/humu.23626

Adalsteinsdottir, B., Teekakirikul, P., Maron, B. J., Burke, M. A., Gudbjartsson, D. F., Holm, H., . . . Gunnarsson, G. T. (2014). Nationwide study on hypertrophic cardiomyopathy in Iceland: evidence of a MYBPC3 founder mutation. *Circulation, 130*(14), 1158-1167. doi:10.1161/CIRCULATIONAHA.114.011207

Aguilo, F., Avagyan, S., Labar, A., Sevilla, A., Lee, D. F., Kumar, P., . . . Snoeck, H. W. (2011). Prdm16 is a physiologic regulator of hematopoietic stem cells. *Blood, 117*(19), 5057-5066. doi:10.1182/blood-2010-08-300145

Akhter, S., Bueltmann, K., Huang, X., & Jin, J. P. (2014). Restrictive cardiomyopathy mutations demonstrate functions of the C-terminal end-segment of troponin I. *Archives of Biochemistry and Biophysics, 552-553*, 3-10. doi:10.1016/j.abb.2013.12.001

Akinrinade, O., Helio, T., Lekanne Deprez, R. H., Jongbloed, J. D. H., Boven, L. G., van den Berg, M. P., . . . Koskenvuo, J. (2019). Relevance of Titin Missense and Non-Frameshifting Insertions/Deletions Variants in Dilated Cardiomyopathy. *Sci Rep, 9*(1), 4093. doi:10.1038/s41598-019-39911-x

Al-Wakeel-Marquard, N., Degener, F., Herbst, C., Kuhnisch, J., Dartsch, J., Schmitt, B., . . . Klaassen, S. (2019). RIKADA Study Reveals Risk Factors in Pediatric Primary Cardiomyopathy. *J Am Heart Assoc, 8*(15), e012531. doi:10.1161/JAHA.119.012531

Alexander, P. M. A., Nugent, A. W., Daubeney, P. E. F., Lee, K. J., Sleeper, L. A., Schuster, T., . . . National Australian Childhood Cardiomyopathy, S. (2018). Long-Term Outcomes of Hypertrophic Cardiomyopathy Diagnosed During Childhood: Results From a National Population-Based Study. *Circulation, 138*(1), 29-36. doi:10.1161/CIRCULATIONAHA.117.028895

Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data. *Babraham Bioinformatics*.

Arndt, A. K., Schafer, S., Drenckhahn, J. D., Sabeh, M. K., Plovie, E. R., Caliebe, A., . . . Klaassen, S. (2013). Fine mapping of the 1p36 deletion syndrome identifies mutation of PRDM16 as a cause of cardiomyopathy. *Am J Hum Genet, 93*(1), 67-77. doi:10.1016/j.ajhg.2013.05.015

Audagnotto, M., & Dal Peraro, M. (2017). Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput Struct Biotechnol J, 15*, 307-319. doi:10.1016/j.csbj.2017.03.004

Baggish, A. L., & Wood, M. J. (2011). Athlete's heart and cardiovascular care of the athlete: scientific and clinical update. *Circulation, 123*(23), 2723-2735. doi:10.1161/CIRCULATIONAHA.110.981571

Baizabal, J. M., Mistry, M., Garcia, M. T., Gomez, N., Olukoya, O., Tran, D., . . . Harwell, C. C. (2018). The Epigenetic State of PRDM16-Regulated Enhancers in Radial Glia Controls Cortical Neuron Position. *Neuron, 98*(5), 945-962 e948. doi:10.1016/j.neuron.2018.04.033

Bedada, F. B., Chan, S. S., Metzger, S. K., Zhang, L., Zhang, J., Garry, D. J., . . . Metzger, J. M. (2014). Acquisition of a quantitative, stoichiometrically conserved ratiometric marker of maturation status in stem cell-derived cardiac myocytes. *Stem Cell Reports, 3*(4), 594-605. doi:10.1016/j.stemcr.2014.07.012

Belkaya, S., Kontorovich, A. R., Byun, M., Mulero-Navarro, S., Bajolle, F., Cobat, A., . . . Casanova, J. L. (2017). Autosomal Recessive Cardiomyopathy Presenting as Acute Myocarditis. *J Am Coll Cardiol, 69*(13), 1653-1665. doi:10.1016/j.jacc.2017.01.043

Bjork, B. C., Turbe-Doan, A., Prysak, M., Herron, B. J., & Beier, D. R. (2010). Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet, 19*(5), 774-789. doi:10.1093/hmg/ddp543

Bollen, I. A. E., & van der Velden, J. (2017). The contribution of mutations in MYH7 to the onset of cardiomyopathy. *Neth Heart J, 25*(12), 653-654. doi:10.1007/s12471-017-1045-5

Bondue, A., Arbustini, E., Bianco, A., Ciccarelli, M., Dawson, D., De Rosa, M., . . . Heymans, S. (2018). Complex roads from genotype to phenotype in dilated cardiomyopathy: scientific update from the Working Group of Myocardial Function of the European Society of Cardiology. *Cardiovasc Res, 114*(10), 1287-1303. doi:10.1093/cvr/cvy122

Brnich, S. E., Abou Tayoun, A. N., Couch, F. J., Cutting, G. R., Greenblatt, M. S., Heinen, C. D., . . . Clinical Genome Resource Sequence Variant Interpretation Working, G. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med, 12*(1), 3. doi:10.1186/s13073-019-0690-2

Buermans, H. P., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta, 1842*(10), 1932-1941. doi:10.1016/j.bbadis.2014.06.015

Burke, M. A., Cook, S. A., Seidman, J. G., & Seidman, C. E. (2016). Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy. *J Am Coll Cardiol, 68*(25), 2871-2886. doi:10.1016/j.jacc.2016.08.079

Carrier, L., Schlossarek, S., Willis, M. S., & Eschenhagen, T. (2010). The ubiquitin-proteasome system and nonsense-mediated mRNA decay in hypertrophic cardiomyopathy. *Cardiovasc Res, 85*(2), 330-338. doi:10.1093/cvr/cvp247

Cecchi, F., Tomberli, B., & Olivotto, I. (2012). Clinical and molecular classification of cardiomyopathies. *Glob Cardiol Sci Pract, 2012*(1), 4. doi:10.5339/gcsp.2012.4

Chen, Q., Huang, L., Pan, D., Zhu, L. J., & Wang, Y. X. (2018). Cbx4 Sumoylates Prdm16 to Regulate Adipose Tissue Thermogenesis. *Cell Rep, 22*(11), 2860-2872. doi:10.1016/j.celrep.2018.02.057

Chuikov, S., Levi, B. P., Smith, M. L., & Morrison, S. J. (2010). Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. *Nat Cell Biol, 12*(10), 999-1006. doi:10.1038/ncb2101

Cirino, A. L., Lakdawala, N. K., McDonough, B., Conner, L., Adler, D., Weinfeld, M., . . . MedSeq, P. (2017). A Comparison of Whole Genome Sequencing to Multigene Panel Testing in Hypertrophic Cardiomyopathy Patients. *Circ Cardiovasc Genet, 10*(5). doi:10.1161/CIRCGENETICS.117.001768

Colan, S. D., Lipshultz, S. E., Lowe, A. M., Sleeper, L. A., Messere, J., Cox, G. F., . . . Towbin, J. A. (2007). Epidemiology and cause-specific outcome of hypertrophic cardiomyopathy in children: findings from the Pediatric Cardiomyopathy Registry. *Circulation, 115*(6), 773-781. doi:10.1161/CIRCULATIONAHA.106.621185

Corrado, D., Wichter, T., Link, M. S., Hauer, R. N., Marchlinski, F. E., Anastasakis, A., . . . Calkins, H. (2015). Treatment of Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia: An International Task Force Consensus Statement. *Circulation, 132*(5), 441-453. doi:10.1161/CIRCULATIONAHA.115.017944

Dadson, K., Hauck, L., & Billia, F. (2017). Molecular mechanisms in cardiomyopathy. *Clinical Science, 131*(13), 1375-1392. doi:10.1042/cs20160170

de Soysa, T. Y., Ranade, S. S., Okawa, S., Ravichandran, S., Huang, Y., Salunga, H. T., . . . Srivastava, D. (2019). Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature, 572*(7767), 120-124. doi:10.1038/s41586-019-1414-x

Dempersmier, J., Sambeat, A., Gulyaeva, O., Paul, S. M., Hudak, C. S., Raposo, H. F., . . . Sul, H. S. (2015). Cold-inducible Zfp516 activates UCP1 transcription to promote browning of white fat and development of brown fat. *Mol Cell, 57*(2), 235-246. doi:10.1016/j.molcel.2014.12.005

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet, 43*(5), 491-498. doi:10.1038/ng.806

Doesch, C., Dierks, D. M., Haghi, D., Schimpf, R., Kuschyk, J., Suselbeck, T., . . . Papavassiliu, T. (2014). Right ventricular dysfunction, late gadolinium enhancement, and female gender predict poor outcome in patients with dilated cardiomyopathy. *Int J Cardiol, 177*(2), 429-435. doi:10.1016/j.ijcard.2014.09.004

Fatkin, D., Lam, L., Herman, D. S., Benson, C. C., Felkin, L. E., Barton, P. J. R., . . . Seidman, C. E. (2016). Titin truncating mutations: A rare cause of dilated cardiomyopathy in the young. *Progress in Pediatric Cardiology, 40*, 41-45. doi:https://doi.org/10.1016/j.ppedcard.2016.01.003

Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics, 30*(17), 2503-2505. doi:10.1093/bioinformatics/btu314

Finsterer, J., Stöllberger, C., & Towbin, J. A. (2017). Left ventricular noncompaction cardiomyopathy: cardiac, neuromuscular, and genetic factors. *Nature Reviews Cardiology, 14*(4), 224-237. doi:10.1038/nrcardio.2016.207

Fog, C. K., Galli, G. G., & Lund, A. H. (2012). PRDM proteins: important players in differentiation and disease. *Bioessays, 34*(1), 50-60. doi:10.1002/bies.201100107

Geisterfer-Lowrance, A. A., Kass, S., Tanigawa, G., Vosberg, H. P., McKenna, W., Seidman, C. E., & Seidman, J. G. (1990). A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell, 62*(5), 999-1006. doi:10.1016/0092-8674(90)90274-i

Gifford, C. A., Ranade, S. S., Samarakoon, R., Salunga, H. T., de Soysa, T. Y., Huang, Y., . . . Srivastava, D. (2019). Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science, 364*(6443), 865-870. doi:10.1126/science.aat5056

Gregorio, N. E., Levine, M. Z., & Oza, J. P. (2019). A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc, 2*(1). doi:10.3390/mps2010024

Haas, J., Frese, K. S., Peil, B., Kloos, W., Keller, A., Nietsch, R., . . . Meder, B. (2015). Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J, 36*(18), 1123-1135a. doi:10.1093/eurheartj/ehu301

Hall, C. L., Sutanto, H., Dalageorgou, C., McKenna, W. J., Syrris, P., & Futema, M. (2018). Frequency of genetic variants associated with arrhythmogenic right ventricular cardiomyopathy in the genome aggregation database. *Eur J Hum Genet, 26*(9), 1312-1318. doi:10.1038/s41431-018-0169-4

Herkert, J. C., Abbott, K. M., Birnie, E., Meems-Veldhuis, M. T., Boven, L. G., Benjamins, M., . . . Jongbloed, J. D. H. (2018). Toward an effective exome-based genetic testing strategy in pediatric dilated cardiomyopathy. *Genet Med, 20*(11), 1374-1386. doi:10.1038/gim.2018.9

Hershberger, R. E., Givertz, M. M., Ho, C. Y., Judge, D. P., Kantor, P. F., McBride, K. L., . . . Guidelines, C. (2018). Genetic evaluation of cardiomyopathy: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med, 20*(9), 899-909. doi:10.1038/s41436-018-0039-z

Hershberger, R. E., Hedges, D. J., & Morales, A. (2013). Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nature Reviews Cardiology, 10*(9), 531-547. doi:10.1038/nrcardio.2013.105

Huang, X., Pi, Y., Lee, K. J., Henkel, A. S., Gregg, R. G., Powers, P. A., & Walker, J. W. (1999). Cardiac troponin I gene knockout: a mouse model of myocardial troponin I deficiency. *Circ Res, 84*(1), 1-8. doi:10.1161/01.res.84.1.1

Hwang, P. M., Cai, F., Pineda-Sanabria, S. E., Corson, D. C., & Sykes, B. D. (2014). The cardiac-specific N-terminal region of troponin I positions the regulatory domain of troponin C. *Proc Natl Acad Sci U S A, 111*(40), 14412-14417. doi:10.1073/pnas.1410775111

IHGSC, I. H. G. S. C. (2004). Finishing the euchromatic sequence of the human genome. *Nature, 431*(7011), 931-945. doi:10.1038/nature03001

Inc., I. (2010). Illumina Sequencing Technology: Highest data accuracy, simple workflow, and a broad range of applications.

