UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Developing cellular models for hypertrophic cardiomyopathy

Marta de Jesus Ribeiro

Supervisor: Doctor Maria do Carmo Salazar Velez Roque da Fonseca

Co-supervisors: Doctor Sandra Cristina Bento Penisga Martins

Doctor Maria Margarida Fonseca Rodrigues Diogo

Thesis approved in public session to obtain the PhD Degree in

Bioengineering

Jury final classification: **Pass with Distinction**

2020

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Developing cellular models for hypertrophic cardiomyopathy

Marta de Jesus Ribeiro

Supervisor: Doctor Maria do Carmo Salazar Velez Roque da Fonseca

Co-supervisors: Doctor Sandra Cristina Bento Penisga Martins

Doctor Maria Margarida Fonseca Rodrigues Diogo

Thesis approved in public session to obtain the PhD Degree in

Bioengineering

Jury final classification: **Pass with Distinction**

Jury:

Chairperson: Doctor Joaquim Manuel Sampaio Cabral, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Maria do Carmo Salazar Velez Roque da Fonseca, Instituto de Biologia Molecular, Faculdade de Medicina, Universidade de Lisboa;

Doctor Joaquim Manuel Sampaio Cabral, Instituto Superior Técnico, Universidade de Lisboa;

Doctor Lino da Silva Ferreira, Faculdade de Medicina, Universidade de Coimbra;

Doctor José António Henriques de Conde Belo, NOVA MEDICAL SCHOOL/ Faculdade de Ciências Médicas, Universidade NOVA de Lisboa;

Doctor Maria Teresa Tenório Figueiredo Carvalho Gonçalves, Instituto de Biologia Molecular, Faculdade de Medicina, Universidade de Lisboa;

Doctor Carlos André Vitorino Rodrigues, Instituto Superior Técnico, Universidade de Lisboa

Funding Institution:

Fundação para a Ciência e a Tecnologia - FCT (SFRH/BD/114054/2015)

2020

RESUMO

A miocardiopatia hipertrófica (MCH) é a doença cardiovascular hereditária mais comum sendo definida por um aumento da espessura do ventrículo esquerdo, não explicado por fatores hemodinâmicos. As suas principais características histopatológicas incluem um aumento de tamanho dos cardiomiócitos, desorganização dos sarcómeros e aumento da fibrose intersticial. Trata-se de uma doença complexa, devido à sua heterogeneidade de apresentação clínica e grande número de mutações causais associadas. Contudo, frequentemente, os testes genéticos não possibilitam a identificação de uma mutação causal e, neste momento, as terapêuticas utilizadas para a MHC procuram o alívio de sintomas e a prevenção de complicações graves, não havendo ainda uma terapêutica farmacológica especifica da doença, que previna ou reverta o seu fenótipo característico. São, portanto, necessárias melhorias não só ao nível do diagnóstico, mas também no desenvolvimento de novas terapêuticas dirigidas às causas moleculares da doença. Para tal é essencial que haja modelos adequados, que permitam o estudo da fisiopatologia da MHC bem como o ensaio de novas terapias dirigidas

Assim, o principal objetivo desta tese foi estabelecer uma nova abordagem experimental que permita a identificação e caracterização de novas variantes genéticas que contribuam para o fenótipo da MHC, investigar os mecanismos moleculares subjacentes às mesmas e testar possíveis abordagens terapêuticas. Primeiramente, fomos avaliar o desempenho de ferramentas de previsão computacional disponíveis e identificamos quais os algoritmos que mostram o melhor desempenho na identificação da patogenicidade de variantes *missense* ou *splice site* associadas a MHC; além disso, esta análise demonstrou que as ferramentas estudadas apresentam uma falha de precisão na previsão de variantes *deep-intronic*. De seguida, e por forma a gerar modelos celulares experimentais que pudessem ser utilizados para estudar o impacto de variantes associadas a MHC, aplicámos o sistema de edição génica CRISPR / Cas9 em células

estaminais pluripotentes induzidas humanas (hiPSCs) para criar uma linha celular isogénica portadora de uma variante associada a MHC. Por fim, desenvolvemos uma estratégia reprodutível e otimizada para gerar populações homogéneas de cardiomiócitos derivados de hiPSCs que fossem capazes de recapitular características celulares específicas da MHC.

O trabalho apresentado serve como base para uma melhoria no diagnóstico de MHC, facilitando não só a identificação de variantes causais e o estudo do seu impacto no fenótipo da doença, bem como o desenvolvimento e teste de novas abordagens terapêuticas.

Palavras-chave: miocardiopatia hipertrófica, células estaminais pluripotentes induzidas, diferenciação cardíaca, ferramentas bioinformáticas, edição génica

ABSTRACT

Hypertrophic cardiomyopathy (HCM) is the most common genetic disease of the heart. It is a complex disease characterized by left ventricular hypertrophy, its anatomical hallmark, abnormal diastolic function, increase in myocyte size with distorted nuclei, myocyte disarray and increased extracellular fibrosis. Several associated molecular mechanisms and a heterogeneous set of clinical manifestations had been reported, ranging from asymptomatic mutation carriers to severe heart failure or even sudden cardiac death as the first manifestation of the disease. Unmet needs in HCM include the need for improved diagnosis since standard genetic testing often fails to identify a causative mutation, as well the need for therapies precisely targeted at the molecular cause of the disease since current treatment options only aim to alleviate symptoms. To address these needs, good disease models that can help understand the pathophysiology of HCM and test new therapeutic approaches are required.

Therefore, the main objective of this thesis was to establish a new experimental approach that allows the identification and characterization of novel genetic variants that contribute to HCM phenotype, investigate the molecular mechanisms underlying it and test possible therapeutic approaches. Firstly, the performance of available computational prediction tools was benchmarked and the algorithms that show the best performance in identification of missense and splice site HCM variants were identified; moreover, this analysis showed lack of prediction power for deep intronic variants of such bioinformatics approaches. Afterwards, and to generate experimental cellular models that could be used to study the impact of HCM-associated variants, the CRISPR/Cas9 gene-editing system was applied in human induced pluripotent stem cells (hiPSCs) to create an isogenic cell line bearing an HCM-associated pathogenic variant. Lastly, a reproducible optimized strategy for generating homogeneous populations of hiPSC-derived

cardiomyocytes capable of recapitulating HCM-specific features at the single-cell level was established.

The presented work sets the foundation for an improvement in HCM diagnosis, by identifying and determining the impact of HCM-causing mutations, and the development and testing of new HCM therapeutic approaches.

Keywords: hypertrophic cardiomyopathy, human induced pluripotent stem cells, cardiac differentiation, *in silico* prediction tools, gene editing

ACKNOWLEDGEMENTS

Se vi mais longe foi porque estava aos ombros de gigantes, Isaac Newton

A todos os “gigantes”, que ao longo destes 4 anos me acompanharam neste caminho, e me ajudaram a desenvolver não apenas o trabalho do meu Doutoramento, mas também a minha personalidade e capacidades, o meu mais profundo obrigada, se agora termino esta etapa a vocês o devo.

Em primeiro lugar, agradeço à Professora Carmo Fonseca e ao Professor Joaquim Cabral pela oportunidade de poder desenvolver este trabalho nos respetivos grupos de investigação, tirando partido do melhor de dois “mundos”, que tão bem se complementam. Ao Professor Cabral agradeço toda a disponibilidade sempre demonstrada. À Professora Carmo agradeço a paixão contagiante pela ciência que nos faz querer ir mais além. Por ter acompanhado de perto todo o trabalho, bem como pelos ensinamentos e discussões científicas que permitiram encontrar caminhos alternativos e interpretar resultados sempre que necessário.

À professora Margarida agradeço a calma transmitida, a disponibilidade e os conselhos ao longo do desenvolvimento deste trabalho.

À Sandra e à Teresa por me terem recebido tão bem e ensinado tanto no laboratório. Pelas diferenças e complementaridades não podia pedir uma melhor equipa “sénior” ao meu lado. À Teresa agradeço a calma e perspetiva analítica sempre oferecidas, bem como a preciosa ajuda, sobretudo na obtenção de imagens fantásticas das “nossas meninas”. À Sandra agradeço a energia, dinamismo e constante disponibilidade para ajudar com qualquer experiência, mas também todo o apoio nesta fase final. A ambas um muito obrigada pela amizade e carinho, assim como pelo apoio que me facultaram durante estes anos, contribuindo, para o meu crescimento pessoal e profissional.

Agradeço também a todos os elementos do SCERG com quem tive o prazer de passar momentos no laboratório, as longas horas a cuidar de células tornam-se mais leves na vossa companhia. Em particular à Mariana e à Claúdia pelos ensinamentos, dicas e ajuda no geral.

Aos colegas e amigos do iMM, os que ainda lá estão e os que a vida lhes mudou o rumo, obrigada pelo suporte ao longo desta fase, tornaram-na sem dúvida mais fácil e alegre.

Aos mais jovens que passaram pela “equipa cardio” agradeço a ajuda no laboratório, mas sobretudo a boa companhia e ouvirem os meus “devaneios”. À Marta Furtado, em particular, obrigada por toda a ajuda do último ano seja no trabalho de laboratório, nas horas tardias de

trabalho de escritório, pelas conversas e por me aturares no geral. À minha “nova sénior”, Patrícia, obrigada pelas conversas, conselhos e por estares sempre disponível para ajudar.

À Noélia, ao Simão e à Geni agradeço por todas as explicações científicas, dicas no laboratório e por terem sempre uma palavra amiga a oferecer.

À Dinora, agradeço todo o apoio técnico, mas também a amizade, a companhia ao almoço, que o torna sempre um momento mais descontraído, e todas as nossas conversas.

À Joana Tavares agradeço a amizade, mas também o ter sido a minha mentora na fase inicial do projeto e ajudar-me a perceber um pouco de área tão distante da que até então tinha trabalhado. Agradecimento que estendo ao Pedro Barbosa por estar sempre disponível para me esclarecer dúvidas de bioinformática que vão surgindo.

Ao Kenny, Ana, Pedro Prudêncio, Rui, João, Maria, Mariana, Beatriz, obrigado pelo companheirismo, boa disposição e incentivo a fazer algum desporto (!!!). À Teresa Silva e Rita Drago obrigado por todo o carinho e apoio, ter-vos como exemplo ajudou nesta jornada.

A todos os amigos que fora do mundo científico e estando mais próximos ou distantes (fisicamente) sempre estiveram presentes. À Teresa obrigada por todas as conversas, viagens e aventuras gastronómicas, no fundo, obrigada por estares sempre aí por mais avessas e desavessas que passemos. Ao Pedro e á Inês obrigada pelo equilíbrio, falar e estar com vocês é sempre, além de divertido, tranquilizador (sim, mesmo com Pandora, Laura, Cho,…). Ao Cotrim, Paulo, João Paulo e afins, obrigada pelos bons momentos á frente de um tabuleiro. À Joana, Petya, Ricardo, Andreia, Daniel, Elisa, Pedro, Melissa, David e Inês obrigada por tornarem, ou terem tornado, a vida em Lisboa melhor. As tardes e jantares na vossa companhia sempre foram “lufadas de ar fresco” (Inês e David, São Pedro é já ali e tem praia, pinhal e bons restaurantes por perto!).

Por fim, um agradecimento muito especial ao Alcides por acreditar nas minhas capacidades, por todo o carinho e abraços de incentivo. Obrigada por estares presente nos momentos bons e menos bons, ao longo destes anos e sempre pronto a fazer-me sorrir. Não podia ter pedido melhor companhia para esta recta final de confinamento, com duplo motivo.

E à minha família, especialmente aos meus pais, por me apoiarem durante este percurso tal como em todas as outras etapas da vida. Obrigada pela transmissão de valores importantes, pela constante valorização do meu potencial, mesmo nos momentos mais difíceis e, por estarem sempre ao meu lado.

E no fim nem sempre o destino é o mais importante, mas sim o caminho percorrido até lá chegar.

“It was the best of times, it was the worst of times,

it was the age of wisdom, it was the age of foolishness,

it was the epoch of belief, it was the epoch of incredulity,

it was the season of light, it was the season of darkness,

it was the spring of hope, it was the winter of despair.”

— Charles Dickens, A Tale of Two Cities

“As for the future, your task is not to foresee it, but to enable it.”

— Antoine de Saint-Exupéry

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ACTC1	Cardiac muscle α-actin
ACTN2	α-Actinin 2
BMP4	Bone morphogenetic protein 4
cDNA	Complementary deoxyribonucleic acid
CMs	Cardiomyocytes
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	Small CRISPR RNAs
DAPI	4’,6-diamidino-2-phenylindole
DCM	Dilated Cardiomyopathy
DSB	Double-strand breaks
EBs	Embryoid bodies
EDTA	Ethylenediamine tetraacetic acid
EJC	Exon junction complex
ESCs	Embryonic stem cells
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
ExAc	Exome Aggregation Consortium
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
GiWi	Gsk3 inhibitor and Wnt inhibitor
GFP	Green fluorescent protein

GSK3	Glycogen synthase kinase-3
HCM	Hypertrophic Cardiomyopathy
HDR	Homologous Directed Recombination
HSF	Human Splicing Finder
iPSCs	Induced pluripotent stem cells
iPSC-CM	Induced pluripotent stem cell-derived cardiomyocytes
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
Klf4	Krüppel-like factor 4
LV	Left ventricle
MYBPC3	Myosin-binding protein C 3
MYH6	Myosin heavy chain 7
MYH7	Myosin heavy chain 7
MYL2	Myosin light chain 2
MYL3	Myosin light chain 3
NHEJ	Non-Homologous End Joining
NMD	Nonsense-mediated decay
OCT4	Octamer-binding transcription factor 4
ORF	Open reading frame
PAM	Protospacer adjacent motif
PBS	Phosphate Buffered Saline
PFA	Paraformaldehyde
PSCs	Pluripotent stem cells
PTC	Premature stop codon
ROCKi	Rho-associated coiled protein kinase inhibitor
RT	Room temperature

SCD	Sudden cardiac death
SIRPA	Signal Regulatory Protein Alpha
SNP	Single nucleotide polymorphisms
sgRNA	Single guide RNA
Sox2	Sex determining region Y-box 2
ssODN	Single-stranded oligonucleotide
TALENs	Transcription activator-like effector nucleases
TGFβ	Transforming growth factor-β
TNNI1	Troponin I 1
TNNI3	Troponin I 3
TNNT2	Troponin T 2
TPM1	Tropomyosin 1
TTN	Titin
tracrRNA	Transactivating crRNA
VCAM-1	Vascular cell adhesion molecule 1
VUS	Variant of uncertain significance
ZFNs	Zinc-finger nucleases

I. GENERAL INTRODUCTION

I.1. THESIS OUTLINE

The central topic of this thesis is the development of a strategy based on cell bioengineering that allows the identification and characterization of novel genetic variants that contribute to the HCM phenotype, to investigate the molecular mechanisms underlying it and test possible therapeutic approaches.

In chapter II we evaluate the performance of several available computational tools in predicting the potential pathogenicity of genetic variants in different contexts by using three datasets, including a specific and curated dataset of HCM associated variants. The assessed tools shown that the combined use of several of them may be needed to correctly assign variant pathogenicity, mainly depending on the variant type (missense or splice-site), and that new approaches for the evaluation of deep intronic variants need to be developed.

In chapter III, we perform gene editing in hiPSCs using a CRISPR/Cas9 based system to establish a cell line bearing a well described HCM associated pathogenic variant. The edited hiPSC line and its isogenic control, upon cardiomyocyte differentiation, would allow testing for a reliable disease model of HCM that can recapitulate the disease phenotype at the cellular level. The challenges and potential approaches for precise introduction of a single nucleotide variant through gene-editing in hiPSCs are further discussed.

Lastly, in chapter IV, we describe an optimized strategy that uses a pre-established GiWi 3D differentiation protocol combined with a purification step by FACS for VCAM1 expressing hiPSC-CMs and further maturation of this cardiomyocytes in a monolayer culture until day 30 of differentiation. The generated hiPSC-CMs represented homogeneous populations of cardiomyocytes with a high degree of maturation that, when used to differentiate HCM patient derived hiPSCs, could recapitulate HCM-specific features at the single-cell level, as demonstrated by the increased cellular size, multinucleation and disorganized sarcomeres in HCM hiPSC-CMs as compared to normal hiPSC-CMs.

I.2. HYPERTROPHIC CARDIOMYOPATHY

Hypertrophic cardiomyopathy (HCM) is a primary disease of the myocardium and the most common genetic cardiovascular disorder. It has an estimated prevalence ranging between 1:200 and 1:500 in the general population affecting people of different ethnicities and races spread by all regions of the 6 continents (Maron, Rowin, and Maron 2018; Semsarian et al. 2015; Maron et al. 1995) and is well known for causing sudden cardiac death (SCD) in young athletes (Maron 2010).

The clinical hallmark of the disease is the increase in heart mass and the left ventricular hypertrophy in the absence of abnormal loading conditions or secondary causes such as other cardiac, metabolic or syndromic diseases namely systemic hypertension or aortic stenosis. Such hypertrophy of the left ventricle leads to decreased chamber volume, resulting in left ventricle outflow obstruction and abnormal diastolic function (Christiaans and Elliott 2016; Viswanathan et al. 2017; Maron 2018; Maron and Maron 2013; P. M. Elliott et al. 2014). Histopathologic manifestations include an increase in cardiomyocytes size (which can be 20–30μm in diameter, in contrast to a diameter of 5–12μm in normal cells) with distorted nuclei, myocyte disarray and increased extracellular fibrosis (Watkins, Ashrafian, and Redwood 2011; Christiaans and Elliott 2016; Tejado and Jou 2018) (Figure I.1).

HCM may present a large spectrum of potential outcomes ranging from asymptomatic with survival to normal life expectancy to heart failure and SCD (Sedaghat-Hamedani et al. 2018; Melacini et al. 2010). The majority of individuals affected by this condition are thought to remain undiagnosed, with approximately two-thirds of patients with HCM being asymptomatic or minimally symptomatic (Klarich et al. 2013; Maron et al. 2014; 2015). Mild manifestations of the disease include symptoms such as chest pain, dyspnea, dizziness and fatigue (Viswanathan et al. 2017; Geske, Ommen, and Gersh 2018). Nevertheless, in up to 25% of HCM-affected individuals, these symptoms can be relentlessly progressive resulting in severe health

complications such as arrhythmic sudden death, end-stage heart failure and atrial fibrillation with embolic stroke (Roma-Rodrigues and Fernandes 2014; Frey, Luedde, and Katus 2012; Ali J. Marian and Braunwald 2017b).

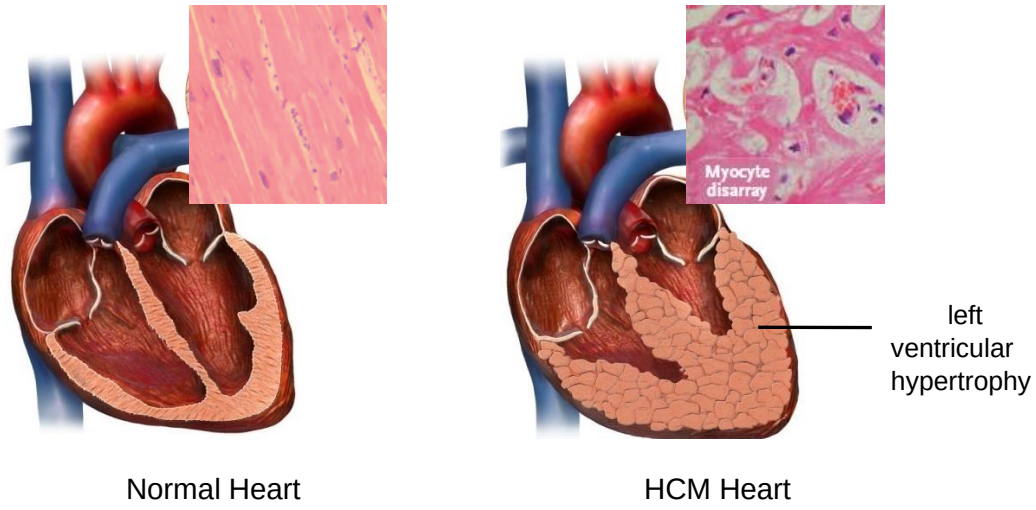


Figure I.1. Schematic representation of a normal heart and a heart with HCM, showing marked hypertrophy of the left ventricle wall. Accompanied with hematoxylin-eosin stained myocardial section images with correspondent histological phenotypes. Adapted from "http://sayostudio.com/health-infographic-art" and (Watkins, Ashrafian, and Redwood 2011).

The heterogeneous phenotype and complex clinical profile of the disease poses a problem to its diagnosis, which is primarily based on a left ventricular wall thickness >15 mm in adults, determined by an echocardiogram or other imaging techniques, such as computed tomography or cardiac magnetic resonance, in the absence of a known secondary cause (Maron, Maron, and Semsarian 2012; Geske, Ommen, and Gersh 2018; P. M. Elliott et al. 2014). Often, such diagnosis is only possible during adulthood, when there is a marked progression of the morphological and functional deficits of the heart (Roma-Rodrigues and Fernandes 2014).

Current treatment options for HCM aim primarily for the reduction of the symptoms, according to their severity and risk for SCD, but do not target the pathophysiology or genetic cause of the disease (Tardiff et al. 2015). Regarding pharmacological approaches, the beta-adrenergic receptor blockers are the most used, followed by calcium channel blockers and disopyramide as

an antiarrhythmic agent. Patients that remain symptomatic despite pharmacological treatment are often referred for surgical strategies such as implantable defibrillators, in order to prevent sudden death, surgical myectomy or percutaneous alcohol septal ablation, for the reduction of the left ventricle outflow obstruction and, in patients with advanced heart failure, implantation of a left ventricular device or cardiac transplantation may be the only option (Christiaans and Elliott 2016; Geske, Ommen, and Gersh 2018; P. M. Elliott et al. 2014).

Patients with a clinical diagnosis of HCM are encouraged to undergo genetic counselling in order to search for a disease-causing variant that would support the diagnosis in a proband. When such pathogenic variant is identified, a cascade screening of the family members is also recommended to identify high-risk individuals whose symptoms are not yet apparent and reassure those with a negative test (Charron et al. 2010). Genetic analysis is also useful to confirm the diagnosis of HCM in ambiguous situations. For example, because HCM is a common cause of SCD in young athletes, who often exhibit left ventricular wall thickness between 13 and 18 mm, it is critical to distinguish the physiological hypertrophy of athletes from the pathological hypertrophy of HCM (Grazioli et al. 2016). On the other hand, a positive test supports the HCM diagnosis in individuals with the septal thickness below the cut point for the clinical diagnosis (Niimura et al. 1998; 2002). Another important contribution of genetic testing is that it allows distinction between HCM and the so-called phenocopy conditions. Namely, autosomal dominant cardiomyopathies caused by *PRKAG2* mutations, X-linked cardiomyopathies such as Fabry's disease and Danon's disease, glycogen storage diseases, lysosomal storage diseases, mitochondrial diseases, which share clinical features with sarcomeric HCM yet are distinct disorders with different natural history and treatment (Ali J. Marian and Braunwald 2017a).

To date, more than 1500 HCM-associated mutations have been reported; however, due to the rarity of individual mutations, the available clinical data are insufficient to support meaningful genotype-phenotype correlations. Thus, for the majority of HCM patients, a positive genetic test is unable to predict the clinical course of the disease or the risk of complications including SCD

and heart failure (Ali J. Marian and Braunwald 2017a). Nevertheless, young age at diagnosis and the presence of a sarcomeric mutation were described as powerful predictors of adverse outcomes (Ho et al. 2018).

In fact, still, a major limitation of current genetic testing is its failure to identify a causative mutation in 50-60% of HCM patients (Gruner et al. 2013). Initially it was thought that patients with a negative test had mutations in genes not yet associated with HCM. Although the full spectrum of HCM genes was expected to be quickly determined using high-throughput DNA sequencing approaches, which allow whole-exome and whole genome analysis, the accumulated results did not improve the effectiveness of HCM gene-based diagnosis. Rather, access to high-throughput DNA sequencing data has vastly expanded the number of variants of uncertain significance (VUS), and currently available *in silico* tools are not always useful to distinguish between benign and deleterious variants leading to many inconclusive results that limit the clinical utility of genetic testing (Sabater-Molina et al. 2018; Walsh, Thomson, et al. 2017). Another issue in this lack of molecular diagnosis is the presence of disease causing variants in poorly studied regions, such as deep within the introns, for which few *in silico* prediction tools are also available (Ward and Kellis 2012).

The insufficiency of clinical diagnosis in cases such as the overlapping between an HCM phenotype and the physiological thickening of the left-ventricular wall due to high physical activity, or in identification of family members at risk, supports the need for a differential diagnosis (Christiaans and Elliott 2016; Viswanathan et al. 2017; Maron 2010). On the other hand, the often-inconclusive results of genetic testing stress the need for better tools to correctly classify the increasing number of HCM-associated variants found by large scale sequencing approaches. Moreover, the relentlessly progressive symptoms that can result in severe health complications underscore the importance of the development of therapeutic strategies that precisely modify HCM pathophysiology to slow, halt, or reverse disease progression (Roma-Rodrigues and Fernandes 2014; Frey, Luedde, and Katus 2012; Ali J. Marian and Braunwald 2017b). These

needs highlight the importance of having a good model to study both genotype-phenotype relations that can aid differential diagnosis, and to test new targeted therapies. To encourage research in this field, HCM has recently been given orphan disease status (Heitner 2018).

Genetics of Hypertrophic Cardiomyopathy

HCM is mostly inherited in an autosomal dominant manner with incomplete penetrance and high phenotypic variability, even within the same family (Greaves et al. 1987; Watkins, Ashrafian, and Redwood 2011). Some sporadic cases of *de novo* mutations and rare cases of X-linked and autosomal recessive modes of inheritance have also been reported (Christiaans and Elliott 2016; Hartmannova et al. 2013; Paranal et al. 2020).

Despite being an inherited cardiomyopathy, the underlying genetic cause of HCM is only found in less than 50% of patients (Gruner et al. 2013). The proportion of mutation carriers with clinically detectable disease tends to increase with age despite of in most persons the hypertrophy starts to manifested in adolescence (Watkins, Ashrafian, and Redwood 2011). However, due to the incomplete penetrance, patients with subtle HCM features are difficult to recognize in the general population and the median age at diagnosis is around 46 years old (Ho et al. 2018).

HCM mutations occur predominantly in genes that encode structural and functional components of the sarcomere, the contractile unit of the cardiomyocytes (P. M. Elliott et al. 2014; Van Driest et al. 2005; Morita et al. 2008; Brito et al. 2012; Kassem et al. 2013; Lopes et al. 2013).

Even though the variable phenotype and clinical outcomes among HCM patients may be associated to the different causative mutations, the heterogeneity between carriers of the same mutation within or between families suggests that the genomic context and environmental modifiers also play an important role (A J Marian 2002; Jodie Ingles et al. 2017). Non-genetic factors, such as hypertension and obesity are reportedly associated to a more severe phenotype (Christiaans and Elliott 2016; Nagueh and Zoghbi 2015). It has also been suggested that the

clinical outcome of HCM is likely to be the sum of multiple genetic mutations, with a subset of HCM patients (~5%) presenting two or more mutations, either in the same gene or in different genes, and these compound or double heterozygosity genotypes being associated with a more severe phenotype (Girolami et al. 2010; J. Ingles et al. 2005; Wessels et al. 2015). In this regard it is poorly understood how multiple mechanisms may be contributing to the heterogeneity of HCM phenotype, including the transcriptional and splicing profile, epigenetic modifications, and protein posttranslational modifications (Roma-Rodrigues et al. 2015; Jellis and Desai 2015).

Given the heterogeneous and complex phenotypes of HCM, a genotype–phenotype correlation has not been well established yet and poses a problem for risk evaluation and preventive care of HCM patients (Maron 2018). Nevertheless, the pathogenicity of a variant will depend on which gene is affected, the site of the mutation in the gene, i.e. if the variant is in a conserved region or in a location that is involved in protein-protein interactions, and the relative amount of wild-type and mutated protein incorporated in the sarcomeric structure (Roma-Rodrigues and Fernandes 2014).

Regarding the gene-disease association, detailed family linkage studies, many of which supported by functional and segregation data, indicate that at least 8 genes, all coding for components of the sarcomere, are associated with HCM, namely myosin-binding protein C (*MYBPC3)*, β-myosin heavy chain (*MYH7*), cardiac troponin T (*TNNT2*), cardiac troponin I (*TNNI3*), α-tropomyosin (*TPM1),* α-actin (*ACTC1*), myosin light chain 2 (*MYL2*), and myosin light chain 3 (*MYL3*). (Walsh, Thomson, et al. 2017; Walsh, Buchan, et al. 2017; Jodie Ingles et al. 2019) (Figure I.2). Among the sarcomeric genes mutated in patients, *MYH7* and *MYBPC3* account for more than half of HCM-causing mutations, while in *TNNT2*, *TNNI3*, *TPM1*, *ACTN1*, *MYL2* and *MYL3* genes the frequency ranges from 1-5% (Christiaans and Elliott 2016; Walsh, Thomson, et al. 2017; Walsh, Buchan, et al. 2017).

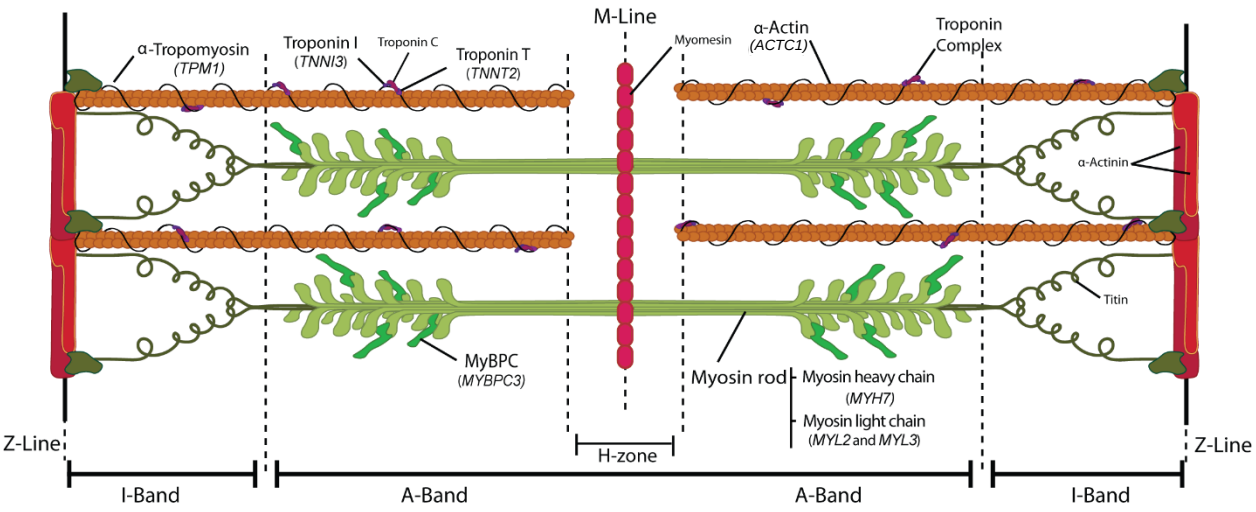


Figure I.2. Molecular structure of the human sarcomere displaying the localization of cardiac sarcomere proteins and genes found to be associated with HCM (*MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3* and *ACTC1*).

Recent studies pointed for distinct molecular mechanisms underlying HCM concerning the two most frequently HCM mutated genes. The vast majority of *MYH7* mutations, as well as most other sarcomere mutations, are classified as missense leading to expression of mutant proteins. This mutant proteins may present a gain of function (either by increasing their activity or lengthen their functional lifespan) or dominant negative properties (causing a decrease in the protein's activity) over the wild-type ones produced by the non-mutated allele (Walsh, Thomson, et al. 2017). In the specific case of *MYH7* mutations, biophysical properties of mutated myosins indicate a gain of function, with enhanced myosin ATPase activity, increased generated force and accelerated actin filament sliding (Tyska et al. 2000). In contrast, most *MYBPC3* mutations are classified as nonsense, frameshift or splice site mutations, and predicted to generate mRNA containing premature termination codons (PTC) that might be degraded by nonsense mediated decay (NMD) causing an haploinsufficiency phenotype (reduced amount of protein due to inactivation of expression of the mutant allele) (Glazier, Thompson, and Day 2019). Indeed, several studies shown that HCM patients with *MYBPC3* gene mutations have an absence of

truncated proteins and reduced levels of the normal one, leading to altered cardiac function (Marston et al. 2009; van Dijk et al. 2009a; Strande 2015; Prondzynski et al. 2017a).

On the other hand, allelic imbalance has also been reported for many missense HCM-related mutations in *MYH7*, *MYBPC3*, *TNNT2* and *MYL2* genes. Allelic imbalance consists of a differential expression of both alleles, at the mRNA or protein level, which can be either for higher or lower expression of the mutant alleles relative to the wild type. This differential expression can be related to different features. Epigenetic modifications such as imprinting or differential methylation combined with sequence variants in regulatory elements, may influence the availability of the DNA to the transcription machinery. Interestingly, studies analyzing heterozygous *MYBPC3* truncating variants show that wild-type *MYBPC3* is expressed at levels higher than 50%, suggesting the existence of a compensation mechanism (Glazier, Thompson, and Day 2019). This compensation from the normal allele is thought to be an attempt of maintaining the normal stoichiometry of the sarcomeric proteins. Despite that, such partial allelic compensation may not be enough to completely overcome the lack of protein being translated, leading to haploinsufficiency (Christiaans and Elliott 2016).

I.3. RNA MIS-SPLICING IN HCM

As it was already discussed, the introduction of high-throughput DNA sequencing approaches for clinical diagnostic has vastly expanded the number of VUS, leading to many inconclusive results that limit the clinical utility of genetic testing. Moreover, the presence of disease-causing variants in poorly studied regions, such as deep within the introns, also contributing to the lack of molecular diagnosis. Recently, developments in RNA analysis are improving the diagnostic outcome by identifying new variants that interfere with splicing; also, and besides improving the precision of genetic diagnostics, the discovery of disease-causing aberrantly spliced mRNAs in HCM patients opens new venues for the development of RNA-targeted therapies. Considering this, we reviewed the current knowledge on how mutations that interfere with splicing cause disease, the recent discoveries of RNA mis-splicing in HCM and discuss emerging strategies for HCM-targeted RNA therapeutics.

The content of the current section was adapted from the publication:

Marta Ribeiro, Marta Furtado, Sandra Martins, Teresa Carvalho, Maria Carmo-Fonseca, “RNA splicing defects in hypertrophic cardiomyopathy: Implications for diagnosis and therapy“, doi: 10.3390/ijms21041329

Disease-causing splicing mutations

There are diverse mechanisms by which mutation-induced defects in RNA splicing act as a primary cause of disease (Faustino and Cooper 2003; Abramowicz and Gos 2018; Scotti and Swanson 2016).

One category of splicing mutations includes single nucleotide substitutions that disrupt constitutive or alternative splice sites. The two highly conserved positions at the 5' and 3' splice

junction (typically GT and AG, respectively) are called the “essential” or “canonical” splice site nucleotides and substitutions at these positions in haploinsufficient autosomal dominant genes are diagnostically classified as pathogenic. However, recognition of a splice site by the spliceosome relies on a larger “splice region” composed of less well-conserved sequences. Although several studies highlight the importance of the near-splice-site region (Rivas et al. 2015; Lek et al. 2016; S. Zhang et al. 2018), variants in this region are often diagnostically classified as variants of unknown significance (VUS).

Splice site mutations result in either exon skipping, activation of a nearby cryptic or alternative 3' or 5' splice site, or intron retention (Figure I.3.A). Use of unnatural splice sites or intron retention frequently changes the reading frame for translation and introduces premature termination codons (PTCs) into the mRNA (Figure I.3.B). Similarly, skipping of some exons alters the reading frame and introduces PTCs. In cells, most abnormal mRNAs harboring PTCs are recognized and degraded by a quality-control mechanism termed nonsense-mediated mRNA decay or NMD (Chang, Imam, and Wilkinson 2007). Degradation of PTC-containing mRNAs by NMD prevents the translation of truncated proteins with potentially deleterious gain-of-function or dominant-negative activity. Thus, in most cases splice site mutations result in inactivation (loss of function) of the mutated allele. However, some mRNAs containing PTCs escape NMD and are translated into truncated proteins. If the skipped exon does not alter the reading frame, then a shorter protein will be synthesized. In other cases, the mutations induce expression of alternatively spliced protein isoforms with different functional properties; rather than create an aberrant (cryptic) splice site usually associated with loss of function, these mutations shift the ratio of expression of natural protein isoforms.

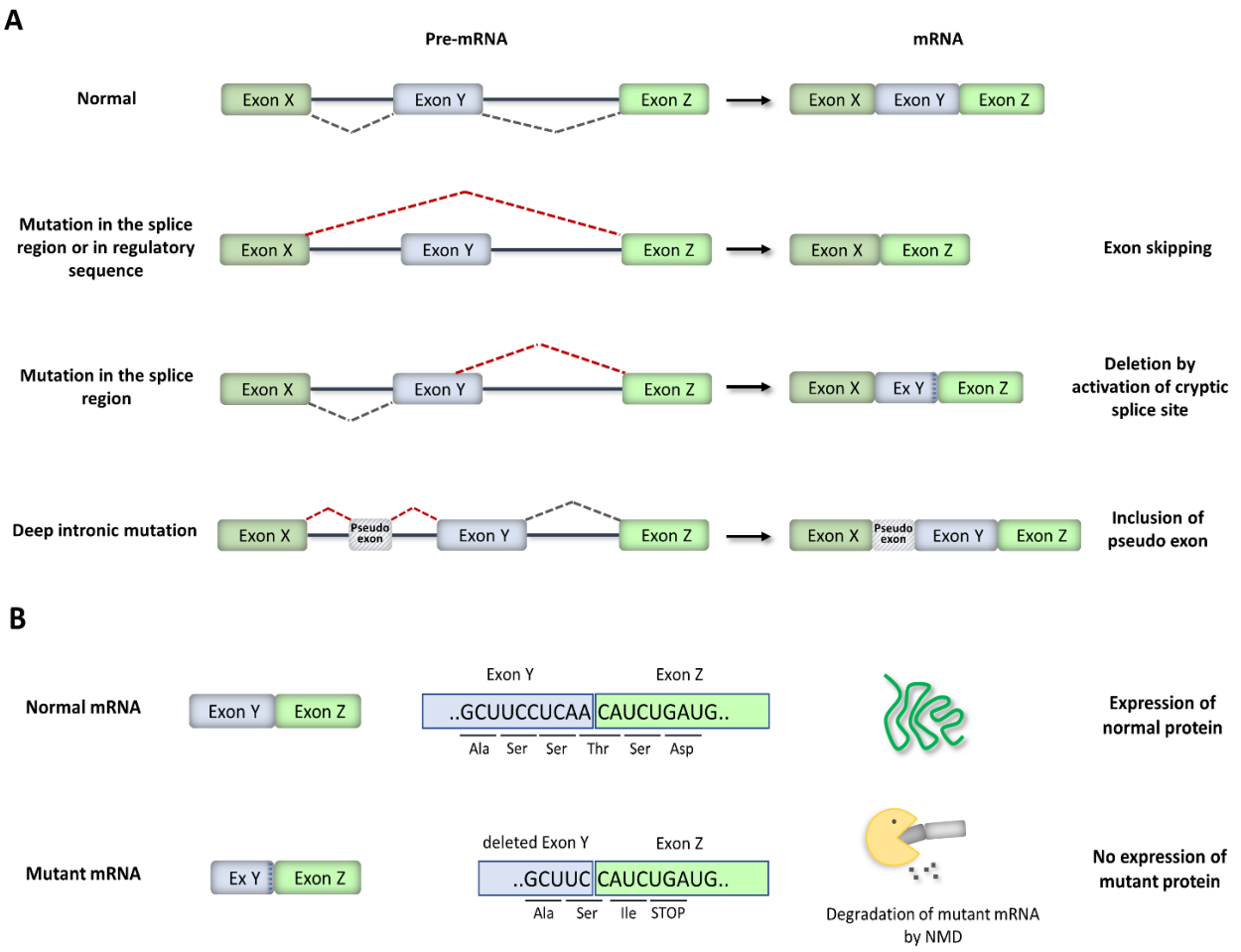


Figure I.3. Splicing mutations. (A) Schematics illustrating the consequences of cis-acting mutations on pre-mRNA splicing. (B) Representation of a frame-shift deletion introduced by activation of a cryptic splice site within an exon, and degradation of the resulting mRNA by nonsense-mediated decay (NMD).

A second category of mutations disrupts exonic splicing regulatory elements. Remarkably, the majority of human exons are predicted to contain exonic splicing elements (Z. Wang et al. 2004; Z. Wang and Burge 2008), suggesting that many disease-causing exonic mutations classified as synonymous, missense or nonsense could be unrecognized splicing mutations (Cartegni, Chew, and Krainer 2002). Exonic mutations that affect RNA splicing may induce skipping of the mutant exons, leading to either synthesis of a shorter protein or inactivation of the mutant allele expression.

Mutations in the third category are located in internal regions of introns (deep-intronic mutations). For a long time, medical genetic testing has focused mainly on sequencing the exons and the exon-intron boundaries, searching for mutations that are more likely to affect the function of the encoded protein. However, in recent years, advances in whole-genome sequencing techniques have resulted in the identification of an increasing number of disease-causing mutations located more than 100 nucleotides away from splice junctions (Vaz-Drago, Custódio, and Carmo-Fonseca 2017). Deep intronic mutations most commonly lead to pseudo exon inclusion due to creation and activation of cryptic splice sites (Figure I.3.A). Alternatively, deep intronic mutations disrupt splicing regulatory elements located within introns. Deep intronic mutations can also interfere with the function of transcription regulatory motifs and non-coding RNA genes (Vaz-Drago, Custódio, and Carmo-Fonseca 2017; Abramowicz and Gos 2018).

In contrast with the previous categories, which all refer to mutations that affect expression of a single gene by disrupting a splicing *cis*-element, mutations in the fourth category have an effect in *trans* on multiple genes by interfering with the function of either core components of the spliceosome or splicing regulatory proteins. For example, some patients with retinitis pigmentosa have mutations in genes that encode protein components of the U4/U6.U5 tri-snRNP (Vithana et al. 2001; Chakarova 2002; McKie 2001). Another example is spinal muscular atrophy, one of the most common genetic causes of childhood mortality, which is caused by loss of function mutations in a gene encoding a protein required for assembly of spliceosomal snRNPs (Paushkin et al. 2002). More recently, mutations in genes encoding either protein or snRNA components of the minor spliceosome have been identified as causative of several diseases (Verma et al. 2018).

HCM-associated splicing mutations

Typically, HCM is inherited in an autosomal dominant manner with incomplete penetrance and high phenotypic variability, even within the same family (Greaves et al. 1987; Watkins,

Ashrafian, and Redwood 2011). The proportion of mutation carriers with clinically detectable disease tends to increase with age and in most persons the hypertrophy is manifested in adolescence (Watkins, Ashrafian, and Redwood 2011). However, due to the incomplete penetrance, patients with subtle HCM features are difficult to recognize in the general population.

HCM can be caused by a single allelic mutation in any of at least 8 genes that encode cardiac sarcomere-associated proteins (Walsh, Thomson, et al. 2017; Konno et al. 2010; Walsh, Buchan, et al. 2017). These are the beta-myosin heavy chain (*MYH7*), the cardiac myosin-binding protein C (*MYBPC3*), cardiac troponin T (*TNNT2*), cardiac troponin I (*TNNI3*), cardiac alpha-actin (*ACTC*), alpha-tropomyosin (*TPM1*), the myosin ventricular essential light chain 1 (*MYL3*) and the myosin ventricular regulatory light chain 2 (*MYL2*) genes. Mutations in *MYH7* and *MYPBC3* occur most often, and together account for approximately half of all HCM cases (Walsh, Thomson, et al. 2017; Walsh, Buchan, et al. 2017).

The vast majority of HCM mutations in MYH7, as well as most other sarcomere mutations (Figure I.4.), have been classified as missense, i.e., they are point mutations that result in the substitution of one amino acid for another in the protein (Walsh, Thomson, et al. 2017). Studies have found normal levels of mutant MYH7 protein, but its function is perturbed. Specifically, biophysical properties of myosins that contain HCM mutations indicate a gain of function, with enhanced myosin ATPase activity, increased generated force and accelerated actin filament sliding (Tyska et al. 2000). In contrast, most MYBPC3 mutations have been classified as nonsense, frameshift or splice site mutations (Figure I.4.). A nonsense mutation is a single nucleotide substitution that results in a PTC in the transcribed mRNA, while a frameshift mutation can be an insertion, deletion or duplication of nucleotides that changes the reading frame of the mRNA and often leads to a PTC. Splice site mutations may also disrupt the normal reading frame and introduce a PTC in the mRNA (Figure I.3.B). It is therefore expected that the most frequent *MYBPC3* mutations result in degradation of PTC-containing mutant mRNAs. Indeed, analysis of myosin binding protein-C expressed in human myectomy samples from HCM patients with

MYBPC3 gene mutations revealed absence of truncated proteins and reduced levels of the normal protein (Marston et al. 2009). Moreover, HCM mutations in the *MYBPC3* gene engineered into mice resulted in reduced expression of myosin binding protein-C and caused altered cardiac function (van Dijk et al. 2009b). These results suggest that MYBPC3 haploinsufficiency (i.e., reduction in the amount of normal protein due to inactivation of expression of the mutant allele) is a pathologic mechanism for HCM.

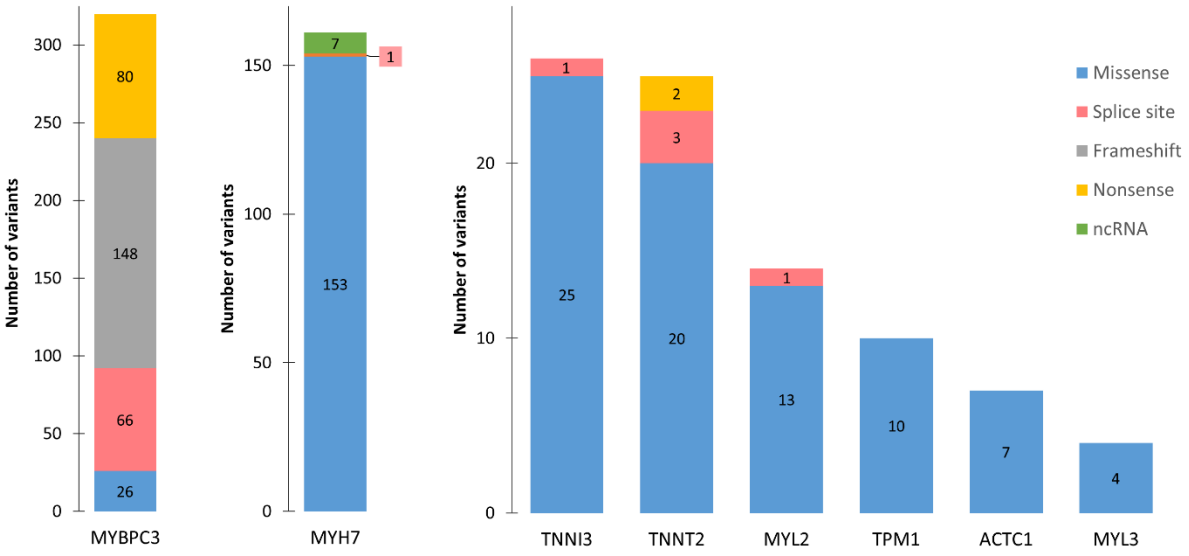


Figure I.4. HCM mutations. The bar graphs depict the number and type of HCM-associated variants in the indicated genes. All variants reported as pathogenic or likely pathogenic in the National Center for Biotechnology Information's ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/) were considered.

As expected, the majority of HCM-associated splice site variants reported in the ClinVar database change the two most highly conserved nucleotides at the splice junction (Figure I.5.). Although several computational algorithms have been developed to predict the effect of single nucleotide variants on splicing (S. Zhang et al. 2018; Barash et al. 2010; Di Giacomo et al. 2013; Erkelenz et al. 2014; Rosenberg et al. 2015; Xiong et al. 2015; Jaganathan et al. 2019), the definitive test of whether a disease-causing mutation affects splicing is by direct analysis of

mRNA. Indeed, it is essential to sequence the mutant mRNA to define its splicing pattern. Moreover, the total levels of mRNA should be measured to determine whether the mutation triggers NMD and therefore reduces the expression of the mutant transcript.

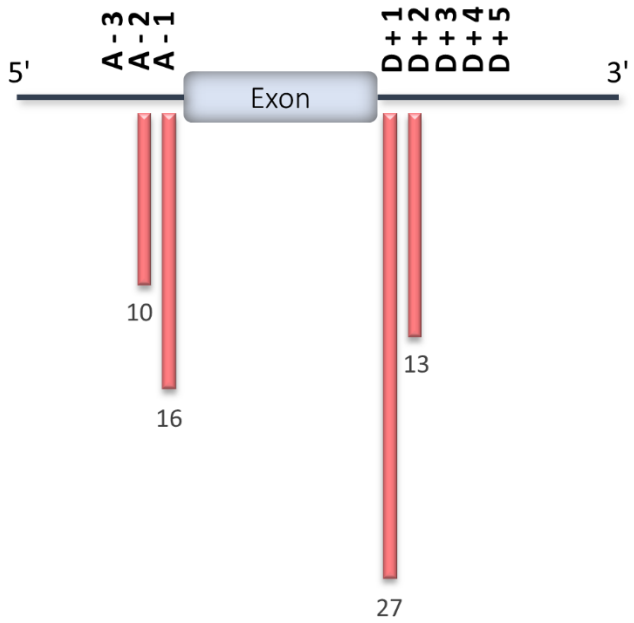


Figure I.5. HCM splice site mutations reported in the ClinVar database. Position of splice site mutations reported in the ClinVar database for the 8 HCM-associated genes shown in Figure 5. The number of variants per position is indicated. The 5' splice site (ss) is also known as the donor (D) splice site, whereas the 3' site is also known as the acceptor (A) splice site. Position D+1 corresponds to the first intronic nucleotide downstream of the depicted exon, D+2 to the second, and so forth. Position A-1 corresponds to the last intronic nucleotide upstream of the depicted exon, A-2 to the previous one, and so forth.

Aberrant splicing caused by HCM mutations has been shown using minigenes transfected into HEK293 cells (Ito et al. 2017). However, this is not an ideal system because mechanisms controlling splicing decisions are known to be influenced by chromatin structure and to be cell-type specific. Thus, RNA from the affected tissue should be studied. One possibility is to analyze mRNA from left ventricular septum samples from patients undergoing septal myectomy to relieve left ventricular outflow tract obstruction (Marston et al. 2009; Helms et al. 2014; Bagnall et al. 2018; Singer et al. 2019). Alternatively, RNA can be studied in cardiomyocytes differentiated *in*

vitro from induced pluripotent stem cells (iPSCs) derived either from patients or from normal iPSCs gene edited to harbor a patient-specific genetic variant (Lan et al. 2013; Ross et al. 2017; Ma et al. 2018a). RNA analysis can also be performed on peripheral blood samples, at least for *MYBPC3* mutations. Indeed, it was recently shown that RNA extracted from fresh venous blood supports amplification of *MYBPC3* transcripts and replicates the splicing patterns found in myocardial tissue (Singer et al. 2019).

RNA analysis of families with *MYBPC3* variants in the near-splice-site region allowed reclassification from uncertain significance to likely pathogenic (Singer et al. 2019). Multiple variants located in the near-splice-site region and predicted to disrupt splicing by computational algorithms were shown to either activate cryptic splice sites or induce exon skipping (Figure I.6.). Some variants resulted in a frameshift and introduction of a PTC, while other variants resulted in shorter mRNAs within frame deletions. RNA analysis further revealed that exonic variants classified as missense but predicted to disrupt splicing by computational algorithms caused exon skipping leading to frameshifts (Singer et al. 2019). Another study used whole genome sequencing and identified deep intronic variants that resulted in inclusion of pseudo exons (Figure I.6) leading to frameshifts (Bagnall et al. 2018). A synonymous exonic variant was further shown to create a novel cryptic splice sequence that truncated the exon (Figure I.6) leading to an in-frame deletion (Bagnall et al. 2018).

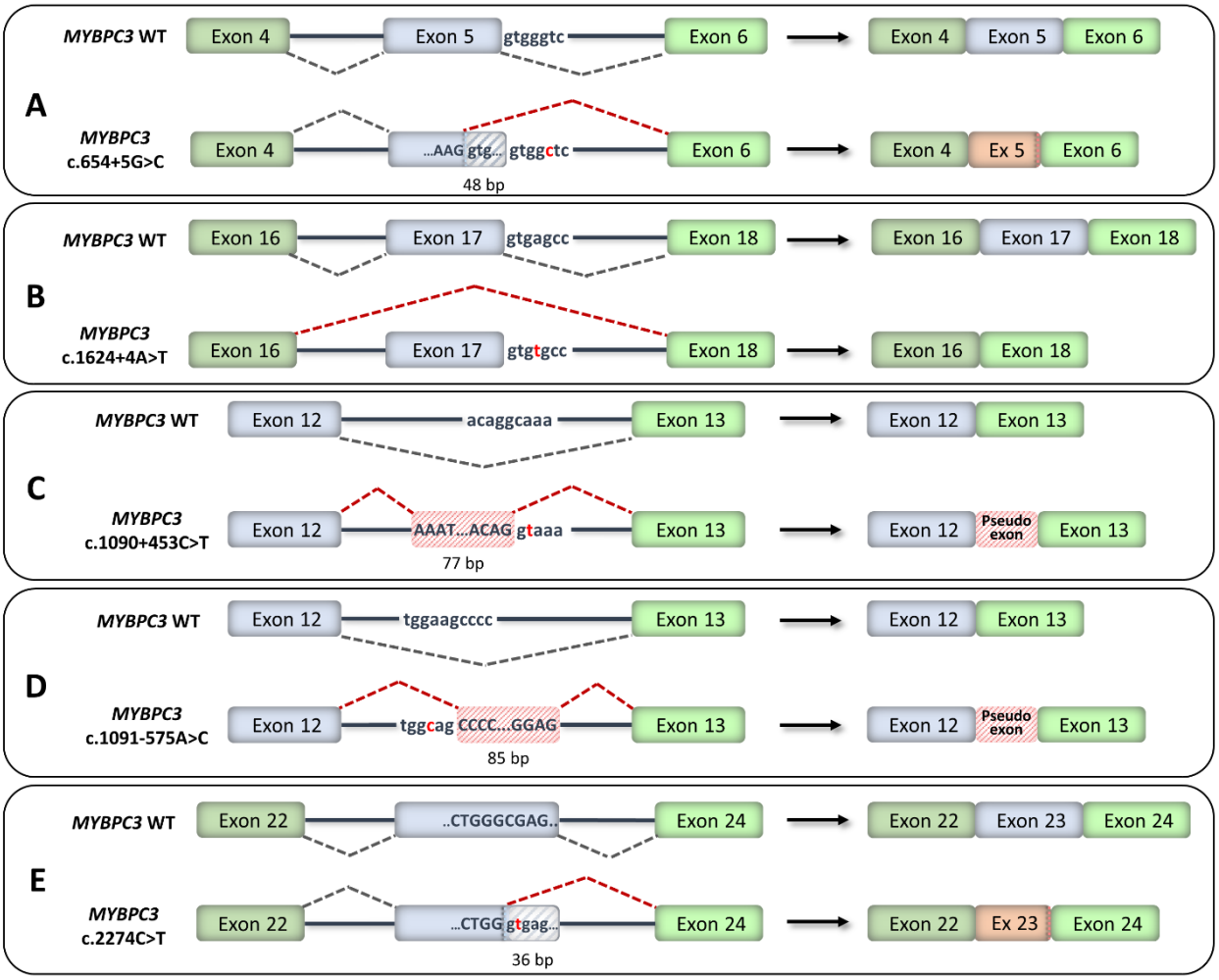


Figure I.6. Recently reclassified MYBPC3 splicing mutations. Exons are represented by boxes and introns by lines. Exonic nucleotides are indicated by capital letters and intronic nucleotides by small-case letters. (A) Variant c.654+5G>C created a cryptic donor splice site within exon 5, leading to a truncated exon with an in-frame deletion of 48 pb (Singer et al. 2019). (B) Variant c.1624+4A>T caused skipping of exon 17, leading to a frameshift (Singer et al. 2019). (C) Deep intronic variant c.1090+453C>T creates a new splice donor sequence, which leads to inclusion of a 77 bp pseudo exon in the mRNA causing a frameshift and introduction of a PTC (Bagnall et al. 2018). (D) Deep intronic variant c.1091-575A>C creates a new splice acceptor sequence, which leads to inclusion of a 85 bp pseudo exon in the mRNA causing a frameshift and introduction of a PTC (Bagnall et al. 2018). (E) Synonymous variant c.2274C>T (p.Gly758Gly) caused truncation of exon 23 by 36 nucleotides, leading to an in-frame deletion of 12 amino acids (Bagnall et al. 2018).

RNA therapeutics for HCM

Besides improving diagnostics, understanding precisely the expression of relevant RNAs in patient cells has the potential to inspire the development of new strategies to specifically treat that individual. Although for years the field of RNA therapeutics had to overcome numerous difficulties in achieving efficacious results without toxic side effects, the first RNA-targeted therapies have recently reached the clinic, and many are advancing to the final phases of clinical trials. In 2016, the antisense oligonucleotides Eteplirsen™ and Nusinersen™ were approved for treatment of Duchenne muscular dystrophy (Charleston et al. 2018) and Spinal muscular atrophy (Michelson et al. 2018), respectively. These oligonucleotides form base pair interactions with the nascent pre-mRNA and alter its splicing pattern. Eteplirsen hybridizes to a site within exon 51 of the DMD pre-mRNA, thereby sterically blocking spliceosome assembly at that site; this results in skipping of exon 51 and correction of the disease-causing frameshift mutation. The corrected mRNA contains exon 50 ligated to exon 52 and generates a shortened but still functional version of the dystrophin protein. Nusinersen hybridizes to an intronic region upstream of exon 7 of SMN2 pre-mRNA and blocks an inhibitory signal located at that site, causing inclusion of exon 7. The SMN2 mRNA with exon 7 encodes a fully functional protein that substitutes for the missing SMN1 protein.

