Marta Isabel Brandão Furtado

BSc in Biochemistry

Developing gene-edited models to study mis-splicing in Hypertrophic Cardiomyopathy

Dissertation to obtain a Master's Degree in
Molecular Genetics and Biomedicine

Supervisor: Professor Sandra Cristina Bento Penisga Martins

Faculdade de Medicina, Instituto de Medicina Molecular, Universidade de Lisboa

Co-supervisor: Professor Maria Teresa Tenório Figueiredo Carvalho Gonçalves

Faculdade de Medicina, Instituto de Medicina Molecular, Universidade de Lisboa

Jury Constitution

President: Maria Alexandra Núncio de Carvalho Ramos Fernandes

Vogals: Vasco Temudo e Melo Cabral Barreto

Sandra Cristina Bento Penisga Martins

September 2019

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Marta Isabel Brandão Furtado

BSc in Biochemistry

Developing gene-edited models to study mis-splicing in Hypertrophic Cardiomyopathy

Dissertation to obtain a Master's Degree in
Molecular Genetics and Biomedicine

Supervisor: Professor Sandra Cristina Bento Penisga Martins

Faculdade de Medicina, Instituto de Medicina Molecular, Universidade de Lisboa

Co-supervisor: Professor Maria Teresa Tenório Figueiredo Carvalho Gonçalves

Faculdade de Medicina, Instituto de Medicina Molecular, Universidade de Lisboa

Jury Constitution

President: Maria Alexandra Núncio de Carvalho Ramos Fernandes

Vogals: Vasco Temudo e Melo Cabral Barreto

Sandra Cristina Bento Penisga Martins

September 2019

Developing gene-edited models to study mis-splicing in Hypertrophic Cardiomyopathy

Acknowledgements

First and foremost, I want to thank Maria do Carmo-Fonseca for giving me the opportunity to develop this work. I'm certain that if I had done my thesis anywhere else, I wouldn't know half of what I know today and wouldn't be facing the opportunities that I am now.

To Sandra Martins and Teresa Carvalho, my incredible supervisors, thank you for all the support, patience and ultimately, for teaching me everything I know. I couldn't have felt more guided and accompanied throughout this process than with the two of you. To Marta Ribeiro, thank you for being my companion in the lab, for putting up with me and always be willing to help. And finally, to every single member of the group that made this journey a bit more joyful and fun.

Importantly, a huge thank you to my family and friends, specially my parents, who supported all my ups and downs this year and listened to all my complaints about hiPSCs and how CRISPR/Cas9 does not work (it does!).

Funding

This work was supported by:

Fundação para a Ciência e a Tecnologia and FEDER through POR Lisboa 2020 - Programa Operacional Regional de Lisboa PORTUGAL 2020 (PAC-PRECISE-LISBOA-01-0145-FEDER-016394); Genomed-SI IDT 24136 (MyCardioGeneTest) and The Discoveries Centre for Regenerative and Precision Medicine H2020-WIDESPREAD-01-2016-2017.

Abstract

Exonic variants that cause abnormal mRNA splicing have been reported in human disease. However, this phenomenon is relatively unstudied in Hypertrophic Cardiomyopathy, a complex genetic cardiac disease with more than 1500 pathogenic mutations identified in over 14 genes. The lack of suitable models to study the causal mechanism of HCM-associated variants poses a drawback to the understanding of the molecular basis underlying this disease. The recent emergence of CRISPR/Cas9 and hiPSCs technologies are an exciting approach to generate disease models that better recapitulate the characteristics of the human heart.

Three gene-edited clones harbouring the *MYBPC3* c.1090G>A variant were generated using the CRISPR/Cas9 system. This variant is predicted to disrupt the recognition of the donor splice-site and lead to the skipping of exon 12. However, throughout the gene-editing process and due to long time in culture, these cells lost expression of pluripotency markers and could not be differentiated into cardiomyocytes to perform further experiments.