Ingles, J., Goldstein, J., Thaxton, C., Caleshu, C., Corty, E. W., Crowley, S. B., . . . Funke, B. (2019). Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ Genom Precis Med, 12*(2), e002460. doi:10.1161/circgen.119.002460

Kajimura, S., Seale, P., Tomaru, T., Erdjument-Bromage, H., Cooper, M. P., Ruas, J. L., . . . Spiegelman, B. M. (2008). Regulation of the brown and white fat gene programs through a PRDM16/CtBP transcriptional complex. *Genes Dev, 22*(10), 1397-1409. doi:10.1101/gad.1666108

Kaski, J. P., Syrris, P., Burch, M., Tome-Esteban, M. T., Fenton, M., Christiansen, M., . . . Elliott, P. M. (2008). Idiopathic restrictive cardiomyopathy in children is caused by mutations in cardiac sarcomere protein genes. *Heart, 94*(11), 1478-1484. doi:10.1136/hrt.2007.134684

Kelly, M. A., Caleshu, C., Morales, A., Buchan, J., Wolf, Z., Harrison, S. M., . . . Funke, B. (2018). Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med, 20*(3), 351-359. doi:10.1038/gim.2017.218

Klaassen, S., Probst, S., Oechslin, E., Gerull, B., Krings, G., Schuler, P., . . . Thierfelder, L. (2008). Mutations in sarcomere protein genes in left ventricular noncompaction. *Circulation, 117*(22), 2893-2901. doi:10.1161/CIRCULATIONAHA.107.746164

Kodo, K., Ong, S. G., Jahanbani, F., Termglinchan, V., Hirono, K., InanlooRahatloo, K., . . . Wu, J. C. (2016). iPSC-derived cardiomyocytes reveal abnormal TGF-beta signalling in left ventricular non-compaction cardiomyopathy. *Nat Cell Biol, 18*(10), 1031-1042. doi:10.1038/ncb3411

Kolanczyk, M., Mautner, V., Kossler, N., Nguyen, R., Kuhnisch, J., Zemojtel, T., . . . Kluwe, L. (2011). MIA is a potential biomarker for tumour load in neurofibromatosis type 1. *BMC Med, 9*, 82. doi:10.1186/1741-7015-9-82

Kolokotronis, K., Kuhnisch, J., Klopocki, E., Dartsch, J., Rost, S., Huculak, C., . . . Gerull, B. (2019). Biallelic mutation in MYH7 and MYBPC3 leads to severe cardiomyopathy with left ventricular noncompaction phenotype. *Hum Mutat, 40*(8), 1101-1114. doi:10.1002/humu.23757

Landrum, M. J., & Kattman, B. L. (2018). ClinVar at five years: Delivering on the promise. *Hum Mutat, 39*(11), 1623-1630. doi:10.1002/humu.23641

Lazzarini, E., Jongbloed, J. D., Pilichou, K., Thiene, G., Basso, C., Bikker, H., . . . van der Zwaag, P. A. (2015). The ARVD/C genetic variants database: 2014 update. *Hum Mutat, 36*(4), 403-410. doi:10.1002/humu.22765

Lee, T. M., Hsu, D. T., Kantor, P., Towbin, J. A., Ware, S. M., Colan, S. D., . . . Lipshultz, S. E. (2017). Pediatric Cardiomyopathies. *Circ Res, 121*(7), 855-873. doi:10.1161/CIRCRESAHA.116.309386

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291. doi:10.1038/nature19057

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv, 1303*.

Lipshultz, S. E., Cochran, T. R., Briston, D. A., Brown, S. R., Sambatakos, P. J., Miller, T. L., . . . Wilkinson, J. D. (2013). Pediatric cardiomyopathies: causes, epidemiology, clinical course, preventive strategies and therapies. *Future Cardiol, 9*(6), 817-848. doi:10.2217/fca.13.66

Lipshultz, S. E., Law, Y. M., Asante-Korang, A., Austin, E. D., Dipchand, A. I., Everitt, M. D., . . . Colan, S. D. (2019). Cardiomyopathy in Children: Classification and Diagnosis: A Scientific Statement From the American Heart Association. *Circulation, 140*(1), e9-e68. doi:10.1161/CIR.0000000000000682

Lipshultz, S. E., Sleeper, L. A., Towbin, J. A., Lowe, A. M., Orav, E. J., Cox, G. F., . . . Colan, S. D. (2003). The incidence of pediatric cardiomyopathy in two regions of the United States. *N Engl J Med, 348*(17), 1647-1655. doi:10.1056/NEJMoa021715

Lu, D. Y., Ventoulis, I., Liu, H., Kudchadkar, S. M., Greenland, G. V., Yalcin, H., . . . Abraham, M. R. (2020). Sex-specific cardiac phenotype and clinical outcomes in patients with hypertrophic cardiomyopathy. *Am Heart J, 219*, 58-69. doi:10.1016/j.ahj.2019.10.004

Mannhardt, I., Breckwoldt, K., Letuffe-Breniere, D., Schaaf, S., Schulz, H., Neuber, C., . . . Hansen, A. (2016). Human Engineered Heart Tissue: Analysis of Contractile Force. *Stem Cell Reports, 7*(1), 29-42. doi:10.1016/j.stemcr.2016.04.011

Marcus, F. I., McKenna, W. J., Sherrill, D., Basso, C., Bauce, B., Bluemke, D. A., . . . Zareba, W. (2010). Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation, 121*(13), 1533-1541. doi:10.1161/CIRCULATIONAHA.108.840827

Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T., & Bild, D. E. (1995). Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation, 92*(4), 785-789. doi:10.1161/01.cir.92.4.785

Maron, B. J., Towbin, J. A., Thiene, G., Antzelevitch, C., Corrado, D., Arnett, D., . . . Prevention. (2006). Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation, 113*(14), 1807-1816. doi:10.1161/CIRCULATIONAHA.106.174287

Mazzarotto, F., Hawley, M. H., Beltrami, M., Beekman, L., de Marvao, A., McGurk, K. A., . . . Walsh, R. (2020). The genetic architecture of left ventricular non-compaction reveals both substantial overlap with other cardiomyopathies and a distinct aetiology in a subset of cases. *medRxiv*, 2020.2001.2003.19015602. doi:10.1101/2020.01.03.19015602

McKenna, W. J., Maron, B. J., & Thiene, G. (2017). Classification, Epidemiology, and Global Burden of Cardiomyopathies. *Circ Res, 121*(7), 722-730. doi:10.1161/CIRCRESAHA.117.309711

Meister-Broekema, M., Freilich, R., Jagadeesan, C., Rauch, J. N., Bengoechea, R., Motley, W. W., . . . Kampinga, H. H. (2018). Myopathy associated BAG3 mutations lead to protein aggregation by stalling Hsp70 networks. *Nat Commun, 9*(1), 5342. doi:10.1038/s41467-018-07718-5

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc, 2010*(6), pdb prot5448. doi:10.1101/pdb.prot5448

Minoche, A. E., Horvat, C., Johnson, R., Gayevskiy, V., Morton, S. U., Drew, A. P., . . . Fatkin, D. (2019). Genome sequencing as a first-line genetic test in familial dilated cardiomyopathy. *Genet Med, 21*(3), 650-662. doi:10.1038/s41436-018-0084-7

Mochizuki, N., Shimizu, S., Nagasawa, T., Tanaka, H., Taniwaki, M., Yokota, J., & Morishita, K. (2000). A novel gene, MEL1, mapped to 1p36.3 is highly homologous to the MDS1/EVI1 gene and is transcriptionally activated in t(1;3)(p36;q21)-positive leukemia cells. *Blood, 96*(9), 3209-3214.

Morita, H., Rehm, H. L., Menesses, A., McDonough, B., Roberts, A. E., Kucherlapati, R., . . . Seidman, C. E. (2008). Shared Genetic Causes of Cardiac Hypertrophy in Children and Adults. *New England Journal of Medicine, 358*(18), 1899-1908. doi:10.1056/NEJMoa075463

Murphy, R. T., Mogensen, J., Shaw, A., Kubo, T., Hughes, S., & McKenna, W. J. (2004). Novel mutation in cardiac troponin I in recessive idiopathic dilated cardiomyopathy. *Lancet, 363*(9406), 371-372. doi:10.1016/s0140-6736(04)15468-8

Nam, J. M., Lim, J. E., Ha, T. W., Oh, B., & Kang, J. O. (2020). Cardiac-specific inactivation of Prdm16 effects cardiac conduction abnormalities and cardiomyopathy-associated phenotypes. *American Journal of Physiology-Heart and Circulatory Physiology, 318*(4), H764-H777. doi:10.1152/ajpheart.00647.2019

Nava, A., Bauce, B., Basso, C., Muriago, M., Rampazzo, A., Villanova, C., . . . Thiene, G. (2000). Clinical profile and long-term follow-up of 37 families with arrhythmogenic right ventricular cardiomyopathy. *Journal of the American College of Cardiology, 36*(7), 2226-2233. doi:10.1016/s0735-1097(00)00997-9

Nugent, A. W., Daubeney, P. E. F., Chondros, P., Carlin, J. B., Cheung, M., Wilkinson, L. C., . . . Weintraub, R. G. (2003). The Epidemiology of Childhood Cardiomyopathy in Australia. *New England Journal of Medicine, 348*(17), 1639-1646. doi:10.1056/NEJMoa021737

Oechslin, E., & Jenni, R. (2011). Left ventricular non-compaction revisited: a distinct phenotype with genetic heterogeneity? *Eur Heart J, 32*(12), 1446-1456. doi:10.1093/eurheartj/ehq508

Oechslin, E., & Klaassen, S. (2019). Left Ventricular Noncompaction: Phenotype in an Integrated Model of Cardiomyopathy? *J Am Coll Cardiol, 73*(13), 1612-1615. doi:10.1016/j.jacc.2018.11.064

Ohno, H., Shinoda, K., Ohyama, K., Sharp, L. Z., & Kajimura, S. (2013). EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. *Nature, 504*(7478), 163-167. doi:10.1038/nature12652

Ohno, H., Shinoda, K., Spiegelman, B. M., & Kajimura, S. (2012). PPARgamma agonists induce a white-to-brown fat conversion through stabilization of PRDM16 protein. *Cell Metab, 15*(3), 395-404. doi:10.1016/j.cmet.2012.01.019

Ouellette, A. C., Mathew, J., Manickaraj, A. K., Manase, G., Zahavich, L., Wilson, J., . . . Mital, S. (2018). Clinical genetic testing in pediatric cardiomyopathy: Is bigger better? *Clin Genet, 93*(1), 33-40. doi:10.1111/cge.13024

Paludan-Muller, C., Ahlberg, G., Ghouse, J., Svendsen, J. H., Haunso, S., & Olesen, M. S. (2017). Analysis of 60 706 Exomes Questions the Role of De Novo Variants Previously Implicated in Cardiac Disease. *Circ Cardiovasc Genet, 10*(6). doi:10.1161/circgenetics.117.001878

Peled, Y., Gramlich, M., Yoskovitz, G., Feinberg, M. S., Afek, A., Polak-Charcon, S., . . . Arad, M. (2014). Titin mutation in familial restrictive cardiomyopathy. *Int J Cardiol, 171*(1), 24-30. doi:10.1016/j.ijcard.2013.11.037

Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics, 93*(2), 105-111. doi:10.1016/j.ygeno.2008.10.003

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F. M., . . . Jenuwein, T. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell, 150*(5), 948-960. doi:10.1016/j.cell.2012.06.048

Pua, C. J., Bhalshankar, J., Miao, K., Walsh, R., John, S., Lim, S. Q., . . . Cook, S. A. (2016). Development of a Comprehensive Sequencing Assay for Inherited Cardiac Condition Genes. *J Cardiovasc Transl Res, 9*(1), 3-11. doi:10.1007/s12265-016-9673-5

Pugh, T. J., Kelly, M. A., Gowrisankar, S., Hynes, E., Seidman, M. A., Baxter, S. M., . . . Funke, B. H. (2014). The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med, 16*(8), 601-608. doi:10.1038/gim.2013.204

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. *Mol Cell, 58*(4), 586-597. doi:10.1016/j.molcel.2015.05.004

Reynolds, J. A., & Tanford, C. (1970). Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proc Natl Acad Sci U S A, 66*(3), 1002-1007. doi:10.1073/pnas.66.3.1002

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). 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*(5), 405-424. doi:10.1038/gim.2015.30

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., . . . Acmg. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med, 22*(2), 245-257. doi:10.1038/s41436-019-0686-8

Roberts, A. M., Ware, J. S., Herman, D. S., Schafer, S., Baksi, J., Bick, A. G., . . . Cook, S. A. (2015). Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med, 7*(270), 270ra276. doi:10.1126/scitranslmed.3010134

Rupp, S., Felimban, M., Schanzer, A., Schranz, D., Marschall, C., Zenker, M., . . . Hahn, A. (2019). Genetic basis of hypertrophic cardiomyopathy in children. *Clin Res Cardiol, 108*(3), 282-289. doi:10.1007/s00392-018-1354-8

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., . . . Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat Genet, 46*(9), 944-950. doi:10.1038/ng.3050

Sasse, S., Brand, N. J., Kyprianou, P., Dhoot, G. K., Wade, R., Arai, M., . . . Barton, P. J. (1993). Troponin I gene expression during human cardiac development and in end-stage heart failure. *Circ Res, 72*(5), 932-938. doi:10.1161/01.res.72.5.932

Schafer, S., de Marvao, A., Adami, E., Fiedler, L. R., Ng, B., Khin, E., . . . Cook, S. A. (2017). Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet, 49*(1), 46-53. doi:10.1038/ng.3719

Schiaffino, S., Gorza, L., & Ausoni, S. (1993). Troponin isoform switching in the developing heart and its functional consequences. *Trends Cardiovasc Med, 3*(1), 12-17. doi:10.1016/1050-1738(93)90022-x

Seeger, T., Shrestha, R., Lam, C. K., Chen, C., McKeithan, W. L., Lau, E., . . . Wu, J. C. (2019). A Premature Termination Codon Mutation in MYBPC3 Causes Hypertrophic Cardiomyopathy via Chronic Activation of Nonsense-Mediated Decay. *Circulation, 139*(6), 799-811. doi:10.1161/CIRCULATIONAHA.118.034624

Seifert, W., Holder-Espinasse, M., Kuhnisch, J., Kahrizi, K., Tzschach, A., Garshasbi, M., . . . Horn, D. (2009). Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of COH1. *Hum Mutat, 30*(2), E404-420. doi:10.1002/humu.20886