In principle, splice-switching antisense oligonucleotides could be valuable for HCM treatment in cases of disease caused by mutations that disrupt normal splicing. A proof-of-concept study demonstrated the feasibility and efficacy of inducing skipping of a mutated *MYBPC3* exon 6 using viral-mediated transfer of antisense oligonucleotides in a mouse model of HCM (Gedicke-Hornung et al. 2013). Transduction of cardiac myocytes or systemic administration of the oligonucleotides reduced aberrantly spliced mRNAs, abolished cardiac dysfunction and prevented left ventricular hypertrophy in newborn mice (Gedicke-Hornung et al. 2013). Although not yet experimentally tested, antisense oligonucleotides designed to hybridize to cryptic splice

sites in mutant pre-mRNAs could sterically block spliceosome assembly at that site, thus preventing mis-splicing.

Another therapeutic strategy is to induce silencing of disease-causing genes by RNA interference (Setten, Rossi, and Han 2019). After the discovery that double-stranded RNAs (often called small interfering RNAs or siRNAs) can silence the expression of proteins encoded by complementary mRNAs in the nematode *Caenorhabditis elegans* (Fire et al. 1998; Montgomery, Xu, and Fire 1998), synthetic exogenous siRNAs were shown to induce sequence specific gene expression knock-down in mammalian cells (Elbashir et al. 2001). RNAi is a fundamental process initiated by the presence of long double-stranded RNA that is cleaved by the enzyme Dicer into shorter fragments of 21-23 nucleotides containing two single-stranded nucleotides on their 3' ends. Synthetic siRNAs are designed to mimic the natural products of Dicer. Each siRNA comprises a sense “passenger” RNA strand and a paired antisense or “guide” RNA strand. The siRNA molecules are loaded onto the RNA-induced silencing complex (RISC), which is composed of Dicer and Argonaute 2 (Ago2). During RISC assembly the siRNA is unwound, the “passenger” strand is removed and the single-stranded antisense guide base-pairs with the mRNA target. The mRNA hybridized to the siRNA is then cleaved by Ago2, which contains an RNase H like domain that functions to cleave one strand of an RNA:RNA duplex. Because siRNAs can in principle down-regulate any human mRNA, they should be ideal to eradicate expression of disease-causing mutant alleles. Yet, despite major efforts, siRNA-based therapies have faced multiple hurdles. Namely, delivery and stability proved difficult. Further, siRNAs were found to trigger innate toll-like immune receptors to initiate inflammation, raising concerns about safety (Setten, Rossi, and Han 2019). Tremendous progress in the field culminated in 2018 with the first-ever siRNA product (Patisiran) approved as a therapy for the rare hereditary disease transthyretin-mediated amyloidosis (J. Yang 2019). More recently, siRNAs targeting mRNA isoforms responsible for expression of the placenta-derived mediators of preeclampsia succeeded in suppressing clinical signs in a primate model (Turanov et al. 2018).

In the case of HCM, selective reduction in the expression of mutant alleles that code for dominant negative proteins would be the most direct therapeutic approach. Proof of concept has been established in a mouse model of HCM that is heterozygous for the R403Q mutation in *Myh6* (Jiang et al. 2013). RNAi constructs delivered by an adeno-associated virus vector preferentially reduced the levels of mutant transcripts and suppressed myocardial hypertrophy and fibrosis (Jiang et al. 2013). Although siRNAs can distinguish between mRNAs that differ by one single nucleotide, transcripts with splicing mutations that lead to in-frame deletions or insertions should be easier to target by RNAi. However, before pursuing any RNAi-based strategy for HCM it is critical to ensure that expression of the normal allele is sufficient to support normal myocardial function.

An alternative modality for correction of mutant mRNAs with potential application in HCM is spliceosome-mediated RNA *trans*-splicing, or SMaRT (Berger et al. 2016). Inspired by the observation that spliceosomes can ligate exons from two distinct pre-mRNAs creating a chimeric mRNA, the SMaRT technology makes use of an exogenous RNA sequence to replace one or several exons of the target mutant pre-mRNA. An artificial pre-mRNA is engineered to contain the coding sequence of substitution next to an intron. The end of this artificial intron consists of a stretch of nucleotides that base pair with the target intron in the endogenous pre-mRNA, bringing the exogenous exon in close proximity to its endogenous mutant counterpart. Efficient substitution of the mutant exon requires that the engineered *trans*-splicing out-competes the physiological *cis*-splicing process, and this currently remains a major challenge. If successful, the SMaRT technology would be particularly appealing to treat HCM because a single engineered RNA construct could be used to repair numerous mutations. Namely, two distinct RNA constructs covering the first and the second half of the *MYBPC3* mRNA should in principle be able to repair all the mutations in this gene and therefore treat 40-60% of all HCM patients. Proof-of-concept studies have been reported using an artificial pre-mRNA that carried the wild type *MYBPC3* cDNA sequence from exon 1 to exon 21 juxtaposed to an intron with a 120-nucleotide binding domain

for base pairing with a complementary sequence in intron 21 of the endogenous *MYBPC3* mRNA (Mearini et al. 2013; Prondzynski et al. 2017b). However, chimeric molecules resulting from *trans*-splicing represented less than 1% of all *MYBPC3* mRNAs in cardiomyocytes, indicating that the efficiency of this approach was too low to be considered as a therapeutic option (Prondzynski et al. 2017b).

Non-coding RNAs, particularly microRNAs (miRNAs), are also attracting much attention as biomarkers of cardiac disease and potential therapeutic targets. Altered expression levels of circulating miRNAs have been reported in association with hypertrophic cardiomyopathy (Sayed et al. 2007; Roncarati et al. 2014; Song et al. 2014; Ming et al. 2018), and forced overexpression of stress-inducible miRNAs was shown to induce cardiomyocyte hypertrophy (Van Rooij et al. 2006). However, it remains to be established whether modulation of miRNA levels is sufficient to revert the HCM phenotype.

I.4. DISEASE MODEL SYSTEMS FOR HCM

Different model systems have been used to model cardiomyopathies and more specifically HCM, providing important information about the pathogenesis of the disease. These models include primary cultures, animal models, immortalized cell lines and human pluripotent stem cell derived cardiomyocytes (hiPSC-CMs).

While primary cell cultures have the advantage of reflecting the disease biology faithfully, since they maintain their morphology, function and protein markers, it is very difficult to obtain cardiac tissue from human patients, as it is limited to extreme clinical situations such as a cardiomyectomy or a biopsy. Moreover, the amount of tissue is often insufficient for extensive molecular analysis and these cells are difficult to maintain in culture due to their limited lifespan and expansion potential. On the other hand, since myectomy or biopsy samples are generally obtained from patients in late stage of HCM development it poses a major difficulty in studying pathophysiological mechanisms leading to disease (Ojala and Aalto-Setälä 2016; Hoes, Bomer, and van der Meer 2019).

Several animal models have been used in the study of HCM, including genetically engineered animals and those with spontaneous mutations. Rodents are the most used model species for HCM research, due to the ease of maintenance and genetic manipulation to generate transgenic or mutant strains and their short gestation time. Moreover, mice short life span allows to follow the natural history of the disease at an accelerated pace, leading to a fast conclusion of the experiments and its potential translation to human research. However, these characteristics can also limit the data's applicability to human cardiovascular function since rodents are phylogenetically farthest from humans compared to other mammals, and there are significant differences between pathophysiological features of HCM phenotypes in mouse models and the human disease presentation (Houser et al. 2012; Purevjav 2019; Milani-Nejad and Janssen 2014). Most mouse models of HCM carry human mutations in sarcomeric protein-encoding genes

such as *MYBPC3* (Q. Yang et al. 1998), *TNNT2* (Ferrantini et al. 2017), *TNNI3* (James et al. 2000) and *TPM1* (Prabhakar et al. 2001). Compared with the mouse, larger animals offer some advantages for cardiovascular research due to the larger size and slower heart rate, which are advantageous for physiological analyses. These include genetically engineered rabbit (Sanbe et al. 2005; Lowey et al. 2018; Nagueh et al. 2004) and pig models (Montag et al. 2018) and some naturally occurring cat (Freeman et al. 2017; Meurs et al. 2005) and pig models (Shyu et al. 2002; J. H. Lin et al. 2002). Considering that the development and investigation of animal models is complex, expensive, time-consuming, and even when natural occurring animal models, larger than rodents, are available, the outcomes can be difficult to translate to humans due to interspecies differences of cardiovascular anatomy, physiology and gene expression, alternative disease models were needed. Nevertheless, the use of animal models had improved our knowledge on the mechanisms of cardiac diseases and promoted an advancement preclinical assessment of drug discovery and development (Hearse and Sutherland 2000; Savoji et al. 2019; Purevjav 2019; Milani-Nejad and Janssen 2014),

Immortalized human cell lines can be derived either from tumors or from the immortalization of other cell lines such as the human ventricular AC16 cell line, which was developed by fusing primary ventricular cardiomyocytes with a SV-40 transformed fibroblast cell line (Davidson et al. 2005). Immortalized cell lines have the advantage of unlimited lifespan, they are easy to grow and maintain with low associated costs. However, this does not reflect accurately normal physiological conditions and can cause genetic and phenotypic variation over time with the accumulating genetic aberrations leading to a different behavior than *in vivo* (Jimenez-Tellez and Greenway 2019; Hoes, Bomer, and van der Meer 2019).

Recently, alternative models have been developed based on the advances in stem cell technology and research, which provide a new human cell source that has demonstrated great potential for disease modeling due to the unlimited *in vitro* expansion potential and differentiation capability that allows them to, in principle, differentiate into any human cell type, including

cardiomyocytes (hiPSC-derived cardiomyocytes, hiPSC-CMs). Encouragingly, it had been shown that patient-derived hiPSC-CMs can recapitulate *in vitro* the HCM phenotype at the cellular level (Lan et al. 2013; van Mil et al. 2018; Ojala and Aalto-Setälä 2016).

Human induced pluripotent stem cells for disease modeling

Stem cells are defined as a population of undifferentiated cells characterized by their self-renewal capability and the ability to differentiate into specialized cell types (De Los Angeles et al. 2015; Mahla 2016). They can be divided in three different types according to their differentiation potential. Totipotent stem cells can generate all cell types, giving rise to an entire organism and the only cells with this classification are the zygote and early blastomeres. Pluripotent stem cells (PSCs) can originate cells from all three embryonic germ layers: ectoderm, endoderm and mesoderm. Lastly, multipotent stem cells, such as hematopoietic stem cells, are already committed to give rise to cells of a restricted spectrum of cell lineages (De Los Angeles et al. 2015; Mahla 2016).

For modeling human diseases, PSCs are the most interesting group since in theory they can be indefinitely expanded maintaining an undifferentiated state, have ability to differentiate into most (if not all) embryonic and adult cell lineages and are easier to obtain than totipotent cells. PSCs can have different origins, the main ones being embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs).

Human ESCs were the first human PSCs obtained. They were first isolated from the inner cell mass of human blastocysts produced *in vitro* in 1998 by James Thomson (Thomson 1998). Even though the isolation and expansion of hESCs opened new avenues for *in vitro* studies of human development and diseases, they are associated with some drawbacks. Human disease-specific ESCs are either obtained by genetic modification, having low efficiency of gene targeting and unknown predisposition for diseases (Urbach 2004; Tulpule and Daley 2009) or isolated from

embryos carrying monogenic disease-associated mutations detectable via pre-implantation genetic diagnosis (Urbach and Benvenisty 2009; Eiges et al. 2007) and often fail to recapitulate disease-associated phenotypes (Halevy and Urbach 2014). Importantly, in most cases, making a hESC line involves the destruction of an embryo raising significant ethical issues.

Human iPSCs were first obtained by Yamanaka & Takahashi in 2007 by overexpressing a cocktail of four transcription factors, (Octamer-binding transcription factor 4 (*OCT4*), Sex determining region Y-box 2 (*SOX2*), Krüppel-like factor 4 (*KLF4*) and v-myc avian myelocytomatosis viral oncogene homolog (c-MYC)) into a human adult fibroblast. The generated hiPSCs showed similar properties to hESCs, such as same morphology and proliferation capacity when cultured *in vitro* (Takahashi et al. 2007). Despite the initial concerns that iPSC could have significant differences from ESCs due to their epigenetic memory, nowadays such concerns largely subsided; in fact, studies comparing isogenic hiPSCs with hESCs show that for all purposes these cell types are molecularly and functionally equivalent, and observed differences are most likely to arise from the genetic background variation (Choi et al. 2015; Young et al. 2012).

Several different somatic cell types can now be reprogramed into hiPSCs. These include the initially used dermal fibroblasts obtained through skin biopsies and other, more convenient, accessible and risk-free cell sources such as peripheral blood cells, typically T cells (Seki et al. 2010; Loh et al. 2009), renal tubular cells collected from urine samples (Zhou et al. 2011) and keratinocytes from plucked hair follicles (Aasen et al. 2008).

Regarding the choice of methods to deliver reprograming factors into human somatic cells, the integrative methods (retrovirus and lentivirus) initially used are still one off the top choices due to their high efficiencies; although, in many cases, they are being progressively substituted by non-integrating methods such as episomal plasmid vectors (Yu et al. 2007; Junying et al. 2009), messenger RNA transfection (Warren et al. 2010) and Sendai virus (Fusaki et al. 2009; Seki et al. 2010). These nonintegrative methods have the advantage of avoiding insertional mutations or

unintentional transgene reactivation, and should consequently result in less variability between the generated hiPSC lines (Yoshida and Yamanaka 2017).

Since the first time they were described, hiPSCs emerged as a powerful tool to study development and human genetic disease. Their use circumvents the ethical issues associated with the manipulation of hESCs and the different possible cell sources facilitate the establishment of new hiPSC lines, including patient-specific cell lines which in turn are the most relevant cell source for drug screening and human disease study since they reproduce the patient's genetic background.

Although patient derived hiPSCs (PD-hiPSCs) are very attractive to be used for disease modeling, there are also several advantages in the use of gene edited hiPSC. One is the case where a given variant of interest is so rare in a population that recruiting a patient with that specific mutation is not feasible. Another is that it enables the study of a chosen variant irrespective of the genetic background. Moreover, gene editing can be very useful in providing isogenic controls that allow for better comparation and genotype-phenotype correlations (Musunuru et al. 2018; Bellin et al. 2012). Therefore, gene editing approaches can be useful either to try to revert the disease phenotype by correcting the causal variant from a PD-hiPSC, or to recapitulate the disease phenotype by inducing a variant thought to be causal in an established hiPSC cell line derived from a healthy donor (Figure I.7) (Musunuru et al. 2018).

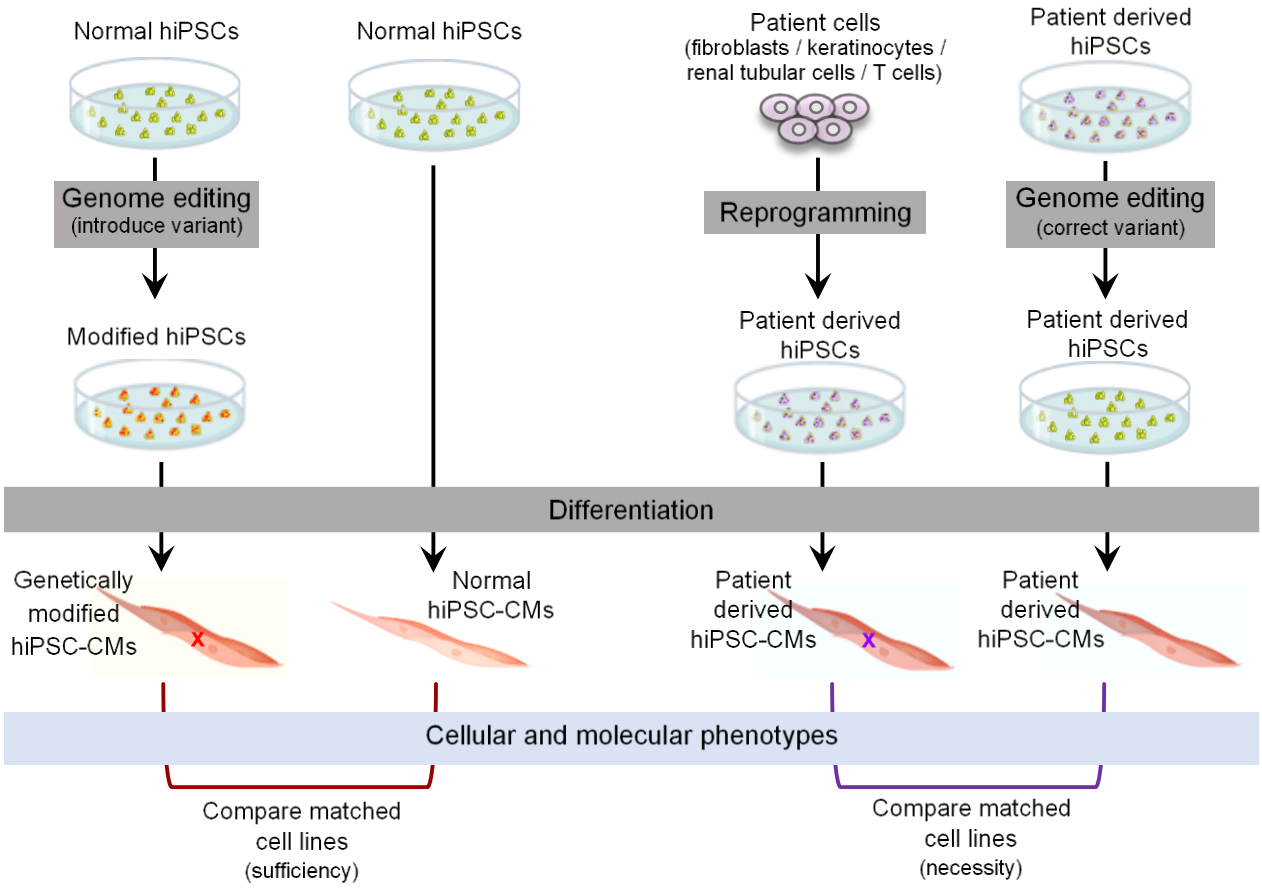


Figure I.7. Disease-modeling applications of iPSC technology. hiPSCs used can be either previously establish cell lines that are gene-edited to carry the variant of interest or generated by reprogramming adult somatic cells derived from patients. The subsequent *in vitro* differentiation, namely in cardiomyocytes, and comparation with matched controls provide a valuable tool for disease modeling and allows to answer different questions of sufficiency or necessity of a given variant to cause a disease.

CRISPR/Cas9 gene editing system

In recent years a variety of genome editing technologies have emerged allowing to insert, delete or modify target DNA sequences in its endogenous context, both *in vitro* and *in vivo.* These genome editing tools include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/associated protein 9 (Cas9) that rely on the use of nucleases to induce double-strand

DNA breaks (DSBs) in desired locations of the genome leading to the activation of endogenous repair mechanisms (Doudna and Charpentier 2014).

ZFNs and TALENs are chimeric nucleases that contain an endonuclease catalytic domain coupled to a modular DNA-binding region that can be programmed according to the genomic target site. On the other hand, in the CRISPR/Cas9 system, site targeting is based on a small RNA (gRNA) that can guide the Cas9 nuclease to the target locus through RNA-DNA base pairing. Since its recent introduction to edit mammalian cells, the CRISPR/Cas9 system became the most widely used gene editing tool due to its ease of use and design, high specificity, more flexible site targeting, and improved efficiency compared to earlier methods (Ran et al. 2013)

CRISPR/Cas systems, found in bacteria and archaea, function as adaptive immune systems protecting them from invading viral or plasmid nucleic acids. The CRISPR/Cas9 system, commonly used for gene editing, is an adaptation of this defense mechanism in the bacteria *Streptococcus pyogenes*. CRISPR *loci* is comprised of conserved repetitive bases intercalated by short fragments of DNA from a previously invading organism (*spacers*) and acts as an "immune memory". In response to a new infection, the bacteria produces RNA segments using these *spacer* sequences as templates, which are then cleaved into small CRISPR RNAs (crRNA). This particular *loci* also encodes an extra small RNA that is complementary to the repeat sequence, known as a trans-activating crRNA (tracrRNA) (Deltcheva et al. 2011). When combined, to form a tracrRNA:crRNA complex, they guide the Cas nuclease to the target sequence of the invader genome. This genomic target sequence (*protospacer)* is complementary to the crRNA, and was found to be adjacent to short (3–5 bp) DNA sequences termed protospacer adjacent motifs (PAM) that are also required for the binding of Cas-RNA complex to the DNA, through Watson-Crick base-pairing. Subsequently, the nuclease domains of the Cas protein cleave the target region in the invading genome (Figure I.8.a) (Jinek et al. 2012; F. Zhang, Wen, and Guo 2014).

For its application as a gene-editing tool, the CRISPR/Cas9 system was engineered in order to be easily adapted to the regions to be edited (Figure I.8.b), by using a single guide RNA

(sgRNA), composed by the fusion between a crRNA and part of the tracrRNA sequence. This sgRNA leads the Cas9 to cleave target DNA sites that are complementary to the 5′ 20 nucleotides of the sgRNA and lie next to a PAM sequence with the canonical form 5′-NGG. Consequently, with this system, the Cas9 endonuclease can be directed to any DNA sequence with the form N-NGG by changing the first 20-nt of the sgRNA to correspond to the target genomic DNA sequence. Upon recognition of the PAM sequence, the Cas9 mediates a DSB 3 bp upstream of the PAM, promoting the activation of host DNA repair pathways (Cong et al. 2013; Mali et al. 2013).

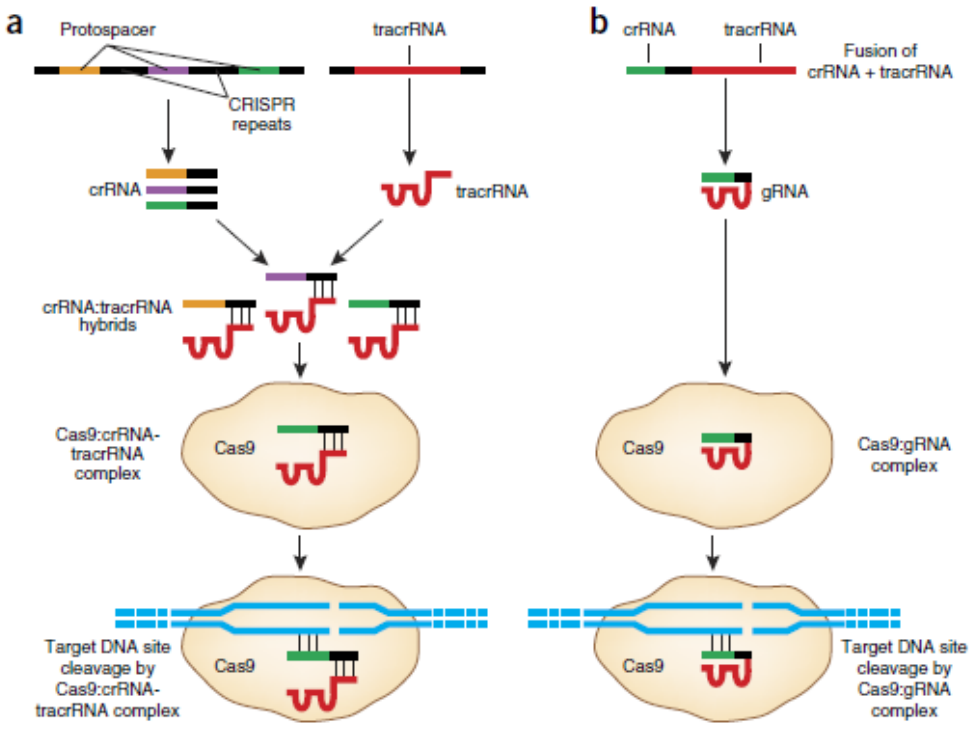


Figure I.8.Naturally occurring and engineered CRISPR-Cas systems. (a) The natural occurring adaptive CRISPR/Cas9 immune system. (b) The engineered CRISPR/Cas9 system for an easier application in gene-editing, requiring only a single guide RNA whose sequence can be modified according to desired target genomic sequence. Adapted from (Sander and Joung 2014).

The repair of the Cas9-induced DSBs can be achieved by two different molecular mechanisms: the error-prone non-homologous end joining (NHEJ) or the homology-directed

repair (HDR), a more high-fidelity and error-free repair mechanism (F. Zhang, Wen, and Guo 2014). The NHEJ pathway is the most typically used since it is faster and active during the entire cell cycle. With this pathway, the two cleaved ends of the DNA can rapidly be ligated often resulting in the introduction of errors such as insertion or deletion of nucleotides at the target site (*indels*). Since these events often lead to frameshifts and generation of premature stop codons it can be useful in gene-editing when attempting gene knockouts (Figure I.9.) (M. Liu et al. 2019). In order to precisely edit DNA sequences, such as to obtain single nucleotide modifications, it is necessary that the DSB is repaired through the HDR pathway. This pathway requires the presence of a repair template, and it is restricted to late S and G2 phases (when DNA replication is completed, and sister chromatids are accessible to serve as templates). Besides the sister chromatids endogenous template, exogenous DNA molecules can also be used to precisely introduce specific nucleotide modifications or knock-ins in the target location (Figure I.9.) (M. Liu et al. 2019).

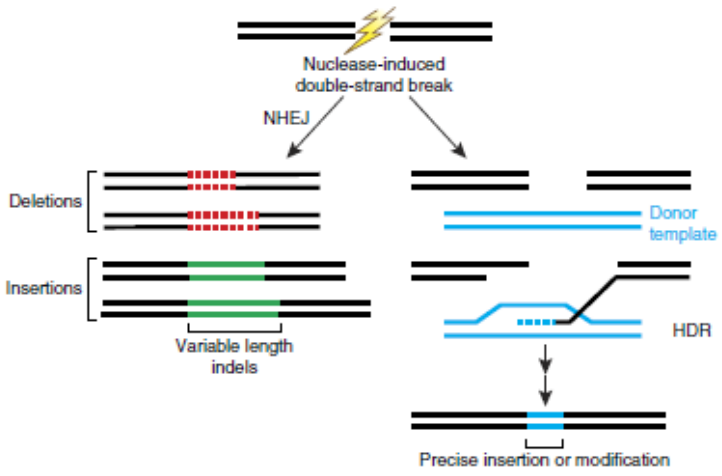


Figure I.9. Nonhomologous end joining (NHEJ) or homology-directed repair (HDR) pathways used in nuclease-induced genome editing. The error-prone NHEJ-mediated repair often results in the insertion or deletion of nucleotides at the target site. HDR-mediated repair can introduce precise point mutations or insertions in the presence of a single or double-stranded DNA donor template. Adapted from (Sander and Joung 2014).

However, even in the presence of a repair template, HDR efficiency is low, reportedly between 0,1-1% (Miyaoka et al. 2016), and frequently detected unwanted mutations can be explained by the high nuclease activity of CRISPR/Cas9, which may continuously re-cut edited *loci*, or by the HDR competition with NHEJ in S and G2 phases and specific down-regulation of HDR at M phase and early G1. To overcome this limitations, several strategies have been proposed to increase HDR frequency or prevent Cas re-cutting such as cell cycle modulation with timely CRISPR/Cas delivery (S. Lin et al. 2014), incorporation of CRISPR-Cas9 blocking mutations along with the mutation of interest (Paquet et al. 2016), modifications to the CRISPR/Cas9 system (Aasen et al. 2008; Howden et al. 2016), or the use of small molecule inhibitors of the NHEJ pathway (Maruyama et al. 2015). Other frequent concerns include the presence of off-target effects, arising from unspecific binding of the sgRNA to other sites in the genome, and high variability between different cell types regarding delivery of the large CRISPR/Cas9 machinery and gene-editing efficiency. These issues can be tackled by reducing non-specific binding of sgRNA sequences and cell line dependent delivery optimization of CRISPR/ Cas9 systems, either in the form of plasmid DNA, RNA, or proteins (Wilbie, Walther, and Mastrobattista 2019; J. J. Liu et al. 2019).

Despite challenges in its applicability, the gene editing CRISPR/Cas9 system significantly enhanced the ability to precisely modify a chosen *locus* with minimal impact on the remaining genome and thereby the utility of hiPSCs for purposes such as disease modeling. Based on that the CRISPR technology has already been used to establish both animal and hiPSC-CMs HCM disease specific models (Motta et al. 2017; Nguyen et al. 2019).

I.5. HIPSC DERIVED CARDIOMYOCYTES

The development of hiPSCs was already proven valuable for different applications, from disease modeling to drug testing, by providing an unlimited supply of differentiated cells from any somatic cell type, particularly cardiomyocytes (hiPSC-CMs). However, such studies present some limitations such as the lack of consistency, purity or maturity presented by the differentiated CMs and approaches to enhance these characteristics are being continuously developed (Figure I.10.).

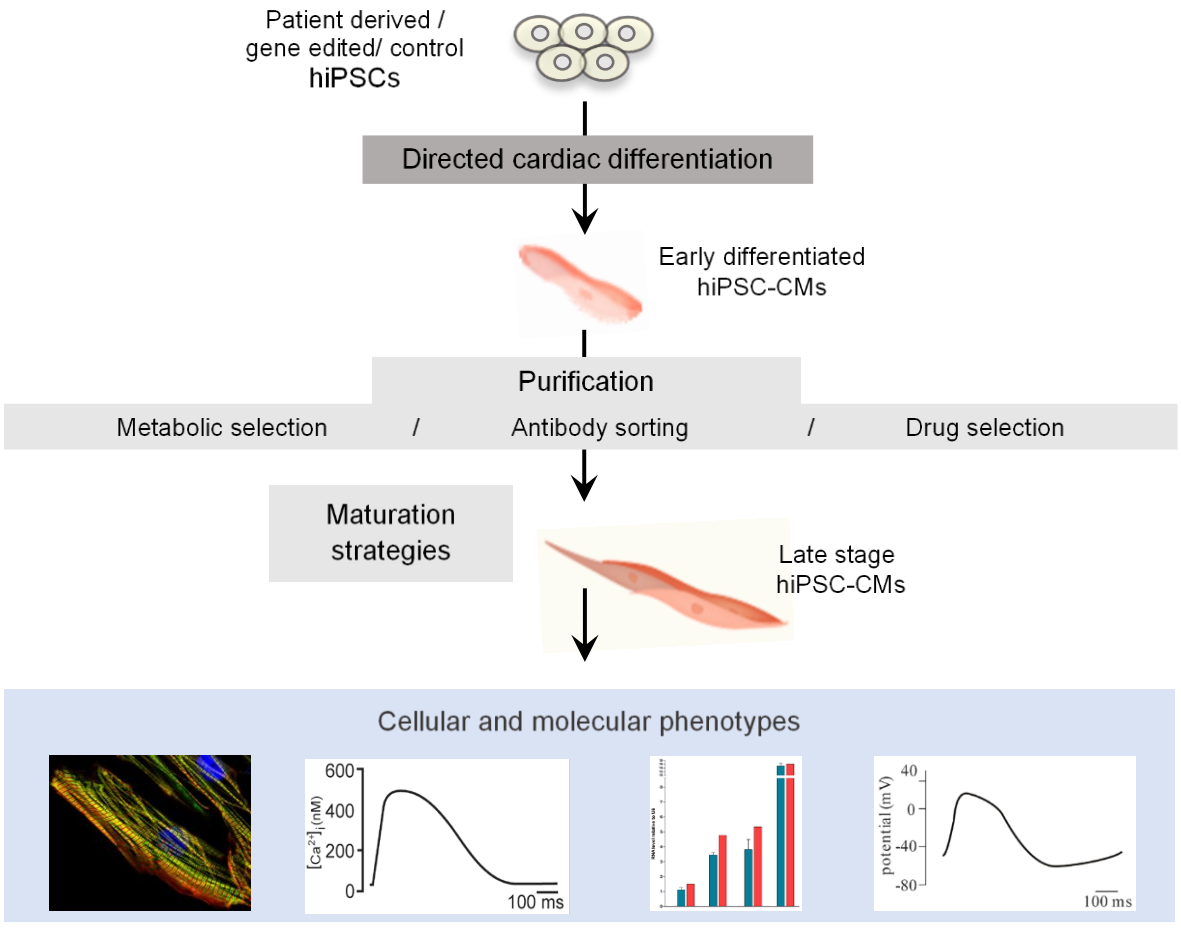


Figure I.10. Cardiomyocyte differentiation approach. Different steps can be performed to enhance the characteristics of the obtained hPSC-CM population.

Differentiation protocols

Initial efforts to obtain cardiomyocytes from human PSCs used hESCs to form embryoid bodies (EBs) in serum containing medium. However, only approximately 8% of the EBs presented spontaneous contraction, indicating the presence of CMs, with CMs typically representing less than 1% of the total culture (Kehat et al. 2001).

Since then more specific and efficient cardiac differentiation methods have emerged, taking advantage of knowledge regarding key signaling pathways associated with embryonic heart development such as activin/nodal/transforming growth factor-β (TGF-β), Wnt, and bone morphogenetic protein (BMP4). Through the appropriate temporal addition of growth factors and/or small molecules, it was possible to recapitulate to some extent the critical stages of cardiac specification leading to a higher cardiac differentiation efficiency (Karakikes et al. 2015).

This information was earlier used to establish a monolayer-based activin A/BMP4-directed differentiation system reported to be at least 50-fold more efficient than the previously used EBs with serum induction (Laflamme et al. 2007). Protocols using EBs in suspension culture could also be improved by the addition of the growth factors BMP4, fibroblast growth factor 2 (βFGF) and Activin A during early stage of differentiation, and the Wnt inhibitor DKK1 and v*ascular endothelial growth factor* (VEGF) during late stage of differentiation, leading to a yield of more than 50% of CMs (L. Yang et al. 2008). Another significant improvement was brought by the 2D differentiation approach proposed by Lian et al., where the sole modulation of the Wnt pathway, by the sub sequential addition of two small molecules, CHIR99021 (a glycogen synthase kinase-3 (GSK3) inhibitor that leads to Wnt pathway activation and it turn to mesoderm commitment of the cells) and IWP2 (an inhibitor of Wnt ligand production that leads to Wnt pathway inhibition, which at this time point can direct the generated mesoderm progenitors to a cardiac fate), generated 80–98% pure populations of CMs (Lian et al. 2012; 2013).

This Wnt modulation approach (commonly termed GiWi approaches – for the GSK3 inhibition, Wnt inhibition) was also used by Burridge et al. to create a chemically defined and xeno-free cardiac differentiation approach, which would reduce the variability in media components and eliminate possible patient immune reactions to animal components in the CM product, as the commonly used B27 supplement (a supplement that contains all factors required to support the survival and differentiation of the cells) was replaced by ascorbic acid and human recombinant albumin, leading to the production of approximately 85% CMs that could further enriched to >95% using chemically defined metabolic selection.

A summary of the mentioned protocols, with the timing at which to induce these pathway changes and the growth factors and small molecules used for such modulations, is seen in Figure I.11.

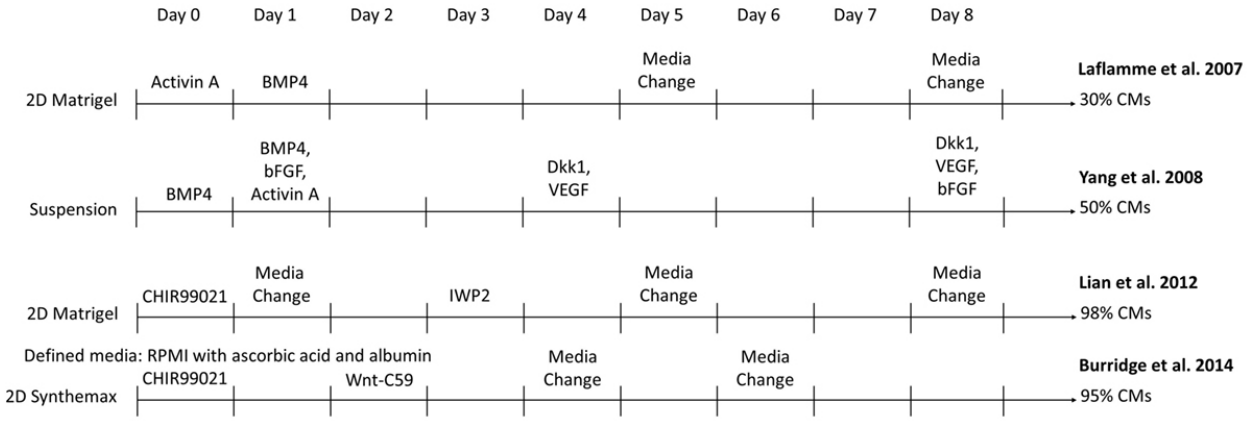


Figure I.11. Summary of cardiac differentiation protocols for differentiating hPSCs into cardiomyocytes. Adapted from (Dunn and Palecek 2018).

Despite the important advances brought by the development of the described cardiac differentiation protocols, with production of higher yields of cardiomyocytes, for some applications it is desirable to have a purer and more mature CMs population than the ones generated with such protocols. In order to address such requirements, several purification and maturation strategies have been proposed.

Purification strategies

One of the most widely used purification strategies takes advantage of the iPSC-CMs ability to metabolize lactate, in contrast to undifferentiated cells. In this metabolic purification, the prolonged incubation with glucose-depleted lactate-supplemented medium has been shown to increase the yield of iPSC-CMs up to 98% (Tohyama et al. 2013; Burridge et al. 2014). Moreover, the combination of glutamine and glucose depletion was reported to further eliminate undifferentiated PSCs (Tohyama et al. 2016). This methodology, if used in a monolayer differentiation protocol, can purify CMs while still in the dish, providing pure hiPSC-CMs with minimal experimental complexity.

Another effective way to separate CMs from other cells is to make use of antibodies recognizing a CM-specific membrane protein for fluorescence-activated cell sorting (FACS) or magnetic activated cell sorting (MACS), followed by replating of the purified hiPSC-CMs population. MACS can be performed under sterile culture conditions but leads to a loss of CM yield following the purification (Dubois et al. 2011). FACS is highly efficient in terms of purity and yield but difficult to scale. Signal-regulatory protein-α (SIRPA) and vascular cell adhesion molecule 1 (VCAM1) have both been identified as such CM markers by two independent studies (Dubois et al. 2011; Uosaki et al. 2011; D. A. Elliott et al. 2011). Elliot *et al.* used expression profiling analyses and found that expression levels of both VCAM1 and SIRPA were significantly upregulated in cardiac progenitor cells expressing NKX2.5. Importantly, SIRPA positive (+)/ VCAM negative (-) cells had higher expression of smooth muscle cell and endothelial cell markers than SIRPA+/VCAM+ cells, indicating that cells sorted solely based on SIRPA expression may not be of pure cardiac lineage (D. A. Elliott et al. 2011). On the other hand, Dubois *et al.* and Uosaki *et al.* screened antibody panels against hPSC-CMs and identified SIRPA and VCAM1 respectively, as specific surface proteins expressed on hPSC-CMs (Dubois et al. 2011; Uosaki et al. 2011). In particular, Uosaki et al. described that 80% of TNNT2+ (an early marker of cardiac

differentiation) cells at day 11 of differentiation were VCAM1+ and 95–98% of VCAM1+ cells at day 11 were positive for TNNT2 (Uosaki et al. 2011). Since their identification, different studies have used VCAM1+ selection to purify cardiomyocyte populations and further establishing VCAM1 up-regulation as a key marker of cardiomyogenic commitment (Schwach and Passier 2016; Skelton et al. 2014; Pontén et al. 2013). Despite cell sorting with antibodies against either of these proteins reportedly result in >95% pure populations of hiPSC-CMs, it appears that these surface markers are not exclusive for CMs as these antibodies also mark other cell types including smooth muscle cells and endothelial cells (Ban, Bae, and Yoon 2017).

Other purification strategies include the use of antibiotics but these typically involve the introduction of antibiotic resistance genes in CMs specific promotors or genes, which is not always feasible or advisable (Musunuru et al. 2018; Schwach and Passier 2016).

Maturation markers

Although the previously mentioned strategies can successfully purify hiPSC-CMs, giving rise to a more homogeneous population, such CMs still present important differences from the ones isolated from adult human heart, and resemble more closely fetal CMs. Some of this differences are summarized in Table I.1. (Veerman et al. 2015; Bedada et al. 2014; Cai et al. 2019; Lundy et al. 2013; Karbassi et al. 2020; Yoshida and Yamanaka 2017; Denning et al. 2016).

Structurally, adult CMs are large and rod-shaped, with high length-to-width ratio and can be bi-nucleated. They are highly organized, with longitudinal alignment and long sarcomeres that display Z-discs, I-, H-, A- and M- bands. In contrast, *in vitro* differentiated CMs are round or multi-angular, mono-nucleated small cells that typically show chaotic alignment with disorganized and short sarcomeres. Commonly, only Z-discs and I-bands can be detected in microscopic analysis and they lack transverse tubules. (Veerman et al. 2015; Lundy et al. 2013; Denning et al. 2016). Furthermore, the less well-developed sarcoplasmic reticulum and the lack of T-tubules of hiPSC-

CMs results in an impaired calcium handling that relies more on flux through the sarcolemma, resembling that of fetal CMs (Veerman et al. 2015; Denning et al. 2016).

Metabolically, while hiPSC-CMs can use fatty acids as an energy source but depend more on glycolysis, metabolizing glucose and lactate, adult CMs rely primarily in β-oxidation of fatty acids for energy and have, accordingly, a higher mitochondria content (Correia et al. 2017; Karbassi et al. 2020).

In general, adult CMs are quiescent but, when stimulated, generate greater force, upstroke and conduction velocities. In contrast, hiPSC-CMs, electrical immaturity is evident from their spontaneous beating; here, the higher resting membrane potential (close to the one found in nodal cells), probably due to low expression of inwardly rectifying potassium current (I_{K1}), triggers the depolarization more easily, leading to spontaneous contraction. Moreover, they present a slow upstroke velocity and small amplitude, which can be related to a lower density of I_{Ks} potassium and I_{Na} sodium channels and inactivation of sodium channels at higher resting membrane potentials. Finally, the conduction velocity is also slower, due the localization of gap junctions around the cell circumference rather than at the intercalated discs (Veerman et al. 2015; Karakikes et al. 2015; Karbassi et al. 2020),

Importantly, the gene expression profile of *in vitro* differentiated CMs is reported to most closely resemble that of fetal CMs, while being quite distinct from that of adult ones. With respect to calcium handling and cardiac ion channel genes as *KCNH2* (potassium voltage-gated channel), *RYR2* (ryanodine receptor 2), *SCN5A* (sodium voltage-gated channel alpha subunit 5) are more expressed in adult-CMs. Moreover, adult CMs express high levels of important structural proteins such as titin, myosins, α-actin and the troponin complex, which are less abundant, or differ in isoform, in fetal and hPSC-CMs. Both myosin heavy chain (MHC), troponin I and myosin regulatory light chain 2 (MLC2) have isoforms that are generated from different genes and regulated in a developmental manner. Myosin heavy chain (MHC) fetal isoform, α-MHC, is encoded by *MYH6* and is characterized by fast ATPase activity and rapid actin binding (features

needed in fetal hearts to support faster beating), while the adult β-isoform (β-MHC) is encoded by the *MYH7* gene that present higher expression levels in adult ventricular CMs. Troponin I slow skeletal muscle isoform (ssTnI) is encoded by *TNNI1* and expressed in CMs during embryonic development causing altered calcium affinities, which result in slowed relaxation; after birth, there is a switch to the cardiac isoform (cTnI) encoded by *TNNI3.* MLC2 cardiac 'atrial' isoform (MLC2A) is encoded by *MYL7* and, during development, is expressed in all heart chambers while in the postnatal heart it becomes restricted to the atrium and pacemaking centers. The cardiac ventricular isoform (MLC2V), encoded by *MYL2*, is only expressed in the ventricle and considered a maturation marker for ventricular CMs (Bedada, Wheelwright, and Metzger 2016; Karbassi et al. 2020; Denning et al. 2016; Yoshida and Yamanaka 2017).

Table I.1. Differences between hiPSC-CMs and human adult cardiomyocytes (Veerman et al. 2015; Bedada et al. 2014; Cai et al. 2019; Lundy et al. 2013; Karbassi et al. 2020; Yoshida and Yamanaka 2017; Denning et al. 2016).

	iPSC-CMs	Adult Cardiomyocytes
Morphology		
Shape	Round or polygonal	Rod and elongated
Size	20–30 μm	150 μm
Nuclei per cell	Mononucleated	≈25% multinucleated
Multicellular organization	Disorganized	Polarized
Sarcomere appearance	Disorganized	Organized
Sarcomere length	Shorter (≈1.6 μm)	Longer (≈2.2 μm)
Sarcomere units	Z lines and I-bands	I-, H-, A-bands, and M- Z- lines
T tubules	Absent	Present
Distribution of gap junctions	Circumferential	Polarized to intercalated disks
Aspect ratio	5–9.5:1	2–3:1
Calcium handling		
Ca^{2+} transient	Inefficient	Efficient
Amplitudes of Ca^{2+} transient	Small; decreases with pacing	Increases with pacing
Adrenergic stimulation response	Lack of inotropic reaction	Inotropic reaction
Metabolism		
Mitochondria	Low; perinuclear	High
Metabolic substrate	Glycolysis (glucose)	Oxidative (fatty acid)
Electrophysiology		
Resting membrane potential	≈ –60 mV	≈ –90 mV
Upstroke velocity	≈50 V/s	≈250 V/s
Amplitude	Small	Large
Contractility		
Contractile force	≈ nN range/cell	≈ μN range/cell
Spontaneous beating	Exhibited	Absent
Gene and protein expression		
Sarcomeric gene or protein expression	Low expression β MHC > α MHC ssTnI (*TNNI1*)	*ACTN2, TNNT2, MYL2, MYH7, MYPC3* β MHC >> α MHC cTnI (*TNNI3*)
Ca^{2+}-handling proteins	Low or absent	*CASQ2, RYR2, PLN*
Ion channels	Lower	SCN5A (I_{Na}); KCNJ2 (I_{K1})

Other important isoforms for maturation assessment are the developmental splicing isoforms of *TTN* and *TNNT2.*

Cardiac alternative splicing developmental isoforms

Over 90% of human genes are known to express several different mRNAs through the process of alternative splicing, which results in proteomic diversity (Zhu, Chen, and Guo 2017). Since alternative splicing exons can be either included or excluded from the mature form of mRNA, the structure of the protein may be altered, with possible impacts on its location, regulation and ultimately, function (Beqqali 2006; Lara-Pezzi et al. 2013; Zhu, Chen, and Guo 2017).

During cardiac development, the alternative splicing pattern of pre-mRNAs leads to an isoform switch that results in the expression of slightly different proteins, in order to adapt to changes in physiological conditions that cardiomyocytes suffer along its maturation, such as drastically different postnatal oxygen and metabolic conditions (Baralle and Giudice 2017). Overall, this programme of alternative splicing and isoform switching supports the postnatal growth and maturation of the heart. In heart disease, and more specifically in cardiomyopathies, several studies have reported that upon stress there is a reversion from adult to fetal alternative splicing patterns, accompanied by a return to neonatal oxygen consumption and metabolic programmes, and by severe alterations in the internal architecture and functionalities of cardiomyocytes (Weeland et al. 2015; Baralle and Giudice 2017).

In the case of titin and troponin T, it has been reported that different protein isoforms are translated from a single pre-mRNA due to alternative splicing (Zhu, Chen, and Guo 2017; Weeland et al. 2015). Titin, one of the most well described cases, is the protein responsible for the maintenance of passive tension in cardiomyocytes. It plays an essential role in sustaining the structural integrity of the sarcomere, mainly due its size and elasticity. Different titin isoforms are produced through alternative splicing of a single *TTN* transcript. In fetal hearts, a large isoform denominated as *N2BA* is expressed. However, after birth and throughout development of the

adult heart, a smaller and stiffer titin isoform, the *N2B*, gradually replaces the large *N2BA* one. The *N2B* is the predominant isoform in the normal adult human heart, being expressed in left ventricle in a ratio of 65% (Zhu, Chen, and Guo 2017; Weeland et al. 2015). The ratio of *N2BA* to *N2B* determine sarcomere length and cardiomyocyte passive tension and, consequently, myocardium wall stiffness during ventricular filling (Baralle and Giudice 2017). Alterations of the *N2B*/*N2BA* expression ratio has been associated to heart disease, and the shift towards the expression of the *N2BA* fetal isoform in adult hearts was found mainly associated with dilated cardiomyopathy (Zhu, Chen, and Guo 2017).

Troponin T (TnT) is part of the troponin complex, a calcium-sensitive regulator of muscle contraction, and mediates the interaction between this complex and tropomyosin. TnT has three muscle type-specific isoforms encoded by different genes. *TNNT2* encodes cardiac troponin T (cTnT), the only present in cardiac muscle (Wei and Jin 2016). In the human heart, cTnT has at least four known alternative splicing isoforms (cTnT1, cTnT2, cTnT3 and cTnT4) and their switch is thought to be regulated by a genetically programmed biological clock (Yin, Ren, and Guo 2015). cTnT1 (all exons present) and cTnT2 (lacking exon 4) isoforms are expressed in fetal heart. During heart development, cTnT1 expression level decreases, whereas expression of cTnT3 (lacking exon 5) increases becoming the predominant isoform in the normal adult heart. cTnT4 isoform (lacking both exon 4 and 5) is expressed in the fetal heart and re-expressed in the failing adult heart (Anderson et al. 1995; Zhu, Chen, and Guo 2017; Yin, Ren, and Guo 2015). The expression of exon 5 in the fetal cTnT isoform helps with the tolerance of cardiac muscle to acidosis since it encodes a segment of the N-terminal variable region that is highly acidic and negatively charged at physiological pH. Moreover, embryonic cTnT's exert higher calcium sensitivity of actomyosin ATPase activity and myofilament force production, compared with the adult cTnT3 (Wei and Jin 2016). Altered splicing patterns of cTnT with a shift toward the fetal isoforms has been described in animal models and adult human heart associated with hypertrophy and cardiac failure. This abnormal combined expression of embryonic and adult cTnT

leads to desynchronized calcium activation of actin filament (Yin, Ren, and Guo 2015; Wei and Jin 2016). Besides, the re-expression of fetal isoforms, with higher Ca^{2+} sensitivity than cTnT3, may increase susceptibility to arrhythmia (Zhu, Chen, and Guo 2017). Possibly in line with this findings, HCM due to *TNNT2* mutations have been clinically characterized as having increased risk of SCD even with modest amounts of LVH (Paranal et al. 2020).

It is also important to be aware that some parameters differ depending on CM subtype (ventricular, atrial or nodal) or can be altered in a disease setting such as the β: α MHC (*MYH7/MYH6)* ratio. To address this issue, a panel of different markers, including some that do not suffer from such fluctuations, such as the *TNNI1* to *TNNI3* switch, should be used (Bedada, Wheelwright, and Metzger 2016; Cai et al. 2019) .

Maturation strategies

In order to achieve *in vitro* differentiated hiPSC-CMs that better resemble adult CMs several maturation strategies have already been attempted, with some demonstrated improvements. Nevertheless, a definitive answer to this issue remains an actively pursued goal in the field.

The long time taken by human CMs to mature *in vivo* lead to the hypothesis that an extended time in culture would also promote hiPSC-CMs maturation and so, this was one of the first tried approaches. If fact, with prolonged culture, hiPSC-CMs become larger, elongated and multinucleated cells with improved myofibril alignment, calcium handling and increased action potential amplitudes (Lundy et al. 2013; Bedada et al. 2014; Lewandowski et al. 2018). Moreover, hiPSC-CMs maintained in culture for one year accumulate structural characteristics of mature CMs demonstrating that despite the fact of their global transcriptome remains mostly unchanged after 8 weeks of differentiation, some improvements in maturation status continue to occur for periods of up to nearly a year (Bedada et al. 2014; Kamakura et al. 2013; Kuppusamy et al. 2015; Piccini et al. 2015; M. Zhang et al. 2015). Although, such long periods of culture are not practical

for routine biomedical application, they offer a useful developmental tool that can provide directions for new possible maturation strategies. In Kuppusamy et al., *let-7* was identified as the most highly up-regulated microRNA in mature hESC-CMs and further analyses demonstrated that, when overexpressed in hESC-CMs, it is sufficient to induce maturation enhancing cell size, sarcomere length and force of contraction (Kuppusamy et al. 2015).

The use of biochemical cues to enhance hiPSC-CMs maturity, either by hormone supplementation (i.e. Triiodothyronine) or by alterations of the energy source available from glucose to fatty acids, are straight forward approaches that were able to induce more adult-like cell size, gene expression profile, as well contractile and electrophysiological function (Bedada et al. 2014; Correia et al. 2017; 2018; Karbassi et al. 2020).

As *in vitro* cell culture generally lacks elements from the *in vivo* environment, different observations regarding this aspect lead to alterative maturation strategies. The human heart development is complex and the three-dimensional (3D) configuration of the embryo, with the appropriate spatial, temporal and mechanical cues, is crucial for the success of this process. Therefore, 3D platforms for *in vitro* cardiac differentiation and maturation that aim at better mimicking the embryonic development of the heart *in vitro* have emerged in the past few years (Correia et al. 2018; Branco et al. 2019; M. Zhang et al. 2015). Their application lead to generation of hiPSC-CMs with a higher degree of structural, functional and metabolic maturation when compared to that in 2D cultures (Correia et al. 2018; Ahmed et al. 2020). Moreover, hiPSC-CMs in 3D culture exhibited a faster and more reproductible source of hiPSC-CMs when compared with similar but 2D approaches (Branco et al. 2019; Correia et al. 2018).

In parallel, the observation that hiPSC-CMs undergo structural and functional maturation when transplanted into the myocardium of model species lead to the hypothesis that CMs can mature when placed in an appropriate milieu (Laflamme et al. 2007; Denning et al. 2016). Therefore, different CMs maturation approaches arose evaluating physical, chemical, genetic and environmental factors in order to facilitate their maturation. Such approaches have mainly relied

in altering the stiffness or morphology of the growth substrate and have shown that polyacrylamide hydrogels with higher degrees of stiffness improved the morphology of CMs and increased contractile stress; also, the use of adhesive micropatterns or nanogrooves, which promotes natural features in a 2D environment, also showed improvements in sarcomere alignment and organization and Ca^{2+} handling (Musunuru et al. 2018). Interestingly, the combination of micropatterned polyacrylamide hydrogels of stiffness similar to that found in myocardial tissue generated differentiated cells with similarities to adult CMs with respect to contractile activity, myofibril alignment, electrophysiology, direction of calcium flow, organization of mitochondria and T- tubules formation (Ribeiro et al. 2015). Lastly, an easy-to-prepare Matrigel mattress could promote rod-shaped morphologies, increased sarcomere length, contractility and expression of maturation markers in iPSC-CMs (Feaster et al. 2015).

Another interesting, although complex, approach that takes advantage of modulating the biophysical cues provided by the cell culture is the generation of engineered heart tissue (EHT). EHTs typically uses a scaffold that mimics some aspects of cardiac extracellular matrix combined with the differentiated CMs and, in some cases, support cells (Denning et al., 2016; Ahmed et al., 2020). This approach is even more powerful when combined with mechanical and electrical stimuli and as demonstrated the possibility to generate cardiac tissues from hPSC-CMs that reach a maturation level with structural and functional characteristics closer to human adult tissue (Ronaldson-Bouchard et al. 2018; 2019).

HCM hiPSC-CM disease models

The ability to culture hiPSC-CMs, combined with improvements in hiPSC reprograming and gene-editing technologies, is rapidly emerging as a central approach to model heart diseases *in*

vitro, with over 90 studies using hiPSCs derived cardiomyocytes (hiPSC-CMs) for cardiac disease modeling being published until date (van Mil et al. 2018). HCM disease modelling using hiPSC-CMs as also been explored in different studies and was already proven to offer a pathophysiologic relevant approach to understand the mechanisms of disease (Musunuru et al. 2018; Giacomelli, Mummery, and Bellin 2017; van Mil et al. 2018; Ojala and Aalto-Setälä 2016).

In 2013, Lan and colleagues demonstrated that iPSC-CMs derived from HCM patients can recapitulate HCM-specific features at the single-cell level, namely cellular enlargement and multinucleation, disarray of sarcomere organization, increased β:αMHC ratio, alterations in contractility, electrophysiological proarrhythmic phenotype and abnormal Ca^{2+}-handling (Lan et al. 2013).

Building on this knowledge, several studies have since been published using hiPSC-CMs for HCM disease modeling, either using PD-hiPSCs (Lan et al. 2013; Liang et al. 2013; Han et al. 2014; Ojala et al. 2016; Tanaka et al. 2014; Birket et al. 2015; Ma et al. 2018b; Prondzynski et al. 2017a) or isogenic gene-edited hiPSCs (Ma et al. 2018b; Mosqueira et al. 2018; L. Wang et al. 2018). Mutations accessed in these studies were mainly located in the *MYBPC3* (Ojala et al. 2016; Birket et al. 2015; Prondzynski et al. 2017a; Ma et al. 2018b) or *MYH7* genes (Lan et al. 2013; Liang et al. 2013; Han et al. 2014; Mosqueira et al. 2018), while two mutations in *TMP1* (Tanaka et al. 2014; Ojala et al. 2016), one in *MYL3* (Ma et al. 2018b) and one in *TNNT2* (L. Wang et al. 2018) were also reported.

In most of these studies, hiPSC-CMs could reflect some characteristics of HCM cardiomyocytes showing that recapitulation of a general HCM phenotype at the cellular level is irrespective of the use of PD-hiPSCs or gene-edited hiPSCs and may be used as an effective disease model.

Despite providing a basis for the possibility of modeling HCM *in vitro* and some insight regarding mechanisms involved in the pathophysiology of the disease, the significant differences amongst these studies regarding the origin of the cells, methods used for differentiation and

parameters analyzed, confound their interpretation and comparison. Importantly, a recent high-throughput phenotyping toolkit for characterizing cellular models of HCM *in vitro* was proposed, in order to tackle this issue of heterogeneity among the different CMs differentiation protocols (Mosqueira, Lis-Slimak, and Denning 2019). Nevertheless, as already discussed, the heterogeneity and immature characteristics of iPSC-CMs obtained from established differentiation protocols remains as a concern around the use of hiPSCs in cardiac disease modeling.