In parallel, a patient derived HCM cell line (Xutl, p.I1250fs) and Gibco and TCLab control lines were differentiated into cardiomyocytes using a 2D protocol. The hiPSCs-CMs obtained expressed high levels of fetal sarcomeric-genes *TNNT2* and *TTN* isoforms, when compared to the adult human heart. Morphologically, they also displayed characteristics that are compatible to those of embryonic cardiomyocytes. Therefore, these hiPSCs-CMs have an immature phenotype and could not fully recapitulate the characteristics of adult cardiomyocytes. For disease-modelling purposes, mature hiPSCs-CMs that are comparable to those found in the adult heart are required and 3D differentiation protocols are some of the approaches being implemented.

Even though optimization of the gene-editing and cardiac differentiation protocols is essential for the success of subsequent experiments, the work performed in this thesis provided important advances in the development of improved cellular models to study complex genetic diseases, such as HCM.

Keywords: cardiac differentiation; CRISPR/Cas9; human induced pluripotent stem cells; hypertrophic cardiomyopathy; splicing.

Sumário

A Cardiomiopatia Hipertrófica (HCM) é uma doença genética complexa, na qual estão identificadas como patogénicas mais de 1500 mutações em genes sarcoméricos. No entanto, a falta de modelos celulares adequados, que reproduzam as características de cardiomiócitos humanos, constitui uma limitação ao estudo do mecanismo molecular que conduz ao desenvolvimento desta patologia. Variantes exónicas que causam o *splicing* anormal do mRNA já foram descritas como estando associadas a doenças humanas, embora este seja um fenómeno pouco estudado em HCM. Neste trabalho foi utilizado o sistema de edição de genoma CRISPR/Cas9 e a diferenciação de células estaminais pluripotentes induzidas (hiPSCs) em cardiomiócitos (hiPSCs-CMs), de forma a gerar modelos celulares portadores de variantes patogénicas específicas.

A variante *MYBPC3* c.1090G>A foi introduzida em hiPSCs utilizando a técnica de CRISPR/Cas9. No entanto, os clones de hiPSCs editados revelaram-se instáveis, perdendo a expressão de genes de pluripotência ao longo do processo de edição e não foi possível diferenciá-los em CMs de forma a realizar as experiências subsequentes previstas.

Em paralelo, uma linha de hiPSCs derivada de um doente com HCM (Xutl, pl1250fs) e as linhas controlo Gibco e TCLab foram diferenciadas em CMs através de um protocolo 2D, adaptado de um previamente descrito na literatura. Os hiPSCs-CMs obtidos têm uma morfologia arredondada e sarcómeros com um reduzido grau de organização. A análise da expressão de genes sarcoméricos revela um elevado nível de isoformas fetais, particularmente dos genes *TNNT2* e *TTN*, em comparação com os níveis de expressão no coração adulto, características compatíveis com as de CMs num estadio embrionário.

O trabalho desenvolvido nesta tese forneceu importantes avanços para o desenvolvimento de modelos celulares para o estudo de doenças genéticas complexas, embora a otimização dos protocolos de edição e de diferenciação cardíaca seja necessária para o sucesso de experiências futuras.

Palavras-chave: Cardiomiopatia hipertrófica; células estaminais pluripotentes induzidas humanas; CRISPR/Cas9; diferenciação cardíaca; *splicing*.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACTC1	Cardiac muscle α-actin
ACTN2	α-Actinin
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
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT3β	DNA methyltransferase 3β
DSB	Double-strand breaks
EBNA-1	Epstein–Barr nuclear antigen-1
EBV	Epstein–Barr virus
EDTA	Ethylenediamine tetraacetic acid
EJC	Exon junction complex
ESC	Embryonic stem cells
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
EtBr	Ethidium Bromide
EtOH	Ethanol
ExAc	Exome Aggregation Consortium
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GSK3β	Glycogen synthase kinase-3β
HCM	Hypertrophic Cardiomyopathy
hCol1a1	Collagen type I α-1
HDR	Homologous Directed Recombination
HGMD	Human Gene Mutation Database
hnRNP	Heterogeneous nuclear ribonucleoproteins
HSF	Human Splicing Finder
HuF	Human Fibroblasts
iPSCs	Induced pluripotent stem cells
iPSC-CM	Induced pluripotent stem cell-derived cardiomyocytes
ISE	Intronic splicing enhancer
ISL1	Insulin gene enhancer protein
ISS	Intronic splicing silencer
IVS	Intraventricular septum
Klf4	Krüppel-like factor 4
LB	Lysogeny broth
LiCl	Lithium chloride
LV	Left ventricle
MEM	Minimum Essential Medium
MYBPC3	Myosin-binding protein C 3
MYH7	β-myosin heavy chain 7
MYL3	Myosin light chain 3
NaCl	Sodium chloride
NHEJ	Non-Homologous End Joining
NMD	Nonsense-mediated decay
ORF	Open reading frame