Semsarian, C., Ingles, J., Maron, M. S., & Maron, B. J. (2015). New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol, 65*(12), 1249-1254. doi:10.1016/j.jacc.2015.01.019

Shah, N., Hou, Y. C., Yu, H. C., Sainger, R., Caskey, C. T., Venter, J. C., & Telenti, A. (2018). Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *Am J Hum Genet, 102*(4), 609-619. doi:10.1016/j.ajhg.2018.02.019

Smith, B. J. (1984). SDS Polyacrylamide Gel Electrophoresis of Proteins. *Methods Mol Biol, 1*, 41-55. doi:10.1385/0-89603-062-8:41

Streff, H., Bi, W., Colon, A. G., Adesina, A. M., Miyake, C. Y., & Lalani, S. R. (2019). Amish nemaline myopathy and dilated cardiomyopathy caused by a homozygous contiguous gene deletion of TNNT1 and TNNI3 in a Mennonite child. *Eur J Med Genet, 62*(11), 103567. doi:10.1016/j.ejmg.2018.11.001

Su, L., Lei, X., Ma, H., Feng, C., Jiang, J., & Jiao, J. (2020). PRDM16 orchestrates angiogenesis via neural differentiation in the developing brain. *Cell Death Differ*. doi:10.1038/s41418-020-0504-5

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell, 131*(5), 861-872. doi:10.1016/j.cell.2007.11.019

Takahata, M., Inoue, Y., Tsuda, H., Imoto, I., Koinuma, D., Hayashi, M., . . . Imamura, T. (2009). SKI and MEL1 cooperate to inhibit transforming growth factor-beta signal in gastric cancer cells. *J Biol Chem, 284*(5), 3334-3344. doi:10.1074/jbc.M808989200

Thiene, G., Corrado, D., & Basso, C. (2004). Cardiomyopathies: is it time for a molecular classification? *Eur Heart J, 25*(20), 1772-1775. doi:10.1016/j.ehj.2004.07.026

Thierfelder, L., Watkins, H., MacRae, C., Lamas, R., McKenna, W., Vosberg, H. P., . . . Seidman, C. E. (1994). Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. *Cell, 77*(5), 701-712. doi:10.1016/0092-8674(94)90054-x

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., . . . Lundberg, E. (2017). A subcellular map of the human proteome. *Science, 356*(6340). doi:10.1126/science.aal3321

Tian, X., Li, Y., He, L., Zhang, H., Huang, X., Liu, Q., . . . Zhou, B. (2017). Identification of a hybrid myocardial zone in the mammalian heart after birth. *Nat Commun, 8*(1), 87. doi:10.1038/s41467-017-00118-1

Toepfer, C. N., Wakimoto, H., Garfinkel, A. C., McDonough, B., Liao, D., Jiang, J., . . . Seidman, C. E. (2019). Hypertrophic cardiomyopathy mutations in MYBPC3 dysregulate myosin. *Sci Transl Med, 11*(476). doi:10.1126/scitranslmed.aat1199

Towbin, J. A. (2014). Inherited cardiomyopathies. *Circ J, 78*(10), 2347-2356. doi:10.1253/circj.cj-14-0893

Towbin, J. A., Lowe, A. M., Colan, S. D., Sleeper, L. A., Orav, E. J., Clunie, S., . . . Lipshultz, S. E. (2006). Incidence, causes, and outcomes of dilated cardiomyopathy in children. *Jama, 296*(15), 1867-1876. doi:10.1001/jama.296.15.1867

van den Wijngaard, A., Volders, P., Van Tintelen, J. P., Jongbloed, J. D., van den Berg, M. P., Lekanne Deprez, R. H., . . . Smeets, B. J. (2011). Recurrent and founder mutations in the Netherlands: cardiac Troponin I (TNNI3) gene mutations as a cause of severe forms of hypertrophic and restrictive cardiomyopathy. *Neth Heart J, 19*(7-8), 344-351. doi:10.1007/s12471-011-0135-z

van der Linde, I. H. M., Hiemstra, Y. L., Bokenkamp, R., van Mil, A. M., Breuning, M. H., Ruivenkamp, C., . . . Barge-Schaapveld, D. (2017). A Dutch MYH7 founder mutation, p.(Asn1918Lys), is associated with early onset cardiomyopathy and congenital heart defects. *Neth Heart J, 25*(12), 675-681. doi:10.1007/s12471-017-1037-5

van Driel, B., Nijenkamp, L., Huurman, R., Michels, M., & van der Velden, J. (2019). Sex differences in hypertrophic cardiomyopathy: new insights. *Current Opinion in Cardiology, 34*(3), 254-259. doi:10.1097/hco.0000000000000612

van Waning, J. I., Caliskan, K., Hoedemaekers, Y. M., van Spaendonck-Zwarts, K. Y., Baas, A. F., Boekholdt, S. M., . . . Majoor-Krakauer, D. (2018). Genetics, Clinical Features, and Long-Term Outcome of Noncompaction Cardiomyopathy. *J Am Coll Cardiol, 71*(7), 711-722. doi:10.1016/j.jacc.2017.12.019

Vasilescu, C., Ojala, T. H., Brilhante, V., Ojanen, S., Hinterding, H. M., Palin, E., . . . Suomalainen, A. (2018). Genetic Basis of Severe Childhood-Onset Cardiomyopathies. *J Am Coll Cardiol, 72*(19), 2324-2338. doi:10.1016/j.jacc.2018.08.2171

Walsh, M. A., Grenier, M. A., Jefferies, J. L., Towbin, J. A., Lorts, A., & Czosek, R. J. (2012). Conduction abnormalities in pediatric patients with restrictive cardiomyopathy. *Circ Heart Fail, 5*(2), 267-273. doi:10.1161/CIRCHEARTFAILURE.111.964395

Walsh, R., Mazzarotto, F., Whiffin, N., Buchan, R., Midwinter, W., Wilk, A., . . . Ware, J. S. (2019). Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Med, 11*(1), 5. doi:10.1186/s13073-019-0616-z

Walsh, R., Thomson, K. L., Ware, J. S., Funke, B. H., Woodley, J., McGuire, K. J., . . . Watkins, H. (2017). Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet Med, 19*(2), 192-203. doi:10.1038/gim.2016.90

Wang, C., Hata, Y., Hirono, K., Takasaki, A., Ozawa, S. W., Nakaoka, H., . . . for, L. S. C. (2017). A Wide and Specific Spectrum of Genetic Variants and Genotype-Phenotype Correlations Revealed by Next-Generation Sequencing in Patients with Left Ventricular Noncompaction. *J Am Heart Assoc, 6*(9). doi:10.1161/JAHA.117.006210

Weill, U., Krieger, G., Avihou, Z., Milo, R., Schuldiner, M., & Davidi, D. (2019). Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast. *J Mol Biol, 431*(3), 636-641. doi:10.1016/j.jmb.2018.12.004

Wilcox, J. E., & Hershberger, R. E. (2018). Genetic cardiomyopathies. *Curr Opin Cardiol, 33*(3), 354-362. doi:10.1097/hco.0000000000000512

Yamanaka, S. (2010). Patient-specific pluripotent stem cells become even more accessible. *Cell Stem Cell, 7*(1), 1-2. doi:10.1016/j.stem.2010.06.009

Zhang, Z., Akhter, S., Mottl, S., & Jin, J. P. (2011). Calcium-regulated conformational change in the C-terminal end segment of troponin I and its binding to tropomyosin. *FEBS J, 278*(18), 3348-3359. doi:10.1111/j.1742-4658.2011.08250.x

Zhou, B., Wang, J., Lee, S. Y., Xiong, J., Bhanu, N., Guo, Q., . . . Dou, Y. (2016). PRDM16 Suppresses MLL1r Leukemia via Intrinsic Histone Methyltransferase Activity. *Mol Cell, 62*(2), 222-236. doi:10.1016/j.molcel.2016.03.010

Supplement

Table S 1: CMP-relevant Genes and classification into functional groups

Functional classification	Gene name	Gen	Protein	Chromo-some	Transcript ID	Exon	Function
Sarcoplasmic reticulum	calreticulin 3	*CALR3*	CALR3	Chr19	ENST00000269881 NM_145046	9	Ca^{2+} binding protein within SR
	dolichol kinase	*DOLK*	DOLK	Chr9	ENST00000372586 NM_014908	1	dolichol kinase within SR; involved in GPI-anchor, N- and O-linked protein glycosylation
	phospholamban	*PLN*	PLN	Chr6	ENST00000357525 NM_002667	2	regulation of cardiac ATP2A2 Ca^{2+} ATPase, critical regulator Ca^{2+} homeostasis
	ryanodine receptor 2	*RYR2*	RYR2	Chr1	ENST00000366574 NM_001035.2	105	Ca^{2+} channel releasing Ca^{2+} from SR into cytoplasm, regulation of cardiomyocyte contractility
Sarcolemma	ATP binding cassette subfamily C member 9	*ABCC9*	ABCC9	Chr12	ENST00000261200 NM_020297 ENST00000261201 NM_005691	38	regulation of KCNJ11 potassium channel
	caveolin 3	*CAV3*	CAV3	Chr3	ENST00000343849 NM_033337	2	scaffolding protein of caveolae, regulation of G-protein alpha subunits and potassium channels, sarcolemma repair
	hyperpolarization activated cyclic nucleotide gated potassium channel 4	*HCN4*	HCN4	Chr15	ENST00000261917 NM_005477.2	8	hyperpolarization-activated ion channel with slow activation and inactivation, exhibiting weak selectivity for potassium over sodium ions, cAMP regulated
	hemochromatosis	*HFE*	HFE	Chr6	ENST00000357618 NM_000410.3	5	transferrin receptor binding, regulation endocytosis of iron loaded transferrin
	junctophilin-2	*JPH2*	JPH2	Chr20	ENST00000372980 NM_020433.4	6	formation of junctional membrane complexes between plasma membrane and SR, regulator of Ca^{2+} signaling
	sodium channel protein type 5 subunit alpha	*SCN5A*	SCN5A	Chr3	ENST00000413689 NM_001099404.1	28	mediates voltage-dependent Na+ permeability of excitable membranes
Z-disc	actinin alpha 2	*ACTN2*	ACTN2	Chr1	ENST00000366578 NM_001103.2	21	F-actin crosslinking protein

	ankyrin repeat domain 1	*ANKRD1*	ANKRD1	Chr10	ENST00000371697 NM_014391.2	9	myofibrillar stretch-sensor Z-disc, transcription factor that negatively regulates cardiac gene expression
	cysteine and glycine rich protein 3	*CSRP3*	CSRP3	Chr11	ENST00000533783 NM_003476	7	scaffolding protein at Z-disc, actin binding, regulation of actin dynamics myogenic transcriptional cofactor
	four and a half LIM domains 1	*FHL1*	FHL1	ChrX	ENST00000394155 NM_001159702.2	8	mechanosensor at Z-disc, transcriptional cofactor
	four and a half LIM domains 2	*FHL2*	FHL2	Chr2	ENST00000344213 NM_201555.1	7	mechanosensor at Z-disc, transcriptional cofactor
	LIM domain binding 3	*LDB3*	LDB3, ZASP	Chr10	ENST00000429277 NM_001171610.1	14	Interaction with actin, stabilization sarcomere during contraction
	myozenin 2	*MYOZ2*	MYOZ2	Chr4	ENST00000307128 NM_016599.4	6	Z-disc organization by interaction with several proteins, myofibrillogenesis, calcineurin signaling
	myopalladin	*MYPN*	MYPN	Chr10	ENST00000358913 NM_032578	20	scaffold at Z-disc; binds nebulette, nebulin, α-actinin
	nexilin F-actin binding protein	*NEXN*	NEXN	Chr1	ENST00000334785 NM_144573.3	13	F-actin binding, maintenance Z-disc and sarcomere integrity
	PDZ and LIM domain 3	*PDLIM3*	PDLIM3	Chr4	ENST00000284770 NM_014476	7	crosslinking of F-actin with α-actinin-2
	titin-cap, telethonin	*TCAP*	TCAP	Chr17	ENST00000309889 NM_003673	2	binds titin at Z-disc, part of mechanosensing system, interaction of Z-disc and T-tubules
	vinculin	*VCL*	VCL	Chr10	ENST00000211998 NM_014000.2	22	F-actin binding, mechanosensing, cell-matrix and cell-cell adhesion, linker of Z-disc proteins to catenin cadherin dystroglycan
Sarcomere thick filament	myosin binding protein C, cardiac	*MYBPC3*	MYBPC3	Chr11	ENST00000545968 NM_000256.3	35	integrates myosin heavy chain and F-actin, modulation of muscle contraction and sarcomere organization
	myosin heavy chain 6	*MYH6*	MYH6, MHC-α	Chr14	ENST00000405093 NM_002471.3	39	major component of cardiac muscle thick filament, muscle contraction, fast myosin isoform, ATPase activity