I.6. REFERENCES

Aasen, Trond, Angel Raya, Maria J. Barrero, Elena Garreta, Antonella Consiglio, Federico Gonzalez, Rita Vassena, et al. 2008. “Efficient and Rapid Generation of Induced Pluripotent Stem Cells from Human Keratinocytes.” *Nature Biotechnology* 26 (11): 1276–84. https://doi.org/10.1038/nbt.1503.

Abramowicz, Anna, and Monika Gos. 2018. “Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation.” *Journal of Applied Genetics* 59 (3): 253–68. https://doi.org/10.1007/s13353-018-0444-7.

Ahmed, Razan Elfadil, Tatsuya Anzai, Nawin Chanthra, and Hideki Uosaki. 2020. “A Brief Review of Current Maturation Methods for Human Induced Pluripotent Stem Cells-Derived Cardiomyocytes.” *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A. https://doi.org/10.3389/fcell.2020.00178.

Anderson, Page A.W., Ann Greig, Trisha M. Mark, Nadia N. Malouf, Annette E. Oakeley, Ross M. Ungerleider, Paul D. Allen, and Brian K. Kay. 1995. “Molecular Basis of Human Cardiac Troponin T Isoforms Expressed in the Developing, Adult, and Failing Heart.” *Circulation Research* 76 (4): 681–86. https://doi.org/10.1161/01.RES.76.4.681.

Bagnall, Richard D, Jodie Ingles, Marcel E Dinger, Mark J Cowley, Samantha Barratt Ross, André E Minoche, Sean Lal, et al. 2018. “Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy.” *Journal of the American College of Cardiology* 72 (4): 419–29. https://doi.org/10.1016/j.jacc.2018.04.078.

Ban, Kiwon, Seongho Bae, and Young-sup Yoon. 2017. “Current Strategies and Challenges for Purification of Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Theranostics* 7 (7): 2067–77. https://doi.org/10.7150/thno.19427.

Baralle, Francisco E., and Jimena Giudice. 2017. “Alternative Splicing as a Regulator of Development and Tissue Identity.” *Nature Reviews Molecular Cell Biology* 18 (7): 437. https://doi.org/10.1038/nrm.2017.27.

Barash, Yoseph, John A. Calarco, Weijun Gao, Qun Pan, Xinchen Wang, Ofer Shai, Benjamin J. Blencowe, and Brendan J. Frey. 2010. “Deciphering the Splicing Code.” *Nature* 465 (7294): 53–59. https://doi.org/10.1038/nature09000.

Bedada, Fikru B., Sunny S.K. Chan, Stefania K. Metzger, Liying Zhang, Jianyi Zhang, Daniel J. Garry, Timothy J. Kamp, Michael Kyba, and Joseph M. Metzger. 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. https://doi.org/10.1016/j.stemcr.2014.07.012.

Bedada, Fikru B., Matthew Wheelwright, and Joseph M. Metzger. 2016. “Maturation Status of Sarcomere Structure and Function in Human IPSC-Derived Cardiac Myocytes.” *Biochimica et Biophysica Acta - Molecular Cell Research* 1863 (7): 1829–38. https://doi.org/10.1016/j.bbamcr.2015.11.005.

Bellin, Milena, Maria C. Marchetto, Fred H. Gage, and Christine L. Mummery. 2012. “Induced Pluripotent Stem Cells: The New Patient?” *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3448.

Beqqali, A. 2006. “Alternative Splicing in Heart Disease” 2700: 1–13.

Berger, Adeline, Séverine Maire, Marie Claude Gaillard, José Alain Sahel, Philippe Hantraye, and Alexis Pierre Bemelmans. 2016. “MRNA Trans-Splicing in Gene Therapy for Genetic Diseases.” *Wiley Interdisciplinary Reviews: RNA*. Blackwell Publishing Ltd. https://doi.org/10.1002/wrna.1347.

Birket, Matthew J., Marcelo C. Ribeiro, Georgios Kosmidis, Dorien Ward, Ana Rita Leitoguinho, Vera van de Pol, Cheryl Dambrot, et al. 2015. “Contractile Defect Caused by Mutation in MYBPC3 Revealed under Conditions Optimized for Human PSC-Cardiomyocyte Function.” *Cell Reports* 13 (4): 733–45.

https://doi.org/10.1016/j.celrep.2015.09.025.

Branco, Mariana A., João P. Cotovio, Carlos A.V. Rodrigues, Sandra H. Vaz, Tiago G. Fernandes, Leonilde M. Moreira, Joaquim M.S. Cabral, and Maria Margarida Diogo. 2019. “Transcriptomic Analysis of 3D Cardiac Differentiation of Human Induced Pluripotent Stem Cells Reveals Faster Cardiomyocyte Maturation Compared to 2D Culture.” *Scientific Reports* 9 (1): 1–13. https://doi.org/10.1038/s41598-019-45047-9.

Brito, Dulce, Gabriel Miltenberger-Miltenyi, Sónia Vale Pereira, Doroteia Silva, António Nunes Diogo, and Hugo Madeira. 2012. “Sarcomeric Hypertrophie Cardiomyopathy: Genetic Profile in a Portuguese Population.” *Revista Portuguesa de Cardiologia* 31 (9): 577–87. https://doi.org/10.1016/j.repc.2011.12.020.

Burridge, Paul W., Elena Matsa, Praveen Shukla, Ziliang C. Lin, Jared M. Churko, Antje D. Ebert, Feng Lan, et al. 2014. “Chemically Defined Generation of Human Cardiomyocytes.” *Nature Methods* 11 (8): 855–60. https://doi.org/10.1038/nmeth.2999.

Cai, Wenxuan, Jianhua Zhang, Willem J. De Lange, Zachery R. Gregorich, Hannah Karp, Emily T. Farrell, Stanford D. Mitchell, et al. 2019. “An Unbiased Proteomics Method to Assess the Maturation of Human Pluripotent Stem Cell-Derived Cardiomyocytes.” *Circulation Research* 125 (11): 936–53. https://doi.org/10.1161/CIRCRESAHA.119.315305.

Cartegni, Luca, Shern L. Chew, and Adrian R. Krainer. 2002. “Listening to Silence and Understanding Nonsense: Exonic Mutations That Affect Splicing.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg775.

Chakarova, C. F. 2002. “Mutations in HPRP3, a Third Member Ofpre-MRNA Splicing Factor Genes, Implicated in Autosomal Dominant Retinitis Pigmentosa.” *Human Molecular Genetics* 11 (1): 87–92. https://doi.org/10.1093/hmg/11.1.87.

Chang, Yao-Fu, J. Saadi Imam, and Miles F. Wilkinson. 2007. “The Nonsense-Mediated Decay RNA Surveillance Pathway.” *Annual Review of Biochemistry* 76 (1): 51–74. https://doi.org/10.1146/annurev.biochem.76.050106.093909.

Charleston, Jay S., Frederick J. Schnell, Johannes Dworzak, Cas Donoghue, Sarah Lewis, Lei Chen, G. David Young, et al. 2018. “Eteplirsen Treatment for Duchenne Muscular Dystrophy.” *Neurology* 90 (24): e2135–45. https://doi.org/10.1212/WNL.0000000000005680.

Charron, Philippe, Michael Arad, Eloisa Arbustini, Cristina Basso, Zofia Bilinska, Perry Elliott, Tiina Helio, et al. 2010. “Genetic Counselling and Testing in Cardiomyopathies: A Position Statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases.” *European Heart Journal* 31 (22): 2715–28. https://doi.org/10.1093/eurheartj/ehq271.

Choi, Jiho, Soohyun Lee, William Mallard, Kendell Clement, Guidantonio Malagoli Tagliazucchi, Hotae Lim, In Young Choi, et al. 2015. “A Comparison of Genetically Matched Cell Lines Reveals the Equivalence of Human IPSCs and ESCs.” *Nature Biotechnology* 33 (11): 1173–81. https://doi.org/10.1038/nbt.3388.

Christiaans, Imke, and Perry M. Elliott. 2016. “Hypertrophic Cardiomyopathy.” In *Clinical Cardiogenetics*, 61–74. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44203-7_4.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. “Multiplex Genome Engineering Using CRISPR/Cas Systems.” *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Correia, Cláudia, Alexey Koshkin, Patrícia Duarte, Dongjian Hu, Madalena Carido, Maria J. Sebastião, Patrícia Gomes-Alves, et al. 2018. “3D Aggregate Culture Improves Metabolic Maturation of Human Pluripotent Stem Cell Derived Cardiomyocytes.” *Biotechnology and Bioengineering* 115 (3): 630–44. https://doi.org/10.1002/bit.26504.

Correia, Cláudia, Alexey Koshkin, Patrícia Duarte, Dongjian Hu, Ana Teixeira, Ibrahim Domian, Margarida Serra, and Paula M. Alves. 2017. “Distinct Carbon Sources Affect Structural and Functional Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Scientific Reports* 7 (1). https://doi.org/10.1038/s41598-017-08713-4.

Davidson, Mercy M., Claudia Nesti, Lluis Palenzuela, Winsome F. Walker, Evelyn Hernandez, Lev Protas, Michio Hirano, and Nithila D. Isaac. 2005. “Novel Cell Lines Derived from Adult Human Ventricular Cardiomyocytes.” *Journal of Molecular and Cellular Cardiology* 39 (1): 133–47. https://doi.org/10.1016/j.yjmcc.2005.03.003.

Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jörg Vogel, and Emmanuelle Charpentier. 2011. “CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III.” *Nature* 471 (7340): 602–7. https://doi.org/10.1038/nature09886.

Denning, Chris, Viola Borgdorff, James Crutchley, Karl S.A. Firth, Vinoj George, Spandan Kalra, Alexander Kondrashov, et al. 2016. “Cardiomyocytes from Human Pluripotent Stem Cells: From Laboratory Curiosity to Industrial Biomedical Platform.” *Biochimica et Biophysica Acta - Molecular Cell Research* 1863 (7): 1728–48. https://doi.org/10.1016/j.bbamcr.2015.10.014.

Dijk, Sabine J. van, Dennis Dooijes, Cris dos Remedios, Michelle Michels, Jos M.J. Lamers, Saul Winegrad, Saskia Schlossarek, et al. 2009a. “Cardiac Myosin-Binding Protein C Mutations and Hypertrophic Cardiomyopathy.” *Circulation* 119 (11): 1473–83. https://doi.org/10.1161/CIRCULATIONAHA.108.838672.

Dijk, Sabine J van, Dennis Dooijes, Cris dos Remedios, Michelle Michels, Jos M J Lamers, Saul Winegrad, Saskia Schlossarek, et al. 2009b. “Cardiac Myosin-Binding Protein C Mutations and Hypertrophic Cardiomyopathy: Haploinsufficiency, Deranged Phosphorylation, and Cardiomyocyte Dysfunction.” *Circulation* 119 (11): 1473–83. https://doi.org/10.1161/CIRCULATIONAHA.108.838672.

Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. “The New Frontier of Genome Engineering with CRISPR-Cas9.” *Science* 346 (6213): 1258096. https://doi.org/10.1126/science.1258096.

Driest, Sara L. Van, Steve R. Ommen, A. Jamil Tajik, Bernard J. Gersh, and Michael J. Ackerman. 2005. “Yield of Genetic Testing in Hypertrophic Cardiomyopathy.” *Mayo Clinic Proceedings* 80 (6): 739–44. https://doi.org/10.4065/80.6.739.

Dubois, Nicole C., April M. Craft, Parveen Sharma, David A. Elliott, Edouard G. Stanley, Andrew G. Elefanty, Anthony Gramolini, and Gordon Keller. 2011. “SIRPA Is a Specific Cell-Surface Marker for Isolating Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Nature Biotechnology* 29 (11): 1011–18. https://doi.org/10.1038/nbt.2005.

Dunn, Kaitlin K., and Sean P. Palecek. 2018. “Engineering Scalable Manufacturing of High-Quality Stem Cell-Derived Cardiomyocytes for Cardiac Tissue Repair.” *Frontiers in Medicine*. Frontiers Media S.A. https://doi.org/10.3389/fmed.2018.00110.

Eiges, Rachel, Achia Urbach, Mira Malcov, Tsvia Frumkin, Tamar Schwartz, Ami Amit, Yuval Yaron, et al. 2007. “Developmental Study of Fragile X Syndrome Using Human Embryonic Stem Cells Derived from Preimplantation Genetically Diagnosed Embryos.” *Cell Stem Cell* 1 (5): 568–77. https://doi.org/10.1016/j.stem.2007.09.001.

Elbashir, Sayda M., Jens Harborth, Winfried Lendeckel, Abdullah Yalcin, Klaus Weber, and Thomas Tuschl. 2001. “Duplexes of 21-Nucleotide RNAs Mediate RNA Interference in Cultured Mammalian Cells.” *Nature* 411 (6836): 494–98. https://doi.org/10.1038/35078107.

Elliott, David A., Stefan R. Braam, Katerina Koutsis, Elizabeth S. Ng, Robert Jenny, Ebba L. Lagerqvist, Christine Biben, et al. 2011. “NKX2-5 EGFP/w HESCs for Isolation of Human Cardiac Progenitors and Cardiomyocytes.” *Nature Methods* 8 (12): 1037–43. https://doi.org/10.1038/nmeth.1740.

Elliott, Perry M, Aris Anastasakis, Michael A. Borger, Martin Borggrefe, Franco Cecchi, Philippe Charron,

Albert Alain Hagege, et al. 2014. “2014 ESC Guidelines on Diagnosis and Management of Hypertrophic Cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC).” *European Heart Journal* 35 (39): 2733–79. https://doi.org/10.1093/eurheartj/ehu284.

Erkelenz, Steffen, Stephan Theiss, Marianne Otte, Marek Widera, Jan Otto Peter, and Heiner Schaal. 2014. “Genomic HEXploring Allows Landscaping of Novel Potential Splicing Regulatory Elements.” *Nucleic Acids Research* 42 (16): 10681–97. https://doi.org/10.1093/nar/gku736.

Faustino, Nuno André, and Thomas A. Cooper. 2003. “Pre-MRNA Splicing and Human Disease.” *Genes and Development*. https://doi.org/10.1101/gad.1048803.

Feaster, Tromondae K., Adrian G. Cadar, Lili Wang, Charles H. Williams, Young Wook Chun, Jonathan E. Hempel, Nathaniel Bloodworth, et al. 2015. “A Method for the Generation of Single Contracting Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes.” *Circulation Research* 117 (12): 995–1000. https://doi.org/10.1161/CIRCRESAHA.115.307580.

Ferrantini, Cecilia, Raffaele Coppini, Josè Manuel Pioner, Francesca Gentile, Benedetta Tosi, Luca Mazzoni, Beatrice Scellini, et al. 2017. “Pathogenesis of Hypertrophic Cardiomyopathy Is Mutation Rather Than Disease Specific: A Comparison of the Cardiac Troponin T E163R and R92Q Mouse Models.” *Journal of the American Heart Association* 6 (7). https://doi.org/10.1161/JAHA.116.005407.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. “Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans.” *Nature* 391 (6669): 806–11. https://doi.org/10.1038/35888.

Freeman, Lisa M, John E Rush, Joshua A Stern, Gordon S Huggins, and Martin S Maron. 2017. “Feline Hypertrophic Cardiomyopathy: A Spontaneous Large Animal Model of Human HCM.” *Cardiology Research* 8 (4): 139–42. https://doi.org/10.14740/cr578w.

Frey, Norbert, Mark Luedde, and Hugo A. Katus. 2012. “Mechanisms of Disease: Hypertrophic Cardiomyopathy.” *Nature Reviews Cardiology*. Nature Publishing Group. https://doi.org/10.1038/nrcardio.2011.159.

Fusaki, Noemi, Hiroshi Ban, Akiyo Nishiyama, Koichi Saeki, and Mamoru Hasegawa. 2009. “Efficient Induction of Transgene-Free Human Pluripotent Stem Cells Using a Vector Based on Sendai Virus, an RNA Virus That Does Not Integrate into the Host Genome.” *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* 85 (8): 348–62. https://doi.org/10.2183/pjab.85.348.

Gedicke-Hornung, Christina, Verena Behrens-Gawlik, Silke Reischmann, Birgit Geertz, Doreen Stimpel, Florian Weinberger, Saskia Schlossarek, et al. 2013. “Rescue of Cardiomyopathy through U7snRNA-Mediated Exon Skipping in Mybpc3-Targeted Knock-in Mice.” *EMBO Molecular Medicine* 5 (7): 1060–77. https://doi.org/10.1002/emmm.201202168.

Geske, Jeffrey B., Steve R. Ommen, and Bernard J. Gersh. 2018. “Hypertrophic Cardiomyopathy: Clinical Update.” *JACC: Heart Failure* 6 (5): 364–75. https://doi.org/10.1016/j.jchf.2018.02.010.

Giacomelli, E., C. L. Mummery, and M. Bellin. 2017. “Human Heart Disease: Lessons from Human Pluripotent Stem Cell-Derived Cardiomyocytes.” *Cellular and Molecular Life Sciences* 74 (20): 3711–39. https://doi.org/10.1007/s00018-017-2546-5.

Giacomo, Daniela Di, Pascaline Gaildrat, Anna Abuli, Julie Abdat, Thierry Frébourg, Mario Tosi, and Alexandra Martins. 2013. “Functional Analysis of a Large Set of Brca2 Exon 7 Variants Highlights the Predictive Value of Hexamer Scores in Detecting Alterations of Exonic Splicing Regulatory Elements.” *Human Mutation* 34 (11): 1547–57. https://doi.org/10.1002/humu.22428.

Girolami, Francesca, Carolyn Y. Ho, Christopher Semsarian, Massimo Baldi, Melissa L. Will, Katia Baldini, Francesca Torricelli, et al. 2010. “Clinical Features and Outcome of Hypertrophic Cardiomyopathy Associated With Triple Sarcomere Protein Gene Mutations.” *Journal of the American College of Cardiology* 55 (14): 1444–53. https://doi.org/10.1016/j.jacc.2009.11.062.

Glazier, Amelia A., Andrea Thompson, and Sharlene M. Day. 2019. “Allelic Imbalance and Haploinsufficiency in MYBPC3-Linked Hypertrophic Cardiomyopathy.” *Pflugers Archiv European Journal of Physiology* 471 (5): 781–93. https://doi.org/10.1007/s00424-018-2226-9.

Grazioli, Gonzalo, Domingo Usín, Emilce Trucco, Maria Sanz, Silvia Montserrat, Bàrbara Vidal, Josep Gutierrez, et al. 2016. “Differentiating Hypertrophic Cardiomyopathy from Athlete’s Heart: An Electrocardiographic and Echocardiographic Approach.” *Journal of Electrocardiology* 49 (4): 539–44. https://doi.org/10.1016/j.jelectrocard.2016.03.005.

Greaves, S. C., A. H.G. Roche, J. M. Neutze, R. M.L. Whitlock, and A. M.O. Veale. 1987. “Inheritance of Hypertrophic Cardiomyopathy: A Cross Sectional and M Mode Echocardiographic Study of 50 Families.” *Heart* 58 (3): 259–66. https://doi.org/10.1136/hrt.58.3.259.

Gruner, Christiane, Joan Ivanov, Melanie Care, Lynne Williams, Gil Moravsky, Hua Yang, Balint Laczay, Katherine Siminovitch, Anna Woo, and Harry Rakowski. 2013. “Toronto Hypertrophic Cardiomyopathy Genotype Score for Prediction of a Positive Genotype in Hypertrophic Cardiomyopathy.” *Circulation: Cardiovascular Genetics* 6 (1): 19–26. https://doi.org/10.1161/CIRCGENETICS.112.963363.

Halevy, Tomer, and Achia Urbach. 2014. “Comparing ESC and IPSC—Based Models for Human Genetic Disorders.” *Journal of Clinical Medicine* 3 (4): 1146–62. https://doi.org/10.3390/jcm3041146.

Han, Lu, Yang Li, Jason Tchao, Aaron D. Kaplan, Bo Lin, You Li, Jocelyn Mich-Basso, et al. 2014. “Study Familial Hypertrophic Cardiomyopathy Using Patient-Specific Induced Pluripotent Stem Cells.” *Cardiovascular Research* 104 (2): 258–69. https://doi.org/10.1093/cvr/cvu205.

Hartmannova, Hana, Milos Kubanek, Marek Sramko, Lenka Piherova, Lenka Noskova, Katerina Hodanova, Viktor Stranecky, et al. 2013. “Isolated X-Linked Hypertrophic Cardiomyopathy Caused by a Novel Mutation of the Four-and-a-Half LIM Domain 1 Gene.” *Circulation: Cardiovascular Genetics* 6 (6): 543–51. https://doi.org/10.1161/CIRCGENETICS.113.000245.

Hearse, David J., and Fiona J. Sutherland. 2000. “Experimental Models for the Study of Cardiovascular Function and Disease.” *Pharmacological Research* 41 (6): 597–603. https://doi.org/10.1006/phrs.1999.0651.

Heitner, Stephen B. 2018. “Novel Medical Therapeutics for Hypertrophic Cardiomyopathy.” In *Hypertrophic Cardiomyopathy*, edited by S S Naidu, 383–88. Springer International Publishing. https://doi.org/doi.org/10.10007/978-3-319-92423-6_27.

Helms, Adam S, Frank M Davis, David Coleman, Sarah N Bartolone, Amelia A Glazier, Francis Pagani, Jaime M Yob, et al. 2014. “Sarcomere Mutation-Specific Expression Patterns in Human Hypertrophic Cardiomyopathy.” *Circulation. Cardiovascular Genetics* 7 (4): 434–43. https://doi.org/10.1161/CIRCGENETICS.113.000448.

Ho, Carolyn Y, Sharlene M Day, Euan A Ashley, Michelle Michels, Alexandre C Pereira, Daniel Jacoby, Allison L Cirino, et al. 2018. “Genotype and Lifetime Burden of Disease in Hypertrophic Cardiomyopathy: Insights from the Sarcomeric Human Cardiomyopathy Registry (SHaRe).” *Circulation* 138 (14): 1387–98. https://doi.org/10.1161/CIRCULATIONAHA.117.033200.

Hoes, Martijn F., Nils Bomer, and Peter van der Meer. 2019. “The Current State of Human In Vitro Cardiac Disease Modeling: A Focus on Gene Editing and Tissue Engineering.” *STEM CELLS Translational Medicine* 8 (1): 66–74. https://doi.org/10.1002/sctm.18-0052.

Houser, Steven R., Kenneth B. Margulies, Anne M. Murphy, Francis G. Spinale, Gary S. Francis, Sumanth D. Prabhu, Howard A. Rockman, et al. 2012. “Animal Models of Heart Failure.” *Circulation Research* 111 (1): 131–50. https://doi.org/10.1161/RES.0b013e3182582523.

Howden, Sara E., Bradley McColl, Astrid Glaser, Jim Vadolas, Steven Petrou, Melissa H. Little, Andrew G. Elefanty, and Edouard G. Stanley. 2016. “A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells.” *Stem Cell Reports* 7 (3): 508–17. https://doi.org/10.1016/j.stemcr.2016.07.001.

Ingles, J., A. Doolan, C. Chiu, J. Seidman, C. Seidman, and C. Semsarian. 2005. “Compound and Double Mutations in Patients with Hypertrophic Cardiomyopathy: Implications for Genetic Testing and Counselling.” *Journal of Medical Genetics*. https://doi.org/10.1136/jmg.2005.033886.

Ingles, Jodie, Charlotte Burns, Richard D. Bagnall, Lien Lam, Laura Yeates, Tanya Sarina, Rajesh Puranik, et al. 2017. “Nonfamilial Hypertrophic Cardiomyopathy: Prevalence, Natural History, and Clinical Implications.” *Circulation: Cardiovascular Genetics* 10 (2): e001620. https://doi.org/10.1161/CIRCGENETICS.116.001620.

Ingles, Jodie, Jennifer Goldstein, Courtney Thaxton, Colleen Caleshu, Edward W. Corty, Stephanie B. Crowley, Kristen Dougherty, et al. 2019. “Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes.” *Circulation: Genomic and Precision Medicine* 12 (2): 57–64. https://doi.org/10.1161/CIRCGEN.119.002460.

Ito, Kaoru, Parth N. Patel, Joshua M. Gorham, Barbara McDonough, Steven R. DePalma, Emily E. Adler, Lien Lam, et al. 2017. “Identification of Pathogenic Gene Mutations in LMNA and MYBPC3 That Alter RNA Splicing.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (29): 7689–94. https://doi.org/10.1073/pnas.1707741114.

Jaganathan, Kishore, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F. McRae, Siavash Fazel Darbandi, David Knowles, Yang I. Li, Jack A. Kosmicki, et al. 2019. “Predicting Splicing from Primary Sequence with Deep Learning.” *Cell* 176 (3): 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015.

James, Jeanne, Yan Zhang, Hanna Osinska, Atsushi Sanbe, Raisa Klevitsky, Timothy E. Hewett, and Jeffrey Robbins. 2000. “Transgenic Modeling of a Cardiac Troponin I Mutation Linked to Familial Hypertrophic Cardiomyopathy.” *Circulation Research* 87 (9): 805–11. https://doi.org/10.1161/01.RES.87.9.805.

Jellis, Christine L, and Milind Y Desai. 2015. “Hypertrophic Cardiomyopathy: Still Connecting the Dots between Genotype and Phenotype.” *Cardiovascular Diagnosis and Therapy* 5 (2): 156–59. https://doi.org/10.3978/j.issn.2223-3652.2015.03.09.

Jiang, Jianming, Hiroko Wakimoto, J. G. Seidman, and Christine E. Seidman. 2013. “Allele-Specific Silencing of Mutant Myh6 Transcripts in Mice Suppresses Hypertrophic Cardiomyopathy.” *Science* 342 (6154): 111–14. https://doi.org/10.1126/science.1236921.

Jimenez-Tellez, Nerea, and Steven C Greenway. 2019. “Cellular Models for Human Cardiomyopathy: What Is the Best Option?” *World Journal of Cardiology* 11 (10): 221–35. https://doi.org/10.4330/wjc.v11.i10.221.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. “A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.” *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Junying, Yu, Hu Kejin, Smuga Otto Kim, Tian Shulan, Ron Stewart, Igor I. Slukvin, and James A. Thomson. 2009. “Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences.” *Science* 324 (5928): 797–801. https://doi.org/10.1126/science.1172482.

Kamakura, Tsukasa, Takeru Makiyama, Kenichi Sasaki, Yoshinori Yoshida, Yimin Wuriyanghai, Jiarong Chen, Tetsuhisa Hattori, et al. 2013. “Ultrastructural Maturation of Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes in a Long-Term Culture.” *Circulation Journal* 77 (5): 1307–14. https://doi.org/10.1253/circj.CJ-12-0987.

Karakikes, Ioannis, Mohamed Ameen, Vittavat Termglinchan, and Joseph C. Wu. 2015. “Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Insights Into Molecular, Cellular, and Functional Phenotypes.” *Circulation Research* 117 (1): 80–88. https://doi.org/10.1161/CIRCRESAHA.117.305365.

Karbassi, Elaheh, Aidan Fenix, Silvia Marchiano, Naoto Muraoka, Kenta Nakamura, Xiulan Yang, and Charles E. Murry. 2020. “Cardiomyocyte Maturation: Advances in Knowledge and Implications for

Regenerative Medicine." *Nature Reviews Cardiology*, no. Box 1. https://doi.org/10.1038/s41569-019-0331-x.

Kassem, Heba Sh, Remon S. Azer, Maha S. Ayad, Sarah Moharem-Elgamal, Gehan Magdy, Ahmed Elguindy, Franco Cecchi, Iacopo Olivotto, and Magdi H. Yacoub. 2013. "Early Results of Sarcomeric Gene Screening from the Egyptian National BA-HCM Program." *Journal of Cardiovascular Translational Research* 6 (1): 65–80. https://doi.org/10.1007/s12265-012-9425-0.

Kehat, Izhak, Dorit Kenyagin-Karsenti, Mirit Snir, Hana Segev, Michal Amit, Amira Gepstein, Erella Livne, Ofer Binah, Joseph Itskovitz-Eldor, and Lior Gepstein. 2001. "Human Embryonic Stem Cells Can Differentiate into Myocytes with Structural and Functional Properties of Cardiomyocytes." *Journal of Clinical Investigation* 108 (3): 407–14. https://doi.org/10.1172/JCI200112131.

Klarich, Kyle W, Christine H Attenhofer Jost, Josepha Binder, Heidi M Connolly, Christopher G Scott, William K Freeman, Michael J Ackerman, Rick A Nishimura, A Jamil Tajik, and Steve R Ommen. 2013. "Risk of Death in Long-Term Follow-up of Patients with Apical Hypertrophic Cardiomyopathy." *American Journal of Cardiology* 111 (12): 1784–91. https://doi.org/10.1016/j.amjcard.2013.02.040.

Konno, Tetsuo, Stephen Chang, Jonathan G. Seidman, and Christine E. Seidman. 2010. "Genetics of Hypertrophic Cardiomyopathy." *Current Opinion in Cardiology* 25 (3): 205–9. https://doi.org/10.1097/HCO.0b013e3283375698.

Kuppusamy, Kavitha T., Daniel C. Jones, Henrik Sperber, Anup Madan, Karin A. Fischer, Marita L. Rodriguez, Lil Pabon, et al. 2015. "Let-7 Family of MicroRNA Is Required for Maturation and Adult-like Metabolism in Stem Cell-Derived Cardiomyocytes." *Proceedings of the National Academy of Sciences of the United States of America* 112 (21): E2785–94. https://doi.org/10.1073/pnas.1424042112.

Laflamme, Michael A., Kent Y. Chen, Anna V. Naumova, Veronica Muskheli, James A. Fugate, Sarah K. Dupras, Hans Reinecke, et al. 2007. "Cardiomyocytes Derived from Human Embryonic Stem Cells in Pro-Survival Factors Enhance Function of Infarcted Rat Hearts." *Nature Biotechnology* 25 (9): 1015–24. https://doi.org/10.1038/nbt1327.

Lan, Feng, Andrew S. Lee, Ping Liang, Veronica Sanchez-Freire, Patricia K. Nguyen, Li Wang, Leng Han, et al. 2013. "Abnormal Calcium Handling Properties Underlie Familial Hypertrophic Cardiomyopathy Pathology in Patient-Specific Induced Pluripotent Stem Cells." *Cell Stem Cell* 12 (1): 101–13. https://doi.org/10.1016/j.stem.2012.10.010.

Lara-Pezzi, Enrique, Jesús Gómez-Salinero, Alberto Gatto, and Pablo García-Pavía. 2013. "The Alternative Heart: Impact of Alternative Splicing in Heart Disease." *Journal of Cardiovascular Translational Research* 6 (6): 945–55. https://doi.org/10.1007/s12265-013-9482-z.

Lek, Monkol, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, Eric Banks, Timothy Fennell, Anne H. O'Donnell-Luria, et al. 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature* 536 (7616): 285–91. https://doi.org/10.1038/nature19057.

Lewandowski, Jarosław, Natalia Rozwadowska, Tomasz J. Kolanowski, Agnieszka Malcher, Agnieszka Zimna, Anna Rugowska, Katarzyna Fiedorowicz, et al. 2018. "The Impact of in Vitro Cell Culture Duration on the Maturation of Human Cardiomyocytes Derived from Induced Pluripotent Stem Cells of Myogenic Origin." *Cell Transplantation* 27 (7): 1047–67. https://doi.org/10.1177/0963689718779346.

Lian, Xiaojun, Cheston Hsiao, Gisela Wilson, Kexian Zhu, Laurie B. Hazeltine, Samira M. Azarin, Kunil K. Raval, Jianhua Zhang, Timothy J. Kamp, and Sean P. Palecek. 2012. "Robust Cardiomyocyte Differentiation from Human Pluripotent Stem Cells via Temporal Modulation of Canonical Wnt Signaling." *Proceedings of the National Academy of Sciences of the United States of America* 109 (27): E1848–57. https://doi.org/10.1073/pnas.1200250109.

Lian, Xiaojun, Jianhua Zhang, Samira M. Azarin, Kexian Zhu, Laurie B. Hazeltine, Xiaoping Bao, Cheston

Hsiao, Timothy J. Kamp, and Sean P. Palecek. 2013. “Directed Cardiomyocyte Differentiation from Human Pluripotent Stem Cells by Modulating Wnt/β-Catenin Signaling under Fully Defined Conditions.” *Nature Protocols* 8 (1): 162–75. https://doi.org/10.1038/nprot.2012.150.

Liang, Ping, Feng Lan, Andrew S. Lee, Tingyu Gong, Veronica Sanchez-Freire, Yongming Wang, Sebastian Diecke, et al. 2013. “Drug Screening Using a Library of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes Reveals Disease-Specific Patterns of Cardiotoxicity.” *Circulation* 127 (16): 1677–91. https://doi.org/10.1161/CIRCULATIONAHA.113.001883.

Lin, Jyh Hung, San Yuan Huang, Wen Chuan Lee, Si Kwang Liu, and Rea Min Chu. 2002. “Echocardiographic Features of Pigs with Spontaneous Hypertrophic Cardiomyopathy.” *Comparative Medicine* 52 (3): 238–42.

Lin, Steven, Brett T. Staahl, Ravi K. Alla, and Jennifer A. Doudna. 2014. “Enhanced Homology-Directed Human Genome Engineering by Controlled Timing of CRISPR/Cas9 Delivery.” *ELife* 3: e04766. https://doi.org/10.7554/eLife.04766.

Liu, Jun Jie, Natalia Orlova, Benjamin L. Oakes, Enbo Ma, Hannah B. Spinner, Katherine L.M. Baney, Jonathan Chuck, et al. 2019. “CasX Enzymes Comprise a Distinct Family of RNA-Guided Genome Editors.” *Nature* 566 (7743): 218–23. https://doi.org/10.1038/s41586-019-0908-x.

Liu, Mingjie, Saad Rehman, Xidian Tang, Kui Gu, Qinlei Fan, Dekun Chen, and Wentao Ma. 2019. “Methodologies for Improving HDR Efficiency.” *Frontiers in Genetics*. Frontiers Media S.A. https://doi.org/10.3389/fgene.2018.00691.

Loh, Yuin Han, Suneet Agarwal, In Hyun Park, Achia Urbach, Hongguang Huo, Garrett C. Heffner, Kitai Kim, Justine D. Miller, Kitwa Ng, and George Q. Daley. 2009. “Generation of Induced Pluripotent Stem Cells from Human Blood.” *Blood* 113 (22): 5476–79. https://doi.org/10.1182/blood-2009-02-204800.

Lopes, Luis R., Anna Zekavati, Petros Syrris, Mike Hubank, Claudia Giambartolomei, Chrysoula Dalageorgou, Sharon Jenkins, William McKenna, Vincent Plagnol, and Perry M. Elliott. 2013. “Genetic Complexity in Hypertrophic Cardiomyopathy Revealed by High-Throughput Sequencing.” *Journal of Medical Genetics* 50 (4): 228–39. https://doi.org/10.1136/jmedgenet-2012-101270.

Los Angeles, Alejandro De, Francesco Ferrari, Ruibin Xi, Yuko Fujiwara, Nissim Benvenisty, Hongkui Deng, Konrad Hochedlinger, et al. 2015. “Hallmarks of Pluripotency.” *Nature*. Nature Publishing Group. https://doi.org/10.1038/nature15515.

Lowey, Susan, Vera Bretton, Peteranne B. Joel, Kathleen M. Trybus, James Gulick, Jeffrey Robbins, Albert Kalganov, Anabelle S. Cornachione, and Dilson E. Rassier. 2018. “Hypertrophic Cardiomyopathy R403Q Mutation in Rabbit β-Myosin Reduces Contractile Function at the Molecular and Myofibrillar Levels.” *Proceedings of the National Academy of Sciences* 115 (44): 11238–43. https://doi.org/10.1073/pnas.1802967115.

Lundy, Scott D., Wei Zhong Zhu, Michael Regnier, and Michael A. Laflamme. 2013. “Structural and Functional Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Stem Cells and Development* 22 (14): 1991–2002. https://doi.org/10.1089/scd.2012.0490.

Ma, Ning, Joe Z. Zhang, Ilanit Itzhaki, Sophia L. Zhang, Haodong Chen, Francois Haddad, Tomoya Kitani, et al. 2018a. “Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells.” *Circulation* 138 (23): 2666–81. https://doi.org/10.1161/CIRCULATIONAHA.117.032273.

———. 2018b. “Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells.” *Circulation* 138 (23): 2666–81. https://doi.org/10.1161/CIRCULATIONAHA.117.032273.

Mahla, Ranjeet Singh. 2016. “Stem Cells Applications in Regenerative Medicine and Disease Therapeutics.” *International Journal of Cell Biology* 2016. https://doi.org/10.1155/2016/6940283.

Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. “RNA-Guided Human Genome Engineering via Cas9.” *Science* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Marian, A J. 2002. “Modifier Genes for Hypertrophic Cardiomyopathy.” *Current Opinion in Cardiology* 17 (3): 242–52. https://doi.org/10.1097/00001573-200205000-00006.

Marian, Ali J., and Eugene Braunwald. 2017a. “Hypertrophic Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy.” *Circulation Research* 121 (7): 749–70.

———. 2017b. “Hypertrophic Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy.” *Circulation Research* 121 (7): 749–70. https://doi.org/10.1161/CIRCRESAHA.117.311059.

Maron, Barry J. 2010. “Contemporary Insights and Strategies for Risk Stratification and Prevention of Sudden Death in Hypertrophic Cardiomyopathy.” *Circulation*. https://doi.org/10.1161/CIRCULATIONAHA.109.878579.

———. 2018. “Clinical Course and Management of Hypertrophic Cardiomyopathy.” *New England Journal of Medicine* 379 (7): 655–68. https://doi.org/10.1056/nejmra1710575.

Maron, Barry J., Julius M. Gardin, John M. Flack, Samuel S. Gidding, Tom T. Kurosaki, and Diane E. Bild. 1995. “Prevalence of Hypertrophic Cardiomyopathy in a General Population of Young Adults: Echocardiographic Analysis of 4111 Subjects in the CARDIA Study.” *Circulation* 92 (4): 785–89. https://doi.org/10.1161/01.CIR.92.4.785.

Maron, Barry J., Martin S. Maron, and Christopher Semsarian. 2012. “Genetics of Hypertrophic Cardiomyopathy after 20 Years: Clinical Perspectives.” *Journal of the American College of Cardiology*. Elsevier Inc. https://doi.org/10.1016/j.jacc.2012.02.068.

Maron, Barry J., and Martin S Maron. 2013. “Hypertrophic Cardiomyopathy.” *The Lancet* 381 (9862): 242–55. https://doi.org/10.1016/S0140-6736(12)60397-3.

Maron, Barry J., Steve R. Ommen, Christopher Semsarian, Paolo Spirito, Iacopo Olivotto, and Martin S. Maron. 2014. “Hypertrophic Cardiomyopathy: Present and Future, with Translation into Contemporary Cardiovascular Medicine.” *Journal of the American College of Cardiology*. Elsevier USA. https://doi.org/10.1016/j.jacc.2014.05.003.

Maron, Barry J., Ethan J. Rowin, Susan A. Casey, Mark S. Link, John R. Lesser, Raymond H.M. Chan, Ross F. Garberich, James E. Udelson, and Martin S. Maron. 2015. “Hypertrophic Cardiomyopathy in Adulthood Associated with Low Cardiovascular Mortality with Contemporary Management Strategies.” *Journal of the American College of Cardiology* 65 (18): 1915–28. https://doi.org/10.1016/j.jacc.2015.02.061.

Maron, Barry J., Ethan J. Rowin, and Martin S. Maron. 2018. “Global Burden of Hypertrophic Cardiomyopathy.” *JACC: Heart Failure* 6 (5): 376–78. https://doi.org/10.1016/j.jchf.2018.03.004.

Marston, Steven, O’Neal Copeland, Adam Jacques, Karen Livesey, Victor Tsang, William J. McKenna, Shapour Jalilzadeh, Sebastian Carballo, Charles Redwood, and Hugh Watkins. 2009. “Evidence from Human Myectomy Samples That MYBPC3 Mutations Cause Hypertrophic Cardiomyopathy through Haploinsufficiency.” *Circulation Research* 105 (3): 219–22. https://doi.org/10.1161/CIRCRESAHA.109.202440.

Maruyama, Takeshi, Stephanie K. Dougan, Matthias C. Truttmann, Angelina M. Bilate, Jessica R. Ingram, and Hidde L. Ploegh. 2015. “Increasing the Efficiency of Precise Genome Editing with CRISPR-Cas9 by Inhibition of Nonhomologous End Joining.” *Nature Biotechnology* 33 (5): 538–42. https://doi.org/10.1038/nbt.3190.

McKie, A. B. 2001. “Mutations in the Pre-MRNA Splicing Factor Gene PRPC8 in Autosomal Dominant Retinitis Pigmentosa (RP13).” *Human Molecular Genetics* 10 (15): 1555–62.

https://doi.org/10.1093/hmg/10.15.1555.

Mearini, Giulia, Doreen Stimpel, Elisabeth Krämer, Birgit Geertz, Ingke Braren, Christina Gedicke-Hornung, Guillaume Précigout, et al. 2013. “Repair of Mybpc3 MRNA by 5′-Trans-Splicing in a Mouse Model of Hypertrophic Cardiomyopathy.” *Molecular Therapy - Nucleic Acids* 2. https://doi.org/10.1038/mtna.2013.31.

Melacini, Paola, Cristina Basso, Annalisa Angelini, Chiara Calore, Fabiana Bobbo, Barbara Tokajuk, Nicoletta Bellini, et al. 2010. “Clinicopathological Profiles of Progressive Heart Failure in Hypertrophic Cardiomyopathy.” *European Heart Journal* 31 (17): 2111–23. https://doi.org/10.1093/eurheartj/ehq136.

Meurs, Kathryn M, Ximena Sanchez, Ryan M David, Neil E Bowles, Jeffrey A Towbin, Peter J Reiser, Judith A. Kittleson, et al. 2005. “A Cardiac Myosin Binding Protein C Mutation in the Maine Coon Cat with Familial Hypertrophic Cardiomyopathy.” *Human Molecular Genetics* 14 (23): 3587–93. https://doi.org/10.1093/hmg/ddi386.

Michelson, David, Emma Ciafaloni, Stephen Ashwal, Elliot Lewis, Pushpa Narayanaswami, Maryam Oskoui, and Melissa J. Armstrong. 2018. “Evidence in Focus: Nusinersen Use in Spinal Muscular Atrophy Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology.” *Neurology* 91 (20): 923–33. https://doi.org/10.1212/WNL.0000000000006502.

Mil, Alain van, Geerthe Margriet Balk, Klaus Neef, Jan Willem Buikema, Folkert W. Asselbergs, Sean M. Wu, Pieter A. Doevendans, and Joost P.G. Sluijter. 2018. “Modelling Inherited Cardiac Disease Using Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Progress, Pitfalls, and Potential.” *Cardiovascular Research* 114 (14): 1828–42. https://doi.org/10.1093/cvr/cvy208.

Milani-Nejad, Nima, and Paul M.L. Janssen. 2014. “Small and Large Animal Models in Cardiac Contraction Research: Advantages and Disadvantages.” *Pharmacology and Therapeutics*. Elsevier Inc. https://doi.org/10.1016/j.pharmthera.2013.10.007.

Ming, Su, Wang Shui-yun, Qiu Wei, Li Jian-hui, Hui Ru-tai, Song Lei, Jia Mei, Wang Hui, and Wang Ji-zheng. 2018. “MIR-139-5p Inhibits Isoproterenol-Induced Cardiac Hypertrophy by Targetting c-Jun.” *Bioscience Reports* 38 (2). https://doi.org/10.1042/BSR20171430.

Miyaoka, Yuichiro, Jennifer R. Berman, Samantha B. Cooper, Steven J. Mayerl, Amanda H. Chan, Bin Zhang, George A. Karlin-Neumann, and Bruce R. Conklin. 2016. “Systematic Quantification of HDR and NHEJ Reveals Effects of Locus, Nuclease, and Cell Type on Genome-Editing.” *Scientific Reports* 6 (March). https://doi.org/10.1038/srep23549.

Montag, J., B. Petersen, A. K. Flögel, E. Becker, A. Lucas-Hahn, G. J. Cost, C. Mühlfeld, T. Kraft, H. Niemann, and B. Brenner. 2018. “Successful Knock-in of Hypertrophic Cardiomyopathy-Mutation R723G into the MYH7 Gene Mimics HCM Pathology in Pigs.” *Scientific Reports* 8 (1): 4786. https://doi.org/10.1038/s41598-018-22936-z.

Montgomery, Mary K., Siqun Xu, and Andrew Fire. 1998. “RNA as a Target of Double-Stranded RNA-Mediated Genetic Interference in Caenorhabditis Elegans.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (26): 15502–7. https://doi.org/10.1073/pnas.95.26.15502.

Morita, Hiroyuki, Heidi L Rehm, Andres Menesses, Barbara McDonough, Amy E Roberts, Raju Kucherlapati, Jeffrey A Towbin, J G Seidman, and Christine E Seidman. 2008. “Shared Genetic Causes of Cardiac Hypertrophy in Children and Adults.” *The New England Journal of Medicine* 358 (18): 1899–1908. https://doi.org/10.1056/NEJMoa075463.

Mosqueira, Diogo, Ingra Mannhardt, Jamie R. Bhagwan, Katarzyna Lis-Slimak, Puspita Katili, Elizabeth Scott, Mustafa Hassan, et al. 2018. “CRISPR/Cas9 Editing in Human Pluripotent Stem Cell-Cardiomyocytes Highlights Arrhythmias, Hypocontractility, and Energy Depletion as Potential

Therapeutic Targets for Hypertrophic Cardiomyopathy." *European Heart Journal* 39 (43): 3879–92. https://doi.org/10.1093/eurheartj/ehy249.

Mosqueira, Lis-Slimak, and Denning. 2019. "High-Throughput Phenotyping Toolkit for Characterizing Cellular Models of Hypertrophic Cardiomyopathy in Vitro." *Methods and Protocols* 2 (4): 83. https://doi.org/10.3390/mps2040083.

Motta, Benedetta M., Peter P. Pramstaller, Andrew A. Hicks, and Alessandra Rossini. 2017. "The Impact of CRISPR/Cas9 Technology on Cardiac Research: From Disease Modelling to Therapeutic Approaches." *Stem Cells International* 2017: 1–13. https://doi.org/10.1155/2017/8960236.

Musunuru, Kiran, Farah Sheikh, Rajat M. Gupta, Steven R. Houser, Kevin O. Maher, David J. Milan, Andre Terzic, and Joseph C. Wu. 2018. "Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling and Precision Medicine: A Scientific Statement From the American Heart Association." *Circulation. Genomic and Precision Medicine*. NLM (Medline). https://doi.org/10.1161/HCG.0000000000000043.

Nagueh, Sherif F., Suetnee Chen, Rajnikant Patel, Natalia Tsybouleva, Silvia Lutucuta, Helen A. Kopelen, William A. Zoghbi, Miguel A. Quiñones, Robert Roberts, and A. J. Marian. 2004. "Evolution of Expression of Cardiac Phenotypes over a 4-Year Period in the β-Myosin Heavy Chain-Q403 Transgenic Rabbit Model of Human Hypertrophic Cardiomyopathy." *Journal of Molecular and Cellular Cardiology* 36 (5): 663–73. https://doi.org/10.1016/j.yjmcc.2004.02.010.

Nagueh, Sherif F, and William A Zoghbi. 2015. "Role of Imaging in the Evaluation of Patients at Risk for Sudden Cardiac Death: Genotype-Phenotype Intersection." *JACC. Cardiovascular Imaging* 8 (7): 828–45. https://doi.org/10.1016/j.jcmg.2015.04.006.

Nguyen, Anh H., Paul Marsh, Lauren Schmiess-Heine, Peter J. Burke, Abraham Lee, Juhyun Lee, and Hung Cao. 2019. "Cardiac Tissue Engineering: State-of-the-Art Methods and Outlook." *Journal of Biological Engineering* 13 (1): 1–21. https://doi.org/10.1186/s13036-019-0185-0.

Niimura, Hideshi, Linda L. Bachinski, Somkiat Sangwatanaroj, Hugh Watkins, Albert E. Chudley, William Mckenna, Arni Kristinsson, et al. 1998. "Mutations in the Gene for Cardiac Myosin-Binding Protein C and Late- Onset Familial Hypertrophic Cardiomyopathy." *New England Journal of Medicine* 338 (18): 1248–57. https://doi.org/10.1056/NEJM199804303381802.

Niimura, Hideshi, Kristen K. Patton, William J. McKenna, Johann Soults, Barry J. Maron, J. G. Seidman, and Christine E. Seidman. 2002. "Sarcomere Protein Gene Mutations in Hypertrophic Cardiomyopathy of the Elderly." *Circulation* 105 (4): 446–51. https://doi.org/10.1161/hc0402.102990.

Ojala, Marisa, and Katriina Aalto-Setälä. 2016. "Modeling Hypertrophic Cardiomyopathy with Human Induced Pluripotent Stem Cells." In *Pluripotent Stem Cells - From the Bench to the Clinic*. InTech. https://doi.org/10.5772/62649.

Ojala, Marisa, Chandra Prajapati, Risto-Pekka Pekka Pölönen, Kristiina Rajala, Mari Pekkanen-Mattila, Jyrki Rasku, Kim Larsson, and Katriina Aalto-Setälä. 2016. "Mutation-Specific Phenotypes in HiPSC-Derived Cardiomyocytes Carrying Either Myosin-Binding Protein C Or *α*-Tropomyosin Mutation for Hypertrophic Cardiomyopathy." Edited by Farah Sheikh. *Stem Cells International* 2016: 1684792. https://doi.org/10.1155/2016/1684792.

Paquet, Dominik, Dylan Kwart, Antonia Chen, Andrew Sproul, Samson Jacob, Shaun Teo, Kimberly Moore Olsen, Andrew Gregg, Scott Noggle, and Marc Tessier-Lavigne. 2016. "Efficient Introduction of Specific Homozygous and Heterozygous Mutations Using CRISPR/Cas9." *Nature* 533 (7601): 125–29. https://doi.org/10.1038/nature17664.

Paranal, Ronald M., Polakit Teekakirikul, Carolyn Y. Ho, Diane Fatkin, and Christine E. Seidman. 2020. "Genetic Cardiomyopathies." In *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics*, Seventh Ed, 77–114. Elsevier. https://doi.org/10.1016/b978-0-12-812532-8.00002-1.

Paushkin, Sergey, Amélie K. Gubitz, Séverine Massenet, and Gideon Dreyfuss. 2002. "The SMN Complex,

an Assemblyosome of Ribonucleoproteins." *Current Opinion in Cell Biology*. https://doi.org/10.1016/S0955-0674(02)00332-0.

Piccini, Ilaria, Jyoti Rao, Guiscard Seebohm, and Boris Greber. 2015. "Human Pluripotent Stem Cell-Derived Cardiomyocytes: Genome-Wide Expression Profiling of Long-Term in Vitro Maturation in Comparison to Human Heart Tissue." *Genomics Data* 4 (June): 69–72. https://doi.org/10.1016/j.gdata.2015.03.008.

Pontén, Annica, Stuart Walsh, Daniela Malan, Xiaojie Xian, Susanne Schéele, Laura Tarnawski, Bernd K. Fleischmann, and Stefan Jovinge. 2013. "FACS-Based Isolation, Propagation and Characterization of Mouse Embryonic Cardiomyocytes Based on VCAM-1 Surface Marker Expression." *PLoS ONE* 8 (12). https://doi.org/10.1371/journal.pone.0082403.

Prabhakar, Rethinasamy, Greg P. Boivin, Ingrid L. Grupp, Brian Hoit, Grace Arteaga, R. John Solaro, and David F. Wieczorek. 2001. "A Familial Hypertrophic Cardiomyopathy α-Tropomyosin Mutation Causes Severe Cardiac Hypertrophy and Death in Mice." *Journal of Molecular and Cellular Cardiology* 33 (10): 1815–28. https://doi.org/10.1006/jmcc.2001.1445.

Prondzynski, Maksymilian, Elisabeth Krämer, Sandra D. Laufer, Aya Shibamiya, Ole Pless, Frederik Flenner, Oliver J. Müller, et al. 2017a. "Evaluation of MYBPC3 Trans-Splicing and Gene Replacement as Therapeutic Options in Human IPSC-Derived Cardiomyocytes." *Molecular Therapy - Nucleic Acids* 7 (June): 475–86. https://doi.org/10.1016/j.omtn.2017.05.008.

———. 2017b. "Evaluation of MYBPC3 Trans-Splicing and Gene Replacement as Therapeutic Options in Human IPSC-Derived Cardiomyocytes." *Molecular Therapy - Nucleic Acids* 7 (June): 475–86. https://doi.org/10.1016/j.omtn.2017.05.008.

Purevjav, Enkhsaikhan. 2019. "Animal Models of Cardiomyopathies." In *Animal Models in Medicine and Biology [Working Title]*, 58:233–39. IntechOpen. https://doi.org/10.5772/intechopen.89033.

Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8 (11): 2281–2308. https://doi.org/10.1038/nprot.2013.143.

Ribeiro, Alexandre J.S., Yen Sin Ang, Ji Dong Fu, Renee N. Rivas, Tamer M.A. Mohamed, Gadryn C. Higgs, Deepak Srivastava, and Beth L. Pruitt. 2015. "Contractility of Single Cardiomyocytes Differentiated from Pluripotent Stem Cells Depends on Physiological Shape and Substrate Stiffness." *Proceedings of the National Academy of Sciences of the United States of America* 112 (41): 12705–10. https://doi.org/10.1073/pnas.1508073112.

Rivas, Manuel A., Matti Pirinen, Donald F. Conrad, Monkol Lek, Emily K. Tsang, Konrad J. Karczewski, Julian B. Maller, et al. 2015. "Effect of Predicted Protein-Truncating Genetic Variants on the Human Transcriptome." *Science* 348 (6235): 666–69. https://doi.org/10.1126/science.1261877.

Roma-Rodrigues, Catarina, and Alexandra R. Fernandes. 2014. "Genetics of Hypertrophic Cardiomyopathy: Advances and Pitfalls in Molecular Diagnosis and Therapy." *Application of Clinical Genetics*. https://doi.org/10.2147/TACG.S49126.

Roma-Rodrigues, Catarina, Luís R. Raposo, Alexandra R. Fernandes, Alexandra R. Fernandes, Catarina Roma-Rodrigues, Lu&#xed Raposo, S R., and Alexandra R. Fernandes. 2015. "MicroRNAs Based Therapy of Hypertrophic Cardiomyopathy: The Road Traveled so Far." *BioMed Research International* 2015: 1–8. https://doi.org/10.1155/2015/983290.

Ronaldson-Bouchard, Kacey, Stephen P. Ma, Keith Yeager, Timothy Chen, Lou Jin Song, Dario Sirabella, Kumi Morikawa, Diogo Teles, Masayuki Yazawa, and Gordana Vunjak-Novakovic. 2018. "Advanced Maturation of Human Cardiac Tissue Grown from Pluripotent Stem Cells." *Nature* 556 (7700): 239–43. https://doi.org/10.1038/s41586-018-0016-3.

Ronaldson-Bouchard, Kacey, Keith Yeager, Diogo Teles, Timothy Chen, Stephen Ma, Lou Jin Song, Kumi Morikawa, et al. 2019. "Engineering of Human Cardiac Muscle Electromechanically Matured to an

Adult-like Phenotype." *Nature Protocols* 14 (10): 2781–2817. https://doi.org/10.1038/s41596-019-0189-8.

Roncarati, Roberta, Chiara Viviani Anselmi, Maria Angela Losi, Laura Papa, Elena Cavarretta, Paula Da Costa Martins, Carla Contaldi, et al. 2014. "Circulating MiR-29a, among Other up-Regulated MicroRNAs, Is the Only Biomarker for Both Hypertrophy and Fibrosis in Patients with Hypertrophic Cardiomyopathy." *Journal of the American College of Cardiology* 63 (9): 920–27. https://doi.org/10.1016/j.jacc.2013.09.041.

Rooij, Eva Van, Lillian B. Sutherland, Ning Liu, Andrew H. Williams, John McAnally, Robert D. Gerard, James A. Richardson, and Eric N. Olson. 2006. "A Signature Pattern of Stress-Responsive MicroRNAs That Can Evoke Cardiac Hypertrophy and Heart Failure." *Proceedings of the National Academy of Sciences of the United States of America* 103 (48): 18255–60. https://doi.org/10.1073/pnas.0608791103.

Rosenberg, Alexander B., Rupali P. Patwardhan, Jay Shendure, and Georg Seelig. 2015. "Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences." *Cell* 163 (3): 698–711. https://doi.org/10.1016/j.cell.2015.09.054.

Ross, Samantha Barratt, Stuart T Fraser, Richard D Bagnall, and Christopher Semsarian. 2017. "Peripheral Blood Derived Induced Pluripotent Stem Cells (IPSCs) from a Female with Familial Hypertrophic Cardiomyopathy." *Stem Cell Research* 20: 76–79. https://doi.org/10.1016/j.scr.2017.02.016.

Sabater-Molina, M., I. Pérez-Sánchez, J. P. Hernández del Rincón, and J. R. Gimeno. 2018. "Genetics of Hypertrophic Cardiomyopathy: A Review of Current State." *Clinical Genetics* 93 (1): 3–14. https://doi.org/10.1111/cge.13027.

Sanbe, Atsushi, Jeanne James, Volkan Tuzcu, Selman Nas, Lisa Martin, James Gulick, Hanna Osinska, et al. 2005. "Transgenic Rabbit Model for Human Troponin I-Based Hypertrophic Cardiomyopathy." *Circulation* 111 (18): 2330–38. https://doi.org/10.1161/01.CIR.0000164234.24957.75.

Sander, Jeffry D., and J. Keith Joung. 2014. "CRISPR-Cas Systems for Editing, Regulating and Targeting Genomes." *Nature Biotechnology* 32 (4): 347–50. https://doi.org/10.1038/nbt.2842.