PAM	Protospacer adjacent motif
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PTC	Premature stop codon
RBPs	RNA binding proteins
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
ROCK	Rho-associated coiled-coil protein kinase
SCD	Sudden cardiac death
SNP	Single nucleotide polymorphisms
sgRNA	Small guide RNA
snRNA	Small nuclear RNA
Sox2	Sex determining region Y-box 2
SSEA3/4	Stage-specific embryonic antigens
ssODN	Single-stranded oligonucleotide
TALENs	Transcription activator-like effector nucleases
TGFβ	Transforming growth factor-β
TNNI1	Cardiac Troponin I 1
TNNI3	Cardiac Troponin I 3
TNNT2	Cardiac 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

1. Introduction

1.1 Hypertrophic Cardiomyopathy

The first comprehensive clinical description of Hypertrophic Cardiomyopathy (HCM) was published in 1964 by Braunwald, *et al.*, [1] where it was initially designated as idiopathic hypertrophic subaortic stenosis. Nowadays, over 50 years after its original report, HCM is known to be the most common genetic heart disease, with a prevalence of 1:200 in the general population and an estimated 20 million people affected worldwide [2]. It is the leading cause of sudden cardiac death (SCD) in young athletes, in which the physiological thickening of the left-ventricular wall due to high physical activity may overlap with the HCM phenotype, supporting the need for a differential diagnosis [3,4]. HCM is a global disease, and its distribution is showcased in **Figure 1.1** [2].

Figure 1.1 – Global distribution of HCM (in red). This estimation includes data from 122 countries, accounting for 88% of the world population. Adapted from [2].

Even though HCM is the leading cause of SCD amongst young people, the majority of individuals affected by this condition are thought to still remain undiagnosed, given that most HCM patients are asymptomatic [3]. Up to 25% of HCM-affected individuals develop severe health complications such as arrhythmic sudden death, end-stage heart failure and atrial fibrillation with embolic stroke [3,5]. The incidence of HCM-related SCD varies from 0.5 – 2% amongst patients [6]. Mild manifestations of the disease include symptoms such as chest pain, dyspnea, dizziness, heart palpitations, fatigue and syncope, often occurring when patients are exposed to stress factors [7,4].

Despite its variable phenotypic characteristics, the HCM hallmark is the distinct hypertrophy of the left ventricle (LV), involving primarily the left ventricular outflow tract and the interventricular septum (IVS) (**Figure 1.2)** [8], when unrelated to a set of possible other cardiac, metabolic or even syndromic diseases, such as systemic hypertension or aortic stenosis [1,3,9].

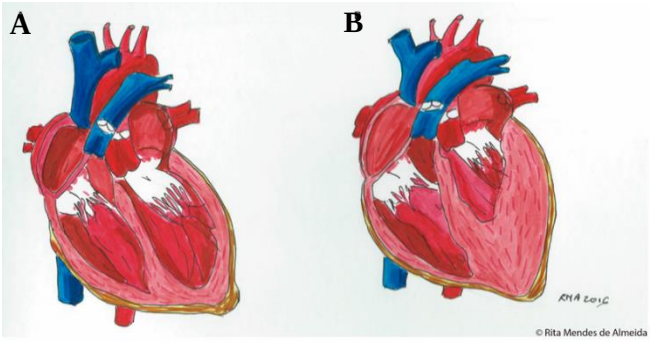


Figure 1.2 - Schematic representation of a **(A)** normal and **(B)** a heart with HCM, showing marked hypertrophy of the left ventricle wall. Adapted from [8].