	myosin heavy chain 7	*MYH7*	MYH7, MHC-β	Chr14	ENST00000355349 NM_000257.2	40	major component of cardiac muscle thick filament, muscle contraction, slow myosin isoform, ATPase activity
	myosin light chain 2	*MYL2*	MYL2	Chr12	ENST00000228841 NM_000432.3	7	stabilizes S1 neck region of myosin heavy chain
	myosin light chain 3	*MYL3*	MYL3	Chr3	ENST00000292327 NM_000258.2	7	stabilizes S1 neck region of myosin heavy chain
	myosin light chain kinase 2	*MYLK2*	MYLK2	Chr20	ENST00000375985 NM_033118.3	13	essential for muscle contraction cycle, phosphorylates specific position of myosin light chain
	titin	*TTN*	TTN	Chr2	ENST00000589042 NM_001267550.1	363	central sarcomere scaffolding protein required for assembly, protein interaction platform, regulation of sarcomere resting length and passive stiffness
Sarcomere thin filament	actin, alpha 1, skeletal muscle	*ACTA1*	ACTA1	Chr1	ENST00000366684 NM_001100	6	globular G-actin form F-actin fibers; essential part of the contractile apparatus thin filament
	actin, alpha, cardiac muscle 1	*ACTC1*	ACTC1	Chr15	ENST00000290378 NM_005159.4	7	globular G-actin form F-actin fibers; essential part of the contractile apparatus thin filament
	troponin C1, slow skeletal and cardiac type	*TNNC1*	TNNC1	Chr3	ENST00000232975 NM_003280.2	6	binds the switch region of troponin I in a Ca^{2+} dependent manner to activate contraction
	troponin I3, cardiac type	*TNNI3*	TNNI3	Chr19	ENST00000344887 NM_000363.4	8	inhibitory subunit of troponin, regulation thin filament Ca^{2+} sensitivity
	troponin T2, cardiac type	*TNNT2*	TNNT2	Chr1	ENST00000236918 NM_001276345	16	binding of troponin complex to tropomyosin
	tropomyosin 1	*TPM1*	TPM1	Chr15	ENST00000403994 NM_001018005.1	9	coiled coil protein that lies along thin filament and blocks myosin binding sites on actin under resting calcium concentrations
Nuclear envelope	emerin	*EMD*	EMD	ChrX	ENST00000369842 NM_000117	6	formation and stabilization of cortical nuclear actin network, cellular signaling
	lamin A/C	*LMNA*	LMNA	Chr1	ENST00000368300 NM_170707.3	12	nuclear lamina and chromatin organization, critical for nuclear dynamics

	transmembrane protein 43	*TMEM43*	TMEM43	Chr3	ENST00000306077 NM_024334.2	12	nuclear envelope structure at inner nuclear membrane, interaction with emerin, mechanotransduction, gene expression
Mitochondria	cytochrome c oxidase assembly homolog	*COX15*	COX15	Chr10	ENST00000370483 NM_001320975	9	heme A biosynthesis, porphyrin synthesis
	DnaJ heat shock protein family (Hsp40) member C19	*DNAJC19*	DNAJC19 TIM14	Chr3	ENST00000382564 NM_145261	6	peptide translocation inner membrane to matrix, subunit of HSP40/HSP70 complex, mitochondrial chaperone
	frataxin	*FXN*	FXN	Chr9	ENST00000377270 NM_000144	5	heme biosynthesis, protection against iron-catalyzed oxidative stress, iron storage
	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit	*HADHA*	HADHA	Chr2	ENST00000380649 NM_000182	20	fatty acid beta-oxidation and in lipid metabolism
	cytochrome c oxidase assembly protein	*SCO2*	SCO2	Chr22	ENST00000252785 NM_001169111	2	copper chaperone, delivering copper to COX2
	succinate dehydro-genase complex flavorprotein subunit A	*SDHA*	SDHA, SDH2	Chr5	ENST00000264932 NM_004168	15	complex II mitochondrial electron transport chain, transfer of electrons from succinate to ubiquinone, involved tricarboxylic acid cycle
	tafazzin	*TAZ*	TAZ	ChrX	ENST00000601016 NM_000116	11	synthesis/remodeling of cardiolipin, phospholipid-lysophospholipid transacylase, important for mitochondrial energy production
Desmosome	desmocollin 2	*DSC2*	DSC2	Chr18	ENST00000280904 NM_024422	16	part of desmosome junctions, cell-cell adhesion, interaction intermediate filaments
	desmoglein 2	*DSG2*	DSG2	Chr18	ENST00000261590 NM_001943.3	15	part of desmosome junctions, cell-cell adhesion, interaction intermediate filaments
	desmoplakin	*DSP*	DSP	Chr6	ENST00000379802 NM_004415.2	24	anchoring of intermediate filaments to desmosomes, essential for desmosomal plaque formation
	plakophilin 2	*PKP2*	PKP2	Chr12	ENST00000070846 NM_004572.3	14	links cadherins to intermediate filaments, mechanical stabilization desmosomes, signaling function

	junction plakoglobin	*JUP*	JUP	Chr17	ENST00000393931 NM_002230.2	14	cytoplasmic component of desmosomes, linker of cadherins to F-actin
Regulation of transcription and splicing	EYA transcriptional coactivator and phosphatase 4	*EYA4*	EYA4	Chr6	ENST00000355167 NM_172105.3	20	tyrosine phosphatase of histone H2AX; impact for DNA repair, stress response and apoptosis
	GATA zinc finger domain containing 1	*GATAD1*	GATAD1	Chr7	ENST00000287957 NM_021167	5	part of chromatin complex recruited to methylated H2K4me
	NK2 homeobox 5	*NKX2-5*	NKX2-5	Chr5	ENST00000329198 NM_004387.3	2	transcriptional activation of myocardial lineage together with GATA4, transcriptionally controlled by PBX1
	PR/SET domain 16	*PRDM16*	PRDM16, MEL1	Chr1	ENST00000270722 NM_022114.3	17	transcriptional regulator of adipocyte development, interaction e.g. PPARγ, C/EBPβ and PGC1α, regulator of TGF-β signaling, has histone methyltransferase activity (H3K9me1, H3K9me3)
	RNA binding motif protein 20	*RBM20*	RBM20	Chr10	ENST00000369519 NM_001134363.1	14	mRNA splicing regulator of a specific target genes, important regulator of TTN slicing
	T-box 20	*TBX20*	TBX20	Chr7	ENST00000408931 NM_001077653.2	8	transcriptional activator/repressor in cardiac development, interacts with GATA4 and NKX2-5, repression of TBX2
Dystrophin complex	desmin	*DES*	DES	Chr2	ENST00000373960 NM_001927.3	9	intermediate filament; connecting Z-discs, sarcomere, sarcolemmal cytoskeleton, nucleus and mitochondria; sarcomeric microtubule anchor
	dystrophin	*DMD*	DMD	ChrX	ENST00000357033 NM_004006.2	79	anchors ECM to F-actin cytoskeleton, ligand dystroglycan
	dystrobrevin alpha	*DTNA*	DTNA	Chr18	ENST00000444659 NM_001390.4	22	interacts with dystrophin
	sarcoglycan beta	*SGCB*	SGCB	Chr4	ENST00000381431 NM_000232.4	6	linker between F-actin and ECM
	sarcoglycan delta	*SGCD*	SGCD	Chr5	ENST00000435422 NM_000337.5	8	linker between F-actin and ECM
	sarcoglycan gamma	*SGCG*	SGCG	Chr13	ENST00000218867 NM_000231	8	linker between F-actin and ECM

Cellular signaling	B-Raf proto-oncogene, serine/threonine kinase	*BRAF*	BRAF	Chr7	ENST00000288602 NM_004333.4	18	phosphorylates MAP2K1, part of MAPK signaling pathway, transmission of mitogenic signals
	CBL proto-oncogene	*CBL*	CBL	Chr11	ENST00000264033 NM_005188.3	16	E3 ubiquitin-protein ligase, ubiquitination → degradation, negative regulator of cell surface receptors e.g. FGFR1/2, EGFR, KIT
	HRAS proto-oncogene, GTPase	*HRAS*	HRAS	Chr11	ENST00000610977 NM_001130442	5	small GTPase, RAS signaling pathway, growth factor signaling
	KRAS proto-oncogene, GTPase	*KRAS*	KRAS	Chr12	ENST00000311936 NM_004985	5	small GTPase, RAS signaling pathway, growth factor signaling, regulating cell proliferation, critical for tumor development
	mitogen-activated protein kinase kinase 1	*MAP2K1*	MAP2K1 MEK1	Chr15	ENST00000307102 NM_002755	11	essential component of MAPK pathway, growth factor signaling, regulating diverse cellular functions, upstream activated by RAF1, downstream activation of ERK1/2
	mitogen-activated protein kinase kinase 2	*MAP2K2*	MAP2K2 MEK2	Chr19	ENST00000262948 NM_030662	11	essential component of MAPK pathway, growth factor signaling, regulating diverse cellular functions, upstream activated by RAF1, downstream activation of ERK1/2
	mindbomb E3 ubiquitin protein ligase 1	*MIB1*	MIB1	Chr18	ENST00000261537 NM_020774	21	ubiquitination of Delta receptors, positively regulates the Delta-mediated Notch signaling
	NRAS proto-oncogene, GTPase	*NRAS*	NRAS	Chr1	ENST00000369535 NM_002524	7	small GTPase, RAS signaling pathway, growth factor signaling, regulating cell proliferation, critical for tumor development
	protein kinase AMP-activated non-catalytic subunit gamma 2	*PRKAG2*	PRKAG2	Chr7	ENST00000287878 NM_016203.3	16	AMP/ATP-binding subunit of AMPK, regulation of cellular energy metabolism by sensing ATP levels

	protein tyrosine phosphatase, non-receptor type 11	*PTPN11*	PTPN11	Chr12	ENST00000635625 NM_001330437	15	regulator unfolded protein response in ER, dephosphorylation EIF2AK3/PERK, tumor development
	Raf-1 proto-oncogene, serine/threonine kinase	*RAF1*	RAF1	Chr3	ENST00000251849 NM_002880.3	17	serine/threonine-protein kinase linking membrane-associated RAS GTPases with MAPK signaling, key regulator of cell fate
	SHOC2, leucine rich repeat scaffold protein	*SHOC2*	SHOC2	Chr10	ENST00000369452 NM_007373.3	9	regulatory subunit of protein phosphatase 1, activation of RAF1 kinase, MAPK signaling pathway activation, growth factor signaling
	SOS Ras/Rac guanine nucleotide exchange factor 1	*SOS1*	SOS1	Chr2	ENST00000402219 NM_005633.3	23	guanine nucleotide exchange factor (GEF) for RAS GTPases, promotion of RAS activation, regulates MAPK3 phosphorylation
	transforming growth factor beta 3	*TGFB3*	TGFB3	Chr14	ENST00000238682 NM_003239.2	7	regulates mesenchymal development, activates TGFBR2, regulates cell adhesion ECM formation
Protein quality control	BCL2 associated athanogene 3	*BAG3*	BAG3	Chr10	ENST00000369085 NM_004281.3	4	activator of autophagy and HSP70 chaperone system, nucleotide exchange factor HSP70 complex, adaptor protein
	crystallin alpha B	*CRYAB*	CRYAB HSPB5	Chr11	ENST00000616970 NM_001885.1	4	detection of misfolded proteins, targeting to the HSP70 complex, prevention of protein aggregate formation, cardioprotective, stabilization of cyto- and nucleoskeleton, stress inducible
	heat shock protein family B (small) member 8	*HSPB8*	HSP22 HSPB8	Chr12	ENST00000281938 NM_014365.2	3	detection of misfolded proteins, induction of autophagy, prevention of protein aggregate formation, chaperone activity, disassembly of stress granules with BAG3/HSP70 complex
ECM components	fibrillin 1	*FBN1*	FBN1	Chr15	ENST00000316623 NM_000138.4	66	structural component of 10-12 nm diameter ECM microfibrils, structural and regulatory function in load-bearing connective tissue, regulation TGF-β and BMP signaling

	laminin subunit alpha 2	*LAMA2*	LAMA2	Chr6	ENST00000421865 NM_000426.3	65	ECM glycoprotein, mediates interaction to dystrophin complex
	laminin subunit alpha 4	*LAMA4*	LAMA4	Chr6	ENST00000230538 NM_001105206.2	39	ECM glycoprotein, constituent of basement membrane
Glycosyl-transferases	fukutin	*FKTN*	FKTN	Chr9	ENST00000223528 NM_006731	10	biosynthesis of phosphorylated O-mannosyl trisaccharide, required for α-dystroglycan synthesis, protein glycosylation
	fukutin related protein	*FKRP*	FKRP	Chr19	ENST00000318584 NM_001039885	4	biosynthesis of phosphorylated O-mannosyl trisaccharide, required for α-dystroglycan synthesis, protein glycosylation
Cytoskeleton	Alstrom syndrome protein 1	*ALMS1*	ALMS1	Chr2	ENST00000613296 NM_015120.4	23	centrosome and basal body associated protein, pericentriolar and cilia transport, interaction α-actinin, endosomal transport
Endo-lysosomal system	glucosidase alpha, acid	*GAA*	GAA	Chr17	ENST00000302262 NM_000152	20	hydrolysis of lysosomal glycogen to glucose
	galactosidase alpha	*GLA*	GLA	ChrX	ENST00000218516 NM_000169	7	hydrolysis of α-D-galactose residues in α-D-galactosides
	lysosomal associated membrane protein 2	*LAMP2*	LAMP2 CD107b	ChrX	ENST00000434600 NM_001122606.1	9	binding of target protein and delivery to lysosomes, protein degradation, required fusion autophagosomes and lysosomes, antigen presentation
Others	transthyretin	*TTR*	TTR	Chr18	ENST00000237014 NM_000371.3	4	thyroid hormone-binding protein, transport of thyroxine from bloodstream to peripheral organs

Table S 2: Scheme to interpret the pathogenicity of genetic variants according to guiedelines of the ACMG given by the publication. For each category (population data, etc.) the strongest argument will be activated for each variant. The sum of arguments in all categories is determining the pathogenicty of a genetic variant (**Figure S 1**).