Savoji, Houman, Mohammad Hossein Mohammadi, Naimeh Rafatian, Masood Khaksar Toroghi, Erika Yan Wang, Yimu Zhao, Anastasia Korolj, Samad Ahadian, and Milica Radisic. 2019. "Cardiovascular Disease Models: A Game Changing Paradigm in Drug Discovery and Screening." *Biomaterials* 198 (April): 3–26. https://doi.org/10.1016/j.biomaterials.2018.09.036.

Sayed, Danish, Chull Hong, Ieng-Yi Chen, Jacqueline Lypowy, and Maha Abdellatif. 2007. "MicroRNAs Play an Essential Role in the Development of Cardiac Hypertrophy." *Circulation Research* 100 (3): 416–24. https://doi.org/10.1161/01.RES.0000257913.42552.23.

Schwach, Verena, and Robert Passier. 2016. "Generation and Purification of Human Stem Cell-Derived Cardiomyocytes." *Differentiation* 91 (4–5): 126–38. https://doi.org/10.1016/j.diff.2016.01.001.

Scotti, Marina M., and Maurice S. Swanson. 2016. "RNA Mis-Splicing in Disease." *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg.2015.3.

Sedaghat-Hamedani, Farbod, Elham Kayvanpour, Oguz Firat Tugrul, Alan Lai, Ali Amr, Jan Haas, Tanja Proctor, et al. 2018. "Clinical Outcomes Associated with Sarcomere Mutations in Hypertrophic Cardiomyopathy: A Meta-Analysis on 7675 Individuals." *Clinical Research in Cardiology* 107 (1): 30–41. https://doi.org/10.1007/s00392-017-1155-5.

Seki, Tomohisa, Shinsuke Yuasa, Mayumi Oda, Toru Egashira, Kojiro Yae, Dai Kusumoto, Hikari Nakata, et al. 2010. "Generation of Induced Pluripotent Stem Cells from Human Terminally Differentiated Circulating t Cells." *Cell Stem Cell*. Cell Press. https://doi.org/10.1016/j.stem.2010.06.003.

Semsarian, Christopher, Jodie Ingles, Martin S. Maron, and Barry J. Maron. 2015. "New Perspectives on the Prevalence of Hypertrophic Cardiomyopathy." *Journal of the American College of Cardiology*.

Elsevier USA. https://doi.org/10.1016/j.jacc.2015.01.019.

Setten, Ryan L., John J. Rossi, and Si-ping ping Han. 2019. “The Current State and Future Directions of RNAi-Based Therapeutics.” *Nature Reviews Drug Discovery*, no. 18: 421–446. https://doi.org/10.1038/s41573-019-0017-4.

Shyu, Jeou Jong, Chiung Hsiang Cheng, Robert A. Erlandson, Jyh Hung Lin, and Si Kwang Liu. 2002. “Ultrastructure of Intramural Coronary Arteries in Pigs with Hypertrophic Cardiomyopathy.” *Cardiovascular Pathology* 11 (2): 104–11. https://doi.org/10.1016/S1054-8807(01)00101-6.

Singer, Emma S., Jodie Ingles, Christopher Semsarian, and Richard D. Bagnall. 2019. “Key Value of RNA Analysis of MYBPC3 Splice-Site Variants in Hypertrophic Cardiomyopathy.” *Circulation. Genomic and Precision Medicine* 12 (1): e002368. https://doi.org/10.1161/CIRCGEN.118.002368.

Skelton, Rhys J.P., Magdaline Costa, David J. Anderson, Freya Bruveris, Ben W. Finnin, Katerina Koutsis, Deevina Arasaratnam, et al. 2014. “SIRPA, VCAM1 and CD34 Identify Discrete Lineages during Early Human Cardiovascular Development.” *Stem Cell Research* 13 (1): 172–79. https://doi.org/10.1016/j.scr.2014.04.016.

Song, Lei, Ming Su, Shuiyun Wang, Yubao Zou, Xiaojian Wang, Yilu Wang, Hongli Cui, Peng Zhao, Rutai Hui, and Jizheng Wang. 2014. “MiR-451 Is Decreased in Hypertrophic Cardiomyopathy and Regulates Autophagy by Targeting TSC1.” *Journal of Cellular and Molecular Medicine* 18 (11): 2266–74. https://doi.org/10.1111/jcmm.12380.

Strande, Jennifer L. 2015. “Haploinsufficiency MYBPC3 Mutations: Another Stress Induced Cardiomyopathy? Let’s Take a Look!” *Journal of Molecular and Cellular Cardiology* 79: 284–86. https://doi.org/10.1016/j.yjmcc.2014.12.008.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, Shinya Yamanaka, et al. 2007. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.” *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Tanaka, Atsushi, Shinsuke Yuasa, Giulia Mearini, Toru Egashira, Tomohisa Seki, Masaki Kodaira, Dai Kusumoto, et al. 2014. “Endothelin-1 Induces Myofibrillar Disarray and Contractile Vector Variability in Hypertrophic Cardiomyopathy-Induced Pluripotent Stem Cell-Derived Cardiomyocytes.” *Journal of the American Heart Association* 3 (6). https://doi.org/10.1161/JAHA.114.001263.

Tardiff, Jil C, Lucie Carrier, Donald M Bers, Corrado Poggesi, Cecilia Ferrantini, Raffaele Coppini, Lars S Maier, Houman Ashrafian, Sabine Huke, and Jolanda van der Velden. 2015. “Targets for Therapy in Sarcomeric Cardiomyopathies.” *Cardiovascular Research* 105 (4): 457–70. https://doi.org/10.1093/cvr/cvv023.

Tejado, Beatriz San Millán, and Cristina Jou. 2018. “Histopathology in HCM.” *Global Cardiology Science and Practice* 2018 (3). https://doi.org/10.21542/gcsp.2018.20.

Thomson, James A. 1998. “Embryonic Stem Cell Lines Derived from Human Blastocysts.” *Science* 282 (5391): 1145–47. https://doi.org/10.1126/science.282.5391.1145.

Tohyama, Shugo, Jun Fujita, Takako Hishiki, Tomomi Matsuura, Fumiyuki Hattori, Rei Ohno, Hideaki Kanazawa, et al. 2016. “Glutamine Oxidation Is Indispensable for Survival of Human Pluripotent Stem Cells.” *Cell Metabolism* 23 (4): 663–74. https://doi.org/10.1016/j.cmet.2016.03.001.

Tohyama, Shugo, Fumiyuki Hattori, Motoaki Sano, Takako Hishiki, Yoshiko Nagahata, Tomomi Matsuura, Hisayuki Hashimoto, et al. 2013. “Distinct Metabolic Flow Enables Large-Scale Purification of Mouse and Human Pluripotent Stem Cell-Derived Cardiomyocytes.” *Cell Stem Cell* 12 (1): 127–37. https://doi.org/10.1016/j.stem.2012.09.013.

Tulpule, Asmin, and George Q. Daley. 2009. “Efficient Gene Knockdowns in Human Embryonic Stem Cells Using Lentiviral-Based RNAi.” *Methods in Molecular Biology* 482: 35–42. https://doi.org/10.1007/978-1-59745-060-7_3.

Turanov, Anton A., Agnes Lo, Matthew R. Hassler, Angela Makris, Ami Ashar-Patel, Julia F. Alterman, Andrew H. Coles, et al. 2018. “RNAi Modulation of Placental SFLT1 for the Treatment of Preeclampsia.” *Nature Biotechnology* 36 (12): 1164–73. https://doi.org/10.1038/nbt.4297.

Tyska, M. J., E. Hayes, M. Giewat, C. E. Seidman, J. G. Seidman, and D. M. Warshaw. 2000. “Single-Molecule Mechanics of R403Q Cardiac Myosin Isolated from the Mouse Model of Familial Hypertrophic Cardiomyopathy.” *Circulation Research* 86 (7): 737–44. https://doi.org/10.1161/01.RES.86.7.737.

Uosaki, Hideki, Hiroyuki Fukushima, Ayako Takeuchi, Satoshi Matsuoka, Norio Nakatsuji, Shinya Yamanaka, and Jun K. Yamashita. 2011. “Efficient and Scalable Purification of Cardiomyocytes from Human Embryonic and Induced Pluripotent Stem Cells by VCAM1 Surface Expression.” *PLoS ONE* 6 (8): e23657. https://doi.org/10.1371/journal.pone.0023657.

Urbach, Achia. 2004. “Modeling for Lesch-Nyhan Disease by Gene Targeting in Human Embryonic Stem Cells.” *Stem Cells* 22 (4): 635–41. https://doi.org/10.1634/stemcells.22-4-635.

Urbach, Achia, and Nissim Benvenisty. 2009. “Studying Early Lethality of 45,XO (Turner’s Syndrome) Embryos Using Human Embryonic Stem Cells.” Edited by Joanna Mary Bridger. *PLoS ONE* 4 (1): e4175. https://doi.org/10.1371/journal.pone.0004175.

Vaz-Drago, Rita, Noélia Custódio, and Maria Carmo-Fonseca. 2017. “Deep Intronic Mutations and Human Disease.” *Human Genetics* 136 (9): 1093–1111. https://doi.org/10.1007/s00439-017-1809-4.

Veerman, Christiaan C., Georgios Kosmidis, Christine L. Mummery, Simona Casini, Arie O. Verkerk, and Milena Bellin. 2015. “Immaturity of Human Stem-Cell-Derived Cardiomyocytes in Culture: Fatal Flaw or Soluble Problem?” *Stem Cells and Development*. Mary Ann Liebert Inc. https://doi.org/10.1089/scd.2014.0533.

Verma, Bhupendra, Maureen V. Akinyi, Antto J. Norppa, and Mikko J. Frilander. 2018. “Minor Spliceosome and Disease.” *Seminars in Cell and Developmental Biology* 79 (July): 103–12. https://doi.org/10.1016/j.semcdb.2017.09.036.

Viswanathan, Shiv Kumar, Heather K. Sanders, James W. McNamara, Aravindakshan Jagadeesan, Arshad Jahangir, A. Jamil Tajik, and Sakthivel Sadayappan. 2017. “Hypertrophic Cardiomyopathy Clinical Phenotype Is Independent of Gene Mutation and Mutation Dosage.” *PLoS ONE* 12 (11): 1–19. https://doi.org/10.1371/journal.pone.0187948.

Vithana, Eranga N., Leen Abu-Safieh, Maxine J. Allen, Alisoun Carey, Myrto Papaioannou, Christina Chakarova, Mai Al-Maghtheh, et al. 2001. “A Human Homolog of Yeast Pre-MRNA Splicing Gene, PRP31, Underlies Autosomal Dominant Retinitis Pigmentosa on Chromosome 19q13.4 (RP11).” *Molecular Cell* 8 (2): 375–81. https://doi.org/10.1016/S1097-2765(01)00305-7.

Walsh, Roddy, Rachel Buchan, Alicja Wilk, Shibu John, Leanne E Felkin, Kate L Thomson, Tang Hak Chiaw, et al. 2017. “Defining the Genetic Architecture of Hypertrophic Cardiomyopathy: Re-Evaluating the Role of Non-Sarcomeric Genes.” *European Heart Journal* 38 (46): 3461–68. https://doi.org/10.1093/eurheartj/ehw603.

Walsh, Roddy, Kate L. Thomson, James S. Ware, Birgit H. Funke, Jessica Woodley, Karen J. McGuire, Francesco Mazzarotto, et al. 2017. “Reassessment of Mendelian Gene Pathogenicity Using 7,855 Cardiomyopathy Cases and 60,706 Reference Samples.” *Genetics in Medicine* 19 (2): 192–203. https://doi.org/10.1038/gim.2016.90.

Wang, Lili, Kyungsoo Kim, Shan Parikh, Adrian Gabriel Cadar, Kevin R. Bersell, Huan He, Jose R. Pinto, Dmytro O. Kryshtal, and Bjorn C. Knollmann. 2018. “Hypertrophic Cardiomyopathy-Linked Mutation in Troponin T Causes Myofibrillar Disarray and pro-Arrhythmic Action Potential Changes in Human IPSC Cardiomyocytes.” *Journal of Molecular and Cellular Cardiology* 114 (January): 320–27. https://doi.org/10.1016/j.yjmcc.2017.12.002.

Wang, Zefeng, and Christopher B. Burge. 2008. “Splicing Regulation: From a Parts List of Regulatory

Elements to an Integrated Splicing Code." *RNA*. https://doi.org/10.1261/rna.876308.

Wang, Zefeng, Michael E. Rolish, Gene Yeo, Vivian Tung, Matthew Mawson, and Christopher B. Burge. 2004. "Systematic Identification and Analysis of Exonic Splicing Silencers." *Cell* 119 (6): 831–45. https://doi.org/10.1016/j.cell.2004.11.010.

Ward, Lucas D, and Manolis Kellis. 2012. "Interpreting Noncoding Genetic Variation in Complex Traits and Human Disease." *Nature Biotechnology* 30 (11): 1095–1106. https://doi.org/10.1038/nbt.2422.

Warren, Luigi, Philip D. Manos, Tim Ahfeldt, Yuin Han Loh, Hu Li, Frank Lau, Wataru Ebina, et al. 2010. "Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified MRNA." *Cell Stem Cell* 7 (5): 618–30. https://doi.org/10.1016/j.stem.2010.08.012.

Watkins, Hugh, Houman Ashrafian, and Charles Redwood. 2011. "Inherited Cardiomyopathies." Edited by Robert S. Schwartz. *New England Journal of Medicine* 364 (17): 1643–56. https://doi.org/10.1056/NEJMra0902923.

Weeland, Cornelis J., Maarten M. van den Hoogenhof, Abdelaziz Beqqali, and Esther E. Creemers. 2015. *Insights into Alternative Splicing of Sarcomeric Genes in the Heart*. *Journal of Molecular and Cellular Cardiology*. Vol. 81. Academic Press. https://doi.org/10.1016/j.yjmcc.2015.02.008.

Wei, Bin, and J.-P. Jin. 2016. "TNNT1, TNNT2, and TNNT3: Isoform Genes, Regulation, and Structure–Function Relationships." *Gene* 582 (1): 1–13. https://doi.org/10.1016/J.GENE.2016.01.006.

Wessels, Marja W., Johanna C. Herkert, Ingrid M. Frohn-Mulder, Michiel Dalinghaus, Arthur Van Den Wijngaard, Ronald R. De Krijger, Michelle Michels, Irenaeus Fm De Coo, Yvonne M. Hoedemaekers, and Dennis Dooijes. 2015. "Compound Heterozygous or Homozygous Truncating MYBPC3 Mutations Cause Lethal Cardiomyopathy with Features of Noncompaction and Septal Defects." *European Journal of Human Genetics* 23 (7): 922–28. https://doi.org/10.1038/ejhg.2014.211.

Wilbie, Danny, Johanna Walther, and Enrico Mastrobattista. 2019. "Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing." *Accounts of Chemical Research* 52 (6): 1555–64. https://doi.org/10.1021/acs.accounts.9b00106.

Xiong, Hui Y., Babak Alipanahi, Leo J. Lee, Hannes Bretschneider, Daniele Merico, Ryan K.C. Yuen, Yimin Hua, et al. 2015. "The Human Splicing Code Reveals New Insights into the Genetic Determinants of Disease." *Science* 347 (6218). https://doi.org/10.1126/science.1254806.

Yang, Junyi. 2019. "Patisiran for the Treatment of Hereditary Transthyretin-Mediated Amyloidosis." *Expert Review of Clinical Pharmacology* 12 (2): 95–99. https://doi.org/10.1080/17512433.2019.1567326.

Yang, Lei, Mark H. Soonpaa, Eric D. Adler, Torsten K. Roepke, Steven J. Kattman, Marion Kennedy, Els Henckaerts, et al. 2008. "Human Cardiovascular Progenitor Cells Develop from a KDR+ Embryonic-Stem-Cell-Derived Population." *Nature* 453 (7194): 524–28. https://doi.org/10.1038/nature06894.

Yang, Qinglin, Atsushi Sanbe, Hanna Osinska, Timothy E. Hewett, Raisa Klevitsky, and Jeffrey Robbins. 1998. "A Mouse Model of Myosin Binding Protein C Human Familial Hypertrophic Cardiomyopathy." *Journal of Clinical Investigation* 102 (7): 1292–1300. https://doi.org/10.1172/JCI3880.

Yin, Zhiyong, Jun Ren, and Wei Guo. 2015. "Sarcomeric Protein Isoform Transitions in Cardiac Muscle: A Journey to Heart Failure." *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier. https://doi.org/10.1016/j.bbadis.2014.11.003.

Yoshida, Yoshinori, and Shinya Yamanaka. 2017. "Induced Pluripotent Stem Cells 10 Years Later." *Circulation Research* 120 (12): 1958–68. https://doi.org/10.1161/CIRCRESAHA.117.311080.

Young, Margaret A., David E. Larson, Chiao Wang Sun, Daniel R. George, Li Ding, Christopher A. Miller, Ling Lin, et al. 2012. "Background Mutations in Parental Cells Account for Most of the Genetic Heterogeneity of Induced Pluripotent Stem Cells." *Cell Stem Cell* 10 (5): 570–82. https://doi.org/10.1016/j.stem.2012.03.002.

Yu, Junying, Maxim A. Vodyanik, Kim Smuga-Otto, Jessica Antosiewicz-Bourget, Jennifer L. Frane, Shulan Tian, Jeff Nie, et al. 2007. “Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells.” *Science* 318 (5858): 1917–20. https://doi.org/10.1126/science.1151526.

Zhang, Feng, Yan Wen, and Xiong Guo. 2014. “CRISPR/Cas9 for Genome Editing: Progress, Implications and Challenges.” *Human Molecular Genetics* 23 (R1): 40–46. https://doi.org/10.1093/hmg/ddu125.

Zhang, Miao, Jan Sebastian Schulte, Alexander Heinick, Ilaria Piccini, Jyoti Rao, Roberto Quaranta, Dagmar Zeuschner, et al. 2015. “Universal Cardiac Induction of Human Pluripotent Stem Cells in Two and Three-Dimensional Formats: Implications for in Vitro Maturation.” *Stem Cells* 33 (5): 1456–69. https://doi.org/10.1002/stem.1964.

Zhang, Sidi, Kaitlin E. Samocha, Manuel A. Rivas, Konrad J. Karczewski, Emma Daly, Ben Schmandt, Benjamin M. Neale, Daniel G. MacArthur, and Mark J. Daly. 2018. “Base-Specific Mutational Intolerance near Splice Sites Clarifies the Role of Nonessential Splice Nucleotides.” *Genome Research* 28 (7): 968–74. https://doi.org/10.1101/gr.231902.117.

Zhou, Ting, Christina Benda, Sarah Duzinger, Yinghua Huang, Xingyan Li, Yanhua Li, Xiangpeng Guo, et al. 2011. “Generation of Induced Pluripotent Stem Cells from Urine.” *Journal of the American Society of Nephrology* 22 (7): 1221–28. https://doi.org/10.1681/ASN.2011010106.

Zhu, Chaoqun, Zhilong Chen, and Wei Guo. 2017. *Pre-MRNA Mis-Splicing of Sarcomeric Genes in Heart Failure*. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Vol. 1863. Elsevier B.V. https://doi.org/10.1016/j.bbadis.2016.11.008.

II. *IN SILICO* PREDICTION TOOLS IN THE EVALUATION OF HCM VARIANTS

The content of this chapter was adapted from the manuscript in preparation for submission:

Pedro Barbosa, **Marta Ribeiro**, Ana Beatriz Garcia, Joana Tavares, Alcides Fonseca, Maria Carmo-Fonseca, *"In silico* evaluation of clinically relevant variants throughout full gene bodies".

II.1. ABSTRACT

Next-generation sequencing has revolutionized molecular diagnostics for many diseases, including HCM, by allowing human genome analysis in a cost-effective way. However, these studies often reveal a vast number of genetic variants of uncertain clinical significance, particularly in intronic regions. *In silico* methods can help predict pathogenicity, but the choice of which tools to use is not straightforward. This work aimed to evaluate the performance of several available computational tools in predicting the potential pathogenicity of genetic variants in different contexts by using three different datasets: one HCM disease specific, one global with all intronic variants described in ClinVar and one of deep intronic variants curated from the literature. Our results show that the clinical interpretation of HCM genetic variants may be a challenge and the use of different tools may be needed to correctly assign variant pathogenicity. We further demonstrate that near splice site prediction is reasonably well covered by several methods from different scopes (including SpliceAI, LINSIGHT and phastCons). Finally, we observed that SpliceAI is the only tool with a good performance in identifying deleterious deep intronic variants. This highlights the existing need to develop new approaches that effectively look at full introns in the context of gene expression regulation and human disease. Knowing the best qualified computational tools to use in each case can greatly improve genetic diagnosis, which in turn may help preventing the disease and/or its progression.

Keywords: Bioinformatics prediction; *in silico* tools, Pathogenicity; Variant of uncertain significance

II.2. INTRODUCTION

Mendelian diseases are considered to be rare, with an incidence of 1 in 500 for a mendelian disease to be considered high. But, collectively, genetic disorders affect a substantial number of people reaching up to 8% of the population (Yang et al. 2013). Knowledge regarding molecular mechanisms and genetic mutations associated with diseases is increasing but many patients remain without a molecular diagnosis, that would be important for the assessment of risk, adequacy of treatment and guidance (Frey, Luedde, and Katus 2012; Yang et al. 2013).

The advent of next-generation sequencing (NGS) technologies has made targeted exome sequencing an exceptional approach to identify both neutral polymorphisms and disease-causing mutations responsible for genetic diseases, providing innovation in the way some diseases are managed and, even guiding drug discovery. Moreover, with their decreasing costs, whole-exome (WES) and whole-genome sequencing (WGS) are emerging as effective alternatives to gene-panel tests, in diagnose and, mainly, in a research setting for identifying causal variants in mendelian disease patients. (Smedley et al. 2016; Yang et al. 2013)

One of the biggest challenges when using NGS approaches is the bioinformatic analyses of massive amounts of data, since a single sequenced exome can yield several thousands of variants (Gilissen et al. 2012) and identifying a single, likely causative, disease mutation is a challenging and time-consuming process. An effective way to distinguish benign from pathogenic variants is through allele frequencies (AF) filtering. If a variant is absent or in a low frequency, similar to the prevalence of the investigated disease, in reference databases such as the Exome Aggregation Consortium (ExAC) or GnomAD (Lek et al. 2016) this is an important criteria for variant pathogenicity, as opposed to variants with high frequency in the general population (Whiffin et al. 2017). However this filtering will only reduce number of candidate pathogenic variants to a couple hundred (Gilissen et al. 2012) and further prioritization of pathogenic variants remains a challenging task (Lelieveld, Veltman, and Gilissen 2016; Whiffin et al. 2017).

Clinical testing laboratories use guidelines to classify variants based upon the American College of Medical Genetics revised criteria (ACMG) (Richards et al. 2015) and others such as their supplementation by the Association for Clinical Genetic Science (ACGS) in 2017 (Ellard et al. 2017). The use of these guidelines intends to help assign variants as pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign in a reproductible way across different labs. However, many patients who undergo genetic sequencing do not receive a clear molecular diagnosis but instead a list of detected changes that are assigned as VUS. These VUS are variants which cannot be discriminated between disease-causing and others of no pathological significance and they are one of the main findings in most genetic sequencing analysis (Hoffman-Andrews 2017).

Even though most research and clinical testing has been focused in the protein-coding exome, it represents less than 2% of the genome and more than 50% of the causes of mendelian disorders are not determined after sequencing affected families (Chong et al. 2015). This raised questions regarding the importance of intronic regions for clinical genetic testing. Historically, introns were considered as “junk DNA” because they are removed from the pre-mRNA by RNA splicing (Gilbert 1978), but throughout the years evidence has shown the opposite: intron-containing genes seem to increase the levels of transcription (Shabalina et al. 2010), they may give rise to functional non-coding RNA by-products (Hubé and Francastel 2015) and are directly involved in alternative splicing (Chen and Manley 2009). Moreover, genome-wide association studies (GWAS) showed that the majority of genetic variants associated with human diseases are noncoding, exhibiting their effects through the regulation of gene expression (Ward and Kellis 2012b). In fact, it was estimated that between 15 and 50% of all monogenic disease-causing mutations affect pre-mRNA splicing, which highlights intronic regions as a potential genetic source for disease discovery (G. S. Wang and Cooper 2007).

Computational approaches that also score in introns were developed throughout the past years to predict impact of genetic variants and empower the identification of novel genotype-

phenotype associations (Table II.1.). Most of the approaches rely on training a classifier using sets of known pathogenic and benign variants, and then apply the fitted model to unseen variants. More recent models were derived by using primary DNA/RNA sequence directly and making use of sequence context. Nevertheless, these tools usually infer a significance cutoff so one can interpret variants in a binary manner as a mean to distinguish putative pathogenic vs benign variants. To ease its application in clinical practice, developers commonly run their model over millions of putative variants occurring in genes and generate pre-computed scores that the researcher/clinician uses to compare with his own set of variants.

Finding the best-fitted method to prioritize likely pathogenic variants to be further studied, especially in non-coding regions remains a difficult problem to solve. These tools can be divided in different groups that vary in their scope, training data and algorithm employed.

The first class of methods do not predict pathogenicity *per se*, rather they rely on the evolutionary conservation of the variant position across multiple species, based on the assumption that evolutionarily conserved positions have essential roles in the structure or function of the encoded protein. These measurements were obtained by constructing genome-wide multiple alignments of dozens of genomes followed by quantification of evolutionary rates either at a single base-pair resolution or by considering neighboring positions (Siepel et al. 2005; Siepel, Pollard, and Haussler 2006; Davydov et al. 2010).

The appearance of large-scale projects aiming to comprehend the functional role of non-coding regions (Dunham et al. 2012; Kundaje et al. 2015) triggered the development of genome wide predictors that use these data as features to train new models. Some of them predict the impact of variants on the fitness of the individual (Huang, Gulko, and Siepel 2017; Rentzsch et al. 2019), while others tell whether variants are likely to be functional relevant (e.g. regulatory variants occurring in promotors or transcription factor binding sites) (Shihab et al. 2015; Ionita-Laza et al. 2016).

The last class of models predict the effect of variants on splicing. Splicing is a process orchestrated by a complex interplay of *cis* and *trans* elements where disease phenotype may be caused by several possible mechanisms. Accordingly, splicing tools may predict different splicing outcomes based on their target function (e.g. effect of variant on exon inclusion or alternative 5' splice site), meaning that some methods are not targeted to predict pathogenicity directly, rather they assess functional effects of variants in the spliced product (Xiong et al. 2015; Cheng et al. 2019). Nevertheless, in theory all splicing tools should be able to accurately predict splice site disruptions.

Despite the diversity and potential of these approaches none can be considered as ideal or complete since all have limitations. Moreover these tools are hardly comparable with each other due to their differences and it is still unclear which one or which combination should be used for prioritizing variants (Salgado et al. 2016; Dong et al. 2015; Lelieveld, Veltman, and Gilissen 2016). Importantly validation studies for these computational prediction tools are also still lacking or give conflicting results (van der Velde et al. 2015; Miosge et al. 2015).

Currently efforts are being made in the review and classification of genetic variants, allowing for a better development of prediction computational tools that aid to prioritize candidate variants. However, there is still an unmet need for a good characterization of these tools and refinement of their prioritization thresholds.

In this study we aimed to assess to what extent available prediction tools can be efficiently applied in clinical genomics by using three clinically relevant datasets to perform the evaluations of a vast number of available bioinformatics tools, and provide some recommendations to the community regarding their performance in variant classification.

Table II.1. Description of the tools evaluated in the study

	Tool	Description	Method	Threshold	Data used	#Features	Feature type	Other scores added	Obtained from
Conservation scores	GERP (Davydov et al. 2010)	Identification of evolutionarily constrained elements	Maximum Likelihood to estimate evolutionary rate and dynamic programming	> 4.4 (Dong et al. 2015)	Genomes of 34 mammals	-	-	No	vcfanno
	phyloP (Siepel, Pollard, and Haussler 2006)	*P*-value that indicates how aligned sequences deviate from the null hypothesis of neutral evolution	Hidden Markov Model	> 1.6 (Dong et al. 2015)	Genomes of 100 vertebrates	-	-	No	vcfanno
	SiPhy (Garber et al. 2009)	Identification of constrained sites as those with a nucleotide substitution pattern significantly deviating from the neutral pattern	Maximum Likelihood and Hidden Markov Model	> 12.7 (Dong et al. 2015)	Genomes of 29 mammals	-	-	No	vcfanno
	Phast Cons (Siepel et al. 2005)	Probability that each nucleotide belongs to a conserved element	Hidden Markov Model	> 0.99 (Li et al. 2018)	Genomes of 100 vertebrates	-	-	No	vcfanno
Genome-wide predictors	FATHMM-MKL (Shihab et al. 2015)	Prediction of functional consequences of coding and non-coding SNVs using genomic annotations from ENCODE and conservation scores	Support Vector Machine based on Multiple Kernel Learning	> 0.5 (Liu et al. 2016)	3,063 disease-implicated SNVs from HGMD; 5,252 negative instances from 1000G project	763 from 4 main feature groups	ENCODE data; Conservation scores;	Yes	vcfanno
	GWAVA (Ritchie et al. 2014)	Genome-wide annotation of the functional impact of DNA variants using ENCODE, conservation, allele frequencies and sequence context data	Random Forest	> 0.4 (Bendl et al. 2016)	1,614 disease-implicated SNVs from HGMD; 3 benign datasets from 1000G project: ~161,400; ~16,140; 5,027	~175	ENCODE data; Conservation scores; Allele frequency; General variant and sequence context	Yes	vcfanno
	Eigen (Ionita-Laza et al. 2016)	Unsupervised learning approach to leverage the functional importance of genetic variants across the whole genome	Principal Component Analysis	> 1 (Jagadeesh et al. 2016)	~76.7 million coding non-synonymous variants from dbNSFP database (other variant set was employed to study the non-coding genome)	29 from 3 main feature groups	ENCODE data; Conservation scores; Allele frequency	Yes	vcfanno
	Funseq2 (Y. Fu et al. 2014)	A genome-wide deleteriousness prediction score designed for non-coding somatic SNVs	Weighted scoring scheme based on feature importance measurement	> 1.5 (Y. Fu et al. 2014)	Functional annotations from ENCODE and Roadmap Epigenomics, PWMs, Conservation scores, Network properties, recurrent sites from COSMIC	11 feature groups	Functional annotations from ENCODE and Roadmap Epigenomics, PWMs, Conservation scores, Network properties, recurrent sites from COSMIC	Yes	vcfanno

Genome-wide predictors	ReMM (Smedley et al. 2016)	Classifier to predict the potential of an arbitrary position in the genome to cause a Mendelian disease if mutated (targeted for non-coding regions).	Random Forest	> 0.984 ("ReMM Threshold " n.d.)	453 disease-implicated by manual curation; 14,755,199 as benign dataset	26	ENCODE data; Conservation scores; FANTOM5 data; Allele frequency	Yes	vcfanno
	CADD v1.4 (Rentzsch et al. 2019)	Deleteriousness prediction using ENCODE, conservation, allele frequencies and sequence context data	Logistic Regressor	> 15 (Dong et al. 2015)	14,695,338 of both proxy-neutral and proxy-deleterious variants	949 representing 63 different annotations from CADD v1.3 + measures of variant density from gnomAD + splice defect prediction from dbscSNV	ENCODE data; Conservation scores; Allele frequency; General variant and sequence context	Yes	VEP plugin
	DANN (Quang, Chen, and Xie 2015)	Same as CADD, but based on a deep learning model	Deep neural network	> 0.9 (Richardson et al. 2016)	Whole CADD proxy-neutral dataset (16,627,775); Sampled the same number from the CADD simulated deleterious dataset	Same as CADD v1.3	ENCODE data; Conservation scores; General variant and sequence context	Yes	vcfanno
	fitCons (Gulko et al. 2015)	Evolution-based measurement of fitness consequence using both divergence data and functional genomics data	Clustering & Approximate Expectation Maximization used by INSIGHT	> 0.4 (Liu et al. 2017)	Genomes of 54 unrelated human individuals	33 functional genomics annotations polymorphism and divergence data used by INSIGHT	ENCODE data from three different cell lines; Evolutionary data	No	vcfanno
	LINSIGHT (Huang, Gulko, and Siepel 2017)	Prediction of noncoding nucleotide sites at which mutations are likely to have deleterious fitness consequences	INSIGHT & Online stochastic gradient-descent	> 0.4 (same as fitCons)	Genomes of 54 unrelated human individuals	48 genomics annotations; polymorphism and divergence data used by INSIGHT	ENCODE, Roadmap Epigenomics and FANTOM5 data; Conservation scores; Predicted binding sites; Evolutionary data	Yes	vcfanno
Splicing	MaxEnt Scan (Yeo and Burge 2004)	Prediction of RNA splice site signal based on the maximum entropy principle	Maximum Entropy distribution	|MaxEnt Scan_diff| > 3 (Desmet et al. 2009)	~8500 real 5' and 3' splice sites; ~180,000 decoy 5' and 3' splice sites from 1.821 set of nonredundant transcripts	-	Estimated from known signal sequences around real and decoy splice sites	No	VEP plugin
	dbscSNV (Jian, Boerwinkle, and Liu 2014)	*In Silico* prediction of splice-altering variants based on an ensemble of individual methods	AdaBoost and Random Forest	> 0.6 (Jian, Boerwinkle, and Liu 2014)	Splice altering variants from HGMD, SpliceDisease and DBASS databases; Negative variants from 1000G	11	Splicing-related prediction scores; Conservation scores; CADD scores	Yes	VEP plugin

Splicing	SPIDEX (Xiong et al. 2015)	Prediction of how much SNVs cause splicing misregulation by measuring differential exon inclusion events	Deep neural network	|deltaPSI_zscore| > 2 (Xiong et al. 2015)	RNA-seq data in 10,700 exons across 16 tissues	1393	Context-dependent RNA sequence properties	No	vcfanno
	HAL (Rosenberg et al. 2015)	Variant efffect prediction (SNPs, indels) on different isoform usage from alternative splicing events (alternative 5'ss and Exon skipping events)	Linear model using hexamer motif frequencies	|deltaPSI| > 5 *	Massively parallel reporter assay (MPRA) containing 265,137 mini-genes in a library of alternative 5' splice donors	4,096 possible 6-mer	6-mer features extracted from sequence information	No	kipoi
	TraP v3 (Gelfman et al. 2017)	Prediction of the damage caused by SNVs at the transcript level by incorporation of splicing-related features	Random Forest	Coding > 0.416 Non-coding > 0.289 (Gelfman et al. 2017)	75 pathogenic synonymous variants; 402 synonymous variants as benign	20	Conservation scores; Splicing-specific measurements; General Variant Annotations	Yes	vcfanno
	S-CAP (Jagadeesh et al. 2018)	Splicing-specific pathogenicity score derived from variant, exon and gene importance measurements	Gradient Boosting tree	Several different thresholds (Jagadeesh et al. 2018) **	17,059 splicing related pathogenic variants from HGMD and ClinVar and 6,760,450 splicing region benign variants from gnomAD	29	Region-specific models derived from gene-level, exon-level, and variant-level features. Incorporates conservation, splicing and functional scores	Yes	vcfanno
	MMSplice (Cheng et al. 2019)	Modular approach to study functional effects of variants on splicing	Linear model that combines coefficient of 5 neural network modules	|deltaLogitPSI| > 1 ("MMSplice Threshold" n.d.)	Vex-Seq high throughput assay containing constructs of 2059 ExAC variants. 957 variants from chr1 to chr8 were used for training	9	Region-specific neural networks that use data from GENCODE and MPRA	No	kipoi
	SpliceAI v1.3 (Jaganathan et al. 2019)	Cryptic splice site prediction from primary sequence	Deep neural network	> 0.2 (Jaganathan et al. 2019)	Primary transcript of 13,384 genes, accounting for 130,796 donor-acceptor pairs, plus novel splice junctions observed in the GTEx cohort	Automatically learned by the model	Context-dependent RNA sequence properties	No	vcfanno
	Kipoi Splice4 (States et al. 2019)	Ensembl method that incorporates predictions from 4 splicing related models	Logistic Regression	> 0.5 (States et al. 2019)	10,715 splice region variants from ClinVar and 2959 variants from dbscSNV	4	Predictions from 4 models (HAL, MaxEntScan5, MaxEntScan5 and LaBranchoR)	Yes	kipoi

* HAL scores percent spliced in (PSI) for the sequence containing alternative splice donor variants, therefore we defined a change in PSI > 5 as the relevant threshold.

** S-CAP authors provide different reference thresholds depending on the location and context of the variant. 3intronic: 0.006; exonic: 0.009; 5intronic: 0.006; 5core_dominant: 0.034; 5core_recessive: 0.367; 5extended: 0.005; 3core_dominant: 0.033; 3core_recessive: 0.264.

II.3. DATASETS CONSTRUCTION AND METHODOLOGY

HCM-associated genetic variants

Over 1,400 mutations spread among at least 11 different genes have been associated with HCM (Richard et al. 2003). Most of the mutations are found in genes normally expressed on the cardiac sarcomere. *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3* and *ACTC1* are the most prevalent causal genes (Lopes, Rahman, and Elliott 2013; Ingles et al. 2019). Genetic counseling and testing is indicated in HCM patients and relatives (Charron et al. 2010). However, not all variants found in HCM patients are confirmed as pathogenic and some can be restricted to one family. Furthermore, only one third of probands with HCM have likely pathogenic/pathogenic variants that can be useful for family screening (Alfares et al. 2015; Maron, Maron, and Semsarian 2012).

To evaluate prediction tools in a disease specific manner, we gathered HCM associated variants assigned as Pathogenic/Likely Pathogenic in the ClinVar database (Landrum et al. 2018), manually curated all articles present in PubMed respecting to each selected variant and applied the ACMG guidelines to ensure that only unequivocal pathogenic mutations were used in the benchmarks. This set of variants included missense and splice site variants spread across the eight sarcomeric genes referred above (Table S II.1.).

Regarding control variants, ACMG guidelines were not fully employed because information about benign variants is scarce as usually labs put their efforts into the classification of disease-causing variants. Therefore, for selecting our set of benign variants we considered the single criteria of variant frequency in the population, as reported in the Genome Aggregation database (gnomAD) v2.1.1tA (Karczewski et al. 2019). According to this criterion, if a variant has a frequency higher than 5% is most likely benign. To generate a homogeneous dataset we wanted to select the same number of benign variants as the pathogenic ones so we used variants found

in 33 genes (Table S II.2.), between missense and splice site variants. The choice of genes was based on the Pan Cardiomyopathy Panel of 62 genes, available at NCBI's website (https://www.ncbi.nlm.nih.gov/gtr/tests/509149/), that has also been used by others (Ouellette et al. 2018).

After this selection we ended up with a "ground truth" dataset composed of 75 pathogenic and 75 benign variants distributed between missense and splice site against which tools were tested.

Intronic variants from ClinVar database

ClinVar (Landrum et al. 2018) is a well stablished and accessible resource that maps the relationship between human genetic variation and disease phenotype. Submissions may contain several levels of supporting evidence and review status are assigned to each call based on how trustworthy an assertion is (e.g. number of submissions with the same interpretation, variant reviewed by expert panel). We downloaded ClinVar v20191202 for the GRCh37 genome build and selected just Single Nucleotide Variants (SNVs) for downstream analysis. To avoid the selection of intronic variants that overlap with exons of alternative transcripts or other genes (e.g. lncRNA) we extracted a set of non-redundant intronic regions by using bedtools (Quinlan and Hall 2010) operations on the Basic GENCODE v31liftt37 gene annotations (Harrow et al. 2012), resulting in a total intronic space of 1,625,325,724bp considering primary assembly units (52.4% of the total genome length). 67,692 ClinVar variants overlapped these genomic intervals.

We employed the Ensembl Variant Effect Predictor (VEP) v98 (Mclaren et al. 2016) to annotate the variants and only one block of consequence per gene was picked per gene affected (--per_gene –pick). We further filtered the dataset to remove consequence blocks without HGVSc expression (e.g. upstream and/or downstream consequences) and variants with conflicting interpretations by selecting variants with "Pathogenic", "Likely Pathogenic", "Benign" and "Likely

Benign” in the Clinical Significance field, which resulted in 45,664 variants that were evaluated. Distance of the variants to the nearest splice junction was calculated based on the HGVS nomenclature that VEP outputs in the HGVSc field. This distance was used to partition variants by the following intronic bins.: 0-10; 10-30; 30-100; 100-200; 200-500; 500+. A bin-based analysis was performed to assess how much prediction performance is influenced by the location depth of the variant within the intron.

Clinically relevant deep intronic variants

A list of previously aggregated disease-causing mutations was selected for testing purposes (Vaz-Drago, Custódio, and Carmo-Fonseca 2017). Additionally, we looked in the literature for recent reports of experimentally confirmed splicing altering variants occurring deep within introns. This list contains 51 novel variants (Table S II.3). We combined these two datasets summing up to 132 deep intronic to be tested. In these papers, authors usually report the effect of the variant in the canonical and/or clinically relevant transcript. Therefore, we did extra validations to ensure variants are effectively deep intronic, regardless of the transcripts considered: from the previously described intronic regions, we subtracted 20bp to obtain intervals representing deeper regions of human introns (>20bp from the nearest annotated junction). Only 7 variants (out of 132) appeared to overlap with non-deep intronic intervals (ENST00000278407.4:c.-22-155G>T; ENST00000270142.6:c.240-188C>G; ENST00000340855.6:c.1007-133A>G; ENST00000311893.9:c.418+382G>C; ENST00000397163.3:c.1782+1072G>C; ENST00000233627.9:c.17-1167C>G, ENST00000321666.5:c.1082-285T>G). However, by manually checking transcript annotations we noticed these 7 variants overlap with shorter transcripts that either represent retained introns, antisense transcripts or alternative exonic UTR of specific isoforms that are hardly expressed on GTEx data (Consortium 2017). Hence, we decided to keep these variants for tool performance analysis.

Control variants were obtained from gnomAD v2.1 by selecting deep intronic SNVs with allele frequency > 5% in the same 93 genes for which positive variants were described. In addition, we removed variants present in ClinVar as well as variants without a valid HGVSc notation, that resulted in 22,338 variants. To create a balanced dataset for evaluation we randomly shuffled and selected 132 variants from this set.

Prediction tools and performance evaluation

Table II.1. describes all the tools that score in introns that were benchmarked. For the missense variants in the HCM dataset we also included protein predictors described in Table S II.4. A local version of the Ensembl v98 GRCh37 cache was installed and pre-computed scores for most of the tools were obtained directly from each tool website or using dbNSFP v4.0b1 (Liu et al. 2017), WGSA v0.76 (a resource that extends dbNSFP database for whole genomes) and UCSC genome browser (conservation scores) (Raney et al. 2014). VCF scores annotation was performed with Ensembl VEP using custom plugins or with vcfanno v0.3.1 (Pedersen, Layer, and Quinlan 2016). Additional splicing related models were run via Kipoi v0.6.24 (States et al. 2019).

The performance of each tool was evaluated upon comparison of the reference thresholds (Table II.1). Depending on the balanced nature of the dataset (number of pathogenic and benign variants), different metrics were used to rank the tools. For balanced data, accuracy and area under Receiving Operating Characteristic (auROC) were applied, while the F1 score, the harmonic mean between precision and recall, was used on unbalanced data. It is common that some methods do not score certain variants (missing data), thus we also considered prediction capacity, referred as coverage, when measuring performance. All the analyzes and plots were produced with VETA (Barbosa et al. 2019), a simple tool to evaluate variant predictors that is available at https://github.com/PedroBarbosa/VETA.

II.4. RESULTS

Ground truth HCM dataset highlights known issues in the field

To test computational tools for predicting the clinical significance of HCM related genetic variants, we firstly re-assessed pathogenicity criteria of 89 previously described pathogenic/likely pathogenic present in ClinVar using ACMG guidelines. We observed that 14 were classified as VUS due to lack of supporting evidence (Table S II.1.). Consequently, our "ground truth" dataset of pathogenic mutations was reduced to 75, from which we carried out evaluations with the control dataset of the same size.

According to our results on the performance of analyzed tools, best performance was achieved when analyzing different variant types (missense or splice site) separately (Figure II.1). When looking at all variants together (Figure II.2.) genome wide predictors perform better, as expected. FATHMM-MKL (weighted accuracy of 0.79) and ReMM (0.78) ranked first, but they did not come close to VEST4 (0.90) or REVEL (0.87) for the missense dataset alone, or to S-CAP (1.0) and SpliceAI (1.0) for splice site data. This is not a surprising observation considering the fact these tools were trained on similar, if not the same data (except of SpliceAI). It highlights a well-known issue in the field, the so-called type 1 circularity (Grimm et al. 2015), which occurs when some variants are used both for training and evaluation, spuriously boosting prediction accuracy. S-CAP ("perfect" classifier in the splice site data) is a clear example of this problem, as it used pathogenic ClinVar and control gnomAD variants to train the model, just like we did to select this HCM associated dataset. Nevertheless, missense predictions were not impressive: the average weighted accuracy score for the top 10 methods was 0.79 and best predictors belong to the same scope (protein predictors). Conversely, splice site variants displayed higher accuracies overall and were well predicted by methods from several scopes (including methods that are not prone to suffer from circularity such as SpliceAI, LINSIGHT or phastCons).

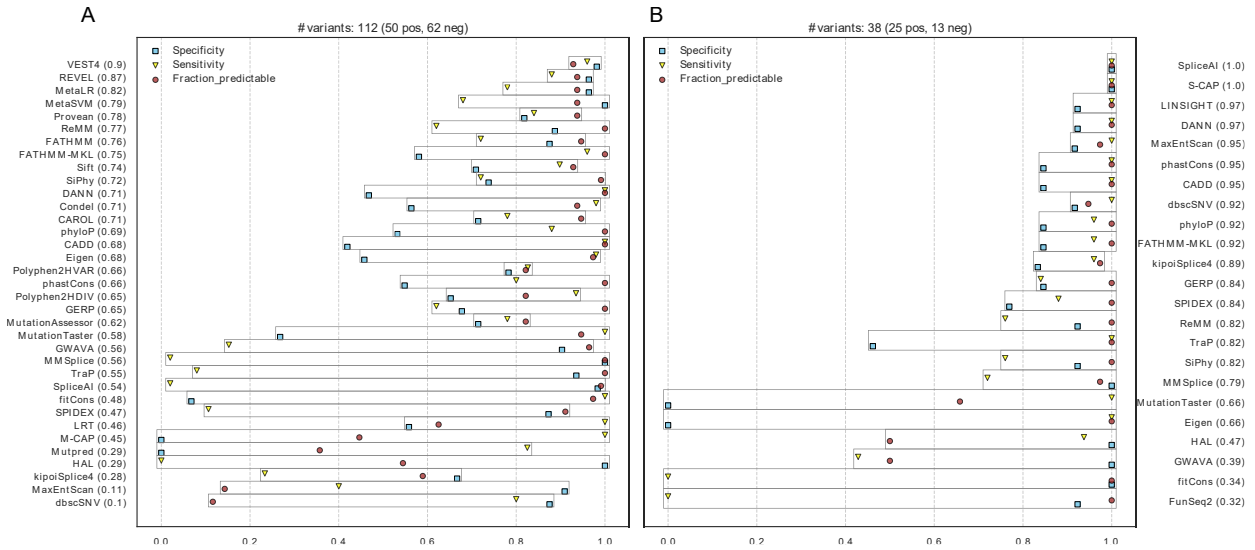


Figure II.1. Prediction performance on the dataset of Missense (A) or Splice site (B) HCM associated variants. Tools with more than 90% of missing scores were discarded. Ranking metric presented is accuracy weighted by the fraction of predictable variants.

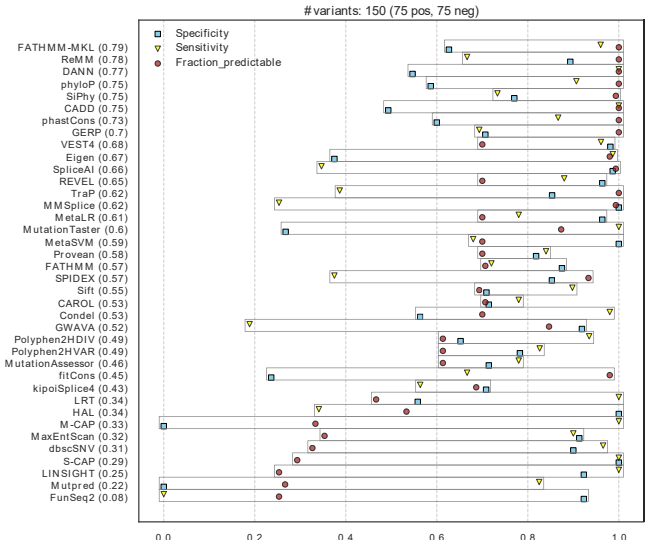


Figure II.2. Prediction performance on the full dataset of HCM associated variants. Tools with more than 90% of missing scores were discarded. Ranking metric presented is accuracy weighted by the fraction of predictable variants.

Additionally, recommended thresholds are vulnerable as they are commonly derived from evaluations performed on some test data. Although clinicians desire a universal score cut-off that represents the best ratio of sensitivity and specificity, this is not usually easy to achieve. Best thresholds are highly dependent of the data at hands. We always tried to apply thresholds

recommended by the authors of each model. If not available, we employed the thresholds suggested by other studies (Table II.1), but by no means this is indicative of the best threshold. FunSeq2 and Eigen are two clear examples of such issues presenting a poor performance (Figure II.3.A and B). However, when looking at the score distribution of each class (Figure II.3.A and B) in splice site variants, it is clear that the reference threshold is not appropriate for this dataset. By adjusting Eigen and FunSeq2 thresholds to 10 and 1, respectively, weighted accuracy score increased substantially (0.87 and 0.89, data not shown). To complicate things further, score statistics differ between variant types. By looking at the Eigen scores distribution on missense variants (Figure II.3.C), an adjusted threshold of 10 is too stringent, pinpointing the fact that threshold decisions for clinical practice may need to be tuned for specific variant types.

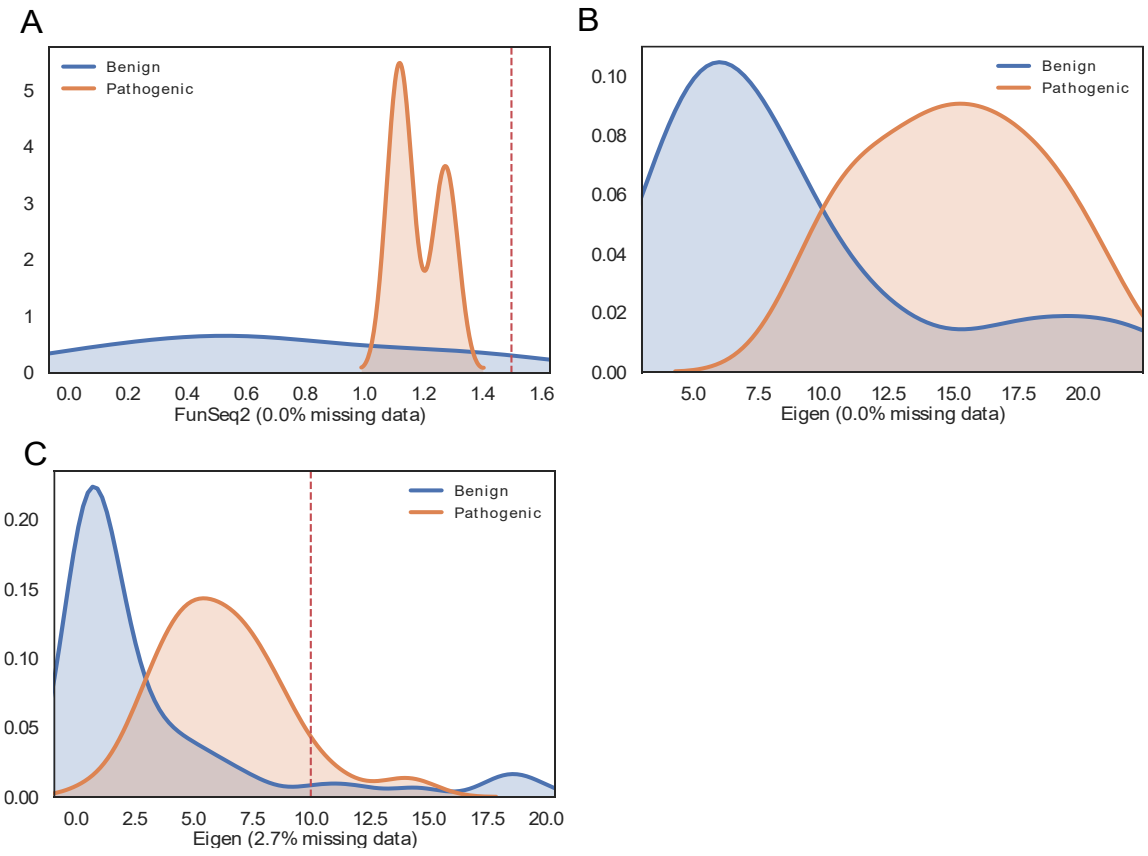


Figure II.3. Scores distribution per each class in the HCM dataset. A) FunSeq2 scores distribution on splice site variants. Vertical read dashed line represents the default threshold. B) Eigen scores distribution on splice site variants. Plot demonstrates how inappropriate default threshold of 1

seems to be. C) Eigen scores distribution on missense variants. Vertical read dashed line represents Eigen adjusted threshold for splice site variants (10)

Predictors start to fail 10bp away from the splice junction

Intronic ClinVar variants were analyzed according the bins previously described. We started to investigate class distribution at each bin, and it was not surprising to observe that the majority of pathogenic intronic variants (96.9%) locate at splice sites or near there (0-10 bin, Figure II.4.). Nonetheless, we confirmed we had enough reported pathogenic variants for analysis in the remaining bins (Figure II.4., zoomed plot). Since we had a rather big and balanced dataset for the 0-10 bin, we also benchmarked the models using ROC curves. As shown before in the HCM splicing variants, tools perform well across different prediction scopes (Figure S II.1). However, from 10bp onwards within the intron tools performance drops dramatically. Most of splicing models can't score variants across whole introns, except for TraP and SpliceAI (Figure II.5.). This is an expected behavior rather than a critical comment on the tools: they were not explicitly designed to predict deep in the introns, be it due to their own formulation or due to the lack of proper data to train. For instance, fraction of unscored variants for MMSplice and SPIDEX significantly increases at the intronic distances (bins) that authors refer to be the recommended limit (100bp and 300 bp, respectively). Because of that, and to perform fair comparisons, models were evaluated using weighted F1 score (albeit conservation-based and whole genome predictors do not suffer from this issue as they give scores across whole introns (Figure S II.2). Overall, SpliceAI was the only model to predict reasonably well deep intronic pathogenic variants reported in ClinVar (Figure II.6.A, B and C). When excluding near splice-site variants from analysis (0-10 bin), weighted F1 score across all the remaining bins was 0.43 for SpliceAI, followed by MMSplice with 0.18 and TraP with 0.16 (Figure S II.3). Apart from having low capacity to reveal pathogenic variants from 10bp onwards (low sensitivity), very low precision was also observed: when a

method predicts a variant to be relevant, it is usually the case the variant is assigned as benign in ClinVar (false positive).

Cautious interpretations should be taken, however. Despite an initial filtering of the ClinVar dataset, this “raw” analysis accommodated variants that lack strong support and evidence for the assigned status. When we tried to restrict the confidence level (e.g. 2 stars), we were left with very few positive instances deep in the introns. Therefore, we blinded trusted disease status, although we are aware that conflicting or wrong ascertainments may be propagated in ClinVar, as we show in this paper and was demonstrated by others (Shah et al. 2018). These errors may slightly skew (either positively or negatively) models’ performance. Having said that, we still believe these results are a clear demonstration of the current limitation on predicting variants within introns.

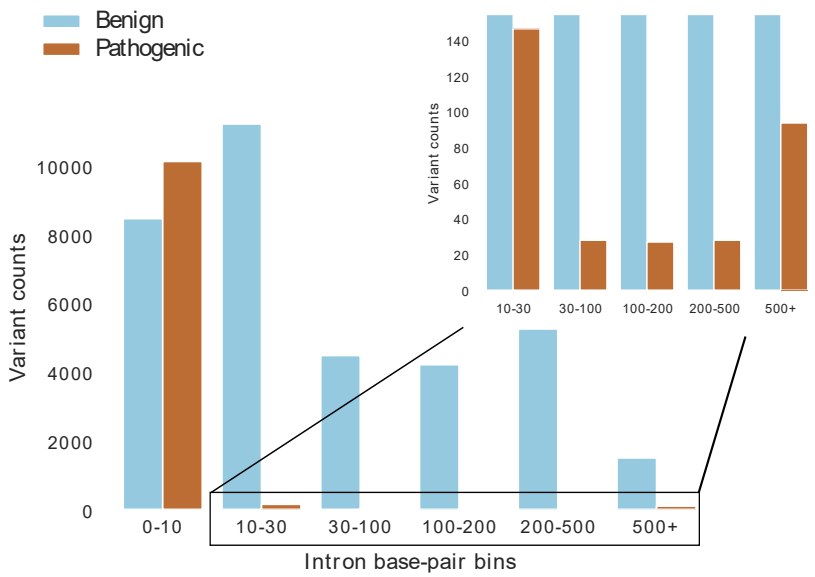


Figure II.4. Number of benign and pathogenic variants at each intronic bin.

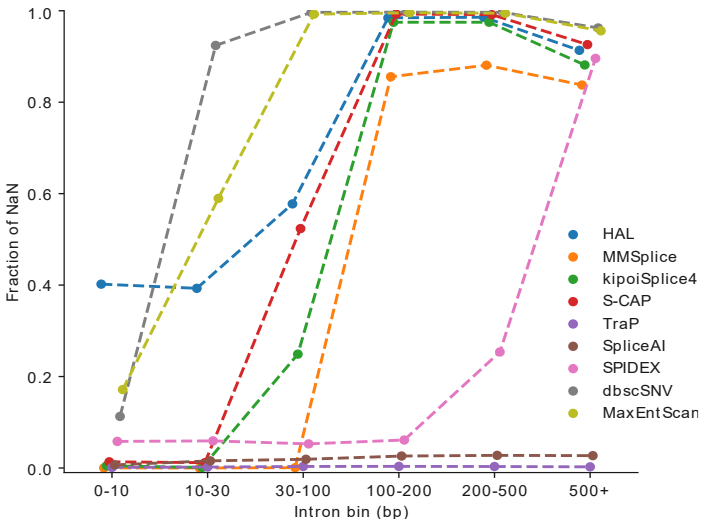


Figure II.5. Fraction of unscored variants by splicing related models at each intronic bin.