Such hypertrophy of the LV leads to a decrease of the left ventricular chamber size, resulting in left ventricle outflow obstruction [3,4].

The heterogeneous phenotype and complex clinical profile of the disease poses a problem to its diagnosis, which is often only possible during adulthood, when there is a marked progression of the morphological and functional deficits of the heart [5]. HCM diagnosis is primarily based on a left ventricular end diastolic wall thickness >15 mm, determined by an echocardiogram or other imaging techniques, such as computed tomography or cardiac magnetic resonance, in the absence of a known secondary cause [10,7].

HCM cardiomyocytes are typically enlarged and the tissue presents myocyte disarray, lacking the normal parallel alignment and organization of the muscle cells. Also at the cellular level, cardiomyocytes may be separated by areas of interstitial fibrosis [11]. The histological phenotypes of HCM are present in **Figure 1.3**, where the disorganized myocyte structure and fibrosis is evident [11].

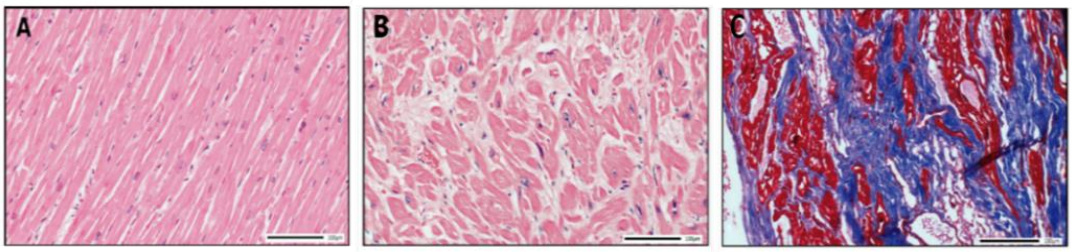


Figure 1.3 – Histological phenotypes of HCM. **(A)** A normal myocardial section stained with hematoxylin (20x magnification). **(B)** Hematoxylin-stained myocardial section from an HCM patient (20x). **(C)** Myocardial section stained with Masson trichrome (in blue*)* showing areas of interstitial fibrosis (20*x*). Adapted from [11].

1.1.1 Genetics of Hypertrophic Cardiomyopathy

HCM is a genetic disease with a Mendelian autosomal dominant pattern of inheritance, in which the offspring has a 50% probability of inheriting the mutation and is at risk of developing the disease. HCM shows variable penetrance and expression. While most cases consist of the familial type of the disease, sporadic cases of *de novo* mutations also occur. Rare cases of X-linked and autosomal recessive modes of inheritance, have also been reported [11,12].

Even though the variable phenotype and clinical outcomes amongst HCM patients may be associated to different causative mutations, the heterogeneity between carriers of the same mutation or within the same family suggests that the genomic context also plays an important role. Non-genetic factors, such as hypertension and obesity are reportedly associated to a more severe phenotype [11].

Over 1500 mutations have been identified in at least 14 genes that encode for sarcomeric proteins, accounting for over 95% of the reported HCM mutations [3,4,13].

Sarcomeres are the basic contractile unit of the cardiac muscle and are organized in a repeated pattern along the cells. They are formed by thick and thin filaments composed of myosin and actin, respectively, the motor proteins responsible for generating the force for muscle contraction. The thin filaments are also formed by regulatory proteins such as cardiac troponin C, I and T, tropomyosin and myosin-binding protein C, which are involved in the maintenance of the sarcomere architecture and its regulation [5, 12]. In **Figure 1.4**, a schematic representation of the sarcomere and the location of each protein in the structure, along with the correspondent percentages of HCM-related mutations identified in the different genes is shown. Z-discs, also represented in **Figure 1.4**, are lateral boundaries of the sarcomere that play a fundamental role in mechanical stretch sensing. The Z-line protein α-actinin (*ACTN2*) links the thin filaments of the sarcomere on both sides of the Z-line, in order to connect the adjacent sarcomeres [12].