Diagnosis	Supporting	Moderate	Strong	Very strong	Conclusion
Population data		Absent or low MAF in gnomAD/population databases **PM2**	Prevalence in affecteds statistically increased over controls **PS4**		**pathogenic**
Computational & predictive data	Multiple lines of computational evidence support a deleterious effect on the gene /gene product **PP3**	Novel missense change at already as pathogenic described amino acid **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	Missense in gene with few benign and lots of path. missense variants **PP2**	Mutational hot spot or well-studied functional domain without benign variation **PM1**	Well-established functional studies **PS3**		
Segregation data	Cosegregation with disease in multiple affected family members **PP1**	Increased segregation data	More increased segregation data		
***De novo* data**		*De novo* (without paternity & maternity confirmed) **PM6**	*De novo* (paternity and maternity confirmed) **PS2**		

Table S 3: Scheme to interpret the pathogenicity of genetic variants according to guiedelines of the ACMG in optimized form (description see Table S 2).

Diagnosis	Supporting	Moderate	Strong	Very strong	Conclusion
Population data		**PM2** check MAF and only include variants with MAF < 0.0001	**PS4** activate with calculated OR > 5, cohorts from literature + ACGV for affected and gnomAD for unaffected		**Pathogenic?**
Computational & predictive data	**PP3 SIFT/Polyphen-2/MT2/NNsplice/MaxEnt/ HSF** → run prediction tools (requirement for splice site variants in PM4)	**PM5** check ClinVar for P/LP variants at same aa position **PM4** coding sequence: be aware of small deletions in aa repeats; splice site: PP3 positive as requirement	**PS1** check ClinVar for P/LP variant with same aa exchange	**PVS1** use o/e ratio (upper CI < 0.35) or pLI (≥ 0.9) from gnomAD; check ClinVar for truncating variants; check literature for LoF	
Functional data	**PP2** use o/e ratio (upper CI < 0.35) or Z-score (> 3.09) from gnomAD; check ClinVar for missense variants; check literature	**PM1** check ClinVar/literature for influence of missense variation	**PS3** check literature for functional studies (HGMD, ClinVar, PubMed, Ensembl, UniProt, .etc.)		
Segregation data	**PP1** Variant segregates in ≥ 3 meioses (every affected member!)	**PP1** Variant segregates in ≥ 5 meioses (every affected member!)	**PP1** Variant segregates in ≥ 8 meioses (every affected member!)		
***De novo* data**		*De novo* (without paternity & maternity confirmed) **PM6**	*De novo* (paternity and maternity confirmed) **PS2**		

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Figure S 1: Counting scheme to interpret the pathogenicity of genetic variants according to the ACMG guidelines. The number of activated arguments (PVS1, PS1, etc.) determines the outcome.

Table S 4: Detection of genetic VOI from 80 index patients with pediatric CMP

Gene	Transcript	cDNA alteration	Protein alteration	Diagnosis	Patient ID complex genotypes	gnomAD allele frequency	Pathogenicity	*de novo*	novel
Pathogenic and likely pathogenic genetic variants									
ACTC1	NM_005159.4	c.328G>A	p.A110T	DCM, HTX	CMP-77	0	Likely pathogenic (PM2, PM6, PP2-3)	yes	no
ACTN2	NM_001103.2	c.574C>T	p.R192*	LVNC, HTX	CMP-10	0.000008122	Pathogenic (PVS1, PM2, PM6)	yes	yes
DSC2	NM_024422.3	c.1034T>C	p.I345T	ARVC	CMP-33	0.000008133	Likely pathogenic (PS3, PM2, PP3)	no	no
DSG2	NM_001943.3	c.1016delA	p.V340*	ARVC	CMP-33	0.00000814	Likely pathogenic (PVS1, PM2)	no	yes
LAMP2	NM_001122606.1	c.222_223delTA	p.Y74*	HCM		0	Pathogenic (PVS1, PS4, PM6)	yes	no
MYBPC3	NM_000256.3	c.772G>A (splice variant)	p.E258K / p.Glu258fsX41	HCM	CMP-42	0.00001851	Pathogenic (PS1, PS4)	?	no
	NM_000256.3	c.927-2A>G	-	HCM	CMP-31	0.00003235	Pathogenic (PS3-4, PM4, PP1)	no	no
	NM_000256.3	c.1504C>T	p.R502W	HCM	CMP-38	0.00005411	Pathogenic (PS1, PS4)	?	no
	NM_000256.3	c.1805C>T	p.T602I	LVNC	CMP-27	0.000009848	Likely pathogenic (PS1, PM2)	no	no
	NM_000256.3	c.2308G>A	p.D770N	HCM	CMP-69	0.00001625	Pathogenic (PS1, PS4, PM1)	no	no
	NM_000256.3	c.2572A>C	p.S858R	DCM, HTX	CMP-09	0	Likely pathogenic (PS2, PM2, PM5)	?	yes
MYH7	NM_000257.2	c.677C>T	p.A226V	HCM		0	Likely pathogenic (PS4, PM1, PP3)	?	no
	NM_000257.2	c.1063G>A	p.A355T	HCM	CMP-22	0	Pathogenic (PS1, PS4, PM1)	?	no
	NM_000257.2	c.1283C>A	p.A428D	LVNC	CMP-27	0	Likely pathogenic (PM1-2, PP1, PP3)	no	yes
	NM_000257.2	c.1357C>T	p.R453C	HCM		0	Pathogenic (PS1, PS3-4)	?	no
	NM_000257.2	c.1987C>T	p.R663C	HCM	CMP-28	0	Pathogenic (PS1, PS4, PM2)	no	no
	NM_000257.2	c.1988G>A	p.R663H	HCM	CMP-43	0.00001443	Pathogenic (PS1, PS4, PM1)	?	no
	NM_000257.2	c.2710C>T	p.R904C	DCM	CMP-44	0.000004061	Pathogenic (PS1, PS4, PM1)	?	no
PKP2	NM_004572.3	c.1716delG	p.R573Efs*4	ARVC		0	Likely pathogenic (PSV1, PM2)	no	yes
PRKAG2	NM_016203.3	c.1199C>A	p.T400N	HCM		0	Pathogenic (PS1, PS3, PM2)	?	no

TAZ	NM_000116.3	c.355G>A	p.V119M	LVNC	CMP-24	0	Likely pathogenic (PM1-2, PM5)	no	yes
TNNI3	NM_000363.4	c.24+2T>A (homozygous)	-	LVNC	CMP-01	0.00001254	Pathogenic (PVS1, PS3, PM2)	?	yes
	NM_000363.4	c.570C>G	p.D190E	HCM		0	Likely Pathogenic (PM1-2, PM5)	?	no
	NM_000363.4	c.574C>T	p.R192C	RCM, HTX	CMP-58	0	Pathogenic (PS1, PS3-4, PM2, PM6)	yes	no
	NM_000363.4	c.582C>A	p.N194K	DCM	CMP-80	0	Pathogenic (PS1, PM1-2, PM6)	yes	no
	NM_000363.4	c.624dupT	p.E209*	RCM, HTX	CMP-29	0	Likely pathogenic (PM1-2, PM4, PM6)	yes	yes
TNNT2	NM_000364	c.620_622delAGA	p.K207del	DCM, HTX	CMP-12	0	Likely Pathogenic (PS1, PM1-2)	?	no
	NM_000364	c.620-622delAGA	p.K207del	DCM, HTX	CMP-74	0	Pathogenic (PS1, PM1-2, PM6)	yes	no
	NM_000364.2	c.812+1G>A	-	HCM		0	Pathogenic (PS1, PS3, PM2)	no	no
TPM1	NM_001018005.1	c.257C>T	p.A86V	LVNC	CMP-34	0	Likely pathogenic (PM1-2, PP1_moderate, PP3)	no	yes
TTN	NM_001267550.1	c.68329+2_68329+3insTT	-	RCM, HTX	CMP-25	0.00009578	Likely pathogenic (PM1-2, PM4)	no	yes
	NM_001267550.1	c.85891delG	p.A28631Lfs*3	DCM, HTX		0	Likely pathogenic (PM1-2, PM4)	no	yes
Genetic variants of uncertain significance (VUS)									
ACTN2	NM_001103.2	c.278G>A	p.R93Q	LVNC	CMP-24	0.00002525	Uncertain significance	no	no
BAG3	NM_004281.3	c.280A>T	p.I94F	DCM		0.0006493	Uncertain significance	?	no
	NM_004281.3	c.881G>A	p.R294H	DCM, HTX		0.00003663	Uncertain significance	?	no
	NM_004281.3	c.1634C>G	p.P545R	DCM		0.00009385	Uncertain significance	no	no
CBL	NM_005188.3	c.805A>G	p.M269V	LVNC	CMP-24	0.00001218	Uncertain significance	no	yes
DMD	NM_004006.2	c.2169-7_2169-4delGTCT	-	HCM	CMP-43	0.00003417	Uncertain significance	?	no
	NM_004006.2	c.2273A>C	p.D758A	RCM, HTX	CMP-25	0.0002632	Uncertain significance	no	yes
	NM_004006.2	c.5723A>T	p.D1908V	HCM, HTX		0.0002101	Uncertain significance	no	yes
	NM_004006.2	c.8996C>T	p.A2999V	RCM, HTX	CMP-29	0.0000112	Uncertain significance	no	yes
DSC2	NM_024422.3	c.304G>A	p.E102K	DCM		0.0007482	Uncertain significance	no	no
	NM_024422.3	c.363G>T	p.K121N	LVNC		0.00001221	Uncertain significance	no	no

	NM_024422.3	c.802A>G	p.T268A	DCM		0.0001444	Uncertain significance	no	no
DSG2	NM_001943.3	c.1003A>G	p.T335A	DCM		0.0005089	Uncertain significance	?	no
	NM_001943.3	c.2001+3C>G	-	LVNC		0.000008136	Uncertain significance	no	yes
DSP	NM_004415.2	c.2774G>A	p.R925Q	RCM, HTX		0.0002309	Uncertain significance	no	no
	NM_004415.2	c.4961T>C	p.L1654P	LVNC, HTX		0.00002849	Uncertain significance	no	no
	NM_004415.2	c.5178C>A	p.N1726K	DCM, HTX		0.0006893	Uncertain significance	no	no
	NM_004415.2	c.7916G>A	p.R2639Q	DCM, HTX	CMP-74	0.000853	Uncertain significance	no	no
	NM_004415.2	c.7994C>T	p.T2665M	DCM	CMP-80	0.00004338	Uncertain significance	no	no
	NM_004415.2	c.8524C>T	p.R2842C	HCM	CMP-69	0.000007384	Uncertain significance	no	no
DTNA	NM_001390.4	c.1757C>T	p.P586L	HCM		0.0001263	Uncertain significance	no	no
EYA4	NM_172105.3	c.59A>G	p.D20G	RCM		0.000008132	Uncertain significance	?	no
	NM_172105.3	c.971-3T>C	-	LVNC, HTX	CMP-10	0.00006461	Uncertain significance	no	yes
FBN1	NM_000138.4	c.902G>T	p.G301V	DCM, HTX		0.0001876	Uncertain significance	?	no
FHL1	NM_001159702.2	c.283C>T	p.R95W	HCM		0.0006232	Uncertain significance	no	no
FHL2	NM_201555.1	c.143G>A	p.G48D	DCM		0	Uncertain significance	no	yes
	NM_201555.1	c.337C>T	p.R113C	DCM		0.000395	Uncertain significance	?	no
JPH2	NM_020433.4	c.572C>G	p.P191R	DCM		0.0005276	Uncertain significance	no	no
	NM_020433.4	c.1306C>T	p.R436C	DCM		0.00003236	Uncertain significance	no	no
	NM_020433.4	c.1896G>C	p.E632D	RCM, HTX	CMP-58	0.00007797	Uncertain significance	no	no
JUP	NM_002230.2	c.1714C>T	p.R572W	DCM, HTX	CMP-09	0.00001219	Uncertain significance	?	no
LAMA4	NM_001105206.2	c.514G>A	p.G172S	HCM	CMP-69	0.0002224	Uncertain significance	?	no
	NM_001105206.2	c.1959T>C	p.D653	ARVC	CMP-33	0.0009604	Uncertain significance	no	yes
	NM_001105206.2	c.2171G>A	p.R724K	DCM, HTX	CMP-12	0.0007326	Uncertain significance	?	yes
	NM_001105206.2	c.4645A>T	p.N1549Y	HCM		0.0001988	Uncertain significance	?	yes
LDB3/ZASP	NM_001171610.1	c.66C>A	p.D22E	HCM	CMP-22	0.00003236	Uncertain significance	no	yes