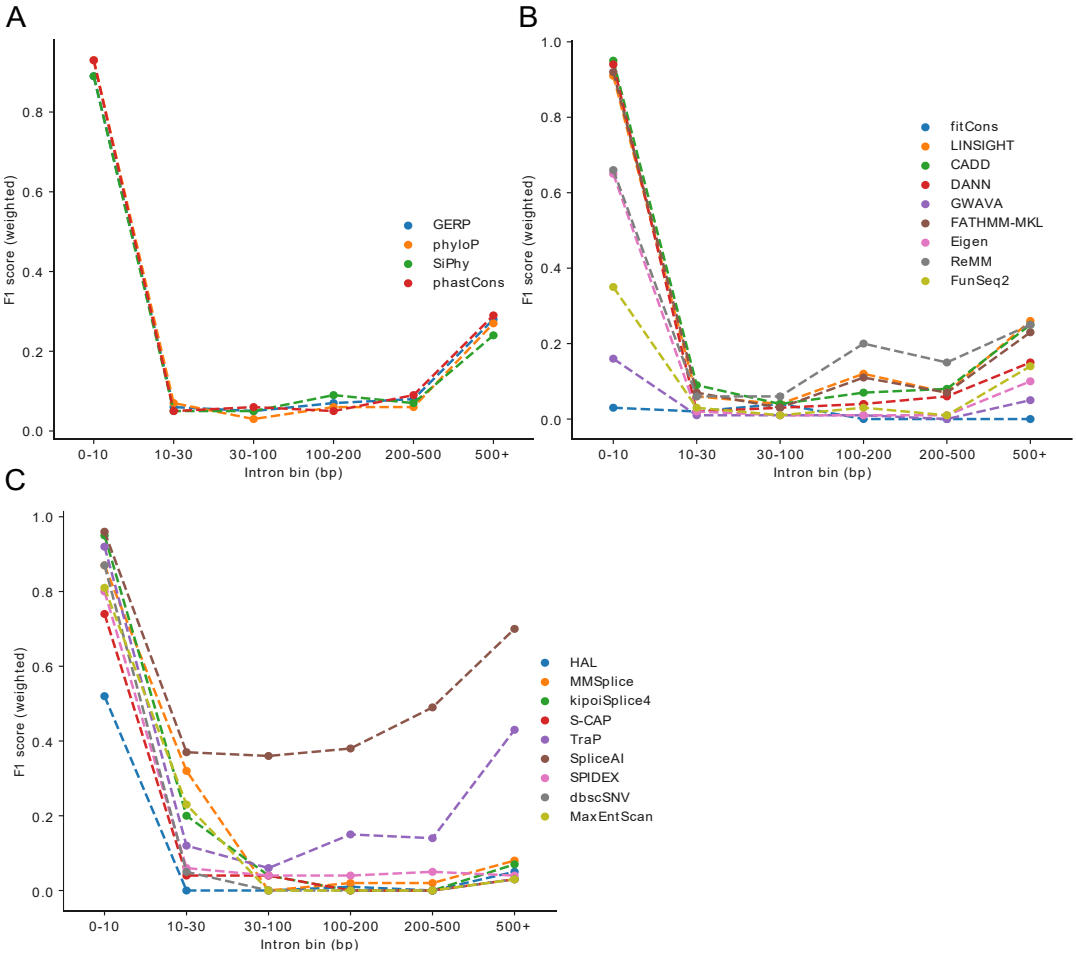


Figure II.6. Prediction tools performance at each intronic bin. A) Conservation scores. B) Whole genome predictors. C) Splicing models

Manual curation identifies a rich and clinically relevant dataset of deep intronic mutations

A literature review of recently reported variants allowed us to extend a previous collection (Vaz-Drago, Custódio, and Carmo-Fonseca 2017) of disease-causing mutations deep within the introns (Table S II.3). This new dataset includes variants from 33 different conditions, and they are mostly involved in splicing aberrations. Variants that create new splice sites or alter the strength of existing cryptic ones were observed at higher frequency. Notwithstanding, no matter the molecular effect of the variant identified, the outcome of the genetic defect ultimately leads to the inclusion of pseudoexons that frequently contain a termination codon, targeting the mRNA for degradation by NMD (Dhir and Buratti 2010). In addition, several of these variants were shown to trigger disease by compound heterozygosity, where other variants located in the coding region of the gene were identified.

Consequently, we merged these variants with the list of previously identified variants by Vaz Drago *et al*. (Vaz-Drago, Custódio, and Carmo-Fonseca 2017). Remarkably, a large fraction of these merged set of variants are absent from gnomAD and ClinVar (Figure II.7.A) and are very rare in the population (Figure II.7.B). From the subset that overlaps with ClinVar, there were 9 SNVs with wrong classifications: 6 VUS, 1 Benign, 1 drug response and one with no criteria provided, which highlights the need for continuous efforts on variant reclassification as new studies are performed. Anyway, we gathered a rich and non-redundant dataset of deep intronic disease-causing SNVs that we used to benchmark the tools along with a control dataset obtained as described in the methods section.

As observed before, most of splicing models did not score deep in the introns (Figure S II.4.). Paradoxically, the two best scoring methods were splicing-based, SpliceAI and TrAP (Figure II.8.A). However, SpliceAI was by far distance the best scoring model. It leveraged a striking weighted accuracy of 0.89, being the only method able to predict most of disease-causing

variants. In fact, all the remaining methods that give a score (conservation and whole genome predictors) were useless as their weighted accuracy lied around 0.5 just because they predict all the variants to be harmless. By doing ROC curve analysis on the methods that score more than 30% of the variants (Figure II.8.A), we observed that the poor performance of the majority of the methods was not due to a wrong pathogenicity threshold that would be inadequate for deep intronic variants (as ROC curves test performance using different thresholds). Rather, it is likely because no model (except for SpliceAI and TraP) was trained with features that explain this data.

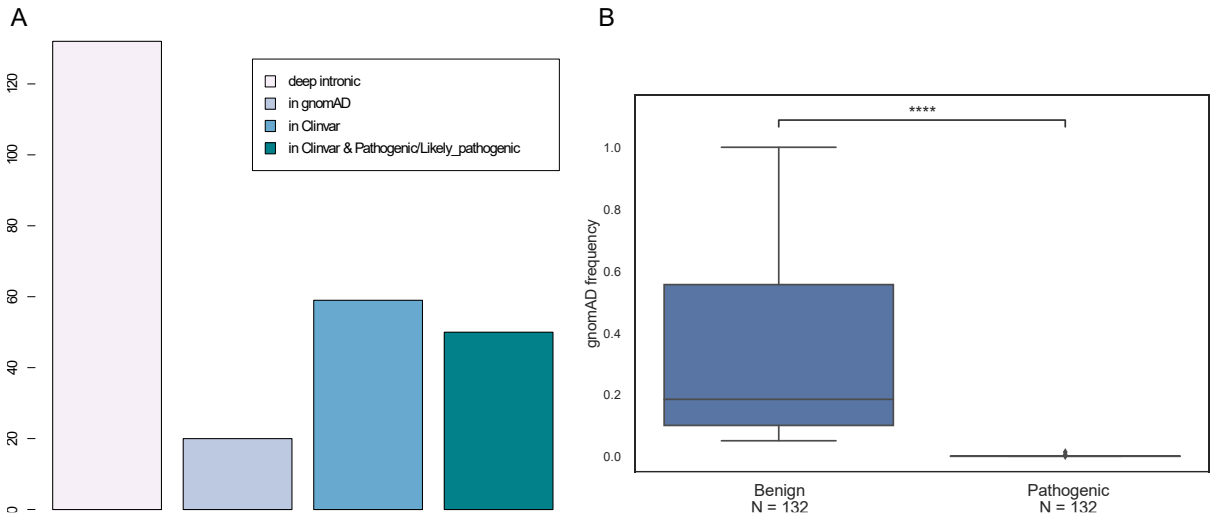


Figure II.7. Description of the deep intronic pathogenic dataset. A) Overlap with existing catalogs of genetic variation. B) gnomAD frequency of the variants compared with the control benign variants used to evaluate tools performance (Wilcoxon rank-sum test, p-value = 6.561e-13)

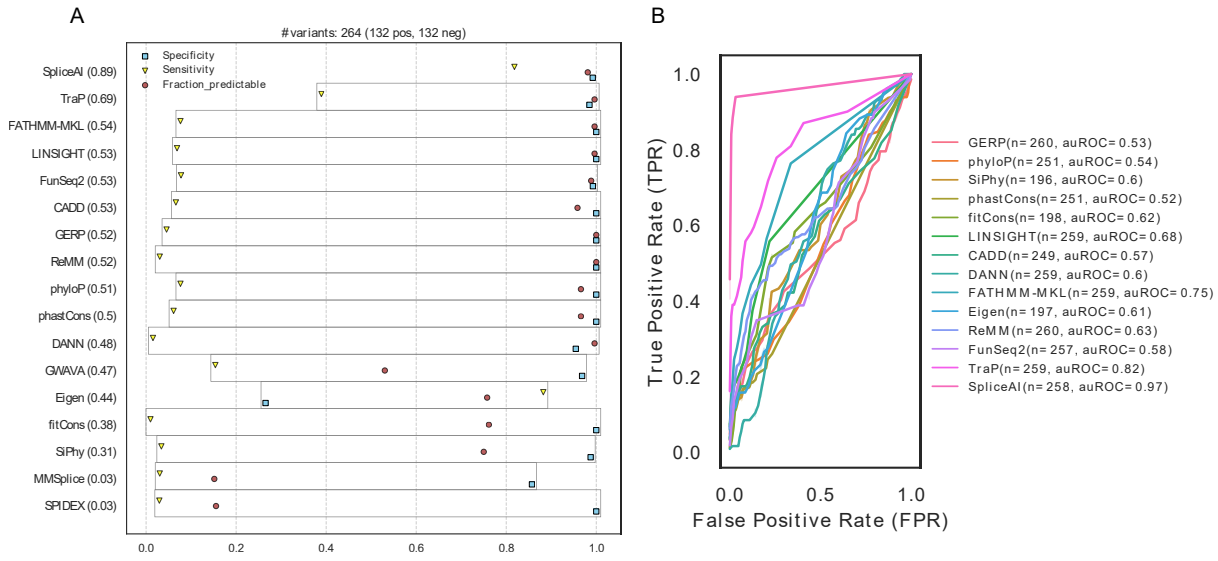


Figure II.8. Tools performance in the deep intronic dataset. A) Analysis using predefined thresholds. B) ROC curve analysis for the methods predicting more than 30% of the variants.

II.5. DISCUSSION

In this study, we performed a comprehensive evaluation of variant effect predictors across the full body of disease related genes, with a focus on intronic regions. We benchmarked the models using 3 datasets that differed in their nature and the way they were collected. The first, a high confidence dataset of missense and splice site variants from a specific mendelian disease, HCM, that was manually reviewed following ACMG guidelines. A second dataset that specifies a subset of the ClinVar public archive representative of intronic variants evaluated on a bin-based approach. And lastly, a dataset of clinically relevant deep intronic variants derived from the literature.

HCM dataset

Ideally, the aim was to find the most appropriate set of tools to prioritize different variant types. Analysis of the HCM variants revealed that despite the known limitations VEST4 or REVEL may be decent choices to score missense variants in this disease. REVEL is an ensemble method that combines predictions from 18 individual scores (including VEST3, a previous version of VEST4 and many other that we also evaluated), thus it was unsurprising REVEL stood out. Curiously, VEST4 and REVEL were also the best performers in a recent benchmark (Li et al. 2018). Importantly, they were both trained on variants from HGMD (Stenson et al. 2017) and Exome Sequencing Project (W. Fu et al. 2013) and used a random forest to build the model (Table II.1.). Additionally, in REVEL paper, authors mention VEST3 scores to be one of the most important features in their model (Ioannidis et al. 2016). Altogether, along with considerations mentioned in the HCM results section, we highlight important concerns in the field of missense variant prediction raised before (Walters-Sen et al. 2015; Mahmood et al. 2017; Grimm et al. 2015), and careful interpretations should always be considered.

Splice site predictions did well (Figure II.1.B), with most of the best performing methods perfectly rescuing pathogenic variants. When they failed, it was mostly due to some false positives (lower specificity). Thus, we would suggest using one (or a combination) of the three following methods: SpliceAI, LINSIGHT and phastCons. They represent different unbiased approaches that did not depend on specific variants to build the model: SpliceAI was trained directly from mRNA sequences, LINSIGHT used the genomes of 54 unrelated human individuals and the widely used phastCons relies on the alignment of 100 vertebrate genomes.

We should also note that the tools were tested using a small panel of variants, as we opted to have a high-quality and manually curated dataset of HCM variants. Therefore, we recognize that the results found in this study may not be entirely reproduced if working with a larger dataset (Schaafsma and Vihinen 2018), or more importantly, with a different mendelian condition. Still, being able to correctly assign the status of a given variant in HCM is crucial for molecular diagnosis and for developing strategies that may predict the occurrence of the disease in a relative and guide management. The possibility of molecular-based interventions to alter or suppress the expression of a given disease-causing mutation by means of gene editing is gaining track and may become a new reality. Thus, all efforts should be done to improve our current capacities to correctly assign variants of unknown significance in HCM.

Prediction across whole intronic sequences

While near splice site, variants were well covered by a great range of methods (Figure S II.1.), analysis of intronic variation further away from splice junctions revealed the overall poor performance of current methods to discern functional variants, both in ClinVar and in the manually derived deep intronic dataset. These analysis disclosed SpliceAI as the best prediction model to score such variants. This model is a deep learning approach that allowed a comprehensive analysis of the biology of splicing and the impact of SNVs in disease (Jaganathan et al. 2019).

Technically, SpliceAI is a 32-layer deep convolutional neural network (CNN) that predicts splice sites directly from the sequence of mRNA. Splicing determinants are learned by evaluating 10,000 flanking base-pairs of the positions of interest. The output of the model consists of three scores (probability of a given position to be a splice acceptor, splice donor and neither). We believe the large amount of training data (Table II.1.) and the use of a large sequence context allowed SpliceAI to accentuate its performance against the other models.

Overall, only two splicing models, SpliceAI and TraP score across full gene body. This reflects a known property of most of these models that are not designed to predict deeper in the introns. In addition, we note that splicing models that were run through Kipoi are limited by the regions defined in the data loader, thus reducing the possible prediction space of these models. One such case is HAL, a very elegant approach to study splicing determinants, that in our evaluation performed very poorly (for now HAL implemented in Kipoi just scores splice donor sites).

Most of the manually derived pathogenic variants by Vaz Drago and all the new variants described in this study (Table S II.3) relate with splicing aberrations. That is a paradoxal observation given the lack of prediction power of most of splicing methods. On the other hand, despite some whole genome predictors include some splicing features (e.g. CADD), they were mostly trained using ENCODE data as features, likely being more suitable to predict non-coding variation affecting transcriptional regulation.

Hence, we recommend the use of SpliceAI in any splicing-related variant prioritization pipeline, with a focus on variants located further away from splice sites (10bp onwards). This agrees with a very recent work where SpliceAI alone was recommended to assess splicing mutations in the context of rare disease (Ellingford et al. 2019). We recognize the fantastic breakthroughs enabled by SpliceAI, but some of the limitations should be pointed out: this model is just predicting splice sites. If splicing regulatory elements are disrupted or created, this model will not perform properly. Additionally, it is very common that variants affecting splicing may be only disease-relevant on specific tissues where the gene is expressed, and it is not straightforward

how much of the aberrant transcript is going to be formed. For example, in RNA-based analysis, small differences in the inclusion of an exon (10% variation) may be functionally relevant but not sufficiently strong to cause disease. Again, SpliceAI does not measure that. Therefore, this study aims to raise awareness of the importance of deep intronic regions on splicing regulation and human disease, and we expect that with all the new available data, forthcoming models can take into consideration deep intronic specificities to improve our own understanding of the splicing code (Barash et al. 2010).

Ultimately, this study highlights the need for caution in the interpretation of likely causative disease mutations prioritized by computational tools. Our data also suggests that some proposed thresholds can be adjusted in order to have more reliable results.

We hope our results can help researchers and clinicians to decide on which tool to use when assigning the status of genetic variants, potentially saving time and resources when investigating the nature of new variants found. To the best of our knowledge this is yet the best effort ever done in evaluating the performance of prediction tools on intronic regions.

II.6. REFERENCES

Adzhubei, Ivan A., Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. 2010. “A Method and Server for Predicting Damaging Missense Mutations.” *Nature Methods* 7 (4): 248–49. https://doi.org/10.1038/nmeth0410-248.

Ajeawung, Norbert F., Thi Tuyet Mai Nguyen, Linchao Lu, Thomas J. Kucharski, Justine Rousseau, Sirinart Molidperee, Joshua Atienza, et al. 2019. “Mutations in ANAPC1, Encoding a Scaffold Subunit of the Anaphase-Promoting Complex, Cause Rothmund-Thomson Syndrome Type 1.” *The American Journal of Human Genetics* 105 (3): 625–30. https://doi.org/10.1016/j.ajhg.2019.06.011.

Albert, Silvia, Alejandro Garanto, Riccardo Sangermano, Mubeen Khan, Nathalie M. Bax, Carel B. Hoyng, Jana Zernant, et al. 2018. “Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease.” *The American Journal of Human Genetics* 102 (4): 517–27. https://doi.org/10.1016/j.ajhg.2018.02.008.

Alfares, Ahmed A., Melissa A. Kelly, Gregory McDermott, Birgit H. Funke, Matthew S. Lebo, Samantha B. Baxter, Jun Shen, et al. 2015. “Results of Clinical Genetic Testing of 2,912 Probands with Hypertrophic Cardiomyopathy: Expanded Panels Offer Limited Additional Sensitivity.” *Genetics in Medicine* 17 (11): 880–88. https://doi.org/10.1038/gim.2014.205.

Alston, Charlotte L., Mike T. Veling, Juliana Heidler, Lucie S. Taylor, Joseph T. Alaimo, Andrew Y. Sung, Langping He, et al. 2020. “Pathogenic Bi-Allelic Mutations in NDUFAF8 Cause Leigh Syndrome with an Isolated Complex I Deficiency.” *The American Journal of Human Genetics* 106 (1): 92–101. https://doi.org/10.1016/j.ajhg.2019.12.001.

Barash, Yoseph, John A. Calarco, Weijun Gao, Qun Pan, Xinchen Wang, Ofer Shai, Benjamin J. Blencowe, and Brendan J. Frey. 2010. “Deciphering the Splicing Code.” *Nature* 465 (7294): 53–59. https://doi.org/10.1038/nature09000.

Barbosa, Pedro, Marta Ribeiro, Ana Beatriz Garcia, Joana Tavares, Maria Carmo-Fonseca, and Alcides Fonseca. 2019. “VETA: A Framework to Assess Computational Tools for Predicting Clinical Significance of Genetic Variation.” Basel, Switzerland: 27th Intelligent Systems for Molecular Biology (ISMB/ECBB 2019).

Bendl, Jaroslav, Miloš Musil, Jan Štourač, Jaroslav Zendulka, Jiří Damborský, and Jan Brezovský. 2016. “PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions.” *PLoS Computational Biology* 12 (5): e1004962. https://doi.org/10.1371/journal.pcbi.1004962.

Boisson, Bertrand, Yoshitaka Honda, Masahiko Ajiro, Jacinta Bustamante, Matthieu Bendavid, Andrew R. Gennery, Yuri Kawasaki, et al. 2018. “Rescue of Recurrent Deep Intronic Mutation Underlying Cell Type–Dependent Quantitative NEMO Deficiency.” *Journal of Clinical Investigation* 129 (2): 583–97. https://doi.org/10.1172/JCI124011.

Caciotti, Anna, Rodolfo Tonin, Matthew Mort, David N. Cooper, Serena Gasperini, Miriam Rigoldi, Rossella Parini, et al. 2018. “Mis-Splicing of the GALNS Gene Resulting from Deep Intronic Mutations as a Cause of Morquio a Disease.” *BMC Medical Genetics* 19 (1): 183. https://doi.org/10.1186/s12881-018-0694-6.

Carter, Hannah, Christopher Douville, Peter D Stenson, David N Cooper, Rachel Karchin, David N Cooper, Christopher Douville, Peter D Stenson, and Hannah Carter. 2013. “Identifying Mendelian Disease Genes with the Variant Effect Scoring Tool.” *BMC Genomics* 14 (Suppl 3): S3. https://doi.org/10.1186/1471-2164-14-S3-S3.

Cassini, Thomas A., Laura Duncan, Lynette C. Rives, John H. Newman, John A. Phillips, Mary E. Koziura, Jennifer Brault, et al. 2019. “Whole Genome Sequencing Reveals Novel *IGHMBP2* Variant Leading

to Unique Cryptic Splice-site and Charcot-Marie-Tooth Phenotype with Early Onset Symptoms." *Molecular Genetics & Genomic Medicine* 7 (6): e00676. https://doi.org/10.1002/mgg3.676.

Catania, Alessia, Anna Ardissone, Daniela Verrigni, Andrea Legati, Aurelio Reyes, Eleonora Lamantea, Daria Diodato, et al. 2018. "Compound Heterozygous Missense and Deep Intronic Variants in NDUFAF6 Unraveled by Exome Sequencing and MRNA Analysis." *Journal of Human Genetics* 63 (5): 563–68. https://doi.org/10.1038/s10038-018-0423-1.

Chang, Chia-Yau, Cherng-Lih Perng, Shin-Nan Cheng, Shu-Hsia Hu, Tzu-Ying Wu, Shyr-Yi Lin, and Yeu-Chin Chen. 2019. "Deep Intronic Variant c.5999-277G>A of *F8* Gene May Be a Hot Spot Mutation for Mild Hemophilia A Patients without Mutation in Exonic DNA." *European Journal of Haematology* 103 (1): 47–55. https://doi.org/10.1111/ejh.13242.

Charron, Philippe, Michael Arad, Eloisa Arbustini, Cristina Basso, Zofia Bilinska, Perry Elliott, Tiina Helio, et al. 2010. "Genetic Counselling and Testing in Cardiomyopathies: A Position Statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases." *European Heart Journal* 31 (22): 2715–28. https://doi.org/10.1093/eurheartj/ehq271.

Chen, Mo, and James L Manley. 2009. "Mechanisms of Alternative Splicing Regulation: Insights from Molecular and Genomics Approaches." *Nature Reviews. Molecular Cell Biology* 10 (11): 741–54. https://doi.org/10.1038/nrm2777.

Chen, Si, Zhidong Cen, Feng Fu, You Chen, Xinhui Chen, Dehao Yang, Haotian Wang, et al. 2019. "Underestimated Disease Prevalence and Severe Phenotypes in Patients with Biallelic Variants: A Cohort Study of Primary Familial Brain Calcification from China." *Parkinsonism & Related Disorders* 64 (July): 211–19. https://doi.org/10.1016/j.parkreldis.2019.04.009.

Cheng, Jun, Thi Yen, Duong Nguyen, Kamil J Cygan, Muhammed Hasan C, and G William. 2019. "MMSplice: Modular Modeling Improves the Predictions of Genetic Variant Effects on Splicing." *Genome Biology*, 1–15. https://doi.org/10.1101/438986.

Chiara, Matteo, Ilaria Primon, Letizia Tarantini, Luca Agnelli, Valentina Brancaleoni, Francesca Granata, Valentina Bollati, and Elena Di Pierro. 2020. "Targeted Resequencing of FECH Locus Reveals That a Novel Deep Intronic Pathogenic Variant and EQTLs May Cause Erythropoietic Protoporphyria (EPP) through a Methylation-Dependent Mechanism." *Genetics in Medicine* 22 (1): 35–43. https://doi.org/10.1038/s41436-019-0584-0.

Choi, Yongwook, and Agnes P. Chan. 2015. "PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels." *Bioinformatics* 31 (16): 2745–47. https://doi.org/10.1093/bioinformatics/btv195.

Chong, Jessica X., Kati J. Buckingham, Shalini N. Jhangiani, Corinne Boehm, Nara Sobreira, Joshua D. Smith, Tanya M. Harrell, et al. 2015. "The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities." *The American Journal of Human Genetics* 97 (2): 199–215. https://doi.org/10.1016/j.ajhg.2015.06.009.

Chun, S., and J. C. Fay. 2009. "Identification of Deleterious Mutations within Three Human Genomes." *Genome Research* 19 (9): 1553–61. https://doi.org/10.1101/gr.092619.109.

Consortium, GTEx. 2017. "Genetic Effects on Gene Expression across Human Tissues." *Nature* 550 (7675): 204–13. https://doi.org/10.1038/nature24277.

Cummings, Beryl B, Jamie L Marshall, Taru Tukiainen, Monkol Lek, Sandra Donkervoort, A Reghan Foley, Veronique Bolduc, et al. 2017. "Improving Genetic Diagnosis in Mendelian Disease with Transcriptome Sequencing." *Science Translational Medicine* 9 (386): eaal5209. https://doi.org/10.1126/scitranslmed.aal5209.

Davydov, Eugene V., David L. Goode, Marina Sirota, Gregory M. Cooper, Arend Sidow, and Serafim Batzoglou. 2010. "Identifying a High Fraction of the Human Genome to Be under Selective Constraint Using GERP++." Edited by Wyeth W. Wasserman. *PLoS Computational Biology* 6 (12): e1001025.

https://doi.org/10.1371/journal.pcbi.1001025.

Desmet, François Olivier, Dalil Hamroun, Marine Lalande, Gwenaëlle Collod-Bëroud, Mireille Claustres, and Christophe Béroud. 2009. “Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals.” *Nucleic Acids Research* 37 (9): 1–14. https://doi.org/10.1093/nar/gkp215.

Dhir, Ashish, and Emanuele Buratti. 2010. “Alternative Splicing: Role of Pseudoexons in Human Disease and Potential Therapeutic Strategies: Minireview.” *FEBS Journal* 277 (4): 841–55. https://doi.org/10.1111/j.1742-4658.2009.07520.x.

Dominov, Janice A., Özgün Uyan, Diane McKenna-Yasek, Babi Ramesh Reddy Nallamilli, Virginie Kergourlay, Marc Bartoli, Nicolas Levy, et al. 2019. “Correction of Pseudoexon Splicing Caused by a Novel Intronic Dysferlin Mutation.” *Annals of Clinical and Translational Neurology* 6 (4): 642–54. https://doi.org/10.1002/acn3.738.

Dong, Chengliang, Peng Wei, Xueqiu Jian, Richard Gibbs, Eric Boerwinkle, Kai Wang, and Xiaoming Liu. 2015. “Comparison and Integration of Deleteriousness Prediction Methods for Nonsynonymous SNVs in Whole Exome Sequencing Studies.” *Human Molecular Genetics* 24 (8): 2125–37. https://doi.org/10.1093/hmg/ddu733.

Dunham, Ian, Anshul Kundaje, Shelley F. Aldred, Patrick J. Collins, Carrie A. Davis, Francis Doyle, Charles B. Epstein, et al. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489 (7414): 57–74. https://doi.org/10.1038/nature11247.

Ellard, Sian, Emma L Baple, Martina Owens, Diana M Eccles, Stephen Abbs, and C Zandra. 2017. “ACGS Best Practice Guidelines for Variant Classification 2017,” 1–12.

Ellingford, Jamie M, Huw B Thomas, Charlie Rowlands, Gavin Arno, Glenda Beaman, Beatriz Gomes-Silva, Christopher Campbell, et al. 2019. “Functional and In-Silico Interrogation of Rare Genomic Variants Impacting RNA Splicing for the Diagnosis of Genomic Disorders.” *BioRxiv*, September, 781088. https://doi.org/10.1101/781088.

Fadaie, Zeinab, Mubeen Khan, Marta Del Pozo-Valero, Stéphanie S. Cornelis, Carmen Ayuso, Frans P. M. Cremers, Susanne Roosing, and The ABCA4 study group. 2019. “Identification of Splice Defects Due to Noncanonical Splice Site or Deep-intronic Variants in *ABCA4*.” *Human Mutation* 40 (12): 2365–76. https://doi.org/10.1002/humu.23890.

Ferraresi, Paolo, Dario Balestra, Caroline Guittard, Delphine Buthiau, Brigitte Pan-Petesh, Iva Maestri, Roula Farah, Mirko Pinotti, and Muriel Giansily-Blaizot. 2019. “Next-Generation Sequencing and Recombinant Expression Characterized Aberrant Splicing Mechanisms and Provided Correction Strategies in Factor VII Deficiency.” *Haematologica*, July. https://doi.org/10.3324/haematol.2019.217539.

Foulquier, François, Mustapha Amyere, Jaak Jaeken, Renate Zeevaert, Els Schollen, Valérie Race, Riet Bammens, et al. 2012. “TMEM165 Deficiency Causes a Congenital Disorder of Glycosylation.” *The American Journal of Human Genetics* 91 (1): 15–26. https://doi.org/10.1016/j.ajhg.2012.05.002.

Frank-Hansen, Rune, Stephen P Page, Petros Syrris, William J McKenna, Michael Christiansen, and Paal Skytt Andersen. 2008. “Micro-Exons of the Cardiac Myosin Binding Protein C Gene: Flanking Introns Contain a Disproportionately Large Number of Hypertrophic Cardiomyopathy Mutations.” *European Journal of Human Genetics* 16 (9): 1062–69. https://doi.org/10.1038/ejhg.2008.52.

Frey, Norbert, Mark Luedde, and Hugo A. Katus. 2012. “Mechanisms of Disease: Hypertrophic Cardiomyopathy.” *Nature Reviews Cardiology*. Nature Publishing Group. https://doi.org/10.1038/nrcardio.2011.159.

Fu, Wenqing, Timothy D. O’Connor, Goo Jun, Hyun Min Kang, Goncalo Abecasis, Suzanne M. Leal, Stacey Gabriel, et al. 2013. “Analysis of 6,515 Exomes Reveals the Recent Origin of Most Human Protein-Coding Variants.” *Nature* 493 (7431): 216–20. https://doi.org/10.1038/nature11690.

Fu, Yao, Zhu Liu, Shaoke Lou, Jason Bedford, Xinmeng Jasmine Mu, Kevin Y Yip, Ekta Khurana, and Mark Gerstein. 2014. “FunSeq2: A Framework for Prioritizing Noncoding Regulatory Variants in Cancer.” *Genome Biology* 15 (10): 480. https://doi.org/10.1186/s13059-014-0480-5.

Garber, Manuel, Mitchell Guttman, Michele Clamp, Michael C. Zody, Nir Friedman, and Xiaohui Xie. 2009. “Identifying Novel Constrained Elements by Exploiting Biased Substitution Patterns.” *Bioinformatics* 25 (12): i54–62. https://doi.org/10.1093/bioinformatics/btp190.

Gelfman, Sahar, Quanli Wang, K. Melodi McSweeney, Zhong Ren, Francesca La Carpia, Matt Halvorsen, Kelly Schoch, et al. 2017. “Annotating Pathogenic Non-Coding Variants in Genic Regions.” *Nature Communications* 8 (1): 1–10. https://doi.org/10.1038/s41467-017-00141-2.

Germenis, Anastasios E., and Marco Cicardi. 2019. “Driving towards Precision Medicine for Angioedema without Wheals.” *Journal of Autoimmunity* 104 (November): 102312. https://doi.org/10.1016/j.jaut.2019.102312.

Gilbert, Walter. 1978. “Why Genes in Pieces?” *Nature* 271 (5645): 501–501. https://doi.org/10.1038/271501a0.

Gilissen, Christian, Alexander Hoischen, Han G. Brunner, and Joris A. Veltman. 2012. “Disease Gene Identification Strategies for Exome Sequencing.” *European Journal of Human Genetics* 20 (5): 490–97. https://doi.org/10.1038/ejhg.2011.258.

González-Pérez, Abel, and Nuria López-Bigas. 2011. “Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel.” *American Journal of Human Genetics* 88 (4): 440–49. https://doi.org/10.1016/j.ajhg.2011.03.004.

Grimm, Dominik G, Chloé-Agathe Azencott, Fabian Aicheler, Udo Gieraths, Daniel G MacArthur, Kaitlin E Samocha, David N Cooper, et al. 2015. “The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity.” *Human Mutation* 36 (5): 513–23. https://doi.org/10.1002/humu.22768.

Gulko, Brad, Melissa J. Hubisz, Ilan Gronau, and Adam Siepel. 2015. “A Method for Calculating Probabilities of Fitness Consequences for Point Mutations across the Human Genome.” *Nature Genetics* 47 (3): 276–83. https://doi.org/10.1038/ng.3196.

Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, et al. 2012. “GENCODE: The Reference Human Genome Annotation for The ENCODE Project.” *Genome Research* 22 (9): 1760–74. https://doi.org/10.1101/gr.135350.111.

Hoffman-Andrews, Lily. 2017. “The Known Unknown: The Challenges of Genetic Variants of Uncertain Significance in Clinical Practice.” *Journal of Law and the Biosciences* 4 (3): 648–57. https://doi.org/10.1093/jlb/lsx038.

Huang, Yi-Fei, Brad Gulko, and Adam Siepel. 2017. “Fast, Scalable Prediction of Deleterious Noncoding Variants from Functional and Population Genomic Data.” *Nature Genetics* 49 (4): 618–24. https://doi.org/10.1038/ng.3810.

Hubé, Florent, and Claire Francastel. 2015. “Mammalian Introns: When the Junk Generates Molecular Diversity.” *International Journal of Molecular Sciences* 16 (12): 4429–52. https://doi.org/10.3390/ijms16034429.

Ingles, Jodie, Jennifer Goldstein, Courtney Thaxton, Colleen Caleshu, Edward W. Corty, Stephanie B. Crowley, Kristen Dougherty, et al. 2019. “Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes.” *Circulation: Genomic and Precision Medicine* 12 (2): 57–64. https://doi.org/10.1161/CIRCGEN.119.002460.

Ioannidis, Nilah M., Joseph H. Rothstein, Vikas Pejaver, Sumit Middha, Shannon K. McDonnell, Saurabh Baheti, Anthony Musolf, et al. 2016. “REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants.” *American Journal of Human Genetics* 99 (4): 877–85.

https://doi.org/10.1016/j.ajhg.2016.08.016.

Ionita-Laza, Iuliana, Kenneth McCallum, Bin Xu, and Joseph D Buxbaum. 2016. “A Spectral Approach Integrating Functional Genomic Annotations for Coding and Noncoding Variants.” *Nature Genetics* 48 (2): 214–20. https://doi.org/10.1038/ng.3477.

Jagadeesh, Karthik A., Joseph M. Paggi, James S. Ye, Peter D. Stenson, David N. Cooper, Jonathan A. Bernstein, and Gill Bejerano. 2018. “S-CAP Extends Clinical-Grade Pathogenicity Prediction to Genetic Variants That Affect RNA Splicing.” *BioRxiv*, 343749. https://doi.org/10.1101/343749.

Jagadeesh, Karthik A, Aaron M Wenger, Mark J Berger, Harendra Guturu, Peter D Stenson, David N Cooper, Jonathan A Bernstein, and Gill Bejerano. 2016. “M-CAP Eliminates a Majority of Variants of Uncertain Significance in Clinical Exomes at High Sensitivity.” *Nature Genetics* 48 (12): 1581–86. https://doi.org/10.1038/ng.3703.

Jaganathan, Kishore, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F. McRae, Siavash Fazel Darbandi, David Knowles, Yang I. Li, Jack A. Kosmicki, et al. 2019. “Predicting Splicing from Primary Sequence with Deep Learning.” *Cell* 176 (3): 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015.

Janin, Alexandre, Claire Bardel, Laurence Faivre, Gilles Millat, Pierre Antoine, Rollat Farnier, Philippe Chevalier, Emma Albert, and Jean Christophe Eicher. 2019. “Whole MYBPC3 NGS Sequencing as a Molecular Strategy to Improve the Efficiency of Molecular Diagnosis of Patients with Hypertrophic Cardiomyopathy.” *Human Mutation*, no. October: 1–11. https://doi.org/10.1002/humu.23944.

Jian, Xueqiu, Eric Boerwinkle, and Xiaoming Liu. 2014. “In Silico Prediction of Splice-Altering Single Nucleotide Variants in the Human Genome.” *Nucleic Acids Research* 42 (22): 13534–44. https://doi.org/10.1093/nar/gku1206.

Kannu, P, M Nour, M Irving, J Xie, D Loder, J Lai, O Islam, J MacKenzie, and L Messiaen. 2013. “Paraspinal Ganglioneuroma in the Proband of a Large Family with Mild Cutaneous Manifestations of NF1, Carrying a Deep *NF1* Intronic Mutation.” *Clinical Genetics* 83 (2): 191–94. https://doi.org/10.1111/j.1399-0004.2012.01882.x.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, et al. 2019. “Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes.” *BioRxiv*, August, 531210. https://doi.org/10.1101/531210.

Khan, Arif O, Elvir Becirovic, Christian Betz, Christine Neuhaus, Janine Altmüller, Lisa Maria Riedmayr, Susanne Motameny, Gudrun Nürnberg, Peter Nürnberg, and Hanno J Bolz. 2017. “A Deep Intronic CLRN1 (USH3A) Founder Mutation Generates an Aberrant Exon and Underlies Severe Usher Syndrome on the Arabian Peninsula.” *Scientific Reports* 7 (1): 1411. https://doi.org/10.1038/s41598-017-01577-8.

Khan, Ranjha, Ayesha Aftab, Sobia Tabassum, Hafiz Muhammad Jafar Hussain, Abdul Hameed, Humera Mahmood, Faiza Munir, and Ihtisham Bukhari. 2019. “Identification of CDKN2A Variants in Breast Cancer Patients in Pakistan.” *Genes & Genetic Systems* 94 (3): 117–22. https://doi.org/10.1266/ggs.18-00053.

Knapp, Karen M, Rosie Sullivan, Jennie Murray, Gregory Gimenez, Pamela Arn, Precilla D’Souza, Alper Gezdirici, et al. 2019. “Linked-Read Genome Sequencing Identifies Biallelic Pathogenic Variants in DONSON as a Novel Cause of Meier-Gorlin Syndrome.” *Journal of Medical Genetics*, November. https://doi.org/10.1136/jmedgenet-2019-106396.

Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. 2009. “Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm.” *Nature Protocols* 4 (7): 1073–82. https://doi.org/10.1038/nprot.2009.86.

Kundaje, Anshul, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, et al. 2015. “Integrative Analysis of 111 Reference Human Epigenomes.” *Nature*

518 (7539): 317–30. https://doi.org/10.1038/nature14248.

Landrum, Melissa J., Jennifer M. Lee, Mark Benson, Garth R. Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, et al. 2018. “ClinVar: Improving Access to Variant Interpretations and Supporting Evidence.” *Nucleic Acids Research* 46 (D1): D1062–67. https://doi.org/10.1093/nar/gkx1153.

Lee, Hane, Alden Y. Huang, Lee-kai Wang, Amanda J. Yoon, Genecee Renteria, Ascia Eskin, Rebecca H. Signer, et al. 2019. “Diagnostic Utility of Transcriptome Sequencing for Rare Mendelian Diseases.” *Genetics in Medicine*, October, 1–10. https://doi.org/10.1038/s41436-019-0672-1.

Lek, Monkol, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, Eric Banks, Timothy Fennell, Anne H. O’Donnell-Luria, et al. 2016. “Analysis of Protein-Coding Genetic Variation in 60,706 Humans.” *Nature* 536 (7616): 285–91. https://doi.org/10.1038/nature19057.

Lelieveld, Stefan H., Joris A. Veltman, and Christian Gilissen. 2016. “Novel Bioinformatic Developments for Exome Sequencing.” *Human Genetics* 135 (6): 603–14. https://doi.org/10.1007/s00439-016-1658-6.

Li, Jinchen, Tingting Zhao, Yi Zhang, Kun Zhang, Leisheng Shi, Yun Chen, Xingxing Wang, and Zhongsheng Sun. 2018. “Performance Evaluation of Pathogenicity-Computation Methods for Missense Variants.” *Nucleic Acids Research* 46 (15): 7793–7804. https://doi.org/10.1093/nar/gky678.

Lionel, Anath C, Gregory Costain, Nasim Monfared, Susan Walker, Miriam S Reuter, S Mohsen Hosseini, Bhooma Thiruvahindrapuram, et al. 2018. “Improved Diagnostic Yield Compared with Targeted Gene Sequencing Panels Suggests a Role for Whole-Genome Sequencing as a First-Tier Genetic Test.” *Genetics in Medicine* 20 (4): 435–43. https://doi.org/10.1038/gim.2017.119.

Liu, Xiaoming, Chunlei Wu, Chang Li, and Eric Boerwinkle. 2016. “DbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs.” *Human Mutation* 37 (3): 235–41. https://doi.org/10.1002/humu.22932.

———. 2017. “DbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-Synonymous and Splice Site SNVs” 37 (3): 235–41. https://doi.org/10.1002/humu.22932.dbNSFP.

Lopes, Luís R., M. Shafiqur Rahman, and Perry M. Elliott. 2013. “A Systematic Review and Meta-Analysis of Genotype-Phenotype Associations in Patients with Hypertrophic Cardiomyopathy Caused by Sarcomeric Protein Mutations.” *Heart* 99 (24): 1800–1811. https://doi.org/10.1136/heartjnl-2013-303939.

Lopes, Margarida C., Chris Joyce, Graham R.S. Ritchie, Sally L. John, Fiona Cunningham, Jennifer Asimit, and Eleftheria Zeggini. 2012. “A Combined Functional Annotation Score for Non-Synonymous Variants.” *Human Heredity* 73 (1): 47–51. https://doi.org/10.1159/000334984.

Lornage, Xavière, Vanessa Schartner, Inès Balbueno, Valérie Biancalana, Tracey Willis, Andoni Echaniz-Laguna, Sophie Scheidecker, et al. 2019. “Clinical, Histological, and Genetic Characterization of PYROXD1-Related Myopathy.” *Acta Neuropathologica Communications* 7 (1): 138. https://doi.org/10.1186/s40478-019-0781-8.

Mahmood, Khalid, Chol-hee Jung, Gayle Philip, Peter Georgeson, Jessica Chung, Bernard J. Pope, and Daniel J. Park. 2017. “Variant Effect Prediction Tools Assessed Using Independent, Functional Assay-Based Datasets: Implications for Discovery and Diagnostics.” *Human Genomics* 11 (1): 10. https://doi.org/10.1186/s40246-017-0104-8.

Maron, Barry J., Martin S. Maron, and Christopher Semsarian. 2012. “Genetics of Hypertrophic Cardiomyopathy after 20 Years: Clinical Perspectives.” *Journal of the American College of Cardiology*. Elsevier Inc. https://doi.org/10.1016/j.jacc.2012.02.068.

Mayer, Anja K, Klaus Rohrschneider, Tim M Strom, Nicola Glöckle, Susanne Kohl, Bernd Wissinger, and Nicole Weisschuh. 2016. “Homozygosity Mapping and Whole-Genome Sequencing Reveals a Deep Intronic PROM1 Mutation Causing Cone–Rod Dystrophy by Pseudoexon Activation.” *European*

Journal of Human Genetics 24 (3): 459–62. https://doi.org/10.1038/ejhg.2015.144.

Mclaren, William, Laurent Gil, Sarah E Hunt, Harpreet Singh Riat, Graham R S Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. 2016. “The Ensembl Variant Effect Predictor.” *Genome Biology*, 1–14. https://doi.org/10.1186/s13059-016-0974-4.

Miosge, Lisa A., Matthew A. Field, Yovina Sontani, Vicky Cho, Simon Johnson, Anna Palkova, Bhavani Balakishnan, et al. 2015. “Comparison of Predicted and Actual Consequences of Missense Mutations.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (37): E5189–98. https://doi.org/10.1073/pnas.1511585112.

“MMSplice Threshold.” n.d. Accessed August 19, 2019. https://github.com/gagneurlab/MMSplice/blob/master/notebooks/example.ipynb.

Ouellette, A. C., J. Mathew, A. K. Manickaraj, G. Manase, L. Zahavich, J. Wilson, K. George, L. Benson, S. Bowdin, and S. Mital. 2018. “Clinical Genetic Testing in Pediatric Cardiomyopathy: Is Bigger Better?” *Clinical Genetics* 93 (1): 33–40. https://doi.org/10.1111/cge.13024.

Pedersen, Brent S., Ryan M. Layer, and Aaron R. Quinlan. 2016. “Vcfanno: Fast, Flexible Annotation of Genetic Variants.” *Genome Biology* 17 (1): 118. https://doi.org/10.1186/s13059-016-0973-5.

Pejaver, Vikas, Jorge Urresti, Jose Lugo-Martinez, Kymberleigh A. Pagel, Guan Ning Lin, Hyun-Jun Nam, Matthew Mort, et al. 2017. “MutPred2: Inferring the Molecular and Phenotypic Impact of Amino Acid Variants.” *BioRxiv*, May, 134981. https://doi.org/10.1101/134981.

Quang, Daniel, Yifei Chen, and Xiaohui Xie. 2015. “DANN: A Deep Learning Approach for Annotating the Pathogenicity of Genetic Variants.” *Bioinformatics* 31 (5): 761–63. https://doi.org/10.1093/bioinformatics/btu703.

Quinlan, Aaron R, and Ira M Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics (Oxford, England)* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Raney, B. J., T. R. Dreszer, G. P. Barber, H. Clawson, P. A. Fujita, T. Wang, N. Nguyen, et al. 2014. “Track Data Hubs Enable Visualization of User-Defined Genome-Wide Annotations on the UCSC Genome Browser.” *Bioinformatics* 30 (7): 1003–5. https://doi.org/10.1093/bioinformatics/btt637.

“ReMM Threshold.” n.d. Accessed May 28, 2018. https://github.com/exomiser/Exomiser/issues/268.

Rentas, Stefan, Komal S. Rathi, Maninder Kaur, Pichai Raman, Ian D. Krantz, Mahdi Sarmady, and Ahmad Abou Tayoun. 2020. “Diagnosing Cornelia de Lange Syndrome and Related Neurodevelopmental Disorders Using RNA Sequencing.” *Genetics in Medicine*, January, 1–10. https://doi.org/10.1038/s41436-019-0741-5.

Rentzsch, Philipp, Daniela Witten, Gregory M. Cooper, Jay Shendure, and Martin Kircher. 2019. “CADD: Predicting the Deleteriousness of Variants throughout the Human Genome.” *Nucleic Acids Research* 47 (D1): D886–94. https://doi.org/10.1093/nar/gky1016.

Reva, Boris, Yevgeniy Antipin, and Chris Sander. 2011. “Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics.” *Nucleic Acids Research* 39 (17): e118–e118. https://doi.org/10.1093/nar/gkr407.

Richard, Pascale, Philippe Charron, Lucie Carrier, Céline Ledeuil, Theary Cheav, Claire Pichereau, Abdelaziz Benaiche, et al. 2003. “Hypertrophic Cardiomyopathy: Distribution of Disease Genes, Spectrum of Mutations, and Implications for a Molecular Diagnosis Strategy.” *Circulation* 107 (17): 2227–32. https://doi.org/10.1161/01.CIR.0000066323.15244.54.

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. 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.” *Genetics in Medicine* 17 (5): 405–24. https://doi.org/10.1038/gim.2015.30.

Richardson, Tom G., Colin Campbell, Nicholas J Timpson, and Tom R. Gaunt. 2016. “Incorporating Non-Coding Annotations into Rare Variant Analysis.” Edited by Junwen Wang. *PLOS ONE* 11 (4): e0154181. https://doi.org/10.1371/journal.pone.0154181.

Riley, Lisa G., Leigh B. Waddell, Roula Ghaoui, Frances J. Evesson, Beryl B. Cummings, Samantha J. Bryen, Himanshu Joshi, et al. 2019. “Recessive DES Cardio/Myopathy without Myofibrillar Aggregates: Intronic Splice Variant Silences One Allele Leaving Only Missense L190P-Desmin.” *European Journal of Human Genetics* 27 (8): 1267–73. https://doi.org/10.1038/s41431-019-0393-6.

Ritchie, Graham R S, Ian Dunham, Eleftheria Zeggini, and Paul Flicek. 2014. “Functional Annotation of Noncoding Sequence Variants.” *Nature Methods* 11 (3): 294–96. https://doi.org/10.1038/nmeth.2832.

Rosenberg, Alexander B., Rupali P. Patwardhan, Jay Shendure, and Georg Seelig. 2015. “Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences.” *Cell* 163 (3): 698–711. https://doi.org/10.1016/j.cell.2015.09.054.

Salgado, David, Jean Pierre Desvignes, Ghadi Rai, Arnaud Blanchard, Morgane Miltgen, Amélie Pinard, Nicolas Lévy, Gwenaëlle Collod-Béroud, and Christophe Béroud. 2016. “UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of Any Human CDNA Substitution.” *Human Mutation* 37 (5): 439–46. https://doi.org/10.1002/humu.22965.

Sangermano, Riccardo, Alejandro Garanto, Mubeen Khan, Esmee H. Runhart, Miriam Bauwens, Nathalie M. Bax, L. Ingeborgh van den Born, et al. 2019. “Deep-Intronic ABCA4 Variants Explain Missing Heritability in Stargardt Disease and Allow Correction of Splice Defects by Antisense Oligonucleotides.” *Genetics in Medicine* 21 (8): 1751–60. https://doi.org/10.1038/s41436-018-0414-9.

Sanz, David J., Jennifer A. Hollywood, Martina F. Scallan, and Patrick T. Harrison. 2017. “Cas9/GRNA Targeted Excision of Cystic Fibrosis-Causing Deep-Intronic Splicing Mutations Restores Normal Splicing of CFTR MRNA.” Edited by Emanuele Buratti. *PLOS ONE* 12 (9): e0184009. https://doi.org/10.1371/journal.pone.0184009.

Schaafsma, Gerard C. P., and Mauno Vihinen. 2018. “Representativeness of Variation Benchmark Datasets.” *BMC Bioinformatics* 19 (1): 461. https://doi.org/10.1186/s12859-018-2478-6.

Schalk, Audrey, Géraldine Greff, Nathalie Drouot, Cathy Obringer, Hélène Dollfus, Vincent Laugel, Jamel Chelly, and Nadège Calmels. 2018. “Deep Intronic Variation in Splicing Regulatory Element of the ERCC8 Gene Associated with Severe but Long-Term Survival Cockayne Syndrome.” *European Journal of Human Genetics* 26 (4): 527–36. https://doi.org/10.1038/s41431-017-0009-y.

Schwarz, Jana Marie, David N. Cooper, Markus Schuelke, and Dominik Seelow. 2014. “Mutationtaster2: Mutation Prediction for the Deep-Sequencing Age.” *Nature Methods* 11 (4): 361–62. https://doi.org/10.1038/nmeth.2890.

Shabalina, Svetlana A., Aleksey Y. Ogurtsov, Alexey N. Spiridonov, Pavel S. Novichkov, Nikolay A. Spiridonov, and Eugene V. Koonin. 2010. “Distinct Patterns of Expression and Evolution of Intronless and Intron-Containing Mammalian Genes.” *Molecular Biology and Evolution* 27 (8): 1745–49. https://doi.org/10.1093/molbev/msq086.

Shah, Naisha, Ying-Chen Claire Hou, Hung-Chun Yu, Rachana Sainger, C. Thomas Caskey, J. Craig Venter, and Amalio Telenti. 2018. “Identification of Misclassified ClinVar Variants via Disease Population Prevalence.” *The American Journal of Human Genetics* 102 (4): 609–19. https://doi.org/10.1016/J.AJHG.2018.02.019.

Shihab, Hashem A., Mark F. Rogers, Julian Gough, Matthew Mort, David N. Cooper, Ian N. M. Day, Tom R. Gaunt, and Colin Campbell. 2015. “An Integrative Approach to Predicting the Functional Effects of Non-Coding and Coding Sequence Variation.” *Bioinformatics* 31 (10): 1536–43. https://doi.org/10.1093/bioinformatics/btv009.

Siepel, Adam, Gill Bejerano, Jakob S Pedersen, Angie S Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram

Clawson, et al. 2005. “Evolutionarily Conserved Elements in Vertebrate, Insect, Worm, and Yeast Genomes.” *Genome Research* 15 (8): 1034–50. https://doi.org/10.1101/gr.3715005.

Siepel, Adam, Katherine S. Pollard, and David Haussler. 2006. “New Methods for Detecting Lineage-Specific Selection.” In *Research in Computational Molecular Biology*, 190–205. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11732990_17.

Smedley, Damian, Max Schubach, Julius O.B. O.B. Jacobsen, Sebastian Köhler, Tomasz Zemojtel, Malte Spielmann, Marten Jäger, et al. 2016. “A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease.” *American Journal of Human Genetics* 99 (3): 595–606. https://doi.org/10.1016/j.ajhg.2016.07.005.

Starokadomskyy, Petro, Terry Gemelli, Jonathan J Rios, Chao Xing, Richard C Wang, Haiying Li, Vladislav Pokatayev, et al. 2016. “DNA Polymerase-α Regulates the Activation of Type I Interferons through Cytosolic RNA:DNA Synthesis.” *Nature Immunology* 17 (5): 495–504. https://doi.org/10.1038/ni.3409.

States, United, United States, I Glenn Cohen, Eli Y Adashi, and Vardit Ravitsky. 2019. “The Kipoi Repository Accelerates Community Exchange and Reuse of Predictive Models for Genomics.” *Nature Biotechnology* 37 (June).

Stenson, Peter D, Matthew Mort, Edward V Ball, Katy Evans, Matthew Hayden, Sally Heywood, Michelle Hussain, Andrew D Phillips, and David N Cooper. 2017. “The Human Gene Mutation Database: Towards a Comprehensive Repository of Inherited Mutation Data for Medical Research, Genetic Diagnosis and next-Generation Sequencing Studies.” *Human Genetics* 136 (6): 665–77. https://doi.org/10.1007/s00439-017-1779-6.

Tozawa, Yusuke, Shimaa Said Mohamed Ali Abdrabou, Natsuko Nogawa-Chida, Ritsuo Nishiuchi, Toshiaki Ishida, Yuichi Suzuki, Hideki Sano, et al. 2019. “A Deep Intronic Mutation of c.1166-285 T > G in SLC46A1 Is Shared by Four Unrelated Japanese Patients with Hereditary Folate Malabsorption (HFM).” *Clinical Immunology* 208 (November): 108256. https://doi.org/10.1016/J.CLIM.2019.108256.

Vaché, Christel, Thomas Besnard, Pauline le Berre, Gema García-García, David Baux, Lise Larrieu, Caroline Abadie, et al. 2012. “Usher Syndrome Type 2 Caused by Activation of an USH2A Pseudoexon: Implications for Diagnosis and Therapy.” *Human Mutation* 33 (1): 104–8. https://doi.org/10.1002/humu.21634.

Vandervore, Laura V, Rachel Schot, Esmee Kasteleijn, Renske Oegema, Katrien Stouffs, Alexander Gheldof, Martyna M Grochowska, et al. 2019. “Heterogeneous Clinical Phenotypes and Cerebral Malformations Reflected by Rotatin Cellular Dynamics.” *Brain* 142 (4): 867–84. https://doi.org/10.1093/brain/awz045.

Vaz-Drago, Rita, Noélia Custódio, and Maria Carmo-Fonseca. 2017. “Deep Intronic Mutations and Human Disease.” *Human Genetics* 136 (9): 1093–1111. https://doi.org/10.1007/s00439-017-1809-4.

Velde, K Joeri van der, Joël Kuiper, Bryony A Thompson, John-Paul Plazzer, Gert van Valkenhoef, Mark de Haan, Jan D H Jongbloed, et al. 2015. “Evaluation of CADD Scores in Curated Mismatch Repair Gene Variants Yields a Model for Clinical Validation and Prioritization.” *Human Mutation* 36 (7): 712–19. https://doi.org/10.1002/humu.22798.

Verbakel, Sanne K., Zeinab Fadaie, B. Jeroen Klevering, Maria M. van Genderen, Ilse Feenstra, Frans P. M. Cremers, Carel B. Hoyng, and Susanne Roosing. 2019. “The Identification of a RNA Splice Variant in *TULP1* in Two Siblings with Early-onset Photoreceptor Dystrophy.” *Molecular Genetics & Genomic Medicine* 7 (6): e660. https://doi.org/10.1002/mgg3.660.

Verdura, Edgard, Agatha Schlüter, Gorka Fernández-Eulate, Raquel Ramos-Martín, Miren Zulaica, Laura Planas-Serra, Montserrat Ruiz, et al. 2020. “A Deep Intronic Splice Variant Advises Reexamination of Presumably Dominant *SPG7* Cases.” *Annals of Clinical and Translational Neurology* 7 (1): 105–11. https://doi.org/10.1002/acn3.50967.

Walters-Sen, Lauren C, Sayaka Hashimoto, Devon Lamb Thrush, Shalini Reshmi, Julie M Gastier-Foster,

Caroline Astbury, and Robert E Pyatt. 2015. “Variability in Pathogenicity Prediction Programs: Impact on Clinical Diagnostics.” *Molecular Genetics & Genomic Medicine* 3 (2): 99–110. https://doi.org/10.1002/mgg3.116.

Wang, Guey Shin, and Thomas A. Cooper. 2007. “Splicing in Disease: Disruption of the Splicing Code and the Decoding Machinery.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2164.

Ward, Lucas D, and Manolis Kellis. 2012. “Interpreting Noncoding Genetic Variation in Complex Traits and Human Disease.” *Nature Biotechnology* 30 (11): 1095–1106. https://doi.org/10.1038/nbt.2422.

Weisschuh, Nicole, Marc Sturm, Britta Baumann, Isabelle Audo, Carmen Ayuso, Beatrice Bocquet, Kari Branham, et al. 2020. “Deep-intronic Variants in *CNGB3* Cause Achromatopsia by Pseudoexon Activation.” *Human Mutation* 41 (1): 255–64. https://doi.org/10.1002/humu.23920.