Mutations in Z-disc genes and in genes that encode proteins involved in calcium regulation are less frequent, comprising only about 5% of the HCM-related variants identified so far [6].

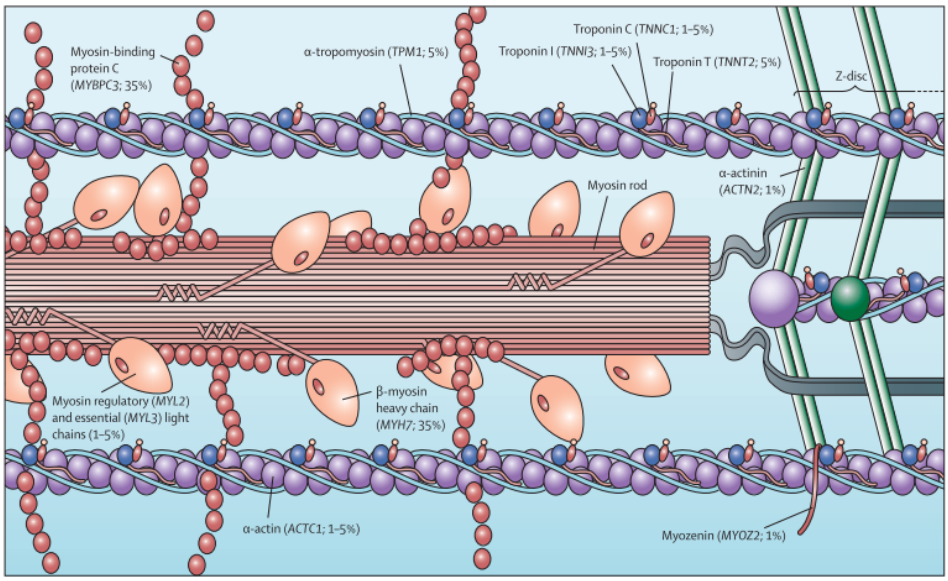


Figure 1.4 – Molecular structure of the human sarcomere displaying the localization of cardiac sarcomere proteins that have been shown to be involved in the pathogenesis of hypertrophic cardiomyopathy. Adapted from [3].

Mutations in the β-myosin heavy chain (*MYH7*) and myosin-binding protein C (*MYBPC3*) genes account for over 70% of the identified mutations in HCM patients. Cardiac troponin T (*TNNT2*), cardiac troponin I (*TNNI3*), tropomyosin (*TPM1*), myosin light chain (*MYL3*) and α-actin (ACTC1) genes have a lower frequency of pathogenic variants identified, ranging from 1-5% each [3]. The genes implicated in HCM and the correspondent percentage of pathogenic variants identified per gene are summarized in **Table 1.1**.

90% of all pathogenic variants associated to HCM are classified as missense. Since these single nucleotide variants lead to changes in the encoded aminoacid, they may alter the structure and function of the mutated protein. Several studies have demonstrated that heterozygous missense variants in sarcomeric genes often lead to the formation of poison polypeptides that are incorporated into the sarcomere, along with the normal protein transcribed from the wild-type allele, acting in a negative dominant manner. Overall, incorporation of these mutated proteins may have consequences for sarcomeres' structure and function [3,14].

Even though missense mutations are the most commonly found alterations in other sarcomeric genes, *MYBPC3* seems to be an exception, with approximately 70% of the mutations leading to frameshifts events. The insertion/deletion events that introduce a frameshift and consequently generate a premature stop codon (PTC) are often responsible for the production of transcripts that are targeted for degradation by the nonsense-mediated decay (NMD) mechanism. Hence, in contrast with what has been primarily characterized in the remaining sarcomeric genes, where mutations reportedly have a dominant negative effect, in the *MYBCP3* gene most variants are thought to lead to the expression of truncated proteins that are degraded, leading to haploinsufficiency [11,14].