	NM_001171610.1	c.664G>A	p.A222T	HCM	CMP-42	0.0003646	Uncertain significance	?	no
	NM_001171610.1	c.778G>A	p.G260S	DCM		0.0000469	Uncertain significance	?	yes
	NM_001171610.1	c.982G>A	p.A328T	LVNC	CMP-34	0.000004062	Uncertain significance	no	yes
	NM_007078	c.1792T>C	p.C598R	HCM	CMP-31	0	Uncertain significance	no	yes
	NM_001171610.1	c.1978C>A	p.P660T	HCM	CMP-22	0.000007219	Uncertain significance	no	yes
LMNA	NM_170707.3	c.986G>C	p.R329P	HCM	CMP-22	0	Uncertain significance	?	yes
MYBPC3	NM_000256.3	c.961G>A	p.V321M	DCM		0.0003261	Uncertain significance	no	no
	NM_000256.3	c.2873C>T	p.T958I	HCM	CMP-42	0.000156	Uncertain significance	?	no
MYH6	NM_002471.3	c.3010G>T	p.A1004S	HCM	CMP-43	0.0009523	Uncertain significance	?	no
	NM_002471.3	c.5348G>A	p.R1783H	HCM		0.00003248	Uncertain significance	no	no
MYH7	NM_000257.2	c.475G>A	p.D159N	LVNC, HTX		0	Uncertain significance	?	yes
	NM_000257.2	c.1425G>T	p.Q475H	LVNC		0	Uncertain significance	no	no
	NM_000257.2	c.2890G>C	p.V964L	DCM		0.0004256	Uncertain significance	no	no
	NM_000257.2	c.3169G>A	p.G1057S	HCM		0.000008121	Uncertain significance	no	no
	NM_000257.2	c.4501G>T	p.E1501*	DCM		0	Uncertain significance	no	yes
	NM_000257.2	c.5767A>G	p.K1923E	DCM		0	Uncertain significance	?	yes
MYL2	NM_000432.3	c.421G>A	p.A141T	DCM, HTX		0	Uncertain significance	?	no
MYLK2	NM_033118.3	c.4G>A	p.A2T	DCM, HTX	CMP-77	0.0009884	Uncertain significance	no	no
	NM_033118.3	c.425G>T	p.G142V	LVNC, HTX	CMP-10	0.00008924	Uncertain significance	no	yes
MYOZ2	NM_016599.4	c.659T>C	p.M220T	HCM	CMP-22	0	Uncertain significance	no	yes
MYPN	NM_001256267.1	c.259C>G	p.P87A	DCM		0.0000433	Uncertain significance	no	no
	NM_001256267.1	c.802C>T	p.P268S	RCM, HTX	CMP-58	0.0004756	Uncertain significance	no	no
	NM_001256267.1	c.970C>T	p.H324Y	HCM	CMP-28	0	Uncertain significance	no	yes
	NM_001256267.1	c.2150C>T	p.T717M	DCM, HTX		0.00004467	Uncertain significance	?	yes
	NM_001256267.1	c.3913A>G	p.M1305V	HCM	CMP-42	0.00005412	Uncertain significance	?	no

NEXN	NM_144573.3	c.893C>G	p.T298R	HCM	CMP-42	0.0001605	Uncertain significance	?	no
	NM_144573.3	c.1572_1574delAGA	p.E525del	DCM, HTX		0.0001556	Uncertain significance	?	yes
	NM_144573.3	c.1572_1574delAGA	p.E525del	DCM	CMP-01	0.0001556	Uncertain significance	?	yes
	NM_144573.3	c.1619T>C	p.M540T	DCM, HTX		0.000004072	Uncertain significance	?	yes
NKX2-5	NM_004387.3	c.355G>T	p.A119S	DCM		0.0009725	Uncertain significance	no	no
PKP2	NM_004572.3	c.1536T>A	p.N512K	DCM, HTX		0.00001446	Uncertain significance	no	yes
	NM_004572.3	c.2200A>G	p.T734A	DCM, HTX		0.00001083	Uncertain significance	?	no
	NM_004572.3	c.2365A>G	p.I789V	DCM		0.0003105	Uncertain significance	?	yes
PRDM16	NM_022114.3	c.1110C>A	p.D370E	LVNC, HTX		0.000004112	Uncertain significance	?	yes
	NM_022114.3	c.2296G>A	p.G766S	HCM	CMP-22	0.0002788	Uncertain significance	?	no
	NM_022114.3	c.2372G>A	p.G791D	RCM, HTX		0	Uncertain significance	no	yes
	NM_022114.3	c.2447A>G	p.N816S	RCM, HTX	CMP-25	0.0002405	Uncertain significance	no	no
RAF1	NM_002880.3	c.974A>C	p.Q325P	DCM	CMP-44	0.00001804	Uncertain significance	?	no
RBM20	NM_001134363.1	c.298C>T	p.L100F	DCM		0	Uncertain significance	no	yes
RYR2	NM_001035.2	c.1699G>C	p.A567P	DCM, HTX		0.000008299	Uncertain significance	?	yes
	NM_001035.2	c.4273A>G	p.T1425A	HCM	CMP-38	0.00005876	Uncertain significance	?	no
	NM_001035.2	c.8162T>C	p.I2721T	HCM		0.0005919	Uncertain significance	no	no
	NM_001035.2	c.9655G>A	p.V3219M	DCM		0.00007581	Uncertain significance	?	no
SCN5A	NM_001099404.1	c.998+5G>A	-	DCM		0.000213	Uncertain significance	?	no
	NM_001099404.1	c.1577G>A	p.R526H	HCM		0.00006255	Uncertain significance	?	no
SOS1	NM_005633.3	c.3841_3843dupAAG	p.E1281dup	DCM	CMP-01	0.000004065	Uncertain significance	?	yes
TAZ	NM_000116.3	c.29C>G	p.P10R	LVNC	CMP-24	0	Uncertain significance	?	yes
TBX20	NM_001077653.2	c.208G>A	p.G70S	DCM		0	Uncertain significance	no	yes
	NM_001077653.2	c.994C>T	p.P332S	DCM, HTX		0.0000366	Uncertain significance	no	yes
TGFB3	NM_003239.2	c.293C>T	p.S98L	DCM		0.0009523	Uncertain significance	no	no

TMEM43	NM_024334.2	c.1177C>T	p.R393W	LVNC		0.000008133	Uncertain significance	no	yes
TNNC1	NM_003280.2	c.25G>A	p.V9I	DCM, HTX		0	Uncertain significance	?	yes
TNNI3	NM_000363.4	c.307C>T	p.R103C	HCM	CMP-22	0.000007216	Uncertain significance	?	no
TNNT2	NM_000364.2	c.808G>A	p.V270I	DCM		0	Uncertain significance	no	yes
TPM1	NM_001018005.1	c.340G>C	p.E114Q	DCM, HTX		0	Uncertain significance	no	yes
TTN	NM_001267550.1	c.25064-4A>G	-	RCM		0.000008577	Uncertain significance	no	yes
	NM_001267550.1	c.39709+1G>T	-	RCM, HTX		0	Uncertain significance	no	yes
VCL	NM_014000.2	c.590C>T	p.T197I	HCM		0.0002526	Uncertain significance	no	no

Table S 5: Selection of genetic variants and complex genotypes

Gene	Transcript	cDNA Position	Protein position	Diagnosis	Patient ID complex genotypes	gnomAD allele frequency	Pathogenicity	*de novo*	novel
Truncating TTN variants									
TTN	NM_001267550.1	c.25064-4A>G*	-	RCM		0.000008577	Uncertain significance	no	yes
	NM_001267550.1	c.39709+1G>T*	-	RCM		0	Uncertain significance	no	yes
	NM_001267550.1	c.68329+2_68329+3insTT*	-	RCM	CMP-25	0.00009578	Likely pathogenic	no	yes
	NM_001267550.1	c.85891delG	p.A28631Lfs*3	DCM		0	Likely pathogenic	no	yes
Homozygous variants									
NEXN	NM_144573.3	c.1572_1574delAGA	p.E525del	DCM, HTX		0.0001556	Uncertain significance	?	yes
TNNI3	NM_000363.4	c.24+2T>A*	p.Ala8_Ala9insGluArgAlaAlaGly*	LVNC	CMP-01	0.00001254	Pathogenic	?	yes
Hemizygous variants									
DMD	NM_004006.2	c.2169-7_2169-4delGTCT*	-	HCM	CMP-43	0.00003417	Uncertain significance	?	no
	NM_004006.2	c.2273A>C	p.D758A	RCM	CMP-25	0.0002632	Uncertain significance	no	yes
	NM_004006.2	c.5723A>T	p.D1908V	HCM		0.0002101	Uncertain significance	no	yes
	NM_004006.2	c.8996C>T	p.A2999V	RCM	CMP-29	0.0000112	Uncertain significance	no	yes
FHL1	NM_001159702.2	c.283C>T	p.R95W	HCM		0.0006232	Uncertain significance	no	no
TAZ	NM_000116.3	c.29C>G	p.P10R	LVNC		0	Uncertain significance	?	yes
	NM_000116.3	c.355G>A	p.V119M	LVNC	CMP-24	0	Likely pathogenic	no	yes
Index patients with compound heterozygous variants									
DSC2	NM_024422.3	c.304G>A	p.E102K	DCM		0.0007482	Uncertain significance	no	no
	NM_024422.3	c.802A>G	p.T268A	DCM		0.0001444	Uncertain significance	no	no
MYBPC3	NM_000256.3	c.2572A>C	p.S858R	DCM	CMP-09	0	Likely pathogenic	?	yes
	NM_000256.3	genomic deletion of *MYBPC3* gene	-	DCM	CMP-09	0	Pathogenic	no	yes

Index patients with >1 pathogenic and/or likely pathogenic variant									
MYBPC3	NM_000256.3	c.1805C>T	p.T602I	LVNC	CMP-27	0.000009848	Likely pathogenic	no	no
MYH7	NM_000257.2	c.1283C>A	p.A428D	LVNC	CMP-27	0	Likely pathogenic	no	yes
DSC2	NM_024422.3	c.1034T>C	p.I345T	ARVC	CMP-33	0.000008133	Likely pathogenic	no	no
DSG2	NM_001943.3	c.1016delA	p.V340*	ARVC	CMP-33	0.00000814	Likely pathogenic	no	yes

Table S 6: Genetic and heterozygous variants in MYH7

Diagnosis	Sex	Age in years	Genomic location	cDNA position	Protein position	gnomAD MAF	ACMG pathogenicity	novel	ClinVar	Age ≤18 years†	Age >18 years‡	PubMed ID
HCM	m	11,8	14:23896042	c.1988G>A	p.R663H	0.000014	P	no	yes	yes	yes	10750581
DCM	f	0,0	14:23893328	c.2710C>T	p.R904C	0.000004	P	no	yes	yes	yes	20573160 29212898
LVNC	m	5,9	14:23897862	c.1425G>T	p.Q475H	0	VUS	no	no	no	yes	21750094
LVNC	m	0,2	14:23901875	c.475G>A	p.D159N	0	VUS	yes	no	-	-	-
DCM	f	0,3	14:23893148	c.2890G>C	p.V964L	0.000426	VUS	no	yes	no	yes	23349452 19412328
HCM	m	7,4	14:23898214	c.1357C>T	p.R453C	0	P	no	yes	yes	yes	1552912 17495353
DCM	m	6,2	14:23882991	c.5767A>G	p.K1923E	0	VUS	yes	no	-	-	-
HCM	m	15,2	14:23899059	c.1063G>A	p.A355T	0	P	no	yes	yes	yes	12707239 20031618
HCM	m	16,2	14:23891465	c.3169G>A	p.G1057S	0.000008	VUS	no	yes	no	yes	15358028
LVNC	f	11,7	14:23898288	c.1283C>A	p.A428D	0	LP	yes	no	-	-	-
HCM	m	5,3	14:23896043	c.1987C>T	p.R663C	0	P	no	yes	yes	yes	11133230 15358028
HCM	f	0,7	14:23900849	c.677C>T	p.A226V	0	LP	no	yes	-	-	-
DCM	f	0.2	14:23886380	c.4501G>T	p.E1501*	0	VUS	yes	no	-	-	-

HTX - heart transplantation, individuals †≤18 years and ‡>18 years with PubMed ID of study; ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)

Table S 7: Genetic and heterozygous variants in MYBPC3

Diagnosis	Sex	Age in years	Genomic location	cDNA position	Protein position	gnomAD MAF	ACMG pathogenicity	novel	ClinVar	Age ≤18 years†	Age >18 years‡	PubMed ID
DCM	f	0,1	11:47358972	c.2572A>C	p.S858R	0	LP	yes	no	-	-	-
HCM	m	2,5	11:47364249	c.1504C>T	p.R502W	0.000054	P	no	yes	yes	yes	18403758 9562578
HCM	m	15,9	11:47369975	c.772G>A	p.E258K / p.Glu258fsX41	0.000019	P	no	yes	yes	yes	17908752 9562578
HCM	m	15,9	11:47356625	c.2873C>T	p.T958I	0.000156	VUS	no	yes	no	yes	18957093
DCM	m	0,2	11:47367887	c.961G>A	p.V321M	0.000326	VUS	no	yes	no	yes	21750094
HCM	m	16,9	11:47360071	c.2308G>A	p.D770N	0.000016	P	no	yes	yes	yes	22555271 15519027
LVNC	f	11,7	11:47362781	c.1805C>A	p.T602I	0.00001	LP	no	yes	-	-	-
HCM	m	15,9	11:47367923	c.927-2A>G	-	0.000032	P	no	yes	no	yes	9562578

HTX - heart transplantation, individuals †≤18 years and ‡>18 years with PubMed ID of study; ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)

Table S 8: Phenotypes and genotypes of families with TNNI3 variants

			Echocardiogram										
ID	**Genotype**	**Age (years)**	**IVSD (mm)**	**LVPWd (mm)**	**LVEDd (mm)**	**LVFS (%)**	**LVEF (%)**	**LVEF plan. (%)**	**CMP, Description**	**NC/C ratio >2:1**	**12 lead ECG**	**Arrhyth-mias**	**Medical History**
Family 1													
1-I:1	heterozygous, *TNNI3* c.204delG, p.Arg69Alafs*8	33	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	no symptoms, no medications
1-I:2	heterozygous, *TNNI3* c.204delG, p.Arg69Alafs*8	29	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	no symptoms, no medications
1-II:1	homozygous, *TNNI3* c.204delG, p.Arg69Alafs*8	1.2	7 (+3.5)	6.9 Z((+5.9)	38 (+4.4)	11	26	25	**DCM**, severe MR and TR, RVSP 46 mmHg	no	SR, iRBB	no	symptomatic with CHD after vaccination; suspected myocarditis; implantation of LVAD at 1.2 y, HTX at 1.8 y
Family 2													
2-II:3	homozygous, *TNNI3* c.24+2T>A, p.(Ala8+GluArgAlaAlaGly)	1	7 (+4.4)	6 (+2.9)	33 (+5.2)	17	37	40	**LVNC**, LV hypertrophy with trabeculations and deep recesses in the apex and lateral wall	yes	SR, normal	no	histology from LV biopsy: deep recesses covered by endocardium highly suspicious of LVNC; see pedigree in Figure S2; death at 1.3 y after LVAD implantation and sepsis
Family 3													
3-I:1	no *TNNI3* variant	33	N/A	N/A	N/A	N/A	N/A	N/A	MRI: mild dilated LA, function and size of LV normal	N/A	SR, normal	no	Arterial hypertension, no symptoms, no medications
3-I:2	no *TNNI3* variant	38	9	N/A	44	35	64	62	normal	no	SR, normal	no	no symptoms, no medications
3-II:1	heterozygous, *TNNI3* c.624T>TT, p.Glu209*	4	4 (-0.4)	N/A	25 (-1.4)	37	69	77	**RCM**, biatrial enlargement, PHT, RVSP 50-80 mmHg	no	SR, biatrial abnormality	no	LVAD implantation at 4 y, HTX at 4.5 y, death at 5 y