Whiffin, Nicola, Eric Minikel, Roddy Walsh, Anne H. O’Donnell-Luria, Konrad Karczewski, Alexander Y. Ing, Paul J.R. Barton, et al. 2017. “Using High-Resolution Variant Frequencies to Empower Clinical Genome Interpretation.” *Genetics in Medicine* 19 (10): 1151–58. https://doi.org/10.1038/gim.2017.26.

Xiong, Hui Y., Babak Alipanahi, Leo J. Lee, Hannes Bretschneider, Daniele Merico, Ryan K.C. Yuen, Yimin Hua, et al. 2015. “The Human Splicing Code Reveals New Insights into the Genetic Determinants of Disease.” *Science* 347 (6218). https://doi.org/10.1126/science.1254806.

Xu, Ying, Tingting Song, Yu Li, Fenfen Guo, Xin Jin, Lu Cheng, Jiao Zheng, et al. 2019. “Identification of Two Novel Insertion Abnormal Transcripts in Two Chinese Families Affected with Dystrophinopathy.” *Journal of Clinical Laboratory Analysis*, December, e23142. https://doi.org/10.1002/jcla.23142.

Yang, Yaping, Donna M. Muzny, Jeffrey G. Reid, Matthew N. Bainbridge, Alecia Willis, Patricia A. Ward, Alicia Braxton, et al. 2013. “Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders.” *New England Journal of Medicine* 369 (16): 1502–11. https://doi.org/10.1056/NEJMoa1306555.

Yeo, Gene, and Christopher B. Burge. 2004. “Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals.” *Journal of Computational Biology* 11 (2–3): 377–94. https://doi.org/10.1089/1066527041410418.

Zernant, Jana, Winston Lee, Takayuki Nagasaki, Frederick T. Collison, Gerald A. Fishman, Mette Bertelsen, Thomas Rosenberg, Peter Gouras, Stephen H. Tsang, and Rando Allikmets. 2018. “Extremely Hypomorphic and Severe Deep Intronic Variants in the *ABCA4* Locus Result in Varying Stargardt Disease Phenotypes.” *Molecular Case Studies* 4 (4): a002733. https://doi.org/10.1101/mcs.a002733.

II.7. SUPPLEMENTAL MATERIAL

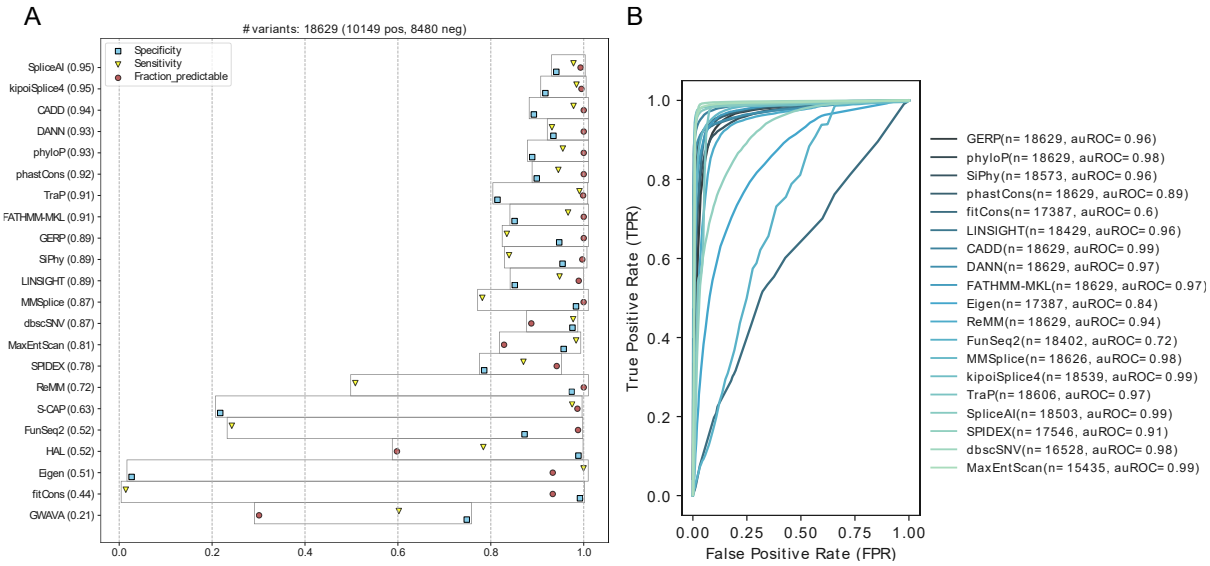


Figure S II.1. Prediction tools performance on Clinvar variants located in the 0-10bp intronic bin. A) Analysis using predefined thresholds. Tools with more than 90% of missing scores were discarded. Ranking metric presented refers to the accuracy weighted by the fraction of predictable variants. B) ROC curve analysis for the methods predicting more than 30% of the variants. S-CAP was not included in this analysis because it contains more than one recommended threshold in intronic regions.

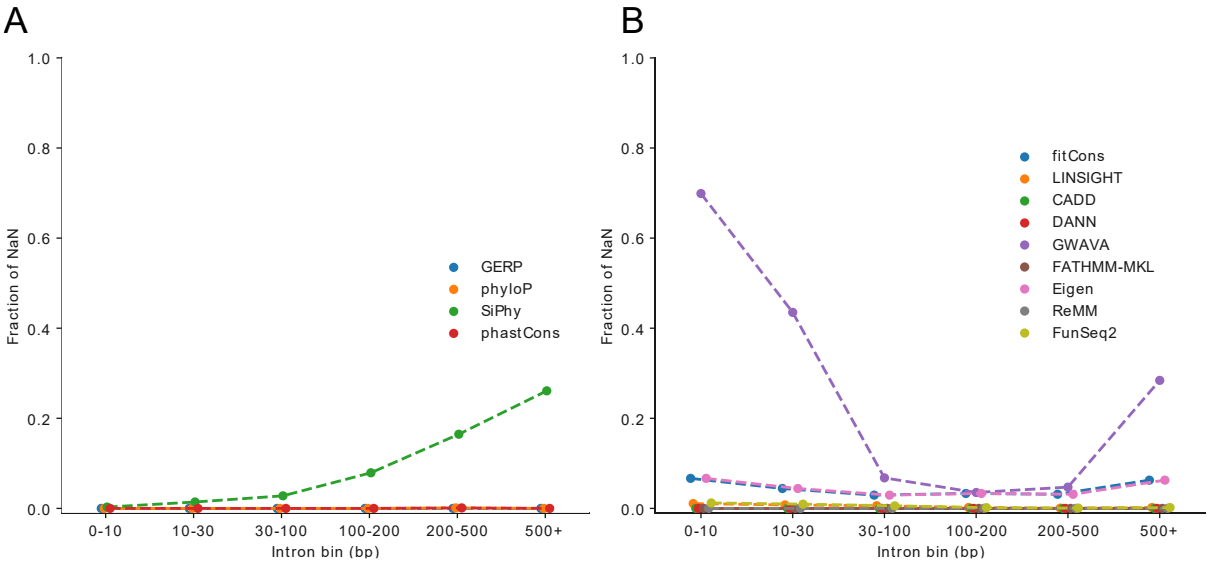


Figure S II.2. Fraction of unscored variants at each intronic bin. A) Conservation scores. B) Whole genome predictors

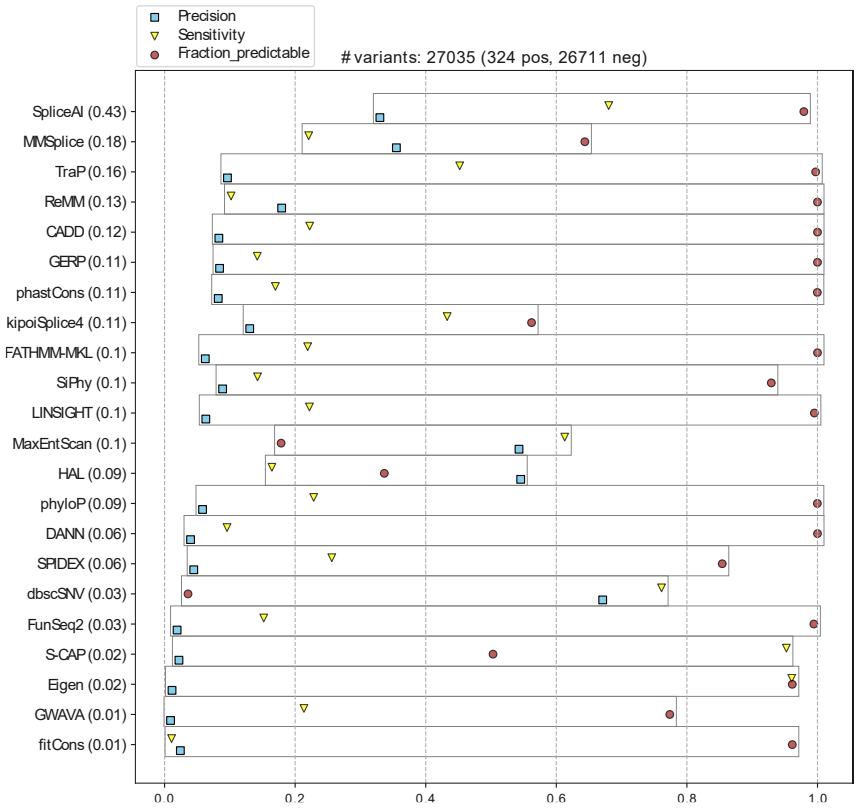


Figure S II.3. Prediction tools performance on Clinvar variants located in all the bins except the 0-10bp one, which shows the large class imbalance. Ranking metric presented is F1 score weighted by the fraction of predictable variants.

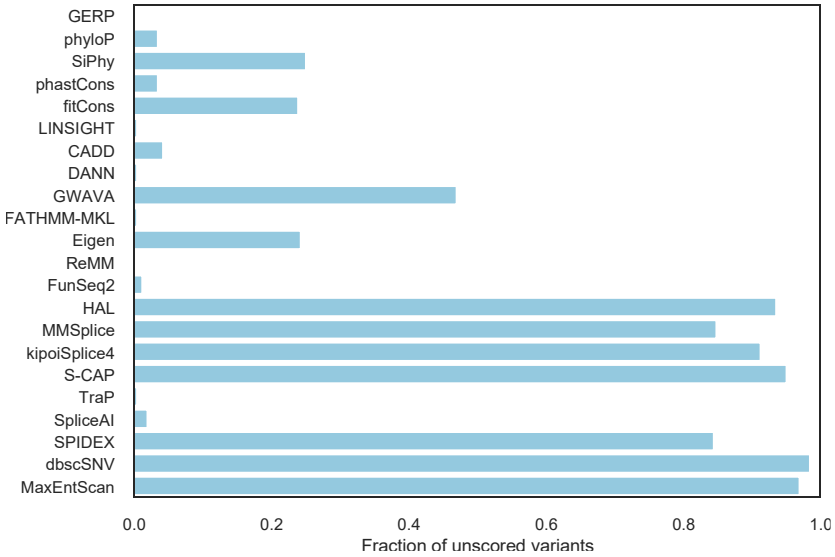


Figure S II.4. Fraction of unscored variants in the deep intronic dataset

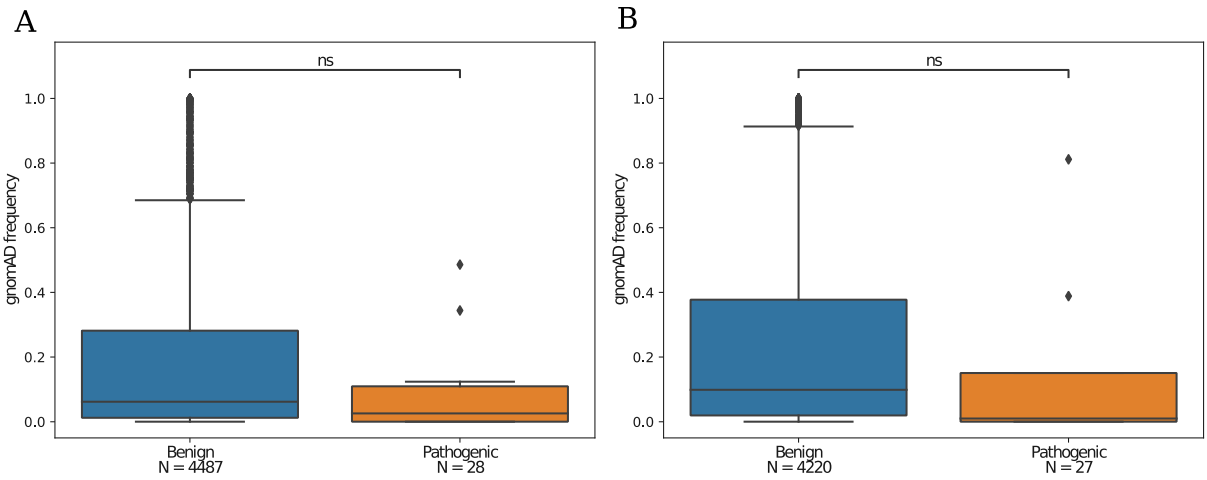


Figure S II.5.: gnomAD frequency of Clinvar variants at two different intronic bins reveal no significant differences between Benign and Pathogenic classifications. A) 30-100bp bin (Wilcoxon rank-sum test, p-value = 9.290e-02). B) 100-200bp bin (Wilcoxon rank-sum test, p-value =6.728e-02)

Table S II.1. HCM associated pathogenic variants evaluated

Chr	Position	Ref	Alt	Gene	HGVSc	Consequence	rsID	Max frequency	Max frequency pop	Frequency gnomAD v2.1.1 (genomes
1	201332458	G	A	TNNT2	ENST00000509001.1: c.536C>T	missense variant	rs727504246; CM002871	None	None	.
1	201332477	G	A	TNNT2	ENST00000509001.1: c.517C>T	missense variant	rs727503512; CM127304	None	None	.
1	201333464	G	A	TNNT2	ENST00000509001.1: c.421C>T	missense variant	rs74315379; CD117175; CM013440	None	None	.
1	201333497	G	A	TNNT2	ENST00000509001.1: c.388C>T	missense variant	rs397516463; CM962587	None	None	.
1	201334425	C	T	TNNT2	ENST00000509001.1: c.275G>A	missense variant	rs121964856; CM951218; CM961373; COSM6805424; COSM6805425; COSM6805426	None	None	.
1	201334426	G	A	TNNT2	ENST00000509001.1: c.274C>T	missense variant	rs397516456; CM971501; COSM4026926; COSM4026927; COSM4026928	8.793e-06	gnomAD NFE	3.1858999136602506e-05
1	201334758	C	T	TNNT2	ENST00000509001.1: c.244G>A	missense variant	rs727504255	None	None	.
1	201334766	A	T	TNNT2	ENST00000509001.1: c.236T>A	missense variant	rs121964855; CM132559; CM951217	8.8e-06	gnomAD NFE	.
3	46901001	T	C	MYL3	ENST00000395869.1: c.445A>G	missense variant	rs104893748; CM961007	None	None	.
3	46902192	C	T	MYL3	ENST00000395869.1: c.281G>A	missense variant	rs199474703; CM082965	2.892e-05	gnomAD AMR	.
3	46902303	G	C	MYL3	ENST00000395869.1: c.170C>G	missense variant	rs139794067; CM014210	0.0002718	gnomAD EAS	.
11	47353433	C	T	MYBPC3	ENST00000545968.1: c.3815-1G>A	splice acceptor variant	rs397516044	6.847e-05	gnomAD AFR	.
11	47354116	C	T	MYBPC3	ENST00000545968.1: c.3627+1G>A	splice donor variant	rs397516031; CS1010363; CS982279	None	None	.
11	47354255	T	A	MYBPC3	ENST00000545968.1: c.3491-2A>T	splice acceptor variant	rs397516022	None	None	.
11	47354526	T	G	MYBPC3	ENST00000545968.1: c.3331-2A>C	splice acceptor variant	rs869025469	0.0001407	gnomAD AFR	.
11	47354743	A	C	MYBPC3	ENST00000545968.1: c.3330+2T>G	splice donor variant	rs387906397; CS034593; CS081937; FHC0086	None	None	3.193500015186146e-05
11	47355106	A	C	MYBPC3	ENST00000545968.1: c.3190+2T>G	splice donor variant	rs113358486; CS135613	1.777e-05	gnomAD NFE	.
11	47355107	C	T	MYBPC3	ENST00000545968.1: c.3190+1G>A	splice donor variant	rs111683277; CS103025	9.97e-05	gnomAD ASJ	.

11	47359280	A	G	MYBPC3	ENST00000545968.1:c.2374T>C	missense variant	rs187830361; CM043543	0.0001203	EA	.
11	47359347	T	C	MYBPC3	ENST00000545968.1:c.2309-2A>G	splice acceptor variant	rs111729952; CS043648	None	None	.
11	47360070	C	T	MYBPC3	ENST00000545968.1:c.2308+1G>A	splice donor variant	rs112738974; CS133963; CS971816; CS982278; FHC0062	0.0001191	EA; EA	.
11	47360070	C	A	MYBPC3	ENST00000545968.1:c.2308+1G>T	splice donor variant	rs112738974; CS133963; CS971816; CS982278; FHC0061	None	None	.
11	47361343	T	C	MYBPC3	ENST00000545968.1:c.1928-2A>G	splice acceptor variant	rs397515937; CS014131	None	None	.
11	47362688	C	T	MYBPC3	ENST00000545968.1:c.1897+1G>A	splice donor variant	rs397515935	None	None	.
11	47364127	A	G	MYBPC3	ENST00000545968.1:c.1624+2T>C	splice donor variant	rs111437311; CS036038; FHC0032	None	None	.
11	47364129	C	G	MYBPC3	ENST00000545968.1:c.1624G>C	missense variant; splice region variant	rs121909374; CM971007	0.0002382	AA	6.374700024025515e-05
11	47364248	C	T	MYBPC3	ENST00000545968.1:c.1505G>A	missense variant	rs397515907; CM981325; FHC0027	None	None	.
11	47364269	C	T	MYBPC3	ENST00000545968.1:c.1484G>A	missense variant	rs200411226; CM981324; COSM5510502; COSM5510503	5.564e-05	gnomAD EAS	.
11	47364296	C	T	MYBPC3	ENST00000545968.1:c.1458-1G>A	splice acceptor variant	rs397515903; CS1110418	None	None	.
11	47364570	A	G	MYBPC3	ENST00000545968.1:c.1351+2T>C	splice donor variant	rs397515897	None	None	.
11	47364571	C	T	MYBPC3	ENST00000545968.1:c.1351+1G>A	splice donor variant	rs727503204; CS0910205	None	None	.
11	47364698	T	C	MYBPC3	ENST00000545968.1:c.1227-2A>G	splice acceptor variant	rs730880531	None	None	3.185500099789351e-05
11	47367757	C	T	MYBPC3	ENST00000545968.1:c.1090+1G>A	splice donor variant	rs727504269; CS068101; CS106129; CS119774	None	None	.
11	47367757	C	A	MYBPC3	ENST00000545968.1:c.1090+1G>T	splice donor variant	rs727504269; CS068101; CS106129; CS119774	8.875e-06	gnomAD NFE	.
11	47367923	T	C	MYBPC3	ENST00000545968.1:c.927-2A>G	splice acceptor variant	rs397516082; CS982277; COSM3383501; COSM3383502	1.164e-05	gnomAD NFE	3.187599941156804 6e-05
11	47369407	C	T	MYBPC3	ENST00000545968.1:c.821+1G>A	splice donor variant	rs397516073; CS034591; CS982276; FHC0005	0.0002127	gnomAD OTH; gnomAD OTH	.
11	47369407	C	G	MYBPC3	ENST00000545968.1:c.821+1G>C	splice donor variant	rs397516073; CS034591; CS982276	None	None	.
11	47369974	C	T	MYBPC3	ENST00000545968.1:c.772+1G>A	splice donor variant	rs397516072	None	None	.
11	47369975	C	T	MYBPC3	ENST00000545968.1:c.772G>A	missense variant; splice region variant	rs397516074; CM981322	3.671e-05	gnomAD NFE	6.372299685608596e-05
11	47370092	C	G	MYBPC3	ENST00000545968.1:c.655G>C	missense variant; splice region variant	rs397516068; CM043535; CM106111	None	None	.
11	47370092	C	A	MYBPC3	ENST00000545968.1:c.655G>T	missense variant; splice region variant	rs397516068; CM043535; CM106111	0.0001695	gnomAD OTH	.
11	47370093	C	T	MYBPC3	ENST00000545968.1:c.655-1G>A	splice acceptor variant	rs397516067; CS106128	None	None	.
11	47373058	T	C	MYBPC3	ENST00000545968.1:c.26-2A>G	splice acceptor variant	rs376395543; CS043791	0.0001204	EA	3.185500099789351e-05
12	111352091	C	T	MYL2	ENST00000228841.8:c.173G>A	missense variant	rs104894369; CM981332	9.24e-05	gnomAD FIN	.
12	111356937	C	T	MYL2	ENST00000228841.8:c.64G>A	missense variant	rs104894368; CM961005; FHC0203; MYL2:c.64G>A; COSM245869	0.0014	AMR; AMR; AMR	.
14	23884861	G	A	MYH7	ENST00000355349.3:c.5134C>T	missense variant	rs121913650; CM050714; COSM4050058	None	None	.
14	23886383	G	A	MYH7	ENST00000355349.3:c.4498C>T	missense variant	rs45544633; CM045547; COSM469829	None	None	.
14	23887453	C	T	MYH7	ENST00000355349.3:c.4135G>A	missense variant	rs397516202; CM020288	None	None	3.185300010954961e-05
14	23893250	C	T	MYH7	ENST00000355349.3:c.2788G>A	missense variant	rs397516171; CM068020; CM952164; FHC0186	None	None	.
14	23893268	C	T	MYH7	ENST00000355349.3:c.2770G>A	missense variant	rs121913628; CM032609; CM920494	None	None	.

14	23893328	G	A	MYH7	ENST00000355349.3:c.2710C>T	missense variant	rs727503253; CM112510; COSM1258462	9.921e-05	gnomAD ASJ	.
14	23894525	C	T	MYH7	ENST00000355349.3:c.2389G>A	missense variant	rs3218716; CM066923; CM950825	0.000227	AA	.
14	23894612	C	T	MYH7	ENST00000355349.3:c.2302G>A	missense variant	rs727503260; CM109192; FHC0161; HM971867	None	None	.
14	23894969	C	T	MYH7	ENST00000355349.3:c.2221G>A	missense variant	rs121913632; CM930505; CM931218; CM952162; FHC0159	None	None	3.185300010954961e-05
14	23894999	G	A	MYH7	ENST00000355349.3:c.2191C>T	missense variant	rs727504299; CM133936; COSM230283	None	None	.
14	23896043	G	A	MYH7	ENST00000355349.3:c.1987C>T	missense variant	rs397516127; CM031273; CM973126; COSM5612930	None	None	.
14	23896451	T	C	MYH7	ENST00000355349.3:c.1954A>G	missense variant; splice region variant	rs727504239; CM025421	8.795e-06	gnomAD NFE	.
14	23896866	C	T	MYH7	ENST00000355349.3:c.1816G>A	missense variant	rs121913627; CM920492; COSM954772	8.79e-06	gnomAD NFE	3.185500099789351e-05
14	23900656	C	T	MYH7	ENST00000355349.3:c.767G>A	missense variant	rs121913633; CM930502; COSM3495248	None	None	.
14	23900677	C	T	MYH7	ENST00000355349.3:c.746G>A	missense variant	rs3218713; CM910268; COSM954783	None	None	.
14	23901912	C	A	MYH7	ENST00000355349.3:c.438G>T	missense variant	rs397516212; CM054005	None	None	.
15	35085599	C	T	ACTC1	ENST00000290378.4:c.301G>A	missense variant	rs193922680; CM003400; COSM960888	8.79e-06	gnomAD NFE	.
15	63349227	T	C	TPM1	ENST00000358278.3:c.284T>C	missense variant	rs104894504; CM010475; COSM3502945; COSM3502946; COSM3502947; COSM3502948; COSM3502949	None	None	.
15	63353098	G	A	TPM1	ENST00000358278.3:c.523G>A	missense variant	rs104894503; CM941333	0.0001389	gnomAD FIN	3.184299930580892e-05
15	63354462	G	A	TPM1	ENST00000358278.3:c.688G>A	missense variant	rs199476317; CM100401	None	None	.
19	55663224	C	T	TNNI3	ENST00000344887.5:c.611G>A	missense variant	rs727504275; CM050764; COSM5699804; COSM7303011	None	None	.
19	55663260	C	T	TNNI3	ENST00000344887.5:c.575G>A	missense variant	rs104894729; CM030288; FHC0415; COSM5183237	None	None	.
19	55663278	C	T	TNNI3	ENST00000344887.5:c.557G>A	missense variant	rs397516357; CM031381	None	None	.
19	55665438	C	T	TNNI3	ENST00000344887.5:c.509G>A	missense variant	rs727503503; CM0910618; COSM4429422	None	None	.
19	55665462	C	T	TNNI3	ENST00000344887.5:c.485G>A	missense variant	rs397516354; CM031380; CM034575; FHC0392; COSM6212838	6.536e-05	gnomAD SAS; gnomAD SAS	.
19	55665477	G	A	TNNI3	ENST00000344887.5:c.470C>T	missense variant	rs397516353; CM031379; COSM6388939	None	None	.
19	55665513	C	T	TNNI3	ENST00000344887.5:c.434G>A	missense variant	rs397516349; CM971497	0.0001113	gnomAD EAS	.
19	55665514	G	A	TNNI3	ENST00000344887.5:c.433C>T	missense variant	rs104894724; CM030285; CM971498; FHC0405; COSM475295	3.268e-05	gnomAD SAS; gnomAD SAS	6.379599653882906e-05
19	55665514	G	C	TNNI3	ENST00000344887.5:c.433C>G	missense variant	rs104894724; CM030285; CM971498; FHC0407; COSM475295	None	None	.
19	55665525	C	T	TNNI3	ENST00000344887.5:c.422G>A	missense variant	rs397516347; CM031378; COSM5056941	None	None	3.187800029991194 6e-05

Table S II.2. Likely benign variants in cardiomyopathy related genes evaluated

Chr	Position	Ref	Alt	Gene	HGVSc	Consequence	rsID	Max frequency	Max frequency pop	Frequency gnomAD v2.1.1 (genomes)
1	3328659	C	T	PRDM16	ENST00000270722.5:c.1898C>T	missense variant	rs2493292	0.2233	gnomAD ASJ	0.13600000739
1	78392446	G	A	NEXN	ENST00000334785.7:c.733G>A	missense variant	rs1166698	0.3925	gnomAD EAS	0.165099993348
1	116283343	A	G	CASQ2	ENST00000261448.5:c.420+6T>C	splice region variant ; intron variant	rs9428083	0.878	EA	0.711799979209
1	116310967	T	C	CASQ2	ENST00000261448.5:c.196A>G	missense variant	rs4074536 ; CM108210	0.5488	gnomAD EAS	0.328500002622

1	156096387	G	A	LMNA	ENST00000504687.1 :c.52G>A	missense variant	rs593987	0.3215	AFR	0.100299999117
1	156096417	T	G	LMNA	ENST00000504687.1 :c.82T>G	missense variant	rs594028	0.6626	AFR	0.20759999752
1	156099669	T	G	LMNA	ENST00000473598.2 :c.29T>G	missense variant	rs513043	0.5408	AFR	0.177399992942
1	156107534	C	T	LMNA	ENST00000368300.4 :c.1698C>T	splice region variant ; synonymous variant	rs4641 ; CM003892	0.3237	gnomAD AMR	0.196199998259
1	236902594	C	G	ACTN2	ENST00000366578.4 :c.877-8C>G	splice region variant ; intron variant	rs2288601	0.7999	gnomAD EAS	0.78539997339
1	237433625	C	T	RYR2	ENST00000360064.6 :c.41+4C>T	splice region variant ; intron variant	rs2275287	0.6954	EAS	0.356000006198
1	237540615	A	C	RYR2	ENST00000366574.2 :c.464-8A>C	splice region variant ; intron variant	rs10925391	0.5484	AFR	0.345299988985
1	237586384	T	C	RYR2	ENST00000366574.2 :c.849-8T>C	splice region variant ; intron variant	rs16835237	0.3135	EAS	0.097599998116
1	237841390	A	G	RYR2	ENST00000366574.2 :c.8873A>G	missense variant	rs34967813 ; CM1110947	0.3069	gnomAD NFE	0.211999997496
1	237957161	A	G	RYR2	ENST00000366574.2 :c.13783-6A>G	splice region variant ; intron variant	rs790901	0.9494	EAS	0.682299971580
2	106013147	C	T	FHL2	ENST00000344213.4 :c.263G>A	missense variant	rs34179780	0.2245	gnomAD SAS	0.145699992775
2	106015412	G	C	FHL2	ENST00000344213.4 :c.142C>G	missense variant	rs76810592	0.0991	gnomAD AMR	0.061299998313
3	14174427	A	T	TMEM43	ENST00000306077.4 :c.504A>T	missense variant	rs4685076	0.4613	EAS	0.296600013971
3	14175262	T	C	TMEM43	ENST00000306077.4 :c.536T>C	missense variant	rs2340917	0.6756	EAS	0.354999989271
3	38645420	T	C	SCN5A	ENST00000413689.1 :c.1673A>G	missense variant	CM031355 ; TMP ESP 3 38645420 38645420 ; rs1805124	0.3094	AFR	0.24510000646
3	38647642	G	T	SCN5A	ENST00000413689.1 :c.1141-3C>A	splice region variant ; intron variant	rs41312433 ; CS057601	0.1933	SAS	0.174300000071
4	186423637	G	A	PDLIM3	ENST00000284770.5 :c.906C>T	splice region variant ; synonymous variant	rs4635850	0.8736	gnomAD FIN	0.71850001811
6	7576527	G	A	DSP	ENST00000379802.3 :c.2631G>A	splice region variant ; synonymous variant	rs1016835	0.8175	gnomAD ASJ	0.73540002107
6	7580958	A	G	DSP	ENST00000379802.3 :c.4535A>G	missense variant	rs2076299 ; COSM3831039	0.3139	AFR	0.128499999642
6	7581636	G	A	DSP	ENST00000379802.3 :c.5213G>A	missense variant	rs6929069	0.379	AFR	0.197099998593
6	112435273	A	C	LAMA4	ENST00000230538.7 :c.5326+6T>G	splice region variant ; intron variant	rs3734289	0.3472	EAS	0.249699994921
6	112457383	G	C	LAMA4	ENST00000230538.7 :c.3356C>G	missense variant	rs1050349	0.3601	EAS	0.21500000357
6	112457390	C	T	LAMA4	ENST00000230538.7 :c.3349G>A	missense variant	rs2032567	0.8926	SAS	0.77310001850
6	112493872	A	G	LAMA4	ENST00000230538.7 :c.1492T>C	missense variant	rs1050348	0.8224	gnomAD EAS	0.65609997510
6	112522852	G	A	LAMA4	ENST00000230538.7 :c.460C>T	missense variant	rs11757455	0.1005	gnomAD SAS	0.0428000018
6	123696766	G	T	TRDN	ENST00000398178.3 :c.1257C>A	missense variant	rs17737379	0.3116	AFR	0.17360000312
6	123699019	A	C	TRDN	ENST00000398178.3 :c.1211T>G	missense variant	rs28494009	0.2371	gnomAD SAS	0.17430000007152557
6	123699042	T	C	TRDN	ENST00000398178.3 :c.1188A>G	splice region variant ; synonymous variant	rs6901953	0.4251	AMR	0.33610001206
6	123869607	G	C	TRDN	ENST00000398178.3 :c.383C>G	missense variant	rs9490809	0.5785	gnomAD ASJ	0.50550001859
7	92077203	G	A	GATAD1	ENST00000287957.3 :c.160G>A	missense variant	rs10281879	0.45	gnomAD ASJ	0.1099999994
10	21112137	A	T	NEBL	ENST00000377122.4 :c.1962T>A	missense variant ; splice region variant	rs4748728	0.2337	AFR	0.0724000036
10	21134282	C	G	NEBL	ENST00000377122.4 :c.1132G>C	missense variant	rs41277370 ; COSM5915956	0.1666	gnomAD ASJ	0.0549999997
10	21139389	T	C	NEBL	ENST00000377122.4 :c.1051A>G	missense variant	rs4025981	0.1716	gnomAD ASJ	0.05530000105
10	69926334	C	G	MYPN	ENST00000358913.5 :c.1884C>G	missense variant	rs10823148	0.544	gnomAD FIN	0.4061999917
10	69933921	G	A	MYPN	ENST00000358913.5 :c.2072G>A	missense variant	rs10997975	0.5298	gnomAD FIN	0.40459999442
10	69933969	G	A	MYPN	ENST00000358913.5 :c.2120G>A	missense variant	rs7916821	0.5277	gnomAD FIN	0.401300013065
10	69934258	C	G	MYPN	ENST00000358913.5 :c.2409C>G	missense variant	rs3814182	0.6307	gnomAD FIN	0.52719998359

10	69959242	C	A	MYPN	ENST00000358913.5 :c.3403C>A	missense variant	rs7079481	0.5268	gnomAD FIN	0.4169999957
10	112544655	C	T	RBM20	ENST00000369519.3 :c.1527+8C>T	splice region variant ; intron variant	rs7077757	0.2288	gnomAD AFR	0.2152000069
10	121429633	T	C	BAG3	ENST00000369085.3 :c.451T>C	missense variant	rs2234962; CM1111343	0.2276	gnomAD FIN	0.14790000021
10	121436286	C	T	BAG3	ENST00000369085.3 :c.1220C>T	missense variant	rs3858340; CM1111345	0.2658	gnomAD EAS	0.12540000677
11	47370041	T	C	MYBPC3	ENST00000545968.1 :c.706A>G	missense variant	rs3729989; CM043536	0.1412	EUR	0.09650000184
11	47371598	C	T	MYBPC3	ENST00000545968.1 :c.472G>A	missense variant	rs3729986	0.1064	EUR	0.0709000006
11	111781047	A	C	CRYAB	ENST00000533971.1 :c.328T>G	missense variant	rs11603779	0.3601	gnomAD ASJ	0.25709998607
12	22068849	G	T	ABCC9	ENST00000261200.4 :c.574-5C>A	splice region variant ; intron variant	rs3759236	0.8026	EAS	0.59820002317
12	33021934	A	G	PKP2	ENST00000070846.6 :c.1097T>C	missense variant	rs1046116	0.2301	EA	0.19769999384
14	23858271	A	G	MYH6	ENST00000405093.3 :c.3979-7T>C	splice region variant ; intron variant	rs535111647	0.08033	gnomAD AMR	0.02180000022
14	23859610	C	T	MYH6	ENST00000405093.3 :c.3388G>A	missense variant	rs28730771	0.1352	EUR	0.0816999971
14	23861811	A	G	MYH6	ENST00000405093.3 :c.3302T>C	missense variant	rs365990	0.6437	AFR	0.41830000281
14	23876267	C	T	MYH6	ENST00000405093.3 :c.166G>A	missense variant	rs28711516; CM122341	0.1105	gnomAD ASJ	0.07050000131
14	23900794	G	A	MYH7	ENST00000355349.3 :c.732C>T	splice region variant ; synonymous variant	rs2069542	0.5976	AFR	0.2556000053
17	39912145	T	A	JUP	ENST00000393931.3 :c.2089A>T	missense variant ; splice region variant	rs1126821	0.7435	EUR	0.6647999882
18	3086065	C	T	MYOM1	ENST00000356443.4 :c.4222G>A	missense variant	rs3765623	0.09502	gnomAD SAS	0.05820000171
18	3100429	G	A	MYOM1	ENST00000356443.4 :c.3576-5C>T	splice region variant ; intron variant	rs7232329	0.5781	gnomAD AMR	0.35469999909
18	3126811	A	G	MYOM1	ENST00000356443.4 :c.2879T>C	missense variant	rs1071600	0.3205	gnomAD EAS	0.19210000336
18	3168816	G	A	MYOM1	ENST00000356443.4 :c.1338C>T	splice region variant ; synonymous variant	rs2230167	0.3313	AFR	0.281899988651
18	3176040	C	G	MYOM1	ENST00000356443.4 :c.1022G>C	missense variant ; splice region variant	rs8099021	0.8081	AA	0.7581999897
18	3188873	G	A	MYOM1	ENST00000356443.4 :c.644C>T	missense variant	rs2230165	0.149	AFR	0.0666000023
18	3188976	A	G	MYOM1	ENST00000356443.4 :c.541T>C	missense variant	rs1962519	0.8363	EAS	0.40259999036
18	3215158	C	G	MYOM1	ENST00000356443.4 :c.64G>C	missense variant	rs1791085	0.1845	EAS	0.10740000009
18	28649042	T	C	DSC2	ENST00000280904.6 :c.2326A>G	missense variant	rs1893963	0.472	AFR	0.17049999535
18	29104714	A	G	DSG2	ENST00000261590.8 :c.877A>G	missense variant	rs2230234; CM1312995	0.1453	gnomAD FIN	0.07169999927
18	29122618	G	A	DSG2	ENST00000261590.8 :c.2137G>A	missense variant	rs79241126; CM068260	0.09818	gnomAD FIN	0.06030000001
18	29122799	G	A	DSG2	ENST00000261590.8 :c.2318G>A	missense variant	rs2278792	0.4666	gnomAD EAS	0.228599995374
18	29172865	G	A	TTR	ENST00000237014.3 :c.76G>A	missense variant	rs1800458	0.09544	gnomAD FIN	0.058499999344
18	32470291	G	A	DTNA	ENST00000598334.1 :c.2062G>A	missense variant	rs9944927	0.5946	AFR	0.29379999637
19	55668509	A	T	TNNI3	ENST00000344887.5 :c.25-8T>A	splice region variant ; intron variant	rs3729836	0.6899	AFR	0.33840000629
20	42747247	C	T	JPH2	ENST00000372980.3 :c.1186G>A	missense variant	rs3810510	0.3933	AFR	0.19730000197
X	31893307	T	G	DMD	ENST00000357033.4 :c.7096A>C	missense variant ; splice region variant	rs1800275	0.5162	gnomAD AMR	0.20270000398
X	32380996	C	T	DMD	ENST00000357033.4 :c.5234G>A	missense variant	rs1801187 ; COSM4999532 ; COSM4999533 ; COSM4999534 ; COSM4999535	0.7348	gnomAD EAS	0.39179998636
X	32503194	T	C	DMD	ENST00000357033.4 :c.2645A>G	missense variant	rs228406	0.8861	gnomAD EAS	0.69510000944

Table S II.3. List of pathogenic deep intronic variants extracted from literature with functional studies performed. Blank fields indicate that specific information was not possible to retrieve from the reading of the paper.

HGVS	Gene ID	Disease	Molecular effect	Functional consequence	Disease expression	Ref
ABCA4 c.1937+435C>G	ENSG00000198691	Stargardt disease	splice silencer disruption			(Sangermano et al. 2019)
ABCA4 c.1938-619A>G	ENSG00000198691	Stargardt disease	alter cryptic splice donor site strength	174bp pseudoexon inclusion & 134bp pseudoexon inclusion		(Fadaie et al. 2019)
ABCA4 c.2919-826T>A	ENSG00000198691	Stargardt disease	alter cryptic splice donor site strength	133bp pseudoexon inclusion		(Fadaie et al. 2019)
ABCA4 c.3050+370C>T	ENSG00000198691	Stargardt disease	new splice donor site	105bp pseudoexon inclusion		(Fadaie et al. 2019)
ABCA4 c.4253+43G>A	ENSG00000198691	Stargardt disease	exonic splicing enhancer creation|exonic splicing silencer disruption	possible pseudoexon inclusion | exon skipping	hypomorph variant; compound heterozygous	(Zernant et al. 2018; Sangermano et al. 2019)
ABCA4 c.4352+61G>A	ENSG00000198691	Stargardt disease	alter cryptic splice donor site strength	59bp exon extension; 59bp partial intron retention		(Fadaie et al. 2019)
ABCA4 c.4539+1100A>G	ENSG00000198691	Stargardt disease	alter cryptic splice donor site			(Sangermano et al. 2019)
ABCA4 c.4539+1106C>T	ENSG00000198691	Stargardt disease	alter cryptic splice donor site			(Sangermano et al. 2019)
ABCA4 c.4539+2001G>A	ENSG00000198691	Stargardt disease	exonic splicing enhancer creation	pseudoexon inclusion		(Albert et al. 2018)
ABCA4 c.4539+2028C>T	ENSG00000198691	Stargardt disease	exonic splicing enhancer creation	pseudoexon inclusion		(Albert et al. 2018)
ABCA4 c.769-784C>T	ENSG00000198691	Stargardt disease	strength increase of intronic splice acceptor	296bp pseudoexon inclusion		(Sangermano et al. 2019)
ABCA4 c.859-506G>C	ENSG00000198691	Stargardt disease	strength cryptic intronic splice acceptor site	56bp pseudoexon inclusion		(Sangermano et al. 2019)
ANAPC1 c.2705-198C>T	ENSG00000153107	Rothmund-Thomson syndrome	new splice donor site	95bp pseudoexon inclusion		(Ajeawung et al. 2019)
CDKN2A c.458-554T>G	ENSG00000147889	Breast Cancer	new splice acceptor site		compound heterozygous	(R. Khan et al. 2019)
CFTR c.1679+1634A>G	ENSG00000001626	Cystic Fibrosis	new splice acceptor site	49bp pseudoexon inclusion		(Sanz et al. 2017)
CLRN1 c.254-649T>G	ENSG00000163646	Usher syndrome	alter cryptic splice donor site	pseudoexon inclusion		(A. O. Khan et al. 2017)
CNGB3 c.1663-1205G>A	ENSG00000170289	Achromatopsia	alter cryptic splice donor site strength	34bp pseudoexon inclusion		(Weisschuh et al. 2020)
CNGB3 c.1663-2137C>T	ENSG00000170289	Achromatopsia	new splice donor site	99bp pseudoexon inclusion		(Weisschuh et al. 2020)
COL6A1 c.930+189C>T	ENSG00000142156	Collagen VI-related dystrophy	new splice donor site	72bp pseudoexon inclusion		(Cummings et al. 2017)
DES c.1289-741G>A	ENSG00000175084	Myopathy with cardiac involvement	new splice acceptor site	118bp pseudoexon inclusion	compound heterozygous	(Riley et al. 2019)
DMD c.5739+362A>G	ENSG00000198947	Duchenne muscular dystrophy	new splice donor site	78bp pseudoexon inclusion		(Xu et al. 2019)

DMD c.9974+175T>A	ENSG00000198947	Duchenne muscular dystrophy	alter cryptic splice donor site strength	88bp pseudoexon inclusion		(Lee et al. 2019)
DONSON c.607-36G>A	ENSG00000159147	Meier-Gorlin syndrome	new splice acceptor site	34bp exon extension		(Knapp et al. 2019)
DYSF c.5668-824C>T	ENSG00000135636	Muscular Dystrophy	new splice donor site	180bp pseudoexon inclusion		(Dominov et al. 2019)
F7 c.571+78G>A	ENSG00000057593	Coagulation factor VII deficiency	alter cryptic splice donor site strength	78bp exon extension; 78bp partial intron retention		(Ferraresi et al. 2019)
F7 c.572-392C>G	ENSG00000057593	Coagulation factor VII deficiency	new splice donor site	282bp pseudoexon inclusion		(Ferraresi et al. 2019)
F8 c.5999-277G>A	ENSG00000185010	Mild hemophilia A	U1snRNP binding creation	95-bp pseudoexon inclusion		(Chang et al. 2019)
FECH c.464-1169A>C	ENSG00000066926	Erythropoietic Protoporphyria	exonic splicing silencer disruption	156bp pseudoexon inclusion		(Chiara et al. 2020)
GALNS c.899-167A>G	ENSG00000141012	Morquio A	new splice donor site	53bp pseudoxeon inclusion		(Caciotti et al. 2018)
GPR143 c.885+748G>A	ENSG00000101850	Ocular albinism				(Lionel et al. 2018)
IGHMBP2 c.1235+894C>A	ENSG00000132740	Charcot-Marie-Tooth disease	new splice acceptor site	182bp pseudoexon inclusion	compound heterozygous	(Cassini et al. 2019)
MYBPC3 c.1224-80G>A	ENSG00000134571	Hypertrophic cardiomyopathy	new splice acceptor site	78bp exon extension; 78bp partial intron retention	haploinsufficiency	(Janin et al. 2019)
MYBPC3 c.1927+600C>T	ENSG00000134571	Hypertrophic cardiomyopathy	new splice donor site	94bp pseudoexon inclusion		(Janin et al. 2019)
MYBPC3 c.906-36G>A	ENSG00000134571	Hypertrophic cardiomyopathy	new splice acceptor site	34bp exon extension; 34bp partial intron retention	haploinsufficiency	(Frank-Hansen et al. 2008)
NDUFAF6 c.420+784C>T	ENSG00000156170	Leigh syndrome	new splice acceptor site	124bp pseudoexon inclusion	compound heterozygous	(Catania et al. 2018)
NDUFAF8 c.195+271C>T	ENSG00000224877	Leigh syndrome	U1snRNP binding creation	exon extension	compound heterozygous	(Alston et al. 2020)
NF1 c.4173+278A>G	ENSG00000196712	Neurofibromatosis	new splice acceptor site	80bp pseudoexon inclusion		(Kannu et al. 2013)
NIPBL c.4560+2069C>T	ENSG00000164190	Cornelia de Lange syndrom	new splice donor site	pseudoexon inclusion		(Rentas et al. 2020)
NM 000082.3 c.173+1119G>C	ENSG00000049167	Cockayne syndrome	exonic splicing enhancer creation	activation cryptic splice sites & pseudoexon inclusion		(Schalk et al. 2018)
NM 001099857.2 c.518+866C>T	ENSG00000269335	EDA-ID	U1snRNP binding creation	44bp pseudoexon inclusion		(Boisson et al. 2018)
POLA1 c.1375-354A>G	ENSG00000101868	X-linked reticulate pigmentary disorder	new splice donor site	76bp pseudoexon inclusion		(Starokadomskyy et al. 2016)
PROM1 c.2077-521A>G	ENSG00000007062	Cone-rod dystrophy	new splice donor site	155bp pseudoexon inclusion		(Mayer et al. 2016)
PYROXD1 c.415-976A>G	ENSG00000121350	Congenital myopathy	exonic splicing silencer disruption & alter cryptic splice donor site strength	110bp pseudoexon inclusion		(Lornage et al. 2019)
RTTN c.2309+1093G>A	ENSG00000176225	Primary microcephaly	new splice acceptor site	85bp pseudoexon inclusion		(Vandervore et al. 2019)
SERPING1 c.-22-155G>T	ENSG00000149131	Hereditary angioedema				(Germenis and Cicardi 2019)

SLC20A2 c.289+937G>A	ENSG00000168575	Primary familial brain calcification	new splice acceptor site	38-bp pseudoexon inclusion		(Chen et al. 2019)
SLC46A1 c.1166-285T>G	ENSG00000076351	Hereditary folate malabsorption	new splice donor site	168bp pseudoexon inclusion		(Tozawa et al. 2019)
SPG7 c.286+853A>G	ENSG00000197912	Hereditary spastic paraplegia	U1snRNP binding creation	75bp pseudoexon inclusion	compound heterozygous	(Verdura et al. 2020)
TMEM165 c.792+182G>A	ENSG00000134851	Congenital Disorder of Glycosylation	new splice donor site & alter cryptic splice donor site strength	117bp pseudoexon inclusion		(Foulquier et al. 2012)
TULP1 c.718+23G>A	ENSG00000112041	Photoreceptor Dystrophy	exon extension	20bp exon extension	compound heterozygous	(Verbakel et al. 2019)
USH2A c.7595-2144A>G	ENSG00000042781	Usher syndrome	alter cryptic splice donor site	pseudoexon inclusion		(Vaché et al. 2012)

Table S II.4. Protein prediction tools evaluated in subset of missense variants within the HCM dataset

Tool	Threshold	Method	Data used	Description	Other scores added	Obtained from
SIFT (Kumar, Henikoff, and Ng 2009)	< 0.05 (Liu et al. 2016)	Sequence homology search using PSI-BLAST to build multiple alignments to further infer a position specific scoring matrix	1,750 deleterious and 2,254 tolerant nsSNVs of E. coli LacI gene	Homology-based prediction of the tolerance of a given aminoacid substitution	No	dbNSFP v4.02
PolyPhen-2 HDIV (Adzhubei et al. 2010)	> 0.5 (Liu et al. 2016)	Naïve Bayes classifier	5,564 damaging alleles from Uniprot; 7,539 neutral mutations	Prediction of impact of an amino acid substitution on the structure and function of a human protein.	No	dbNSFP v4.02
PolyPhen-2 HVAR (Adzhubei et al. 2010)	> 0.5 (Liu et al. 2016)	Naïve Bayes classifier	22,196 damaging alleles from Uniprot; 21,119 neutral mutations	Prediction of impact of an amino acid substitution on the structure and function of a human protein.	No	dbNSFP v4.02
LRT (Chun and Fay 2009)	< 0.01(Liu et al. 2016)	Likelihood Ratio Test	Protein coding regions of 32 vertebrates and 3 human genomes	Identification of deleterious mutations within protein-coding sequences. Obtained via dbNSFP v4.02	No	dbNSFP v4.02
Mutation Assessor (Reva, Antipin, and Sander 2011)	> 1.935 (Liu et al. 2016)	Combinatorial entropy approach	Multiple alignments of sequence homologs from the Uniprot database	Functional impact of protein mutations based on evolutionary information	No	dbNSFP v4.02
FATHMM (Shihab et al. 2015)	< -1.5 (Shihab et al. 2015)	Sequence homology and Hidden Markov Models combined with protein domain annotations	SNVs from HGMD and UniProt	Prediction of the functional impact of protein missense mutations.	No	dbNSFP v4.02
PROVEAN (Choi and Chan 2015)	< -2.5 (Liu et al. 2016)	Sequence homology search and BLAST hits clustering followed by a delta alignment score calculation	NCBI nr protein database and ~58k variants from Uniprot	Predicts whether an amino acid substitution or indel has an impact on the biological function of a protein.	No	dbNSFP v4.02

Mutation Taster2 (Schwarz et al. 2014)	> 0.5 (Liu et al. 2016)	Naïve Bayes classifier	Disease associated mutations from HGMD and Clinvar + neutral SNVs from 1000G	Prediction of the functional consequences of protein-coding substitutions, splice site and intronic alterations.	No	dbNSFP v4.02
MutPred2 (Pejaver et al. 2017)	> 0.5 (Pejaver et al. 2017)	Random Forest	Protein substitutions from HGMD, Swissprot and cancer-related datasets	Based upon aminoacid sequence, it models the changes in the structure and function of proteins. It incorporates SIFT predictions.	Yes	dbNSFP v4.02
Condel (González-Pérez and López-Bigas 2011)	> 0.98 (González-Pérez and López-Bigas 2011)	Weighted average of the normalized scores of several tools	Variants from HumVar and HumDiv datasets	Unified classification of Nonsynonymous SNVs based on the integration of several individual tools	Yes	VEP plugin
CAROL (Lopes et al. 2012)	> 0.468 (Lopes et al. 2012)	Weighted Z method	SIFT and PolyPhen2 scores	Combination of SIFT and PolyPhen2 into a functional annotation score for non-synonymous variants	Yes	VEP plugin
M-CAP (Jagadeesh et al. 2016)	> 0.025 (Jagadeesh et al. 2016)	Gradient boosting trees	12,418 pathogenic variants from HGMD; 3,137,919 benign dataset retrieved from ExAC v0.3	Pathogenicity likelihood score that aims to aggressively reduce the number of variants of uncertain significance (VUS)	Yes	dbNSFP v4.02
REVEL(Ioannidis et al. 2016)	> 0.5 (Ioannidis et al. 2016)	Random Forest	6,182 disease variants from HGMD ; 123,706 putative neutral missense variants from the Exome Sequencing Project (ESP)	An Ensemble method for predicting the pathogenicity of rare missense variants	Yes	dbNSFP v4.02
VEST4 (Carter et al. 2013)	> 0.67 (Carter et al. 2013)	Random Forest	47,724 disease variants from HGMD ; 45,818 putative neutral missense variants from ESP	Prioritization of rare missense variants with likely involvement in human disease.	Yes	dbNSFP v4.02
MetaSVM (Dong et al. 2015)	> 0.5 (Dong et al. 2015)	Support Vector Machine	14,191 deleterious and 22,001 neutral variants from Uniprot	An Ensemble method for predicting nonsynonymous variants using 9 independent scores and allele frequencies	Yes	dbNSFP v4.02
MetaLR (Dong et al. 2015)	> 0.5 (Dong et al. 2015)	Logistic Regressor	Same as MetaSVM	Same as MetaSVM	Yes	dbNSFP v4.02

III. GENE EDITING BY CRISPR TO CREATE ISOGENIC HCM DISEASE MODELS

III.1. ABSTRACT

Hypertrophic cardiomyopathy (HCM) is an inherited heart disease with an estimated prevalence higher than 1:500 caused by mutations in sarcomeric proteins. Disease presentation is extremely heterogeneous even in patients bearing the same mutation.

Recent advances in stem cell technologies, with the discovery of induced pluripotent stem cells (iPSCs) and their differentiation into cardiomyocytes, provided a pathophysiologic relevant approach to understand the mechanisms of disease. Nonetheless, the creation of reliable disease models for HCM that allow to overcome the disease heterogeneity and establish trustworthy genotype-phenotype relations requires the use of isogenic controls. Here we apply the CRISPR/Cas9 gene-editing system to introduce an HCM-associated variant in a healthy donor hiPSCs in order to help build such a model.

Development of CRISPR/Cas9 technology greatly improved the ability to precisely modify a chosen locus with minimal impact on the remaining genome and consequently the usefulness of hiPSCs for disease modeling purposes. However, its application still poses important challenges namely low homology direct repair (HDR) efficiency, frequently detected unwanted mutations or issues with cell viability, which are further discussed.

Key words: CRISPR; hypertrophic cardiomyopathy; human pluripotent stem cells; cellular disease models

III.2. INTRODUCTION

HCM is a primary disease of the myocardium and the most common hereditary disease of the heart, occurring in more than 1:500 individuals and often leading to sudden cardiac death in young adults (Maron, Rowin, and Maron 2018; Maron 2018). Most HCM is caused by inherited autosomal dominant gene mutations with variable expressivity and penetrance. These mutations occur predominantly in genes that code for structural and functional components of the sarcomere and until date, more than 1000 mutations have been identified in at least 15 different genes (Elliott et al. 2014; Marian and Braunwald 2017; Skelton et al. 2014). The majority of identified HCM disease causing mutations are missense; however, with the increasing use of NGS approaches in clinical diagnosis, experimental evidences pointing for the role of non-coding variants in human disease, namely in HCM, are growing. Among these, splicing mutations appear as an important subgroup, highlighting splicing has a relevant and overlooked mechanism in human genetic diseases, mainly when considering variants found in non-canonical splice-sites, such as deep intronic variants, for each is particularly challenging establish a genotype-phenotype relationship, proving them as disease causing variants.

Gene-editing approaches such as the CRISPR/Cas9 technology, by allowing introduction of a single variant in any desired genomic region, are emerging as a powerful tool in generation of isogenic human disease models, where the impact of any variant of interest may be successfully addressed. In particular, the combination of CRISPR/Cas9 approaches with the recent emergence of human induced pluripotent stem cells (hiPSCs) represents a promising tool for disease modeling, since this cell type have unlimited expansion potential and can be differentiated into most cell lineages of the human body including cardiomyocytes (Takahashi et al. 2007; Yoshida and Yamanaka 2017). Since their discovery hiPSCs have been widely used for the study of inherited cardiac diseases, with over 90 studies using hiPSCs derived cardiomyocytes (hiPSC-CMs) for cardiac disease modeling being published until date (van Mil et al. 2018). Even though

disease modelling of HCM using hPSC-CMs has been proven to offer a pathophysiologic relevant approach to understand the mechanisms of disease, most of these studies lack isogenic controls which confounds the establishment of reliable genotype-phenotype correlations (Ma et al. 2018; Wang et al. 2018; Mosqueira et al. 2018), raising the need to combine them with gene-editing methodologies in order to obtain isogenic controls and trustworthy disease models (Yoshida and Yamanaka 2017).

The use or CRISPR/Cas9 to introduce any given variant of interest in healthy donor hiPSCs allows to determine if such variant is causing the disease upon differentiation of gene-edited hiPSCs into cardiomyocytes that recapitulate the disease phenotype. Moreover, simultaneous differentiation of the healthy donor hiPSCs provides a isogenic control that allows to access if the variant sufficient to cause the disease phenotype, regardless of the genetic background (Musunuru et al. 2018).

Despite being a powerful genome-editing tool, CRISPR/Cas9 accuracy can severely be reduced by its tendency to re-edit previously modified loci, causing unwanted mutations (indels) along with intended changes (Doudna and Charpentier 2014). Importantly, the selective and scarless introduction of mono-allelic single nucleotide sequence modifications is particularly critical for the study of variants expected to have a non-coding effect since, if variants other than the study variant are present, it is not possible to prove that any effect seen at the cellular level is solely due to the variant of interest .

For this reason, in this study, we decided to use a recently reported CRISPR-Cas9 strategy termed “CORRECT re-Cas” approach, which involves two rounds of CRISPR and where silent CRISPR-Cas9 blocking mutations are incorporated along with the pathogenic mutation of interest. Briefly, in a first round of CRISPR, the variant of interest “X” is introduced in a target locus along with a blocking mutation, “B”, by an HDR repair template (XB template). This blocking mutation prevents the recognition of the PAM sequence by the Cas9 thus preventing repeated cutting and subsequent unwanted mutations. Then, in a second round of CRISPR, the altered PAM sequence

is targeted by a Cas9 mutant (Cas9-VRER) and the template used preserves the intended mutation while removing the blocking mutation (X template) (Kwart et al. 2017). Moreover, to obtain the desired mono-allelic modifications, the chosen guide RNA was designed to provide a cut-to-mutation distance between 10-12bps, described as the ideal cut-to-variant distance to favor heterozygous incorporation of the intended mutation (Paquet et al. 2016). An overview of the CORRECT re-Cas editing strategy in the creation of a hiPSC line with the mono-allelic modification for the *MYBPC3* c.927-2A>G variant is depicted in Figure III.1.