Table 1.1 – Mutation frequency in the genes implicated in HCM. Adapted from [6].

Gene	Protein	Frequency in HCM (%)
Thick Filament		
MYH7	B-myosin heavy chain	25-35
MYL2	Regulatory myosin light chain	<5
MYL3	Essential myosin light chain	1
MYH6	α-myosin heavy chain	<1
TTN	Titin	<1
Thin Filament		
TNNT2	Cardiac troponin T	3-5
TNNI3	Cardiac troponin I	<5
TNNC1	Cardiac troponin C	<1
TPM1	α-tropomyosin	<5
ACTC	α-cardiac actin	<1
Intermediate Filament		
MYBPC3	Cardiac myosin-binding protein C	20-30
Z-disk		
ACTN2	α-actinin 2	<1
MYOZ2	Myozenin 2	<1
CSRP3	Muscle LIM protein	<1
TCAP	Telethonin	<1
VCL	Vinculin	<1
Calcium handling		
CASQ2	Calsequestrin	<1
JPH2	Junctiphilin 2	<1

Additionally, it has been reported that HCM patients bearing *MYBPC3* mutations that cause protein truncations have a more severe phenotype when compared to patients with missense and in-frame deletions. Compound heterozygous cases of HCM have also been identified and associated with a more critical state [4]. However, given the heterogeneous and complex phenotypes and clinical outcomes of HCM, a genotype–phenotype correlation has not been well established yet. This lack of consistency in genotype-phenotype association poses a problem for risk evaluation and preventive care of HCM patients [9]. Nevertheless, the pathogenicity of a variant will depend on which gene involved in the sarcomere structure or Ca^{2+} flux 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 [5].

To date, the available treatments aim primarily for the reduction of the symptoms of the disease, according to their severity and risk for SCD. Regarding pharmacological approaches, the beta-adrenergic receptor blockers are the most commonly used, followed by calcium channel blockers and disopyramide as an antiarrhythmic agent. In more severe cases, surgical strategies are also available, including implantable defibrillators, in order to prevent sudden death, and surgical myectomy or alcohol septal ablation, for the reduction of the left ventricle outflow obstruction [7,11].

1.2 Splicing in Heart Disease

RNA splicing is the process responsible for the removal of introns from immature pre-mRNAs and the linkage of exons in order to give rise to the mature form of mRNA [15]. This process is catalysed by a dynamic ribonucleoprotein complex (RNP) called the spliceosome. Over 100 core proteins and five small nuclear RNAs (the U1, U2, U4, U5, and U6 snRNAs) are associated to the spliceosome [16,17].

Throughout the gene, there are splicing sequences that are recognized by snRNAs of the spliceosome and that are essential for the binding of the RNP and its catalytic activity. In the pre-mRNA, three elements constitute the core splicing signal: the 5' and the 3' splice site, which define the exon-intron boundary, and the branch point sequence. The consensus 5' and 3' splice site motifs, GT and AG, respectively, are present in 98.7% of genes. The 3' splice site also contains the polypyrimidine tract [16,17,18].

The spliceosome assembly begins with the recognition of the 5' (donor) splice site by the U1 snRNP (U1 snRNA+ associated proteins). Simultaneously, the splicing factor 1 (SF1) binds to the branch point region, and the U2AF65 protein binds to the polypyrimidine tract and the 3' (acceptor) splice site, forming the E complex (early complex). The U2 snRNP is then recruited to the branch point and is responsible for the displacement of SF1. At this point, the A complex (ATP-dependent) is formed. The RNA helicases Prp5 and Sub2 stabilize the U2-branch point interaction, signalling for the recruitment of the U4, 5 and 6 tri-snRNP and the formation of the B complex, also called the pre-catalytic spliceosome. The catalytic active B complex is formed through a sequence of conformational changes, leading to the release of U1, U4 snRNPs and the interaction between U6-U2, and the formation of the C complex. The 5' splice site is cleaved, and the pre-mRNA loop (lariat) is formed, followed by the cleavage of the 3' splice site and consequent intron removal and ligation of the exons together [18,19,20]. In **Figure 1.5** an overview of the splicing process is depicted.