Family 4													
4-I:1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	died from SCD at age 30 with pathology of HCM at autopsy
4-I:2	no *TNNI3* variant	39	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	SR, normal	N/A	chronic obstructive pulmonary disease, NYHA II
4-II:1	heterozygous, *TNNI3* c.570C>G, p.Asp190Glu	15	37 (+24)	16	29 (-3,6)	43	76	77	**HCM**, PHT	no	pacemaker: atrial stimulation, WPW	WPW, AV-reentry tachycardia	multiple family members with SCD< 40 y, LGE: septal and inferoseptal; ICD implantation at 11 y, NYHA II-III
Family 5													
5-I:1	no *TNNI3* variant	32	13	N/A	53	22	43	57	normal	no	discordant T-waves in III, slow R progression	2 PVCs	no symptoms, no medications
5-I:2	no *TNNI3* variant	30,4	8	N/A	50	38	68	N/A	normal	no	SR, normal	normal	no symptoms, no medication
5-II:1	heterozygous, *TNNI3* c.574C>T, p.Arg192Cys	2,7	4 (-0.3)	N/A	24 (-1.8)	25	52	44	**RCM**, severe biatrial enlargement, TI °II, RVSP 50 mmHg	no	SR, LVH, right atrial abnormality	no	Cardiac decompensation, postcapillary PHT, biventricular VAD implantation at 2.7 y, HTX at 3.4 y
Family 6													
6-II:1	heterozygous, *TNNI3* c.582C>A, p.Asn194Lys	9	5,5 (-0.3)	5,9 (0)	54,2 (+4.6)	18 (<28)	45 (>55)	30 (>55)	**DCM**, mild MR, LA slightly enlarged	no	SR, normal	no	no symptoms; twin brother died of acute heart failure and DCM at 9 y

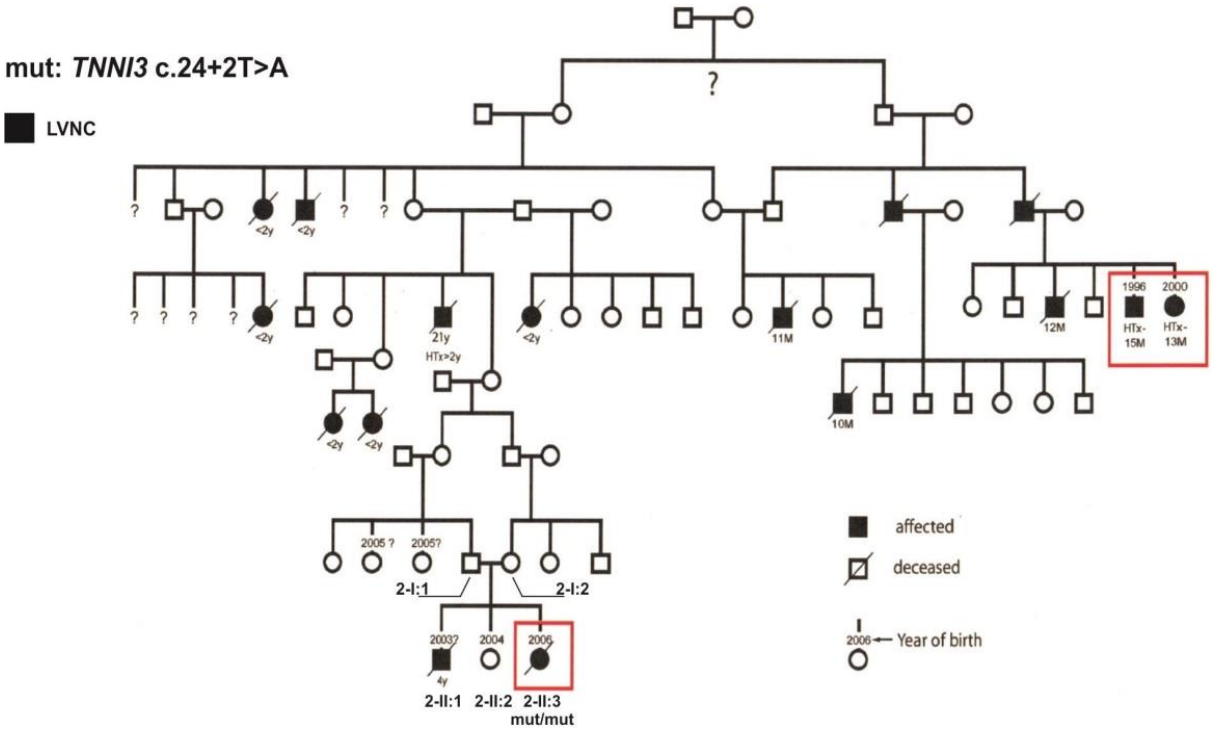


Figure S 2: Extended pedigree of family 2. Genotypes are shown as wt or mut, representing the wildtype or altered (mutated) allele. For individuals with no genotype no samples for sequencing were available. Phenotypes and the variants on coding sequence and protein level are indicated. Males are represented with squares and females with circles. Affected individuals have a filled form and deceased family members are marked with a diagonal line

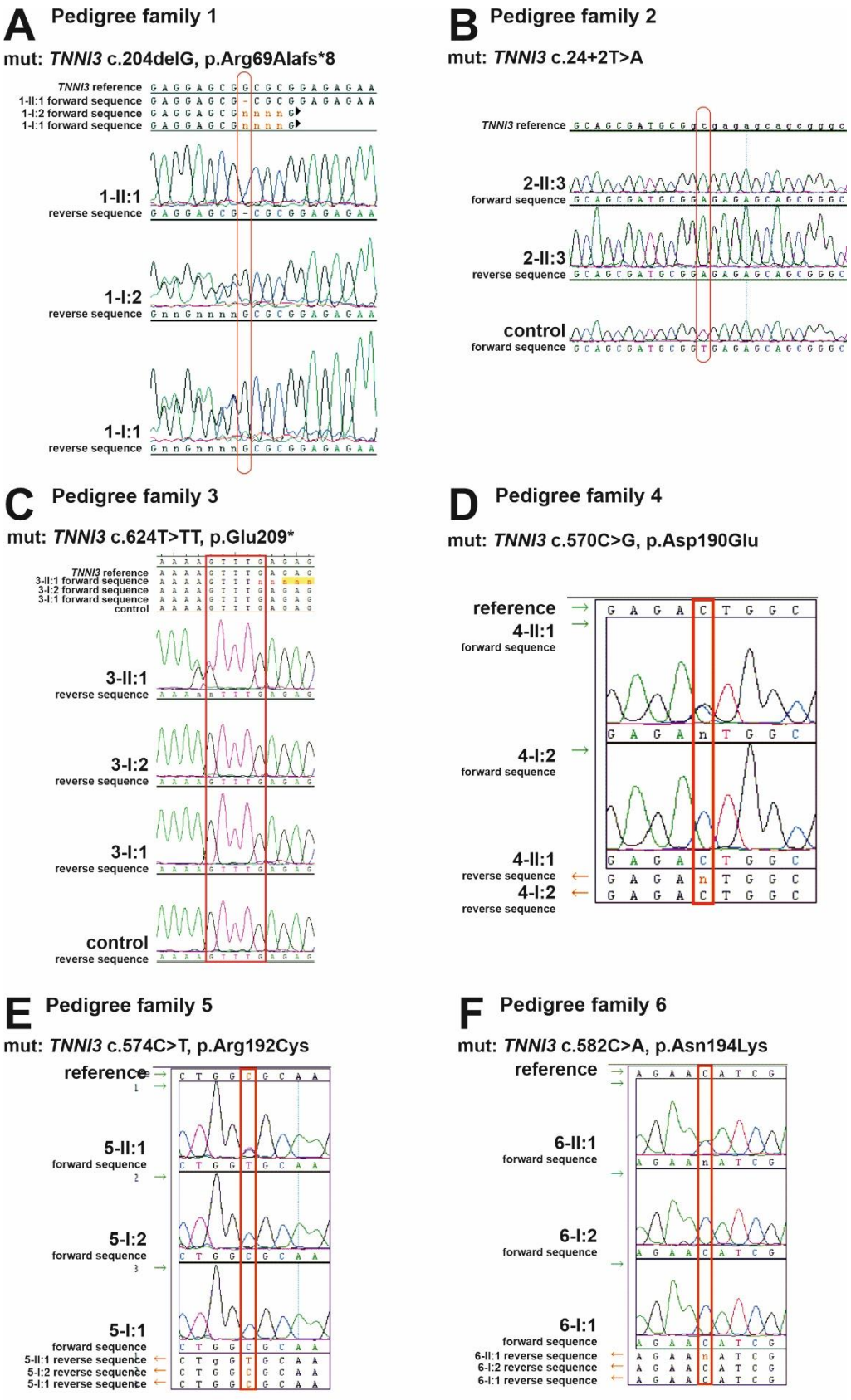


Figure S 3: Sequences of index patients and family members carrying a TNNI3 variant. Shown are the DNA sections from sequencing variants in all available family members of patients carrying a P or LP TNNI3 VOI.

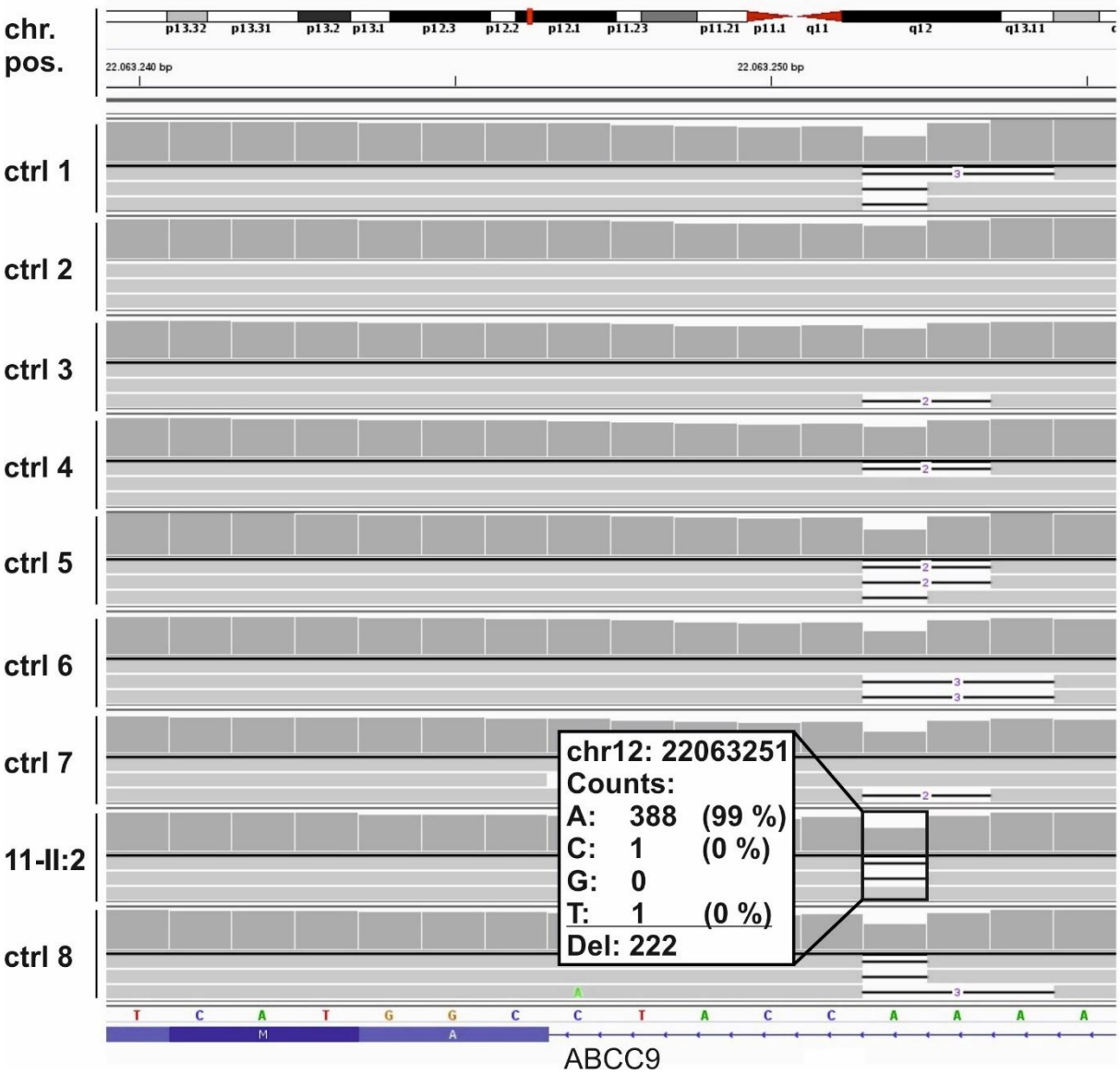


Figure S 4: Sequences from controls and a CMP-affected individual of NGS-generated data analyzed with the program IGV for a variant detected in ABCC9.