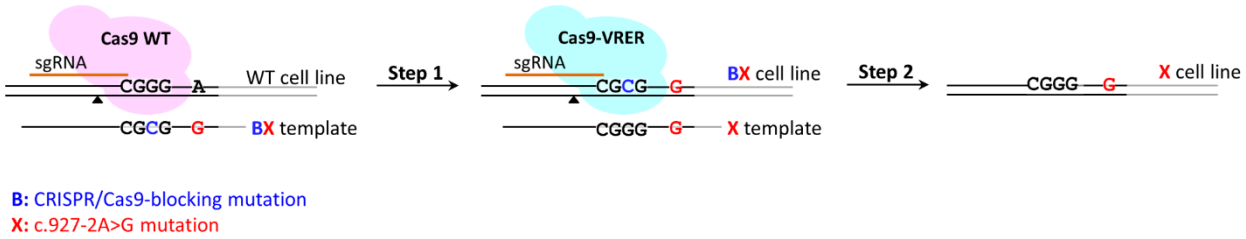


Figure III.1. CORRECT re-Cas strategy. In the first round a PAM altering mutation "B" (NGG > NGCG) is introduced by HDR along with the MYBPC3 c.927-2A>G mutation "X" using the "XB template". In the second round the altered PAM sequence is targeted by the Cas9-VRER and "B" is removed, resulting in the introduction of the c.927-2A>G variant alone.

The aim of this work was to use the CRISPR/Cas9 system to generate hiPSC lines carrying candidate genetic variants predicted to affect splicing, previously identified in HCM patients. This hiPSC lines, by being an isogenic cellular model, upon cardiomyocyte differentiation, will allow to study the effect of each individual genetic variant in cardiomyocyte differentiation and function and, by that, their potential clinical relevance.

III.3. METHODS AND RESULTS

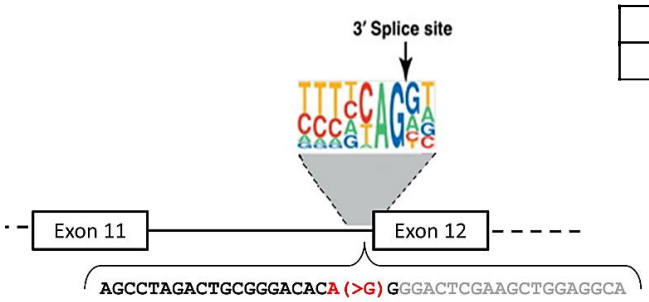
Chosen variant

In order to select a variant that could be used as a proof of concept for our work model, the ClinVar database (Landrum et al. 2014) was searched for well-described, pathogenic, splice-site variants associated with HCM. Selected variants were then analyzed by bioinformatic tools to predict their impact on splicing, and literature search on these variants was performed concerning familial segregation pattern and functional impact.

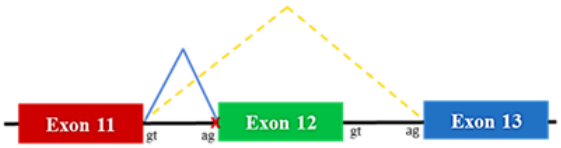
The computational prediction tools used were Human Splicing Finder (HSF) and MaxEntScan. HSF predicts the impact that a variant might have on splicing by employing algorithms to evaluate the strength of cryptic donor and acceptor splice sites as well as *cis*-acting elements such as enhancer and silencer motifs, created by the presence of the variant, comparing their strength to the consensus sites and issuing a score that reflects the probability of splicing alterations due to that specific mutation (Desmet et al. 2009). Similarly, the MaxEntScan method assigns a score to the 5' or 3' splicing sequences and compares reference and mutated motifs scores to predict the usage of the splice site. Given two splicing sequences, i.e. wild-type and mutated, the higher scoring sequence has a superior likelihood of being used in detriment of the other (Eng et al. 2004; Leman et al. 2018).

Considering the splice-altering scores obtained from the computational tools and the available literature, the *MYBPC3* c.927-2A>G was selected as the first variant to be tested (Figure III.2).

MYBPC3 c.927-



GRCh37.p13	GRCh38.p12	rs_id
11:47367923	11:47346372	rs397516082



MaxEnt 3' Score: 431.14 % **HSF Score:** -Ref Sequence: -1.83

Figure III.2. Scheme of MYBPC3 c.927-2A>G variant location, identification and predicted consequences. The variant is located in the canonical acceptor site and serves as a positive control for a variant that may alters splicing. Scheme depicts the normal intron splicing (blue) and the hyphotesized splicing alteration caused by the variant (orange, dashed lines). HSF and Max Ent Scores are also presented.

This variant is predicted to destroy the canonical splice acceptor site in intron 11, leading to skipping of *MYBPC3* exon 12. Such abnormal *MYBPC3* RNAs can either give rise to a truncated protein (if in frame exon skipping occurs) or be subjected to NMD leading to a haploinsufficiency phenotype (due to generation of PTC). Studies shown that this variant associates with cardiomyopathy, heart failure, ventricular tachycardia and atrial fibrillation (Adalsteinsdottir et al. 2014; Norland et al. 2019). It has been reported in multiple unrelated individuals with HCM (Niimura et al. 1998; Richard et al. 2003; Ehlermann et al. 2008; Yiu et al. 2012; Adalsteinsdottir et al. 2014) and segregates with the disease within families (Bonne et al. 1995; Ehlermann et al. 2008; Christiaans et al. 2010). Moreover, mRNA from lymphoblastoid cells derived from affected family members with this variant showed evidence of aberrant splicing (Bonne et al. 1995; Niimura et al. 1998) and it was observed in large population cohorts with an allele frequency as low as $8.7x10^{-6}$ (Karczewski et al. 2019).

Maintenance of human iPSCs

We used the DF6.9.9 T.B cell line (WiCell®), a vector-free commercially available human iPS cell line reprogrammed from foreskin fibroblasts (Junying et al. 2009).

hiPSCs were cultured on Matrigel™(Corning)-coated plates with mTeSR™1 Medium (StemCell Technologies). Medium was changed daily. Cells were passaged every three to four days (at a confluency of approximately 85% of the surface area of the culture dish) using 0.5mM EDTA dissociation buffer (Life Technologies). Two to three passages were performed before the transfection.

Validation of target genomic sequences

Genomic DNA was isolated from the DF6 hiPSC line using a standard protocol. Then target region was amplified by PCR. PCR products were purified using the PCR Clean protocol from the NZYGelpure Kit (NZYTech) and sent for sequencing by Sanger at STAB Vida or Eurofins Genomics, according to the conditions defined by the designated company. Produced chromatograms were aligned to the genomic sequence to confirm the absence of target variant, as shown in Figure III.3. The DF6 hiPSC line was confirmed to be WT for the chosen variant.

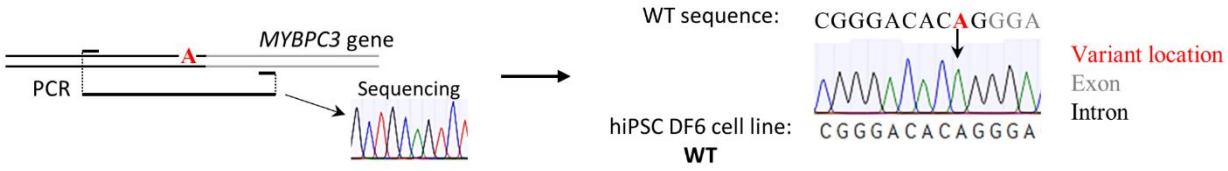


Figure III.3. **Validation of target genomic sequence. Absence of the *MYBPC3* c.927-2A>G variant in the hiPSC DF6 cell line.**

gRNAs design and activity test

Candidate sgRNAs for the target variants were initially chosen using the available bioinformatics tool website from Zhang laboratory (http://crispr.mit.edu/) according to their scores and predicted off-target effects as previously described (Ran et al. 2013). It was also considered the cut-to-mutation distance since, as described by Paquet *et. al.*, to introduce a mutation in heterozygoty the sgRNA should mediate the Cas9 cut at a distance of 2-26 bp from the mutation site, with the optimal being between 10-12 bp (Paquet et al. 2016; Kwart et al. 2017). Guide RNAs that fulfilled these criteria were chosen and designed according to what was previously described.

MLM3636 U6-sgRNA expression vector backbone (Addgene #43860) was digested for 1 hour at 37ºC with *BsmBI*. The linearized plasmid was then dephosphorylated using the Fast AP Enzyme (Thermo Scientific) for 10 minutes at 37 ºC followed by 5 minutes at 75ºC, to avoid linearized plasmid religation. Synthetized sgRNAs were annealed and cloned into linearized plasmid using T4 DNA ligase (Thermo Scientific) overnight at 4ºC. Ligated MLM3636 vectors were transformed into DH5α chemically competent *E.Coli*, plasmid DNA was extracted with the Mini Prep Protocol for isolation and purification of plasmid DNA (NZYTech) and analyzed by Sanger sequencing to ensure correct cloning of the sgRNA in the vector.

The efficiency of the chosen gRNAs was tested using the GeneArt® Genomic Cleavage Detection Kit (Thermo Fisher), according to the manufacturer's instructions. Briefly, HEK293T cells at 70-90% confluency were transfected with pCas9_GFP (Addgene #44719) and MLM3636 with sgRNA, using Lipofectamine® 2000 Reagent (Invitrogen). Cells were incubated for 48 hours at 37°C with 5% CO_2 before being collected and lysed. Cell lysates were then used as template for PCR amplification of the target genomic region, in parallel with a control reaction to ensure the good performance of the detection kit. Afterwards, these PCR products were run in a thermal cycler program to randomly anneal the PCR fragments with and without indels to form heterogeneous DNA duplexes. The heteroduplex DNAs that contain mismatches, meaning that

Cas9 introduced a DSB that, in absence of an HDR template, was repaired by NHEJ resulting in the incorporation of indel mutations, were cleaved by the Detection Enzyme for 1 hour at 37°C and resulting samples were subsequently ran on a 2% agarose gel (0.5x TBE, EtBr).

The process of sgRNA design and testing for the *MYBPC3* c.927-2A>G variant is depicted in Figure III.4. and shows efficient in the guidance of the Cas9 to the intended genomic target site by the chosen sgRNA. The chosen sgRNA sequence is shown in

Table III.1.

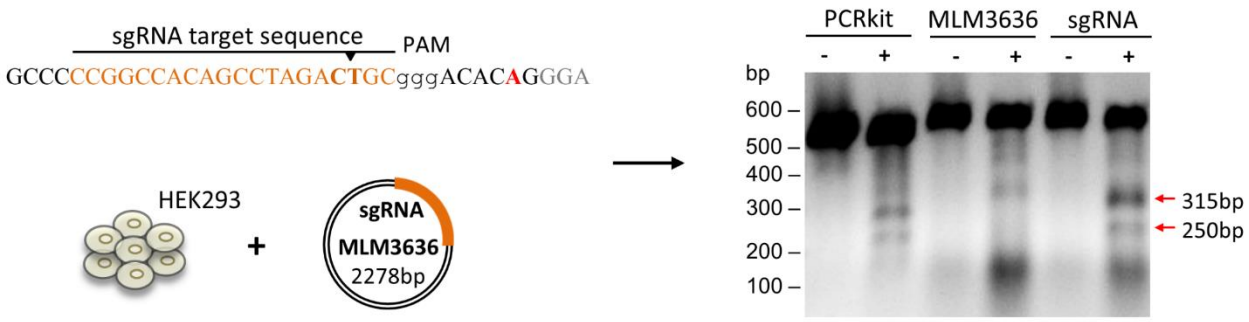


Figure III.4. Design and activity test of sgRNA. Design of a gRNA with cut site 10bp upstream of variant, which inside the ideal described cut-to-variant distance to favor heterozygous. Activity test shows that the correct genomic locus is being targeted by the presence of two bands corresponding to the cleaved DNA (arrows in red) with sizes consistent with the distance to the cut site, results from the PCR Kit sample are positive control and from MLM3636 with no sgRNA a negative control.

Table III.1. Chosen sgRNAs for *MYBPC3* c.927-2A>G target variant.

sgRNA	Sense	Antisense
MYBPC3	5´- ACACCgCCGGCCACAGCCTAGACTGCG- 3´	5´- AAACGCAGTCTAGGCTGTGGCCGGCcG - 3´

HDR template design

The donor template to be used in the first round of edition (BX template) should contain the intended variant and a blocking mutation that prevents PAM detection by mutating the NGG to

NGCG, which is recognized by a Cas9 variant (Cas9 VRER), to be used in the second round of edition. The donor template for the second round of edition (X template) should contain only the intended variant (Kwart et al. 2017).

Both a plasmid containing homology arms (HA) and a single-stranded donor oligonucleotide (ssODN) were designed to be used as HDR repair templates.

The donor plasmid was created amplifying by PCR the region that comprised target variant flanked by ~1kb HA, from DF6 hiPSC genomic DNA, with primers containing the appropriate restriction enzyme sequence, in order to get “sticky ends” for the ligation to the plasmid (Table III.2). PCR products were purified from gel with the Gel Extraction Protocol from the NZYGelpure Kit (NZYTech). The generated HA constructs were cloned into pPur plasmid by digestion of the plasmid with the indicated (BamHI and EcoRI) restriction enzymes. Dephosphorylation of the pPur plasmid, ligation, bacterial transformation and plasmid DNA purification were performed as above. Sanger sequencing was used to confirm that the HA insert was correctly cloned in the vector and site directed mutagenesis was performed in order to introduce the chosen variant and PAM altering nucleotide (Figure III.5.A.).

Table III.2. Primers for HA design.

Primer	Prod. Size	Sequence
MYBPC3_Fwd	1742	**GCA**GGATCCGTCCCGTCAACAGTCATCCT
MYBPC3_Rev		**GCA**GAATTCCCTCCATTCAGTCGGTGTTT

BamHI	GGATCC
EcoRI	GAATTC
+3 nucleotides	**GCA**

The ssODN was designed to have 100-nt centered around the cut site promoted by the chosen gRNA and using the sequence of the nontargeted strand in order to avoid base-pairing between the gRNA and repair oligo (Figure III.5.B.).

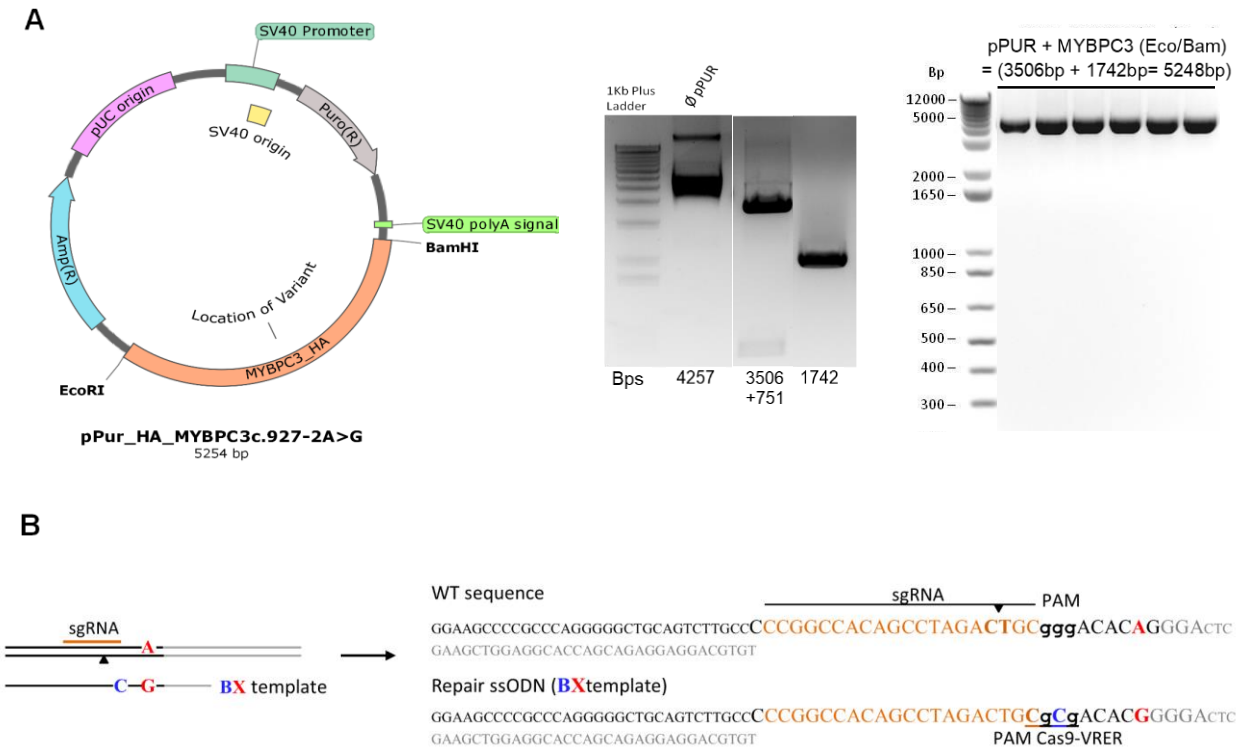


Figure III.5. HDR templates design. Design of the BX template to be used in the first CRISPR round, bearing the PAM-altering CRISPR/Cas9-blocking mutation and our desired mutation *MYBPC3* c.927-2A>G. A- pPur plasmid containing the HA with the intended mutation plus the NGG>NGCG. B- 100-nt ssODN centered around the cut site, using the sequence of the same strand as the gRNA containing the PAM-altering mutation and our desired mutation.

Edition steps

As already mentioned, the chosen variant was the *MYBPC3* c.927-2A>G variant since it would work as a proof of concept for our work model. Edition steps described below were performed for this variant and are illustrated in Figure III.6.

Transfection

For the transfection two different strategies were used: electroporation, using the Neon® Transfection System (Invitrogen) and lipofection with Lipofectamine® 3000 Reagent (Invitrogen).

Electroporation

In brief, for electroporation transfection, cells were treated with ROCKi (Y-27632, 10µM, StemCell Technologies) for 24h at 37ºC prior to electroporation and then dissociated using Accutase (Sigma) for 7 min at 37ºC. After dissociation, $2x10^6$ cells were carefully resuspended in 150µL of cold R buffer (Neon Transfection Kit) and kept on ice. Just before electroporation cells were added and mixed with the DNAs to be transfected (pCas9_GFP; MLM3636_sgRNA; donor plasmid or ssODN). This mixture was then carefully aspirated into the 100-µl Neon tip and placed on the pipette station, as per manufacturer instructions, and electroporation was performed using selected conditions. Electroporated cells were subsequently plated in a pre-warmed Matrigel-coated 6-well plate with mTeSR™1 + 10µM ROCKi (~$1x10^6$ cells/well) and incubated at 37ºC 5% CO_2 for 48h, replacing the media with fresh mTeSR™1 each 24h.

To select electroporation parameters to be used, considering cell viability and transfection efficiency, DF6 hiPSCs were first transfected only with 20 µg of pCas9_GFP (Table III.3). After analysis of GFP+ transfected cells by flow cytometry, 72h post transfection, the parameters selected for further experiments were 1400V / 5ms / 3 pulses since they result in less cell death and a reasonable transfection efficiency. With the 1150V / 30ms / 2 pulses parameters there was more cell death for the same transfection efficiency, and with 1600V / 20ms / 1 pulse the efficiency was the highest but almost all cells died.

In edition experiments, cells were treated as described and electroporated using the chosen parameters and DNA amounts selected according to the literature and protocols previously established in the lab (Table III.3.) (Kwart et al. 2017). For the conditions using donor plasmid constructs as templates for HDR, a mix of template with the mutation and without was used in order to favor heterozygous.

Table III.3. Amounts of DNA to be tansfected and electroporation parameters to be used.

Electroporation parameters		1400 Volts / 5miliseconds / 3 pulses	1150V / 30ms / 2 pulses	1600V / 20ms / 1 pulse
DNA amounts	Electroporation parameters test	20 µg Cas9_GFP	20 µg Cas9_GFP	20 µg Cas9_GFP
	Kwart et.al., 2017 conditions	20 µg Cas9_GFP 5µg sgRNA 3 µg ssODN		
	Kwart et.al., 2017 conditions adapted to HA	20 µg Cas9_GFP 5µg sgRNA 15 µg pPUR_HA_BX 15 µg pPUR_HA		
	HA conditions	5 µg Cas9_GFP 2,5µg sgRNA 2 µg pPUR_HA_BX 2 µg pPUR_HA		

Lipofection

For lipofection, hiPSCs were treated and dissociated to obtain single-cell suspensions as previously described. After counting, $1x10^5$ cells were plated per 24 well-plate coated with Matrigel. The following day, medium was changed, and transfection of the gene-editing reagents was performed by lipofection, using Lipofectamine® 3000 Reagent (Invitrogen) according to manufacturer's protocol. The DNA amounts used in the transfection were: 50 ng sgRNA, 200 ng pCas9-GFP and 250ng ssODN or 100ng sgRNA, 400ng pCas9_GFP and 600ng ssODN.

Selection of GFP+ hiPSCs by FACS

72h after transfection cells were dissociated into single-cell suspension, as previously described. Cell pellets were resuspended in 200-1000 µL of mTeSR™1 with 10 µM ROCKi and filtered through a 70µm strainer before proceeding to the sorting, in a BD FACS Aria IIu (BD Biosciences). Between $3×10^4$ to $5×10^4$ GFP^+ sorted cells were plated in 10 cm culture dishes coated with Matrigel with mTeSR™1 supplemented with 10µM ROCKi, in order to form individualized clones. Cells were incubated at 37°C, 5% CO_2 for 48 hours after which half of the

medium was replaced with fresh mTeSR™1 with 10µM ROCKi. 72h after sorting, full feeds with mTeSR™1 were performed every day to every other day until single-cell clone colonies emerge (typically 12 days after sorting).

hiPSCs clone picking

hiPSC clones were picked with the help of a microscope (JuLI™, NanoEnTek) and a micropipette, inside the flow chamber to a 96 or 48-well plate with mTeSR1+ROCKi. 48 hours after picking, the ROCKi was removed and clones were maintained in culture with daily medium changes until splitting.

Splitting of hiPSCs clones

Clones were passaged using Accutase, as previously described. A fraction of the cell suspension was maintained in culture in a 96 or 48-well plate. The remaining cells were used for DNA isolation and sequencing of the target region to identify correctly gene-edited clones.

Confirmation of the XB clones by Sanger sequencing

After clone splitting, ~ half of cells from each clone were pelleted and genomic DNA was isolated as previously described (Kwart et al. 2017). Briefly, cells were lysed overnight at 65°C with a Nuclear Lysis Buffer containing Proteinase K. DNA was precipitated with an EtOH/NaCl solution (1.5% 5M NaCl in absolute EtOH) and DNA pellets were washed with EtOH 70%, air-dried and resuspended in RNase/DNase free H_2O.

From the first 25 clones screened by Sanger sequencing, 2 (#14 and #17) had the desired mutations in heterozygosity, as shown in Figure III.6., suggesting a high efficiency of the edition protocol.

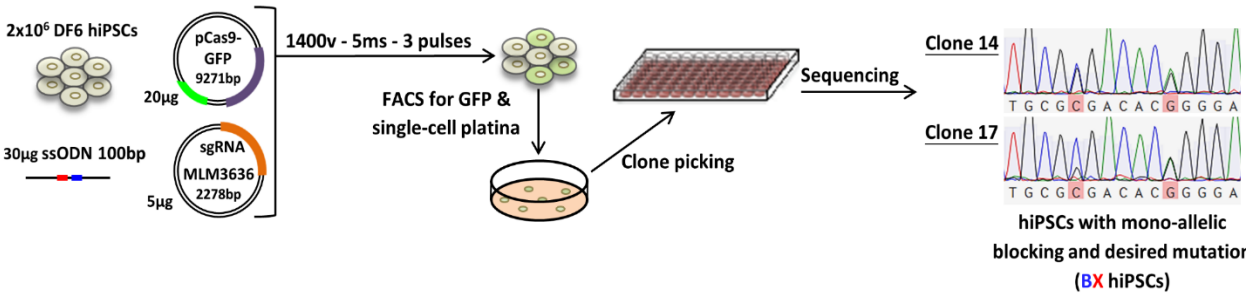


Figure III.6. Edition steps followed to obtain the hiPSCs with mono-allelic blocking and desired mutation (BX hiPSCs). Electroporation conditions used; selection of transfected cells by FACS for GFP; clone picking and confirmation of the XB clones by Sanger sequencing.

Gene-edited clones were further expanded and frozen. However, upon thawing, and prior to second round of edition, cells show low viability and proliferating capability. This led to the impossibly to maintain these cells in culture and perform further experiments.

Subsequently, two more attempts to obtain clones with the intended mutation were made using the same parameters and the ssODN as repair template, but no correctly edited clones could be obtained, despite some had indel mutations in the target region.

In order to promote HDR and enhance cell viability some derivations of the protocol were also tried. These attempts included using different electroporation parameters or choosing the lipofection method, that causes less cell death, combined with a “cold shock” step to promote HDR, in which in the first 48 hours after transfection cells were incubated at 32°C (Guo et al. 2018) (Table III.4.).

Table III.4. Variations to the edition protocol.

Electroporation		Lipotransfection
• 1200V /10ms / 2pulses • 1400V / 5ms / 2 pulses	• 20 µg Cas9_GFP • 5µg sgRNA • 3 µg ssODN	• 400ng Cas9_GFP • 100ng sgRNA • 600ng ssODN • 1,5 uL Lipo 3000

We also tested the transfection of Cas9-Gem mRNA instead of GFP-Cas9 plasmid DNA. Cas9-Gem mRNA was shown to promote HDR due to its transitory and cell-cycle-tailored expression resulting in low levels of Cas9 in G1 (when the HDR mechanism is absent), but high expression in S/G2/M, the cell cycle phases where HDR mechanism is active (Gutschner et al. 2016).

In total at least 8 edition experiments were attempted, with 120 to 230 clones screened from each one, but none lead to the obtention of correctly edited viable clones.

III.4. DISCUSSION

New experimental tools are needed for the study of HCM in order to improve molecular diagnosis and aid the development of new targeted therapies.

Since HCM is extremely heterogeneous, with patients bearing the same mutation presenting entirely different phenotypes and clinical outcomes, and establishing reliable genotype-phenotype correlations is challenging, the use of isogenic controls is of uttermost importance.

Development of CRISPR/Cas9 system greatly improved gene-editing approaches rendering more feasible to create isogenic sets of “diseased” and control hiPSCs. Even though edited hiPSCs lines have already been established in several studies, the low efficiency of gene-editing in these cells remains a serious challenge in the field, with different laboratories reporting HDR efficiencies that range from 0.1 – 2% (Miyaoka et al. 2016; Paquet et al. 2016).

The primary protocol followed for this work was adapted from *Kwart et al*. who developed this gene-editing approach based on the observation that in many of the cells, where the DSBs had been repaired through HDR, there were also indel mutations, due to a subsequent re-cutting by the Cas9 of a correctly edited locus. Using CRISPR/Cas-blocking mutations at the PAM site would minimize re-cutting by the Cas9 nuclease and so subsequent indels (Kwart et al. 2017).

After the first round of edition using this protocol, we obtained two correctly edited clones, containing the blocking mutation and our intended variant. The edited clones were amplified and did not show evident morphological changes or a marked decrease in replication rate that could indicate spontaneous differentiation or foresee future culturing difficulties. Despite that, upon thawing the clones to perform the second round of editing, their viability was very low, and it was impossible to recover them. It is possible that, although no morphologic changes were noted, our edited clones had lost pluripotency and were not in ideal conditions when cryopreserved. Despite the use of experimental strategies to improve the rate of cells edited by the HDR pathway, the edition efficiency is still very low, and to overcome this low efficiency, several hundred clones

must be isolated, expanded and screened, which translates in a multistep process with prolonged periods of cells in culture, eventually under suboptimal conditions. One important issue should be optimization of the clone picking and screening techniques used in order to reduce the time clones are kept in culture and enhance their viability.

An alternative approach for reducing the time spent with clones in culture could be their cryopreservation after splitting for screening. In this work we did not attempt this approach; however previous experience from our lab indicates that freezing hiPSCs in plates leads to the death of many cells, which could mean losing the correctly edited clones. This seams a reasonable and cost-effective alternative but to be used it would be essential to optimize freezing and thawing protocols for this specific condition.

After two subsequent and unsuccessful gene-editing attempts, using the same protocol and conditions, and since several clones screened presented indel mutations in the target region, we hypothesized that the sgRNA was efficiently guiding the Cas9 to its target, but the HDR efficiency would to be too low to obtain correctly edited clones. It is also important to note that, in addition to the inherent difficulties of HDR mediated gene-editing efficiency, hiPSCs are particularly sensitive cell lines and some difficulties posed in gene-editing are related to low transfection efficiencies and cell viability after transfection of the CRISPR/Cas9 system components. One of the most effective transfection approaches is electroporation; however, it typically leads to major cell death, which is even more problematic in the case of hiPSCs. Other methods, such as lipofection, cause less cell death but are also less effective.

Taking this into account some derivations of the protocol, intended to enhance cell viability or promote HDR, were also tried. In these modified attempts we combined the use of lipofection or different electroporation parameters (to lower cell death) with a “cold shock” step to promote HDR. Even though the mechanism is not completely understood this “cold shock” step has been reported to substantially increase HDR rates, especially when HDR efficiency is lower, and it is hypothesized that this is due to the influence of the lower temperature (32ºC) in the cell cycle,

leading cells to accumulate in the G2/M stage, which concurs with the phase of the cycle when HDR is active (G2) (Guo et al. 2018).

Lastly, since the primary protocol devised two consecutive rounds of editing and accomplishing the first round successfully was already proving to be a challenge, we decided to test an alternative strategy to prevent the re-editing and unwanted mutations without making the gene-editing process lengthier. For this new approach it was decided to use lipofection and the cell-cycle regulated Cas9-Gem mRNA. This approach relies on the fact that the HDR pathway is only active in the S and G2 phases of the cell cycle. The fusion of the Cas9 with the Geminin peptide leads to their degradation during the G1 phase reducing Cas9 presence when only the NHEJ pathway is active and by that promoting a 2 to 3-fold decrease of NHEJ activity. Moreover, since mRNA is transfected, instead of a plasmid, the expression of the Cas9 is more transient, also contributing to minimize re-cutting activity and the introduction of unwanted mutations in the target region (Howden, Thomson, and Little 2018; Howden et al. 2016).

In addition to the modifications to the gene-editing method already discussed, several other approaches have been reported in the literature. Tube electroporation was described to enhance transfection efficiency in the hard-to-transfect hiPSCs while present a low cytotoxicity when compared to conventional electroporation methods (Xu et al. 2018). Inhibition of the NHEJ pathway with small-molecule inhibitors of the DNA ligase IV combined with cell cycle regulation of the Cas9 presence, is also reported to result in a higher HDR efficiency (Tang et al. 2019).

Although the primary objective of this work could not be achieved, this study highlights the challenges that the precise introduction of a single nucleotide variant through gene-editing in hiPSCs still poses. Moreover, it provides important insights regarding strategies that can be applied or should be avoided to improve future gene-edited assays, drawing attention to the fact that additional care should be taken regarding time and culture conditions to ensure hiPSCs viability and pluripotency after edition.

III.5. REFERENCES

Adalsteinsdottir, Berglind, Polakit Teekakirikul, Barry J. Maron, Michael A. Burke, Daniel F. Gudbjartsson, Hilma Holm, Kari Stefansson, et al. 2014. “Nationwide Study on Hypertrophic Cardiomyopathy in Iceland Evidence of a MYBPC3 Founder Mutation.” *Circulation* 130 (14): 1158–67. https://doi.org/10.1161/CIRCULATIONAHA.114.011207.

Bonne, Gisèle, Lucie Carrier, Josiane Bercovici, Corinne Cruaud, Pascale Richard, Bernard Hainque, Mathias Gautel, et al. 1995. “Cardiac Myosin Binding Protein–C Gene Splice Acceptor Site Mutation Is Associated with Familial Hypertrophic Cardiomyopathy.” *Nature Genetics* 11 (4): 438–40. https://doi.org/10.1038/ng1295-438.

Christiaans, Imke, Erwin Birnie, Irene M. Van Langen, Karin Y. Van Spaendonck-Zwarts, J. Peter Van Tintelen, Maarten P. Van Den Berg, Douwe E. Atsma, et al. 2010. “The Yield of Risk Stratification for Sudden Cardiac Death in Hypertrophic Cardiomyopathy Myosin-Binding Protein C Gene Mutation Carriers: Focus on Predictive Screening.” *European Heart Journal* 31 (7): 842–48. https://doi.org/10.1093/eurheartj/ehp539.

Desmet, François Olivier, Dalil Hamroun, Marine Lalande, Gwenaëlle Collod-Bëroud, Mireille Claustres, and Christophe Béroud. 2009. “Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals.” *Nucleic Acids Research* 37 (9): 1–14. https://doi.org/10.1093/nar/gkp215.

Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. “The New Frontier of Genome Engineering with CRISPR-Cas9.” *Science* 346 (6213): 1258096. https://doi.org/10.1126/science.1258096.

Ehlermann, Philipp, Dieter Weichenhan, Jörg Zehelein, Henning Steen, Regina Pribe, Raphael Zeller, Stephanie Lehrke, Christian Zugck, Boris T Ivandic, and Hugo A Katus. 2008. “Adverse Events in Families with Hypertrophic or Dilated Cardiomyopathy and Mutations in the MYBPC3 Gene.” *BMC Medical Genetics* 9: 1–11. https://doi.org/10.1186/1471-2350-9-95.

Elliott, Perry M, Aris Anastasakis, Michael A. Borger, Martin Borggrefe, Franco Cecchi, Philippe Charron, Albert Alain Hagege, et al. 2014. “2014 ESC Guidelines on Diagnosis and Management of Hypertrophic Cardiomyopathy.” *European Heart Journal* 35 (39): 2733–79. https://doi.org/10.1093/eurheartj/ehu284.

Eng, Laura, Gabriela Coutinho, Shareef Nahas, Gene Yeo, Robert Tanouye, Mahnoush Babaei, Thilo Dörk, Christopher Burge, and Richard A. Gatti. 2004. “Nonclassical Splicing Mutations in the Coding and Noncoding Regions of the ATM Gene: Maximum Entropy Estimates of Splice Junction Strengths.” *Human Mutation* 23 (1): 67–76. https://doi.org/10.1002/humu.10295.

Guo, Q., G. Mintier, M. Ma-Edmonds, D. Storton, X. Wang, X. Xiao, B. Kienzle, D. Zhao, and John N. Feder. 2018. “‘Cold Shock’ Increases the Frequency of Homology Directed Repair Gene Editing in Induced Pluripotent Stem Cells.” *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-20358-5.

Gutschner, Tony, Monika Haemmerle, Giannicola Genovese, Giulio F. Draetta, and Lynda Chin. 2016. “Post-Translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair.” *Cell Reports* 14 (6): 1555–66. https://doi.org/10.1016/j.celrep.2016.01.019.

Howden, Sara E., Bradley McColl, Astrid Glaser, Jim Vadolas, Steven Petrou, Melissa H. Little, Andrew G. Elefanty, and Edouard G. Stanley. 2016. “A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells.” *Stem Cell Reports* 7 (3): 508–17. https://doi.org/10.1016/j.stemcr.2016.07.001.

Howden, Sara E., James A. Thomson, and Melissa H. Little. 2018. “Simultaneous Reprogramming and Gene Editing of Human Fibroblasts.” *Nature Protocols* 13 (5): 875–98. https://doi.org/10.1038/nprot.2018.007.

Junying, Yu, Hu Kejin, Smuga Otto Kim, Tian Shulan, Ron Stewart, Igor I. Slukvin, and James A. Thomson.

2009. “Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences.” *Science* 324 (5928): 797–801. https://doi.org/10.1126/science.1172482.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, et al. 2019. “Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes.” *BioRxiv*, August, 531210. https://doi.org/10.1101/531210.

Kwart, Dylan, Dominik Paquet, Shaun Teo, and Marc Tessier-Lavigne. 2017. “Precise and Efficient Scarless Genome Editing in Stem Cells Using CORRECT.” *Nature Protocols* 12 (2): 329–34. https://doi.org/10.1038/nprot.2016.171.

Landrum, Melissa J., Jennifer M. Lee, George R. Riley, Wonhee Jang, Wendy S. Rubinstein, Deanna M. Church, and Donna R. Maglott. 2014. “ClinVar: Public Archive of Relationships among Sequence Variation and Human Phenotype.” *Nucleic Acids Research* 42 (D1): 980–85. https://doi.org/10.1093/nar/gkt1113.

Leman, Raphäel, Pascaline Gaildrat, Gérald L. Gac, Chandran Ka, Yann Fichou, Marie Pierre Audrezet, Virginie Caux-Moncoutier, et al. 2018. “Novel Diagnostic Tool for Prediction of Variant Spliceogenicity Derived from a Set of 395 Combined in Silico/in Vitro Studies: An International Collaborative Effort.” *Nucleic Acids Research* 46.

Ma, Ning, Joe Z. Zhang, Ilanit Itzhaki, Sophia L. Zhang, Haodong Chen, Francois Haddad, Tomoya Kitani, et al. 2018. “Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells.” *Circulation* 138 (23): 2666–81. https://doi.org/10.1161/CIRCULATIONAHA.117.032273.

Marian, Ali J., and Eugene Braunwald. 2017. “Hypertrophic Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy.” *Circulation Research* 121 (7): 749–70. https://doi.org/10.1161/CIRCRESAHA.117.311059.

Maron, Barry J. 2018. “Clinical Course and Management of Hypertrophic Cardiomyopathy.” *New England Journal of Medicine* 379 (7): 655–68. https://doi.org/10.1056/nejmra1710575.

Maron, Barry J., Ethan J. Rowin, and Martin S. Maron. 2018. “Global Burden of Hypertrophic Cardiomyopathy.” *JACC: Heart Failure* 6 (5): 376–78. https://doi.org/10.1016/j.jchf.2018.03.004.

Mil, Alain van, Geerthe Margriet Balk, Klaus Neef, Jan Willem Buikema, Folkert W. Asselbergs, Sean M. Wu, Pieter A. Doevendans, and Joost P.G. Sluijter. 2018. “Modelling Inherited Cardiac Disease Using Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Progress, Pitfalls, and Potential.” *Cardiovascular Research* 114 (14): 1828–42. https://doi.org/10.1093/cvr/cvy208.

Miyaoka, Yuichiro, Jennifer R. Berman, Samantha B. Cooper, Steven J. Mayerl, Amanda H. Chan, Bin Zhang, George A. Karlin-Neumann, and Bruce R. Conklin. 2016. “Systematic Quantification of HDR and NHEJ Reveals Effects of Locus, Nuclease, and Cell Type on Genome-Editing.” *Scientific Reports* 6 (March). https://doi.org/10.1038/srep23549.

Mosqueira, Diogo, Ingra Mannhardt, Jamie R. Bhagwan, Katarzyna Lis-Slimak, Puspita Katili, Elizabeth Scott, Mustafa Hassan, et al. 2018. “CRISPR/Cas9 Editing in Human Pluripotent Stem Cell-Cardiomyocytes Highlights Arrhythmias, Hypocontractility, and Energy Depletion as Potential Therapeutic Targets for Hypertrophic Cardiomyopathy.” *European Heart Journal* 39 (43): 3879–92. https://doi.org/10.1093/eurheartj/ehy249.

Musunuru, Kiran, Farah Sheikh, Rajat M. Gupta, Steven R. Houser, Kevin O. Maher, David J. Milan, Andre Terzic, and Joseph C. Wu. 2018. “Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling and Precision Medicine: A Scientific Statement From the American Heart Association.” *Circulation. Genomic and Precision Medicine*. NLM (Medline). https://doi.org/10.1161/HCG.0000000000000043.

Niimura, Hideshi, Linda L. Bachinski, Somkiat Sangwatanaroj, Hugh Watkins, Albert E. Chudley, William

Mckenna, Arni Kristinsson, et al. 1998. “Mutations in the Gene for Cardiac Myosin-Binding Protein C and Late- Onset Familial Hypertrophic Cardiomyopathy.” *New England Journal of Medicine* 338 (18): 1248–57. https://doi.org/10.1056/NEJM199804303381802.

Norland, Kristjan, Gardar Sveinbjornsson, Rosa B. Thorolfsdottir, Olafur B. Davidsson, Vinicius Tragante, Sridharan Rajamani, Anna Helgadottir, et al. 2019. “Sequence Variants with Large Effects on Cardiac Electrophysiology and Disease.” *Nature Communications* 10 (1): 1–10. https://doi.org/10.1038/s41467-019-12682-9.

Paquet, Dominik, Dylan Kwart, Antonia Chen, Andrew Sproul, Samson Jacob, Shaun Teo, Kimberly Moore Olsen, Andrew Gregg, Scott Noggle, and Marc Tessier-Lavigne. 2016. “Efficient Introduction of Specific Homozygous and Heterozygous Mutations Using CRISPR/Cas9.” *Nature* 533 (7601): 125–29. https://doi.org/10.1038/nature17664.

Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. 2013. “Genome Engineering Using the CRISPR-Cas9 System.” *Nature Protocols* 8 (11): 2281–2308. https://doi.org/10.1038/nprot.2013.143.

Richard, Pascale, Philippe Charron, Lucie Carrier, Céline Ledeuil, Theary Cheav, Claire Pichereau, Abdelaziz Benaiche, et al. 2003. “Hypertrophic Cardiomyopathy: Distribution of Disease Genes, Spectrum of Mutations, and Implications for a Molecular Diagnosis Strategy.” *Circulation* 107 (17): 2227–32. https://doi.org/10.1161/01.CIR.0000066323.15244.54.

Skelton, Rhys J.P., Magdaline Costa, David J. Anderson, Freya Bruveris, Ben W. Finnin, Katerina Koutsis, Deevina Arasaratnam, et al. 2014. “SIRPA, VCAM1 and CD34 Identify Discrete Lineages during Early Human Cardiovascular Development.” *Stem Cell Research* 13 (1): 172–79. https://doi.org/10.1016/j.scr.2014.04.016.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, Shinya Yamanaka, et al. 2007. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.” *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Tang, Xi Dian, Fei Gao, Ming Jie Liu, Qin Lei Fan, De Kun Chen, and Wen Tao Ma. 2019. “Methods for Enhancing Clustered Regularly Interspaced Short Palindromic Repeats/Cas9-Mediated Homology-Directed Repair Efficiency.” *Frontiers in Genetics* 10 (JUN). https://doi.org/10.3389/fgene.2019.00551.

Wang, Lili, Kyungsoo Kim, Shan Parikh, Adrian Gabriel Cadar, Kevin R. Bersell, Huan He, Jose R. Pinto, Dmytro O. Kryshtal, and Bjorn C. Knollmann. 2018. “Hypertrophic Cardiomyopathy-Linked Mutation in Troponin T Causes Myofibrillar Disarray and pro-Arrhythmic Action Potential Changes in Human IPSC Cardiomyocytes.” *Journal of Molecular and Cellular Cardiology* 114 (January): 320–27. https://doi.org/10.1016/j.yjmcc.2017.12.002.

Xu, Xiaoyun, Dongbing Gao, Ping Wang, Jian Chen, Jinxue Ruan, Jie Xu, and Xiaofeng Xia. 2018. “Efficient Homology-Directed Gene Editing by CRISPR/Cas9 in Human Stem and Primary Cells Using Tube Electroporation.” *Scientific Reports* 8 (1): 1–11. https://doi.org/10.1038/s41598-018-30227-w.

Yiu, Kai Hang, Douwe E Atsma, Victoria Delgado, Arnold C T Ng, Tomasz G Witkowski, Hooi Ewe, Dominique Auger, et al. 2012. “Myocardial Structural Alteration and Systolic Dysfunction in Preclinical Hypertrophic Cardiomyopathy Mutation Carriers” 7 (5): 1–10. https://doi.org/10.1371/journal.pone.0036115.

Yoshida, Yoshinori, and Shinya Yamanaka. 2017. “Induced Pluripotent Stem Cells 10 Years Later.” *Circulation Research* 120 (12): 1958–68. https://doi.org/10.1161/CIRCRESAHA.117.311080.

IV. A 3D2D DIFFERENTIATION AND PURIFICATION APPROACH RECAPITULATES HCM CELLULAR PHENOTYPE IN HUMAN INDUCED PLURIPOTENT STEM CELL DERIVED CARDIOMYOCYTES

IV.1. ABSTRACT

A limitation in cardiac research has been the lack of adequate disease models. As an alternative to animal models, or primary cells, cardiomyocytes have been derived from human induced pluripotent stem cells (hiPSC-CMs). However, many differentiation protocols are only able to produce heterogeneous populations of cardiomyocytes, with a low degree of maturation.

Here we describe a novel combinatory approach, taking advantage of an initial GiWi 3D differentiation strategy from which VCAM1 expressing hiPSC-CMs can be purified by FACS and further matured in a monolayer culture until day 30 of differentiation. This process allowed the generation of homogeneous populations of cardiomyocytes with a higher degree of maturation when compared to hiPSC-CMs generated through a standard GiWi monolayer protocol.

To ensure the capability of this approach to be used as an HCM disease modeling, we also differentiated HCM patient-derived hiPSCs and accessed if the generated hiPSC-CMs could recapitulate the HCM phenotype at the cellular level, namely increased cellular size, multinucleation, and disorganized sarcomeres.

We anticipate that this approach can be used to better understand mechanisms underlying the pathophysiology of HCM and test targeted therapeutic approaches.

Keywords: cardiomyocytes, hypertrophic cardiomyopathy, human pluripotent stem cells, cardiac differentiation; disease models

IV.2. INTRODUCTION

Cardiac diseases are the most prevalent cause of death in the worldwide (World Health Organization 2018) and understanding molecular mechanisms underlying many cardiac afflictions and identifying individual drug sensitivities is an active field of research.

The recent emergence of hiPSCs provided an important human cell source to develop alternative disease models, based on their *in vitro* expansion potential and capability to differentiate into any somatic cell-type, including cardiomyocytes (Takahashi et al. 2007; Yoshida and Yamanaka 2017).

Cardiac differentiation protocols were greatly improved by the modulation of key signaling pathways associated with embryonic heart development that enable to recapitulate to some extent the critical stages of cardiac specification, leading to a higher differentiation efficiency /yield. One of the most widely applied is based on the temporal modulation of Wnt signaling pathway (GiWi protocols) (Karakikes et al. 2015; Lian et al. 2013; Burridge et al. 2014). However, the heterogeneity and immature characteristics of the obtained hiPSC-CMs, which resemble more the characteristics of fetal CMs than adult ones, may limit their application in generating reliable cellular disease models to study genotype-phenotype relations, the impact of VUS or testing new therapeutic approaches (Hoes, Bomer, and van der Meer 2019; Purevjav 2019; Musunuru et al. 2018). To address these concerns different strategies have been proposed for the purification and/or maturation of hiPSC-CMs.

Since the previously mentioned differentiation protocols have been mostly applied in monolayer cultures, they do not consider the 3D configuration of the embryo, which provides appropriate spatial, temporal and mechanical cues that are crucial for the human heart development. Recently, cardiac differentiation and maturation strategies that use a 3D culture to mimic the embryonic development of the heart *in vitro* have emerged. These approaches

demonstrated a faster and more reproducible way to generate hiPSC-CMs with higher levels of maturity than the ones obtain from similar 2D strategies (Branco et al. 2019; Correia et al. 2018).

To purify hiPSC-CMs from other cells types that may be present in the differentiated cardiac cell cultures, several methods have been proposed. An effective method to do so is by FACS, using an antibody for a CM-specific membrane protein, the VCAM-1. VCAM-1 has been identified as a CM markers in previous studies, with 80% of TNNT2-expressing cells, at day 11 of differentiation, being VCAM+, and 95–98% of VCAM1+ cells being positive for TNNT2 (Uosaki et al. 2011; D. A. Elliott et al. 2011).

Here, we describe a novel combinatory *in vitro* cardiac differentiation approach, taking advantage of an initial GiWi 3D differentiation strategy, from which VCAM-1 expressing hiPSC-CMs are purified by FACS and further matured in a monolayer culture until day 30 of differentiation.

Using this approach, we aimed to obtain a pure and mature population of hiPSC-CMs that could be further used for HCM disease modeling, as standard genetic testing often fails to identify causative mutation for this disease and currently available treatments do not target the pathophysiology or genetic cause of the disease (Gruner et al. 2013; P. M. Elliott et al. 2014).

To ensure the capability of the proposed approach to be used in HCM disease modeling, we used it to differentiate HCM patient-derived hiPSCs and access if generated hiPSC-CMs could recapitulate the HCM phenotype at the cellular level, namely increased cellular size, multinucleation and disorganized sarcomeres.

IV.3. METHODS

Maintenance of human iPSCs

Three commercially available human iPSC lines were used as wild-type (WT) control cell lines, namely DF6.9.9 T.B cell line (DF6, WiCell®), reprogrammed from foreskin fibroblasts (Junying et al. 2009), F002.1A.13 (TCLab, Portugal), derived from female donor fibroblasts (Takahashi et al. 2007) and iPSC6.2 (Gibco®), derived from $CD34^+$ cord blood (Burridge et al. 2011). Six HCM patient-derived hiPSC lines, reprogrammed from fibroblasts using Sendai virus, were also analyzed (Table IV.1.).

Table IV.1. hiPSC lines derived from HCM patients used in the study. AA - amino acid; Ref – reference nucleotide; Alt – HCM altered nucleotide.

Name	Mutated gene	AA change	Start coordinate	Ref	Alt
Miov	*TTN*	p.F17690L	179430596	A	G
		p.V20731M	179418346	C	T
		p.I22692T	179410693	A	G
		p.I10273T	179538425	A	G
	MYH7	p.R663C	23896043	G	A
iooi	*TTN*	p.K6071N	179486037	C	A
		p.E7030K	179582913	C	T
	MYBPC3	p.R502W	47364249	G	A
Ruah	*TTN*	p.A13755P	179442784	C	G
		p.S14161G	179441295	T	C
		p.R6843H	179482089	C	T
		p.K4646E *intronic	179613191	T	C
Jewl	*TTN*	p.S14161G	179441295	T	C
		p.R6843H	179482089	C	T
Xutl	*MYBPC3*	p.I1250fs	47353690	C	insC
Newl	*ACTN2*	p.Q741R	236920853	A	G

All human iPSCs were cultured on Matrigel™(Corning)-coated plates with mTeSR™1 Medium (StemCell Technologies). Medium was changed daily, and cells were passaged every three to four days (at a confluence of approximately 85% of the surface area of the culture dish) using 0.5mM EDTA dissociation buffer (Life Technologies). Two to three passages were performed before starting the differentiation process.

Cardiac differentiation

For 2D culture, hiPSCs were seeded onto Matrigel-coated 12-well tissue culture plates and cultured in mTeSR™1. Medium was changed daily until a confluence of around 90% was attained. For 3D aggregates formation, before seeding, hiPSCs were incubated with ROCKi (Y-27632, 10μM, StemCell Technologies) for 1 h at 37ºC and then treated with Accutase (Sigma) for 7 min at 37ºC. After dissociation, cells were forced to aggregate using microwell plates (AggreWell™800, StemCell Technologies) according to the manufacturer's instructions. Cells were plated at a density of 0.9 x 10^6cells/well (3,000 cells/microwell) in 1.5 mL/well of mTeSR™1 supplemented with 10μM ROCKi. The day on which the cell seeding was performed was defined as day -3. After 24 hours, full volume of the medium was replaced, and aggregates were maintained in mTeSR™1 without ROCKi for an additional 48h.

For hiPSCs differentiation into cardiomyocytes, for both 2D and 3D2D culture conditions, an adapted GiWi protocol was used (Lian et al. 2013; Branco et al. 2019). From day 0 to day 6, cells were cultured in RPMI 1640 (Thermo Fisher Scientific) supplemented with 2%(v/v) B-27 minus insulin (Thermo Fisher Scientific). Then, from day 7 until the end of differentiation, cells were cultured in RPMI supplemented with 2%(v/v) B-27 (Thermo Fisher Scientific). At day 0 of differentiation, the Wnt signaling pathway was activated using the GSK3 inhibitor CHIR99021 (Stemgent) at a final concentration of 6 μM, in 2D conditions, and 11 μM, in the 3D aggregates. After 24 hours (day 1), full-volume medium replacement with RPMI + B27 minus insulin was

performed. At day 3, half of the medium in each well was replaced and cells were supplemented with Wnt inhibitor IWP-4 (Stemgent) at a final concentration of 5 μM, for two days. At day 5, a total volume of medium change was performed. On day 7, medium was changed for RPMI + B27 and, in the case of the 3D culture, aggregates were flushed from the AggreWell™800 plate and transferred to ultra-low attachment 6-well plates (Costar, Corning). At day 10, full-volume medium replacement was performed, and for the 2D differentiation, total volume of RPMI + B27 medium was replaced at each 2–3 days, until cell harvest at day 30 of differentiation. For the 3D2D differentiation, at day 12, aggregates were dissociated using 0.25% Trypsin-EDTA (Gibco) for 7 min at 37ºC. After dissociation, cells were washed with 2% fetal bovine serum in phosphate buffered saline (1xPBS, 0.1M), resuspended in VCAM1 antibody (BioLegend, 1:50) diluted in 1xPBS/2% FBS and incubated for 30 min at 37ºC. Afterwards, cells were washed and resuspended in with 1xPBS/2% FBS prior to FACS sorting. VCAM1 positive cells were plated on wells coated with Matrigel, at a seeding density between 20,000 – 40,000 cells/cm^2. For the remaining time in culture, two-thirds of the total volume of RPMI + B27 medium was replaced at every 2–3 days, until cell harvest at day 30 of differentiation.

Quantitative real time (qRT)-PCR

Expression profiles of sarcomeric genes were assessed by qRT-PCR analysis in hiPSC-CMs at day 30 of differentiation. As a positive cardiac control, a commercially available human heart RNA was used (cat # AM7966, Ambicon, Invitrogen). Briefly, total RNA was extracted using NZYol (NZYTech®) with a standard protocol, and contaminating DNA was removed by DNase I (Roche®) treatment. Complementary cDNA synthesis was achieved with the Transcriptor High Fidelity cDNA Synthesis Kit (Roche®). qRT-PCR was performed using the Universal SYBR Green Supermix (Bio-Rad) and specific primers for each gene and/or developmental isoform (Table IV.2.) All PCR reactions were run in triplicate, using the ViiA™7 RT-PCR Systems (Applied

BioSystems). Data was normalized against the *U6* housekeeping gene (or cardiac specific *TNNT2*) and depicted in histograms with mean and standard deviation from at least three independent experiments**.**

Table IV.2. Primers used for qRT-PCR

Gene	Foward	Reverse
U6	CGTTCGGCAGCACATATAC	AAATATGGAACGCTTCACGA
MYBPC3	TAGCAAGAAGCCACGGTCAG	CCAGCAATGACTGCGTAAGA
MYH7	TCGTGCCTGATGACAAACAGGAGT	ATACTCGGTCTCGGCAGTGACTTT
MYH6	GATAGAGAGACTCCTGCGGC	CCGTCTTCCCATTCTCGGTT
ACTN2	ACTCCCGAGAAGACCATGC	CAGGACGGTTGCTGATCC
***TNNT2* total**	AGAGCGGAAAAGTGGGAAGA	GCTGATCTTCATTCAGGTGGT
***TNNT2* fetal**	AGAGGAGGACTGGAGAGAGG	CTGGGCTTTGGTTTGGACTC
***TTN* total**	TTCAGAAGCAACCTTGGGCT	GTAACGGCTGCGTAAACGTC
***TTN* N2BA**	GCCTGGAATGAGCCTCACAT	ATGTTGCATGACTCCCCAGG
***TTN* N2B**	GGTTGACTGCGGCGAGTATA	ACAACTTCTTCCTTTGGTTCAGG
TNNI1	CAGCTCCACGAGGACTGAAC	CTCTTCAGCAAGAGTTTGCG
TNNI3	CCTCAAGCAGGTGAAGAAGG	CAGTAGGCAGGAAGGCTCAG

Immunofluorescence (IF) assays and morphologic characterization of hiPSC-CMs

IF assays were performed using different antibodies against several sarcomeric proteins. Briefly, at day 30 of differentiation, hiPSC-CMs were fixed with 3.7% PFA/1xPBS and permeabilized with 0.5% Tx100/1XPBS. Next, a double stained with (1) phalloidin conjugated with TxRed (Thermo Fisher, T7471) (for detection of actin) and (2) a specific antibody against a given sarcomeric protein (Table IV.3), detected by an anti-mouse Alexa Fluor 488-conjugated (A-11018

Thermo Fisher Scientific) was performed overnight at 37ºC, in a moist chamber. Nuclei counterstaining was performed using 4',6-diamidino-2-phenylindole (DAPI, 1µg/mL; Enzo Life Sciences). After brief drying, coverslips were mounted in VECTASHIELD® Mounting Medium. Fluorescence images were acquired with Zeiss LSM 710 Confocal Laser Point-Scanning Microscope.

Table IV.3. List of primary antibodies and dilutions used for IF assays

Antibody	Company	Host species	Dilution
α-Actinin (Actinin)	Abcam	mouse	1:75
cMyBP-C (MYBPC3)	Santa Cruz Biotechnology	mouse	1:50
cTnT (Troponin T)	Neomarkers	mouse	1:50

hiPSC-CMs cell morphology parameters, namely area and shape descriptors (circularity and aspect ratio), were determined in Image J software using standard analysis plugins. Sarcomere length and percentage of binucleated cells were also accessed. For each cell line, at least 100 cells from 2 independent experiments were examined. Cell morphology parameters of control and PD hiPSC-CMs were compared using Student's t-test at a 95% confidence interval. The statistical analysis of the data was performed using GraphPad Prim software, version 5.03.

RNA sequencing

Total RNA was extracted from hiPSC-CMs at day 30 of differentiation using NZYol (NZYTech®), as previously described. Strand-specific RNA libraries were prepared for Illumina sequencing using standard protocols. Raw data was preprocessed with TrimGalore v0.4.4. Splicing isoform quantification was performed with Salmon v0.13.1.

IV.4. RESULTS

Cardiac differentiation of hiPSCs using a combined 3D/2D approach

Human iPSCs were differentiated into cardiomyocytes using a 3D/2D combined approach of differentiation, purification and maturation steps. The initial 3D differentiation steps were adapted from a GiWi protocol previously published (Branco et al. 2019), where hiPSCs were forced to aggregate in AggreWell™800 plates and then induced to differentiate into cardiomyocytes (CMs), as size-controlled aggregates, via temporal modulation of the Wnt signaling pathway, by sequential addition of a GSK3 inhibitor (CHIR99021) and a Wnt inhibitor (IWP-4). After this step, on day 12 of differentiation, aggregates were dissociated and cells expressing VCAM-1 were isolated by FACS (Uosaki et al. 2011). These purified CMs were replated on wells coated with Matrigel and further maintained in 2D culture until cell harvest at day 30 of cardiac differentiation Figure IV.1.

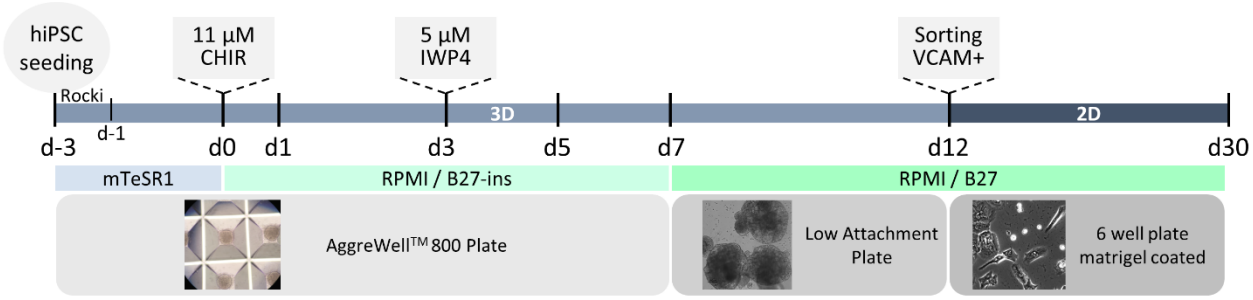


Figure IV.1. Schematics illustrating the culture conditions used to induce differentiation of iPSCs to cardiomyocytes with the 3D2D differentiation protocol. DF6, TCLab and Gibco cells were forced to aggregate in AggreWell™800 plates. CM differentiation was performed following temporal modulation of Wnt signalling pathway. On day 7 of differentiation, aggregates were transferred to ultra-low attachment plates; then, on day 12, aggregates were dissociated and VCAM1 positive cells, sorted by FACS, were replated on wells coated with Matrigel, and kept in 2D monolayer culture until day 30 of cardiac differentiation.

3D/2D cardiac differentiation of hiPS cells gives rise to a population of pure CMs with a high degree of maturation

The three WT hiPSC lines were differentiated using two different approaches, an established GiWi monolayer protocol, adapted from Lian et al., and the 3D/2D approach just described (Lian et al. 2013; Branco et al. 2019). hiPSC-CMs at day 30 of differentiation were then collected and further analyzed. To access the impact of the introduced FACs purification step in our combined 3D/2D approach, a fraction of dissociated 3D aggregates not subjected to FACS purification (unsorted), as well as cells that were negative for VCAM1 (VCAM1 -) were also collected and further analyzed.

Comparative IF analysis of cells generated through each of the protocols and stained for sarcomeric proteins (cMYBPC, cTnT and a-actinin) revealed a marked improvement in cell elongation and sarcomere organization, which was evidenced by higher density and more visible striation alignment across the cells, in the 3D/2D hiPSC-CMs. In turn, cells differentiated through the 2D protocol are rounder, smaller and present less organized and aligned sarcomeres (Figure IV.2.).

The length of sarcomeres within the contractile apparatus was measured, for one of the differentiated cell lines, using cells stained with α-actinin (present in the Z-lines of sarcomeres) (Figure IV.3.A.). The statistically significant elongation of sarcomere length, from 1.59±0.008 mm in 2D hiPSC-CMs to 1.765 ± 0.008 mm in 3D/2D hiPSC-CMs, further confirms a more mature phenotype of the generated 3D/2D hiPSC-CMs (Figure IV.3.B.).

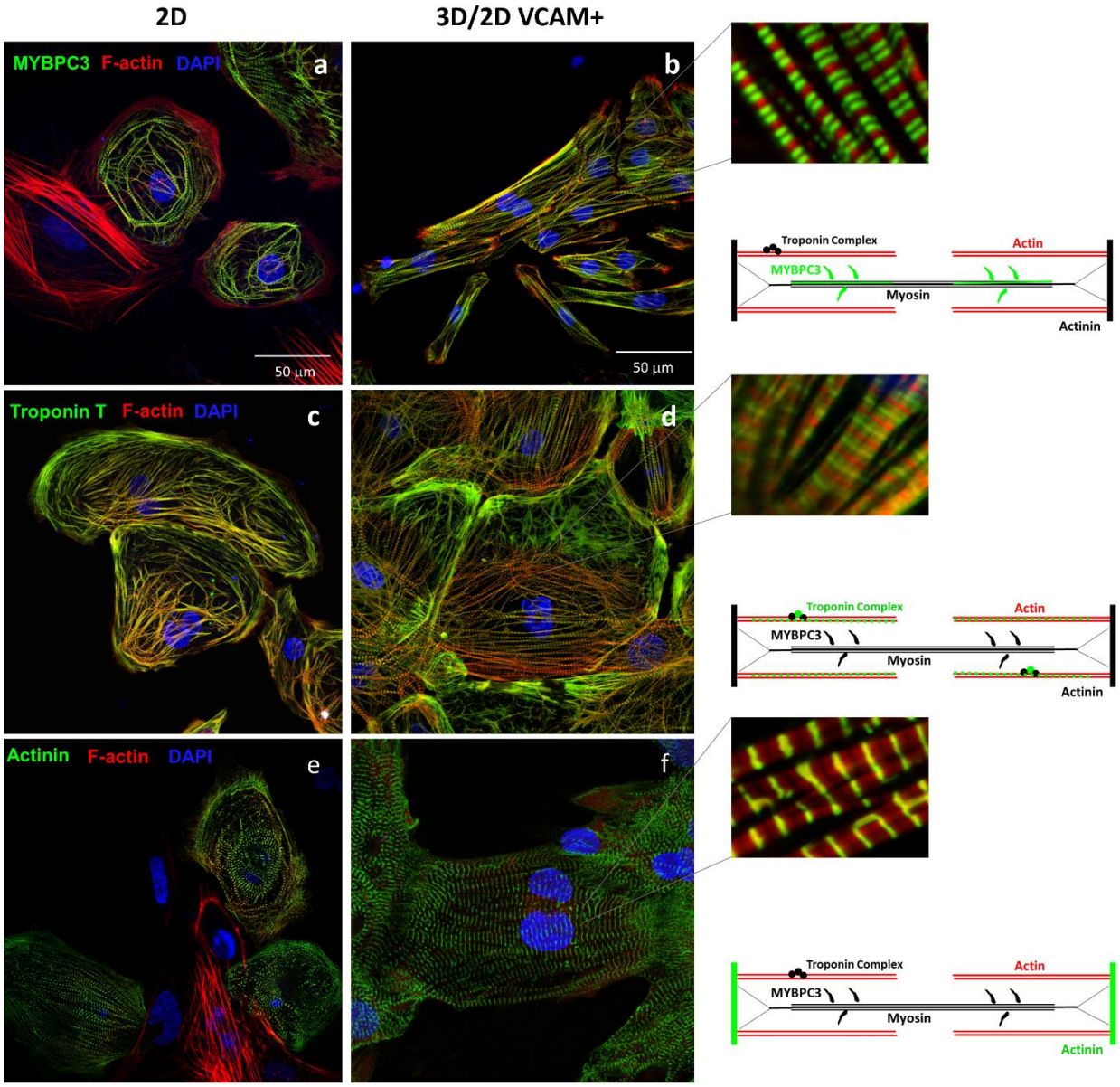
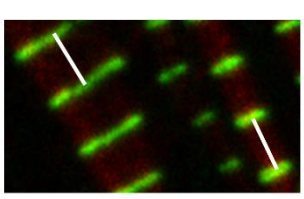


Figure IV.2. Representative IF image of hiPSC-CMs generated using the 2D or the 3D/2D protocol. Cells were stained for: MYBPC3 (a and b), Troponin T (c and d) or α-Actinin (e and f) (green) and F-actin (red). Nuclei are stained with DAPI (in blue). On the right side, schematic illustrations depict the location of the analyzed sarcomeric proteins, with a magnification of the sarcomeric pattern formed among them.



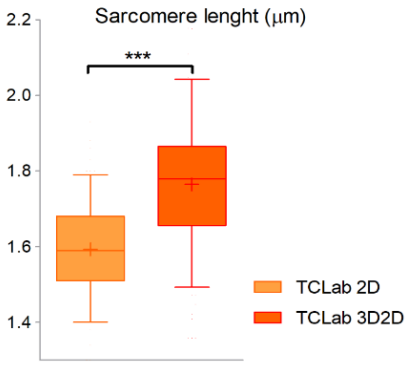
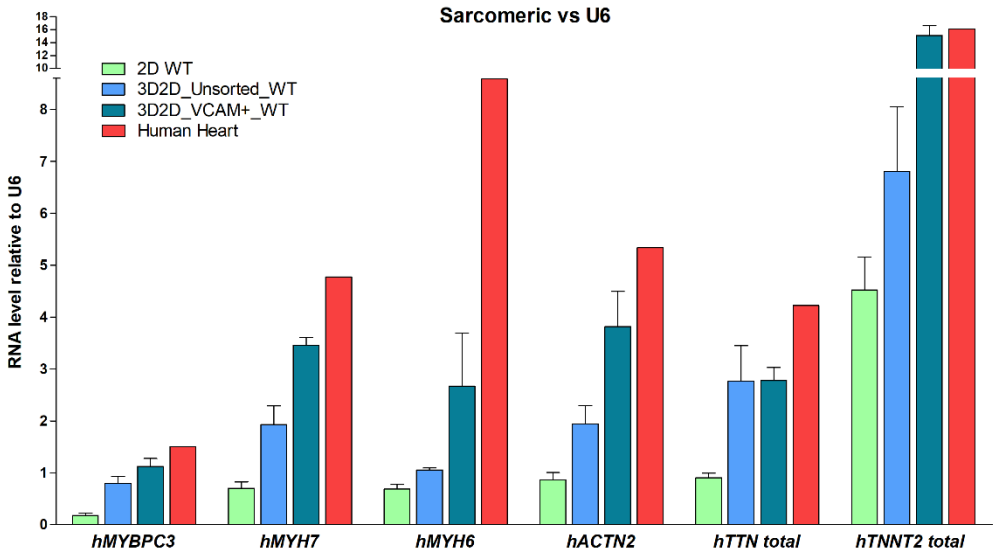


Figure IV.3. Sarcomere length measurement using IF against α-Actinin. (A) Example of a sarcomere measurement. (B) Quantitative plot of the data collected from 200 and 330 sarcomere measurements in 2D or 3D/2D TCLab hiPSC-CMs, respectively. Data represented as mean ± SD. ***p<0.001,

To evaluate the performance of the two cardiac differentiation protocols used, the expression of sarcomeric genes was accessed by qRT-PCR analysis. Results show that all analyzed sarcomeric genes *MYBPC3, MYH7, MYH6, ACTN2, TTN, TNNT2* are expressed in the different experimental conditions (2D protocol; 3D/2D protocol without purification; and 3D/2D protocol VCAM1 positive cells); however, in the 3D2D VCAM+ condition, expression levels were significantly higher (Figure IV.4.A.). Interestingly, when normalizing our data to TNNT2 expression, the differences amongst the expression of the remaining sarcomeric genes were attenuated, suggesting the presence of a more homogeneous hiPSC-CM population in 3D/2D VCAM1+ cells (Figure IV.4.B.). In contrast, VCAM1 negative cells shown very low to inexistent expression levels of cardiac sarcomeric genes (qRT-PCR analysis), and absence of sarcomere formation (IF assays) (data not shown), confirming that hiPSC-CMs were positively selected by FACs.

A



B

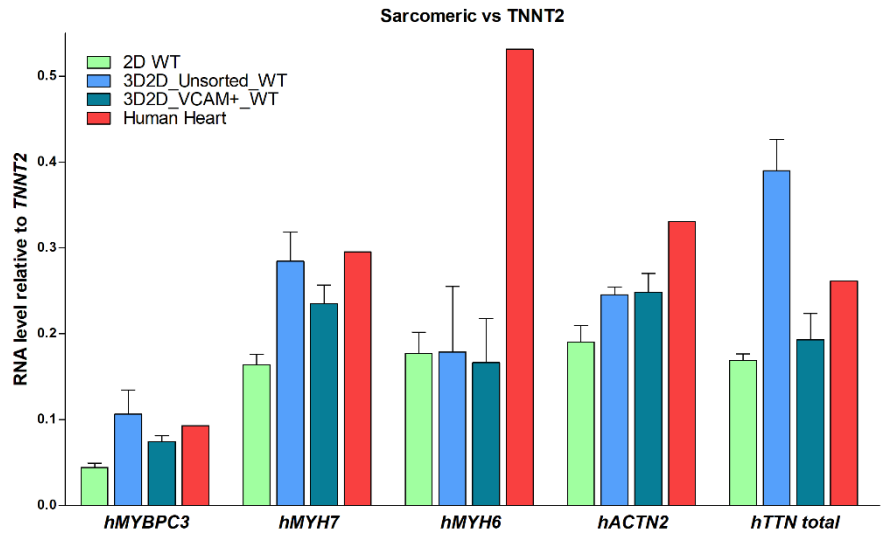


Figure IV.4. qRT-PCR analysis of sarcomeric genes expression in WT hiPSC-CMs collected on day 30 of differentiation (average of 3 independent cell lines), under the different experimental conditions: 2D protocol (2D_WT); 3D/2D protocol without sorting (3D2D_Unsorted_WT) and 3D/2D protocol VCAM1 positive cells (3D2D_VCAM+_WT), compared to human heart as a positive reference. (A) Data normalized against housekeeping *U6* gene. (B) Data normalized against *TNNT2* (n= 3 independent experiments, data represented as mean ± SD).

Since all tested conditions were able to produce hiPSC-CMs, although in different yields, we further characterized the degree of maturation of such hiPSC-CMs by accessing the expression

levels of different developmental cardiac isoforms of selected sarcomeric genes. Specifically, here we evaluated the levels of fetal troponin T (*TNNT2*), fetal and adult titin (*TTN*) and fetal and adult troponin I (*TNNI*), by qRT-PCR. Results show that there is an increased expression of the adult cardiac isoforms in the 3D/2D VCAM+ hiPSC-CMs population (Figure IV.5.A.). This can be better visualized by looking to the ratios between fetal / adult (or total) isoforms, which are higher in the 2D hiPSC-CMs, demonstrating their tendency to be more immature (Figure IV.5.B).

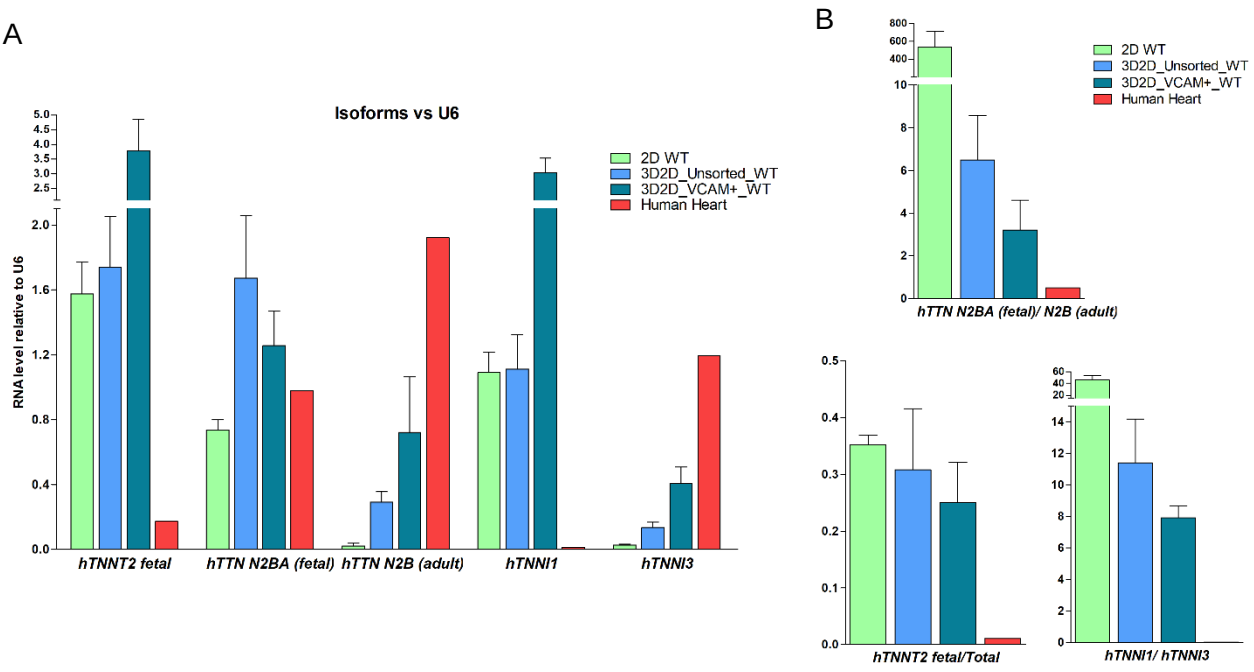


Figure IV.5. (A) qRT-PCR analysis of sarcomeric splicing isoforms (*TNNT2*, *TTN and TNNI*) in WT hiPSC-CMs (average of 3 independent cell lines), under the different experimental conditions: 2D protocol (2D_WT); 3D/2D protocol without sorting (3D2D_Unsorted_WT) and 3D/2D protocol VCAM1 positive cells (3D2D_VCAM+_WT), compared to human heart as a positive reference. (B) Ratios between fetal/adult (or total) sarcomeric splicing isoforms. (n= 3 independent experiments, data represented as mean ± SD).

RNA-seq data analysis further confirmed that 3D/2D VCAM+ hiPSC-CMs are expressing splicing isoforms characteristic of the adult heart in greater levels than the fetal ones (Figure IV.6.).

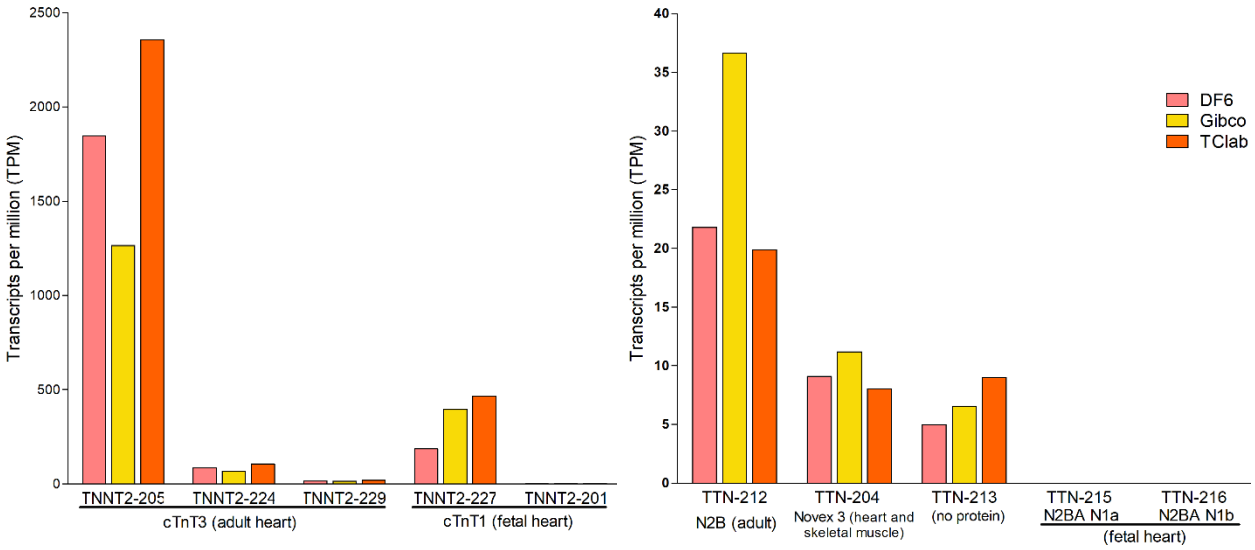


Figure IV.6. Analysis of *TTN* and *TNNT2* developmental splicing isoforms expressed in hiPSC-CMs differentiated from the three WT hiPS cell lines (DF6, Gibco and TCLab).

3D/2D patient derived hiPSC-CMs recapitulates HCM phenotype at the cellular level

The expression of cardiac sarcomeric genes in hiPSC-CMs, generated through the previously described 3D/2D protocol, either from WT or HCM patient derived hiPSC lines, was further evaluated by qRT-PCR. Both HCM and WT hiPSC-CMs show similar levels of expression of the analyzed sarcomeric genes, confirming that we were able to successfully generate hiPSC-CMs from all tested cell lines (Figure IV.7).

When evaluating the maturity of such WT and HCM hiPSC-CMs through the expression of developmental cardiac splicing isoforms, it can be observed that fetal splicing *TNNT2* isoform appears enriched in some HCM iPSC-CMs relative to WT ones; also, the majority of HCM iPSC-CMs show reduced levels of the adult splicing *N2B TTN* isoform relative to control, WT hiPSC-CMs. Nevertheless, no clear differences regarding fetal *TNNI1* and adult *TNNI3* expression levels were observed among WT and HCM hiPSC-CMs (Figure IV.8.).

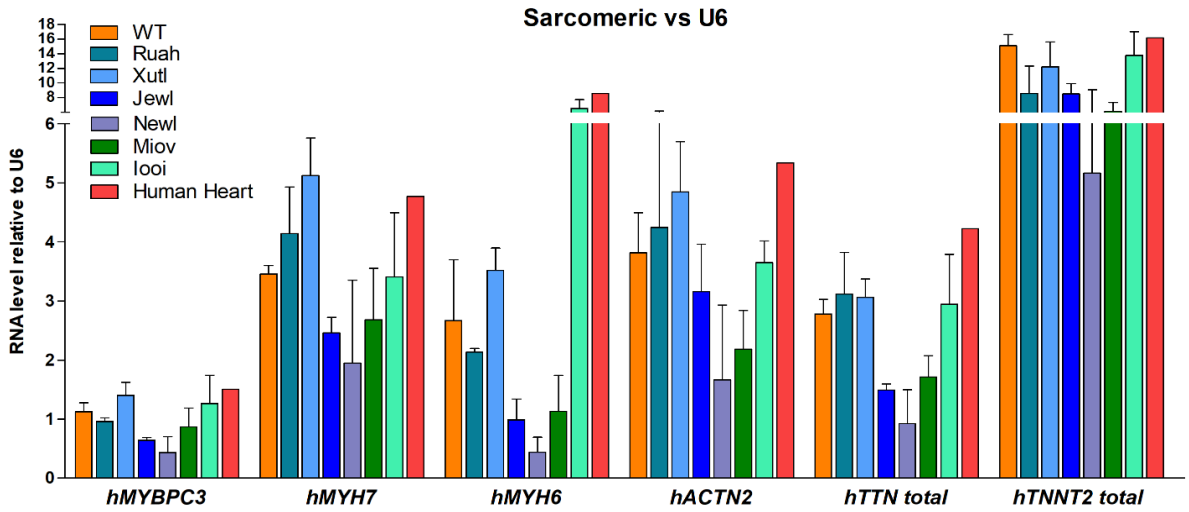


Figure IV.7. qRT-PCR analysis of sarcomeric genes expression in WT hiPSC-CMs VCAM+ (average of 3 independent cell lines) and HCM PD hiPSC-CMs VCAM+ (n= 3 independent experiments, data represented as mean ± SD).

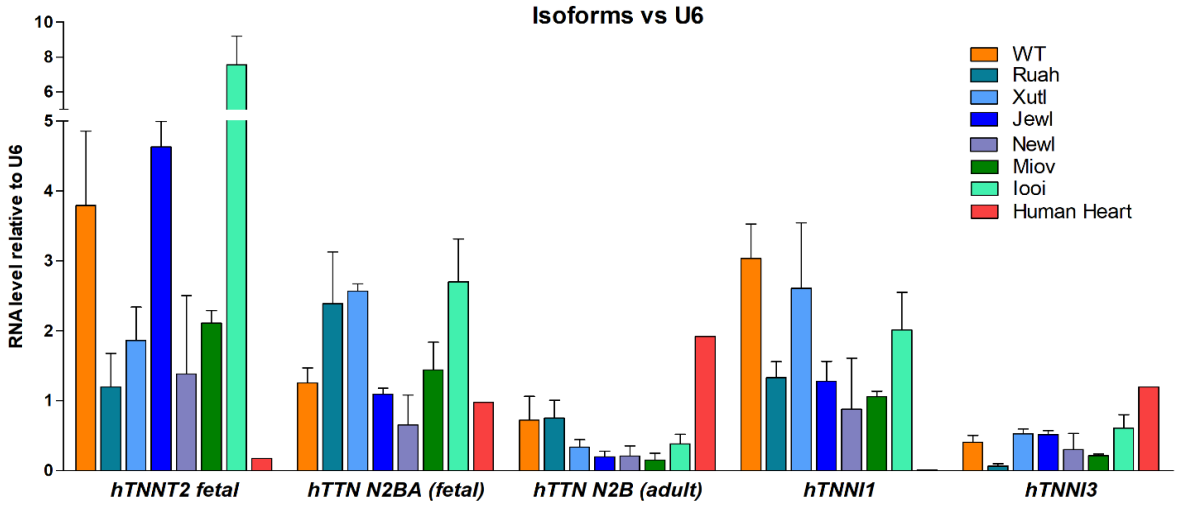


Figure IV.8. qRT-PCR analysis of sarcomeric isoforms (*TNNT2*, *TTN and TNNI*) in WT hiPSC-CMs VCAM+ (average of 3 independent cell lines), and HCM iPSC-CMs VCAM+. (n= 3 independent experiments, data represented as mean ± SD).

Comparative IF analysis of WT and HCM hiPSC-CMs generated through the proposed 3D/2D protocol show that HCM hiPSC-CMs are larger, rounder and have a less organized sarcomeric structure than the WT ones (Figure IV.9).

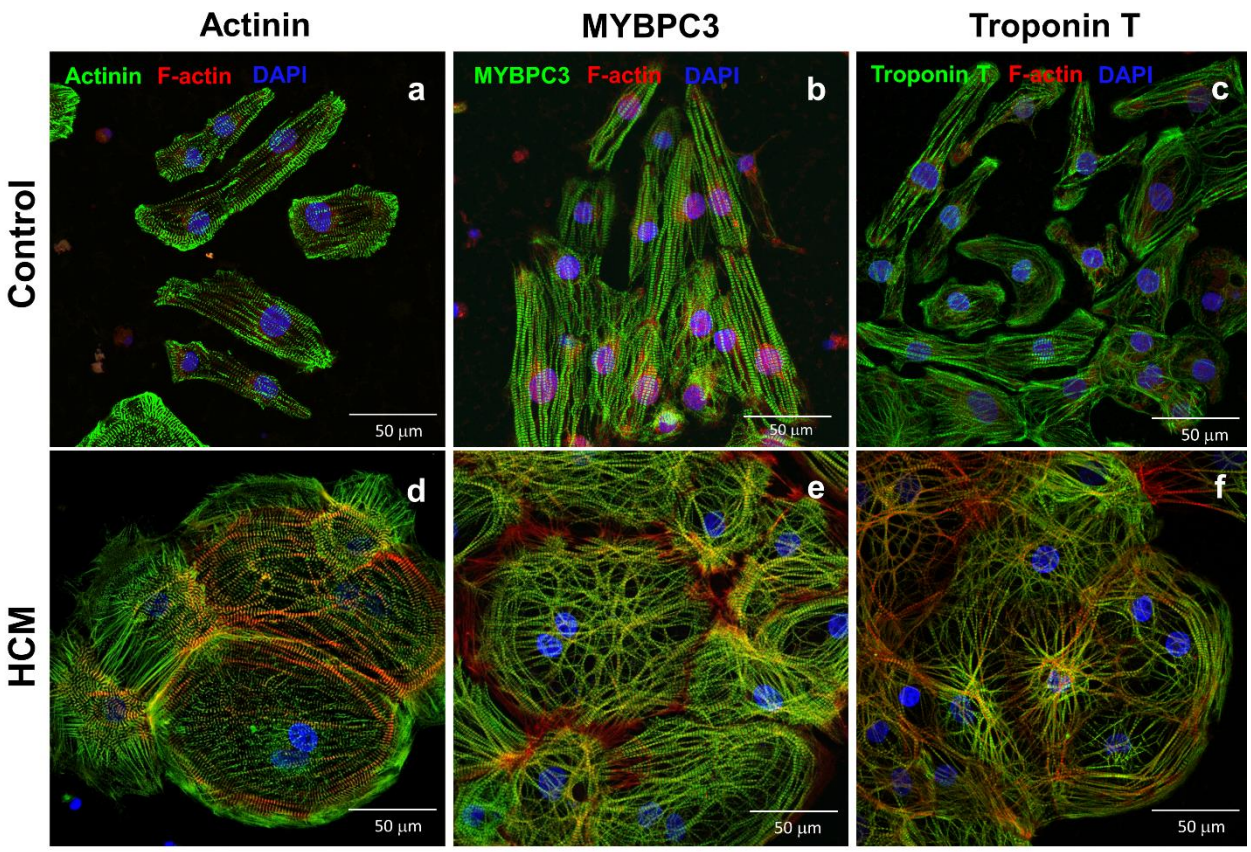


Figure IV.9. Representative IF image of WT (Control; a, b, c) and HCM (d-Xutl, e- Ruah, f- Ruah) hiPSC-CMs generated with the proposed differentiation approach and stained for α-actinin, cMYBPC3 or Troponin T (green) and F-actin (red). Nuclei are depicted in blue (DAPI staining).

To further characterize hiPSC-CMs morphology in a quantitative manner, several cell parameters were analyzed, namely shape descriptors (circularity and aspect ratio) and area. The percentage of binucleated cells was also analyzed ((Figure IV.10.A,B and C). Results show that HCM hiPSC-CMs are larger (area) and rounder than the normal ones (circularity and aspect ratio) (Figure IV.10.A, B and C). Moreover, the percentage of binucleated was found to be elevated in the HCM hiPSC-CMs population when compared with the WT ones (Figure IV.10.D).

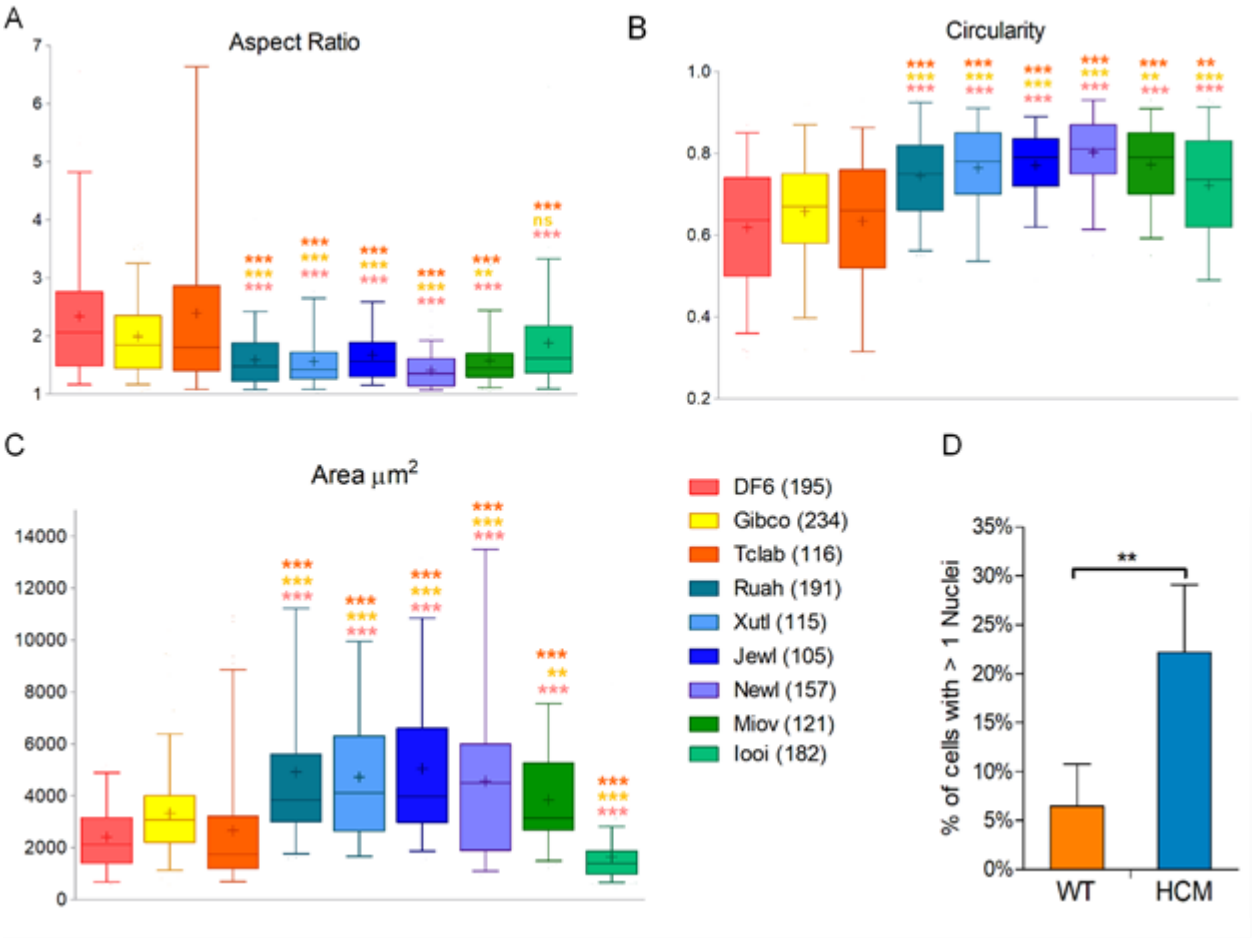


Figure IV.10. Quantification of morphologic parameters and percentage of binucleated cells in WT (DF6, Gibco, TCLab) and HCM (Ruah, Xutl, Jewl, Newl, Miov and Iooi) hiPSC-CMs using the 3D/2D approach. (A) Aspect Ratio. (B) Circularity. (C) Cell size, as Area. (D) Percentage of binucleated cells. Min. n=2 independent experiments (105-234 cells/cell line). Data represented as mean ± SD. ***p<0.001, **p<0.01.

IV.5. DISCUSSION

To date, in vitro differentiated hiPSC-CMs have not yet fully resembled the functional phenotype of an adult cardiomyocyte and it is arguable if it ever will in a practical laboratory time frame, since *in vivo* human cardiomyocytes undergo developmental changes even postnatally. However, that might not be completely necessary for hiPSC-CMs to function as a valid disease model provided they can recapitulate the studied disease phenotype well enough to answer the research questions (Pohjoismäki et al. 2013; Veerman et al. 2015).

This work describes the establishment of an optimized method for the generation of a pure and mature population of CMs from hPSCs that can be successfully used for HCM disease modeling. This approach combines the use of a 3D GiWi differentiation protocol with purification of hPSCs-derived CMs using VCAM1 staining for FACS and a subsequent period of culture in 2D conditions.

This is a fitting combination since during the first step of differentiation, 3D conditions potentiate a faster maturation, while being highly reproducible across cell lines, since cell density and number can be better controlled when compared with 2D cultures. In fact, this initial phase of differentiation, by mimicking the 3D configuration of the embryo, was described to provide a better mesoderm commitment and faster maturation of the CMs (Branco et al. 2019). Then, at day 12 of differentiation, when aggregates are dissociated, hiPSC-CMs are selected with high purity by FACS for VCAM1 and further replated at an appropriate density in Matrigel coated plates. We chose this selection strategy since our main concern was related with the purity of the obtained hiPSC-CMs, regardless of the scalability of the method, moreover the handling and replating of cells for some further analysis as already required. VCAM-1 has previously been identified as a CM marker with 95–98% of VCAM1+ cells being positive for TNNT2 (Uosaki et al. 2011; D. A. Elliott et al. 2011). Nevertheless, some studies described this surface marker as not exclusive for CMs as it could also be found in other cell types including smooth muscle and endothelial cells

(Ban, Bae, and Yoon 2017). So, in the future, a combined metabolic selection of the sorted CMs may be considered. Lately, hiPSC-CMs are further matured in monolayer culture for 18 more days after which they can be readily analyzed. Such extended period in culture (~4 weeks) proved to favor the development of additional structural features of mature, adult CMs such as cell elongation, higher degree of sarcomere organization and increased action potential amplitudes (Lundy et al. 2013; Bedada et al. 2014; Lewandowski et al. 2018). In summary, when compared with standard 2D cell cultures, our proposed 3D/2D protocol allows the generation of a more mature cardiomyocyte population promoted not only by the initial 3D cues but also by the subsequent extended period in monolayer culture. Supporting this, qRT-PCR and RNA-seq analysis of the splicing pattern of pre-mRNAs that code for sarcomeric proteins (namely titin and cardiac troponin T) that changes during cardiac development shows that our iPSC-CMs are expressing splicing isoforms characteristic of the adult heart. Moreover, FACs selection of cells expressing the VCAM1 cardiac surface marker leads to a much more homogeneous population of CMs, as demonstrated from the higher expression of several sarcomeric genes in 3D/2D VCAM1+ hiPSC-CMs when compared with the ones generated by 2D standard approaches or our approach without this purification step. Finally, in order to evaluate if our proposed 3D/2D protocol could be used for HCM disease modeling, WT and HCM hiPSC lines were differentiated under the same conditions until day 30 of differentiation, and further analyzed. Our results show increased cellular size, multinucleation, and disorganized sarcomeres in HCM iPSC-CMs as compared to WT iPSC-CMs, confirming that patient iPSC-CMs recapitulate HCM-specific features at the single-cell level.

Therefore, the use of a newly established combined approach to differentiate, purify and further mature hiPSC-CMs allowed us to obtain populations of cardiomyocytes that, after 30 days of differentiation, presented less heterogeneity and higher degree of maturation when compared to a standard monolayer protocol. Moreover, it was possible to differentiate six HCM patient derived hiPSC lines that, upon characterization of the generated HCM iPSC-CMs, demonstrate

increased cellular size, multinucleation, and disorganized sarcomeres as compared to normal hiPSC-CMs. Overall, these results confirm that patient hiPSC-CMs obtained through this approach recapitulate HCM-specific features at the single-cell level, which can further be used to study HCM genotype-phenotype relations, the impact of VUS or testing new therapeutic approaches.

IV.6. REFERENCES

Ban, Kiwon, Seongho Bae, and Young-sup Yoon. 2017. “Current Strategies and Challenges for Purification of Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Theranostics* 7 (7): 2067–77. https://doi.org/10.7150/thno.19427.

Bedada, Fikru B., Sunny S.K. Chan, Stefania K. Metzger, Liying Zhang, Jianyi Zhang, Daniel J. Garry, Timothy J. Kamp, Michael Kyba, and Joseph M. Metzger. 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. https://doi.org/10.1016/j.stemcr.2014.07.012.

Branco, Mariana A., João P. Cotovio, Carlos A.V. Rodrigues, Sandra H. Vaz, Tiago G. Fernandes, Leonilde M. Moreira, Joaquim M.S. Cabral, and Maria Margarida Diogo. 2019. “Transcriptomic Analysis of 3D Cardiac Differentiation of Human Induced Pluripotent Stem Cells Reveals Faster Cardiomyocyte Maturation Compared to 2D Culture.” *Scientific Reports* 9 (1): 1–13. https://doi.org/10.1038/s41598-019-45047-9.

Burridge, Paul W., Elena Matsa, Praveen Shukla, Ziliang C. Lin, Jared M. Churko, Antje D. Ebert, Feng Lan, et al. 2014. “Chemically Defined Generation of Human Cardiomyocytes.” *Nature Methods* 11 (8): 855–60. https://doi.org/10.1038/nmeth.2999.

Burridge, Paul W., Susan Thompson, Michal A. Millrod, Seth Weinberg, Xuan Yuan, Ann Peters, Vasiliki Mahairaki, Vassilis E. Koliatsos, Leslie Tung, and Elias T. Zambidis. 2011. “A Universal System for Highly Efficient Cardiac Differentiation of Human Induced Pluripotent Stem Cells That Eliminates Interline Variability.” Edited by Martin Pera. *PLoS ONE* 6 (4): e18293. https://doi.org/10.1371/journal.pone.0018293.

Correia, Cláudia, Alexey Koshkin, Patrícia Duarte, Dongjian Hu, Madalena Carido, Maria J. Sebastião, Patrícia Gomes-Alves, et al. 2018. “3D Aggregate Culture Improves Metabolic Maturation of Human Pluripotent Stem Cell Derived Cardiomyocytes.” *Biotechnology and Bioengineering* 115 (3): 630–44. https://doi.org/10.1002/bit.26504.

Elliott, David A., Stefan R. Braam, Katerina Koutsis, Elizabeth S. Ng, Robert Jenny, Ebba L. Lagerqvist, Christine Biben, et al. 2011. “NKX2-5 EGFP/w HESCs for Isolation of Human Cardiac Progenitors and Cardiomyocytes.” *Nature Methods* 8 (12): 1037–43. https://doi.org/10.1038/nmeth.1740.

Elliott, Perry M, Aris Anastasakis, Michael A. Borger, Martin Borggrefe, Franco Cecchi, Philippe Charron, Albert Alain Hagege, et al. 2014. “2014 ESC Guidelines on Diagnosis and Management of Hypertrophic Cardiomyopathy.” *European Heart Journal* 35 (39): 2733–79. https://doi.org/10.1093/eurheartj/ehu284.

Gruner, Christiane, Joan Ivanov, Melanie Care, Lynne Williams, Gil Moravsky, Hua Yang, Balint Laczay, Katherine Siminovitch, Anna Woo, and Harry Rakowski. 2013. “Toronto Hypertrophic Cardiomyopathy Genotype Score for Prediction of a Positive Genotype in Hypertrophic Cardiomyopathy.” *Circulation: Cardiovascular Genetics* 6 (1): 19–26. https://doi.org/10.1161/CIRCGENETICS.112.963363.

Hoes, Martijn F., Nils Bomer, and Peter van der Meer. 2019. “The Current State of Human In Vitro Cardiac Disease Modeling: A Focus on Gene Editing and Tissue Engineering.” *STEM CELLS Translational Medicine* 8 (1): 66–74. https://doi.org/10.1002/sctm.18-0052.

Junying, Yu, Hu Kejin, Smuga Otto Kim, Tian Shulan, Ron Stewart, Igor I. Slukvin, and James A. Thomson. 2009. “Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences.” *Science* 324 (5928): 797–801. https://doi.org/10.1126/science.1172482.

Karakikes, Ioannis, Mohamed Ameen, Vittavat Termglinchan, and Joseph C. Wu. 2015. “Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Insights Into Molecular, Cellular, and Functional Phenotypes.” *Circulation Research* 117 (1): 80–88.

https://doi.org/10.1161/CIRCRESAHA.117.305365.

Lewandowski, Jarosław, Natalia Rozwadowska, Tomasz J. Kolanowski, Agnieszka Malcher, Agnieszka Zimna, Anna Rugowska, Katarzyna Fiedorowicz, et al. 2018. “The Impact of in Vitro Cell Culture Duration on the Maturation of Human Cardiomyocytes Derived from Induced Pluripotent Stem Cells of Myogenic Origin.” *Cell Transplantation* 27 (7): 1047–67. https://doi.org/10.1177/0963689718779346.

Lian, Xiaojun, Jianhua Zhang, Samira M. Azarin, Kexian Zhu, Laurie B. Hazeltine, Xiaoping Bao, Cheston Hsiao, Timothy J. Kamp, and Sean P. Palecek. 2013. “Directed Cardiomyocyte Differentiation from Human Pluripotent Stem Cells by Modulating Wnt/β-Catenin Signaling under Fully Defined Conditions.” *Nature Protocols* 8 (1): 162–75. https://doi.org/10.1038/nprot.2012.150.

Lundy, Scott D., Wei Zhong Zhu, Michael Regnier, and Michael A. Laflamme. 2013. “Structural and Functional Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells.” *Stem Cells and Development* 22 (14): 1991–2002. https://doi.org/10.1089/scd.2012.0490.

Musunuru, Kiran, Farah Sheikh, Rajat M. Gupta, Steven R. Houser, Kevin O. Maher, David J. Milan, Andre Terzic, and Joseph C. Wu. 2018. “Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling and Precision Medicine: A Scientific Statement From the American Heart Association.” *Circulation. Genomic and Precision Medicine*. NLM (Medline). https://doi.org/10.1161/HCG.0000000000000043.

Pohjoismäki, Jaakko L.O., Marcus Krüger, Natalie Al-Furoukh, Anssi Lagerstedt, Pekka J. Karhunen, and Thomas Braun. 2013. “Postnatal Cardiomyocyte Growth and Mitochondrial Reorganization Cause Multiple Changes in the Proteome of Human Cardiomyocytes.” *Molecular BioSystems* 9 (6): 1210–19. https://doi.org/10.1039/c3mb25556e.

Purevjav, Enkhsaikhan. 2019. “Animal Models of Cardiomyopathies.” In *Animal Models in Medicine and Biology [Working Title]*, 58:233–39. IntechOpen. https://doi.org/10.5772/intechopen.89033.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, Shinya Yamanaka, et al. 2007. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.” *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Uosaki, Hideki, Hiroyuki Fukushima, Ayako Takeuchi, Satoshi Matsuoka, Norio Nakatsuji, Shinya Yamanaka, and Jun K. Yamashita. 2011. “Efficient and Scalable Purification of Cardiomyocytes from Human Embryonic and Induced Pluripotent Stem Cells by VCAM1 Surface Expression.” *PLoS ONE* 6 (8): e23657. https://doi.org/10.1371/journal.pone.0023657.

Veerman, Christiaan C., Georgios Kosmidis, Christine L. Mummery, Simona Casini, Arie O. Verkerk, and Milena Bellin. 2015. “Immaturity of Human Stem-Cell-Derived Cardiomyocytes in Culture: Fatal Flaw or Soluble Problem?” *Stem Cells and Development*. Mary Ann Liebert Inc. https://doi.org/10.1089/scd.2014.0533.

World Health Organization. 2018. “The Top 10 Causes of Death.” 2018. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Yoshida, Yoshinori, and Shinya Yamanaka. 2017. “Induced Pluripotent Stem Cells 10 Years Later.” *Circulation Research* 120 (12): 1958–68. https://doi.org/10.1161/CIRCRESAHA.117.311080.

V. CONCLUDING REMARKS

During this project, we aimed to generate a new bioengineering approach to model HCM.

Being the most common hereditary disease of the heart, HCM often leads to sudden cardiac death, particularly in young athletes (Maron, Rowin, and Maron 2018; Watkins, Ashrafian, and Redwood 2011). It is a complex disease characterized by left ventricular hypertrophy, abnormal diastolic function, increase in myocyte size with distorted nuclei, myocyte disarray and increased extracellular fibrosis (Christiaans and Elliott 2016; Watkins, Ashrafian, and Redwood 2011). While clinical diagnosis is primarily based on left ventricular hypertrophy, the anatomic hallmark of the disease, many individuals die without ever being diagnosed (Semsarian et al. 2015; Elliott et al. 2014). Moreover, in more than 50% of the HCM clinical cases, standard genetic testing fails to identify a causative mutation (Elliott et al. 2014; Gruner et al. 2013). Later advances in next-generation sequencing have revolutionized molecular diagnostics for many diseases, including HCM, by allowing human genome analysis in a cost-effective way. However, these approaches have also reveal a vast number of genetic variants of uncertain clinical significance (Hoffman-Andrews 2017), particularly in intronic regions. Therefore, our first aim was to address the potential pathogenicity of missense and splice site variants, taking advantage of already available *in silic*o prediction tools, as described in Chapter II of this thesis. However, the choice of which tools to use is not straightforward, as different computational tools perform differently depending on the genomic context. Our analysis demonstrated that the combined use of several of the assessed tools may be needed to correctly assign variant pathogenicity, mainly depending on the variant type (missense or splice-site). Additionally, new approaches for the evaluation of deep intronic variants need to be developed, with splicing being an overlooked mechanism in what concerns to disease pathophysiology, as we have reviewed.

Despite the usefulness of *in silico* prediction tools, the existence of mechanistic and functional studies is important to establish variants pathogenicity. In HCM, such studies have been largely restricted to animal models, which do not recapitulate the physiological and mechanistic characteristics of the human heart. The recent emergence of hiPSCs with their unlimited

expansion potential and capability to be differentiated into CMs that can recapitulate the HCM phenotype at single-cell level, hold great promise for disease modeling (Takahashi et al. 2007; Yoshida and Yamanaka 2017; Lan et al. 2013). In fact, since their discovery, hiPSCs have been widely used for the study of inherited cardiac diseases, with over 90 studies using hiPSC-CMs for cardiac disease modeling being published until date (van Mil et al. 2018). Despite providing a basis for the possibility of modeling HCM *in vitro* and important insight regarding mechanisms involved in the pathophysiology of the disease, the significant differences amongst these studies specially regarding the origin of the cells, methods used for differentiation and parameters analyzed confound their interpretation and comparison (Ma et al. 2018; Musunuru et al. 2018). Furthermore, the frequent lack of adequate isogenic controls in most of these studies difficult the establishment of reliable genotype-phenotype correlations. To overcome this issue, we used the CRISPR/Cas technology to gene edit hiPSCs, enabling the creation of isogenic controls and trustworthy disease models to study the impact of a well-described HCM associated pathogenic splice variant. But, the precise insertion of a given single nucleotide variant into the genome of human iPSCs has proved to be particularly challenging, as discussed on Chapter IV of this thesis. In fact, even when state-of-art approaches are used, the editing efficiencies remain low and successfully genotyping of the edited cell clones while still maintaining their pluripotency is a work-intensive task. In the future, the optimization of CRISPR screening and genotyping protocols may allow the generation of personalized HCM disease models, either by the insertion of study variants in a WT cell lines or the correction of PD-hiPSCs, where the contribution of a given variant to the establishment of the HCM phenotype can be properly addressed.

Another of the biggest concerns around the use of hiPSCs in cardiac disease modeling is the heterogeneity and immature characteristics of iPSC-CMs obtained from established differentiation protocols which can compromise the reliability of such models (Ma et al. 2018; Wang et al. 2018; Mosqueira et al. 2018; Musunuru et al. 2018). In Chapter V of this thesis, we describe an optimized strategy that uses a combined 3D/2D GiWi differentiation protocol, where

VCAM1 positive hiPSC-CMs are purified from 3D aggregates by FACS at day 12 of differentiation and further maturated in a 2D monolayer culture until day 30. The generated hiPSC-CMs represented homogeneous populations of cardiomyocytes with a high degree of maturation when compared to a standard monolayer based protocol. When this 3D/2D approach was used to differentiate HCM patient derived hiPSCs, the originated hiPSC-CMs could recapitulate HCM-specific features at the single-cell level, demonstrated by the increased cellular size, multinucleation, and disorganized sarcomeres in HCM hiPSC-CMs as compared to normal ones. In the future we intend to further characterize and confirm the ability of these hiPSC-CMs to recapitulate the HCM phenotype using image analysis of light microscopy time-lapse videos to quantify changes in contractility. And, if this is insufficient to reveal differences between normal and patient-derived iPSC-CMs, we will try a hydrogel-based engineered heart tissue protocol that has been shown to replicated canonical response to different physiological and pharmacological regulators in a systematic contractility analysis (Mannhardt et al. 2016). By developing a highly reproducible strategy for generating an homogeneous population of cardiomyocytes that can recapitulate the HCM phenotype at the cellular level, we expected to contribute with a valuable approach that can be further used in the establishment of reliable HCM genotype-phenotype correlations and test of new targeted therapies.

Overall, the work presented in this thesis represents a contribution to the advancement of HCM disease modeling using hiPSC technology. In a near future, integrating all steps of this work would allow us to use accurate *in silico* prediction tools to evaluate WGS of HCM patients with no pathogenic variant identified and choose a VUS predicted with high confidence to be disease causing. Produce matched hiPSC lines harboring the chosen variant, either patient derived with isogenic controls through its correction by CRISPR, or through its introduction in a normal hiPSC line. And use these hiPSCs to, upon differentiation with the previously proposed 3D/2D approach, obtain homogenous and mature populations of hiPSC-CMs to access the impact of the chosen

variant, infer on its pathogenicity, study associated molecular mechanisms and eventually test new therapies.

V.1. REFERENCES

Christiaans, Imke, and Perry M. Elliott. 2016. “Hypertrophic Cardiomyopathy.” In *Clinical Cardiogenetics*, 61–74. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44203-7_4.

Elliott, Perry M, Aris Anastasakis, Michael A. Borger, Martin Borggrefe, Franco Cecchi, Philippe Charron, Albert Alain Hagege, et al. 2014. “2014 ESC Guidelines on Diagnosis and Management of Hypertrophic Cardiomyopathy.” *European Heart Journal* 35 (39): 2733–79. https://doi.org/10.1093/eurheartj/ehu284.

Gruner, Christiane, Joan Ivanov, Melanie Care, Lynne Williams, Gil Moravsky, Hua Yang, Balint Laczay, Katherine Siminovitch, Anna Woo, and Harry Rakowski. 2013. “Toronto Hypertrophic Cardiomyopathy Genotype Score for Prediction of a Positive Genotype in Hypertrophic Cardiomyopathy.” *Circulation: Cardiovascular Genetics* 6 (1): 19–26. https://doi.org/10.1161/CIRCGENETICS.112.963363.

Hoffman-Andrews, Lily. 2017. “The Known Unknown: The Challenges of Genetic Variants of Uncertain Significance in Clinical Practice.” *Journal of Law and the Biosciences* 4 (3): 648–57. https://doi.org/10.1093/jlb/lsx038.

Lan, Feng, Andrew S. Lee, Ping Liang, Veronica Sanchez-Freire, Patricia K. Nguyen, Li Wang, Leng Han, et al. 2013. “Abnormal Calcium Handling Properties Underlie Familial Hypertrophic Cardiomyopathy Pathology in Patient-Specific Induced Pluripotent Stem Cells.” *Cell Stem Cell* 12 (1): 101–13. https://doi.org/10.1016/j.stem.2012.10.010.

Ma, Ning, Joe Z. Zhang, Ilanit Itzhaki, Sophia L. Zhang, Haodong Chen, Francois Haddad, Tomoya Kitani, et al. 2018. “Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells.” *Circulation* 138 (23): 2666–81. https://doi.org/10.1161/CIRCULATIONAHA.117.032273.

Mannhardt, Ingra, Kaja Breckwoldt, David Letuffe-Brenière, Sebastian Schaaf, Herbert Schulz, Christiane Neuber, Anika Benzin, et al. 2016. “Human Engineered Heart Tissue: Analysis of Contractile Force.” *Stem Cell Reports* 7 (1): 29–42. https://doi.org/10.1016/j.stemcr.2016.04.011.

Maron, Barry J., Ethan J. Rowin, and Martin S. Maron. 2018. “Global Burden of Hypertrophic Cardiomyopathy.” *JACC: Heart Failure* 6 (5): 376–78. https://doi.org/10.1016/j.jchf.2018.03.004.

Mil, Alain van, Geerthe Margriet Balk, Klaus Neef, Jan Willem Buikema, Folkert W. Asselbergs, Sean M. Wu, Pieter A. Doevendans, and Joost P.G. Sluijter. 2018. “Modelling Inherited Cardiac Disease Using Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Progress, Pitfalls, and Potential.” *Cardiovascular Research* 114 (14): 1828–42. https://doi.org/10.1093/cvr/cvy208.

Mosqueira, Diogo, Ingra Mannhardt, Jamie R. Bhagwan, Katarzyna Lis-Slimak, Puspita Katili, Elizabeth Scott, Mustafa Hassan, et al. 2018. “CRISPR/Cas9 Editing in Human Pluripotent Stem Cell-Cardiomyocytes Highlights Arrhythmias, Hypocontractility, and Energy Depletion as Potential Therapeutic Targets for Hypertrophic Cardiomyopathy.” *European Heart Journal* 39 (43): 3879–92. https://doi.org/10.1093/eurheartj/ehy249.

Musunuru, Kiran, Farah Sheikh, Rajat M. Gupta, Steven R. Houser, Kevin O. Maher, David J. Milan, Andre Terzic, and Joseph C. Wu. 2018. “Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling and Precision Medicine: A Scientific Statement From the American Heart Association.” *Circulation. Genomic and Precision Medicine*. NLM (Medline). https://doi.org/10.1161/HCG.0000000000000043.

Semsarian, Christopher, Jodie Ingles, Martin S. Maron, and Barry J. Maron. 2015. “New Perspectives on the Prevalence of Hypertrophic Cardiomyopathy.” *Journal of the American College of Cardiology*. Elsevier USA. https://doi.org/10.1016/j.jacc.2015.01.019.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda,

Shinya Yamanaka, et al. 2007. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.” *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Wang, Lili, Kyungsoo Kim, Shan Parikh, Adrian Gabriel Cadar, Kevin R. Bersell, Huan He, Jose R. Pinto, Dmytro O. Kryshtal, and Bjorn C. Knollmann. 2018. “Hypertrophic Cardiomyopathy-Linked Mutation in Troponin T Causes Myofibrillar Disarray and pro-Arrhythmic Action Potential Changes in Human IPSC Cardiomyocytes.” *Journal of Molecular and Cellular Cardiology* 114 (January): 320–27. https://doi.org/10.1016/j.yjmcc.2017.12.002.

Watkins, Hugh, Houman Ashrafian, and Charles Redwood. 2011. “Inherited Cardiomyopathies.” Edited by Robert S. Schwartz. *New England Journal of Medicine* 364 (17): 1643–56. https://doi.org/10.1056/NEJMra0902923.

Yoshida, Yoshinori, and Shinya Yamanaka. 2017. “Induced Pluripotent Stem Cells 10 Years Later.” *Circulation Research* 120 (12): 1958–68. https://doi.org/10.1161/CIRCRESAHA.117.311080.