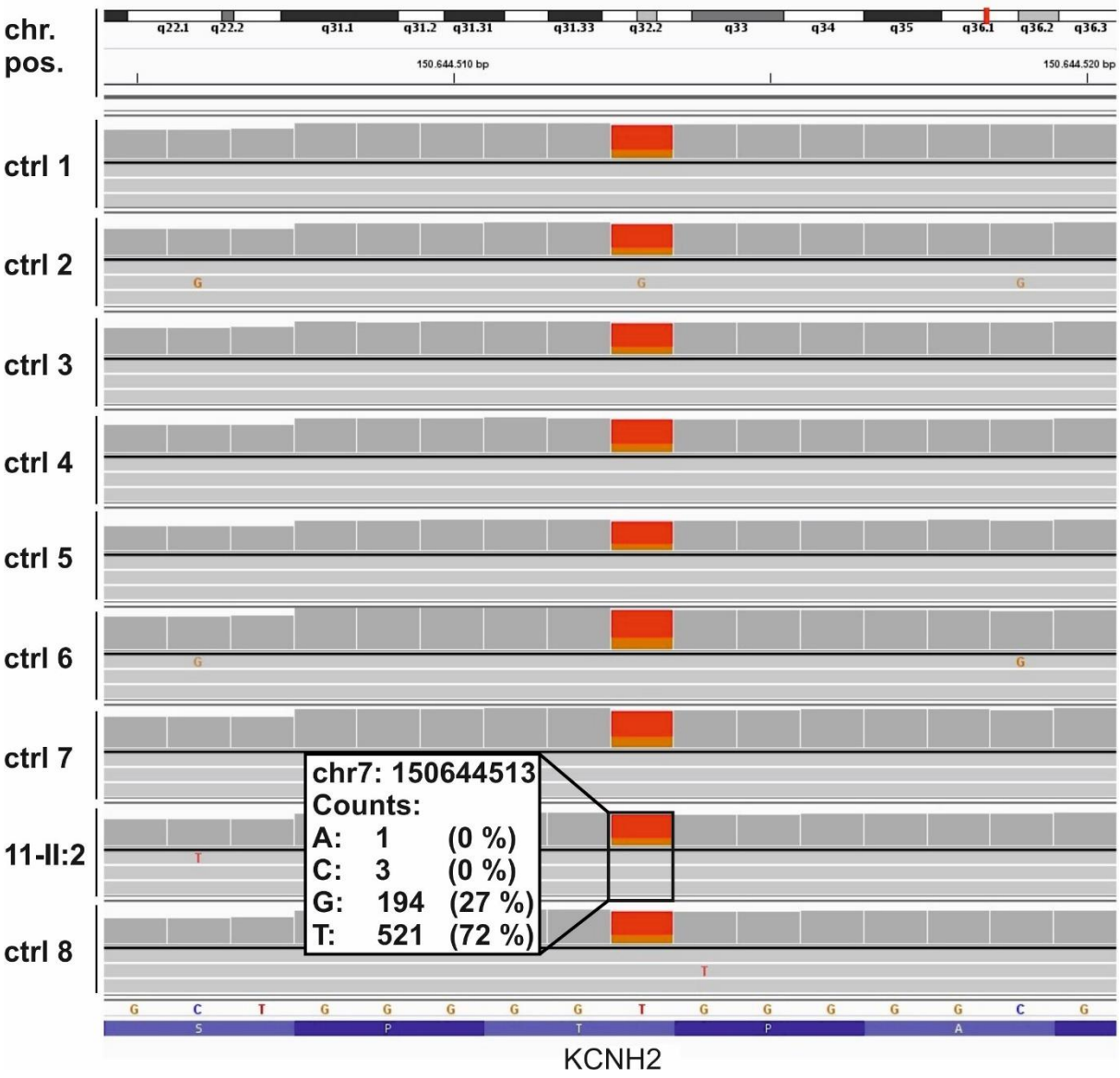


Figure S 5: Sequences from controls and a CMP-affected individual of NGS-generated data analyzed with the program IGV for a variant detected in KCNH2.

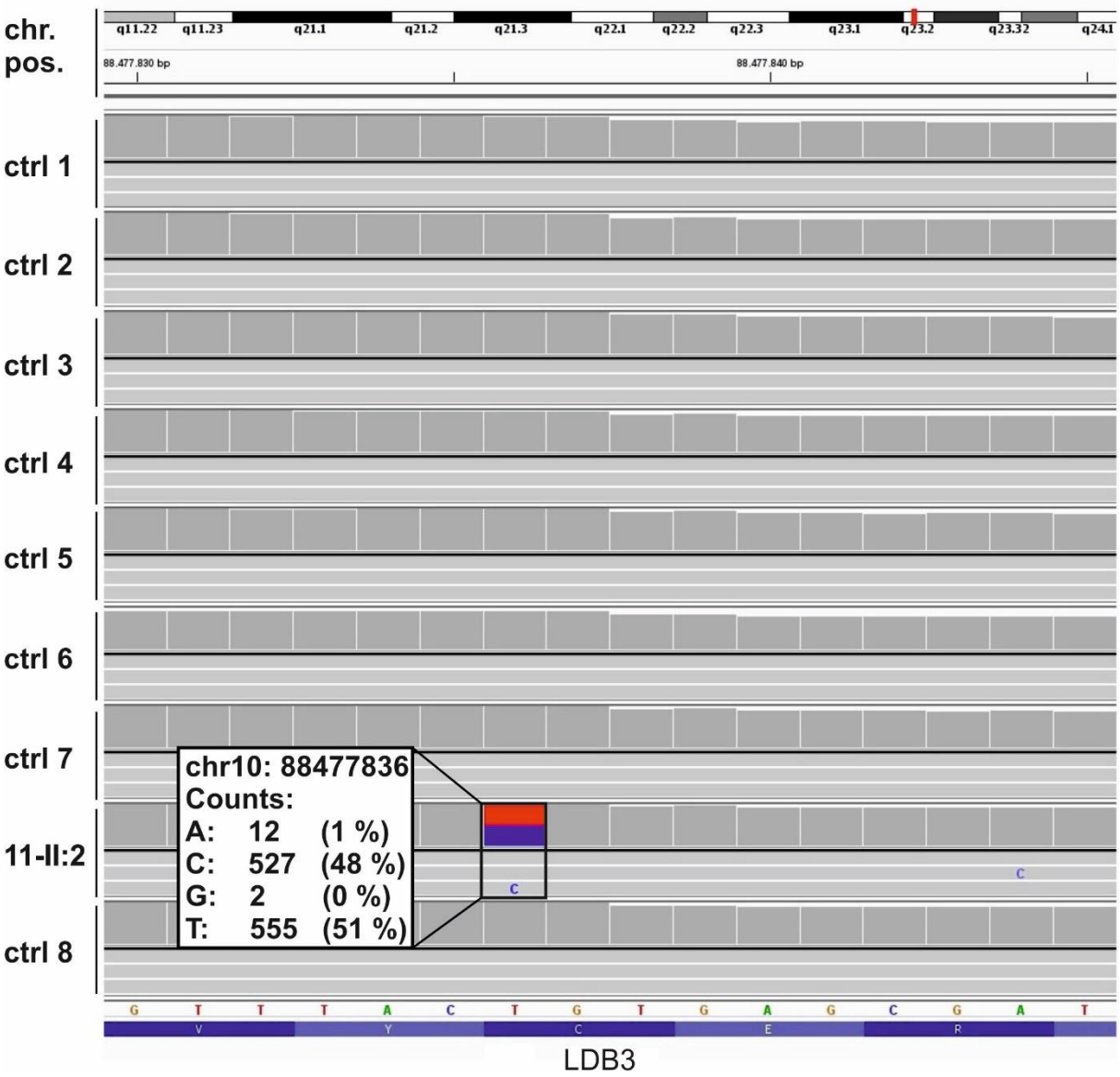


Figure S 6: Sequences from controls and a CMP-affected individual of NGS-generated data analyzed with the program IGV for a variant detected in LDB3.

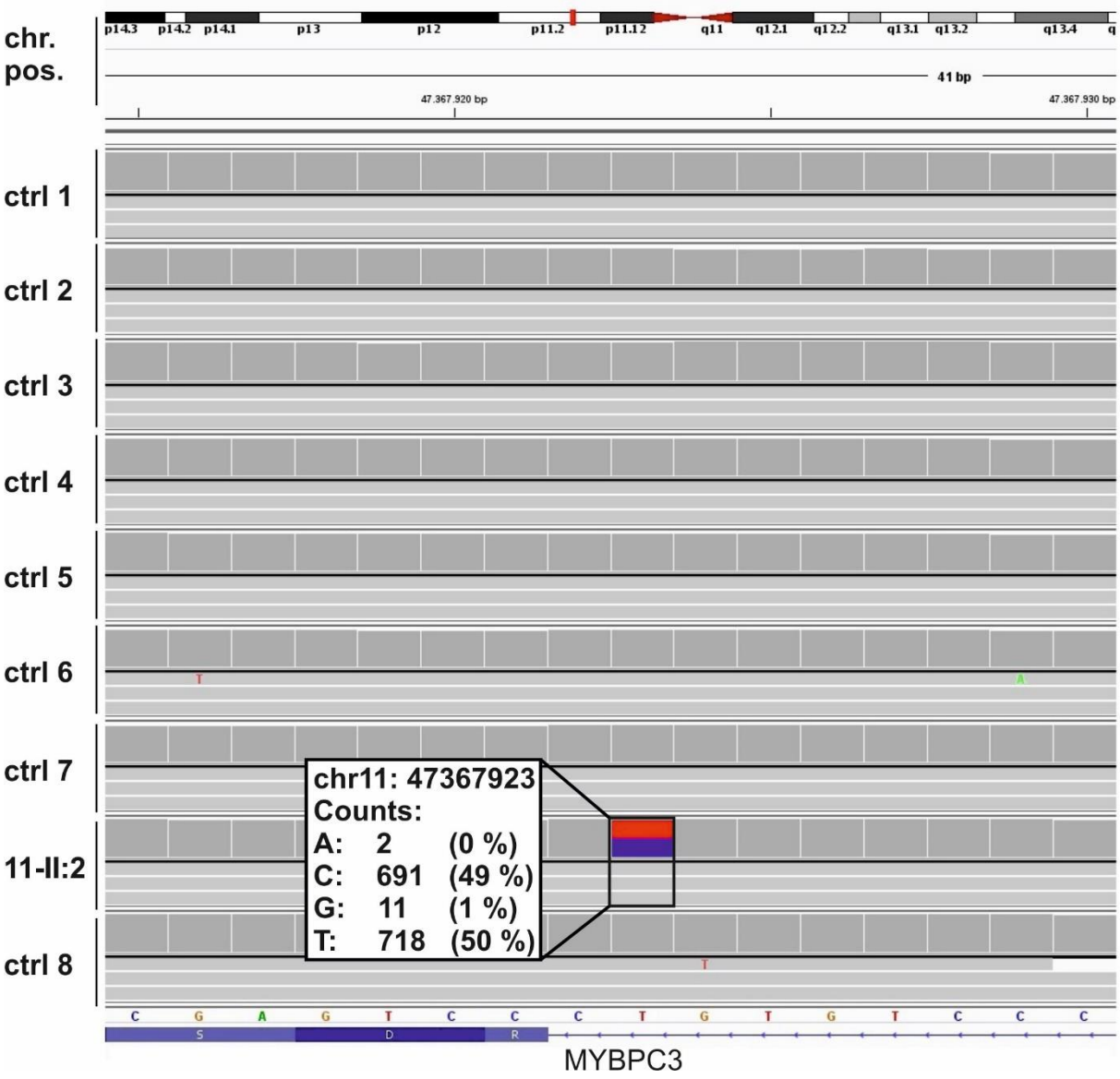


Figure S 7: Sequences from controls and a CMP-affected individual of NGS-generated data analyzed with the program IGV for a variant detected in MYBPC3.

List of Publications

Al-Wakeel-Marquard, N., Degener, F., Herbst, C., Kühnisch, J., Dartsch, J., Schmitt, B., Kuehne, T., Messroghli, D., Berger, F. and Klaassen, S. (2019). "RIKADA Study Reveals Risk Factors in Pediatric Primary Cardiomyopathy." J Am Heart Assoc **8**(15): e012531.

Kühnisch*, J., Herbst*, C., Al-Wakeel-Marquard, N., Dartsch, J., Holtgrewe, M., Baban, A., Mearini, G., Hardt, J., Kolokotronis, K., Gerull, B., Carrier, L., Beule, D., Schubert, S., Messroghli, D., Degener, F., Berger, F. and Klaassen, S. (2019). "Targeted panel sequencing in pediatric primary cardiomyopathy supports a critical role of TNNI3." Clin Genet **96**(6): 549-559.

*Jirko Kühnisch and Christopher Herbst contributed equally to the work.

Statement of Authorship

I hereby declare that I am the sole author of this thesis with the title “Characterization of Genetic Causes of Pediatric Cardiomyopathies – Analysis of PRDM16 as a Major Disease Gene” and that I have not used any sources other than those listed in the references and identified as references. I further declare that I have not submitted this thesis at any other institution in order to obtain a degree.

Berlin, 20th of April 2020	Christopher Herbst
(Place, Date)	(Name)