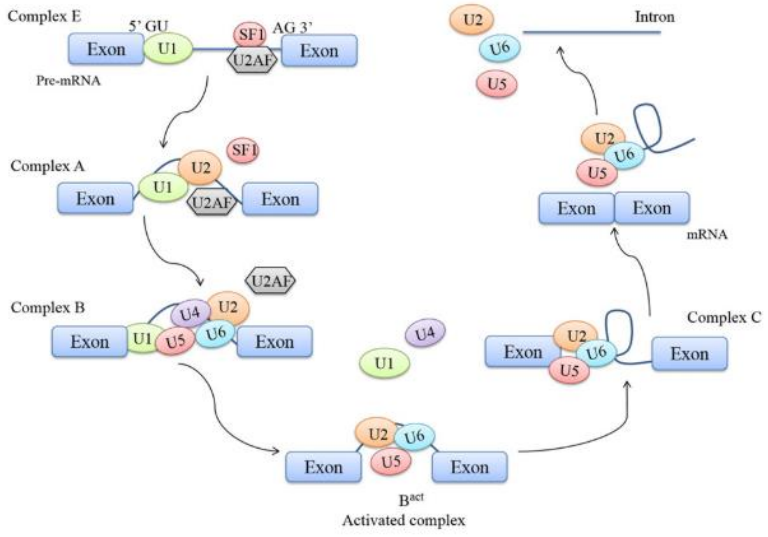


Figure 1.5 – The splicing process. Splicing of an intron depends on the recognition of the splice sites at intron-exon junctions. Four complexes between the pre-mRNA and the spliceosome machinery are formed in order to remove introns from the immature mRNA in the lariat form, which are subsequently degraded. Adapted from [18].

Other motifs similar to the splice sites may be recognized in detriment of the canonical ones, depending on its strength. These sequences are denominated as noncanonical or cryptic splice sites [18].

1.2.1 Alternative Splicing

Constitutive splicing is the default pattern of intron removal from the pre-mRNA. However, over 90% of human genes are known to express several different mRNAs through the process of alternative splicing, which results in proteomic diversity [20]. Since alternative spliced 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. Hence, the use of alternative splice sites allows for the production of different RNA transcripts and proteins from the same gene [16,17,20].

The regulation of alternative splicing is mediated by *cis*-regulatory sequences that can be located in exons or neighbouring introns. These sequences promote the recruitment of splicing regulatory factors that can either facilitate or prevent spliceosome assembly to a specific splice site. If these sequences promote the inclusion of the exon in the mRNA, they are called exonic splicing enhancer, ESE or intronic splicing enhancers, ISE. If they prevent spliceosome assembly and lead to the exclusion of the exon they are exonic splicing silencers, ESS or intronic splicing silencers, ISS [16,20]. Therefore, the recognition of the exon-intron boundaries is achieved through the presence of multiple signals across the pre-mRNA, that in combination with each other define the splice sites to be used at a given time [21].

Trans-regulatory splicing factors are recruited by the *cis*-regulatory sequences and are responsible for the regulation of the splicing process. Serine and Arginine-rich proteins (SR proteins) bind to the splicing enhancers sequences, promoting the recruitment of the spliceosome. In contrast, splicing silencers recruit heterogeneous nuclear ribonucleoproteins (hnRNPs), among others, that affect spliceosome's assembly, resulting in the exclusion of an exon or exon skipping [16,20]. The localization of the *cis* and *trans*-splicing elements within the pre-mRNA is shown in **Figure 1.6**, along with the core splicing signals.

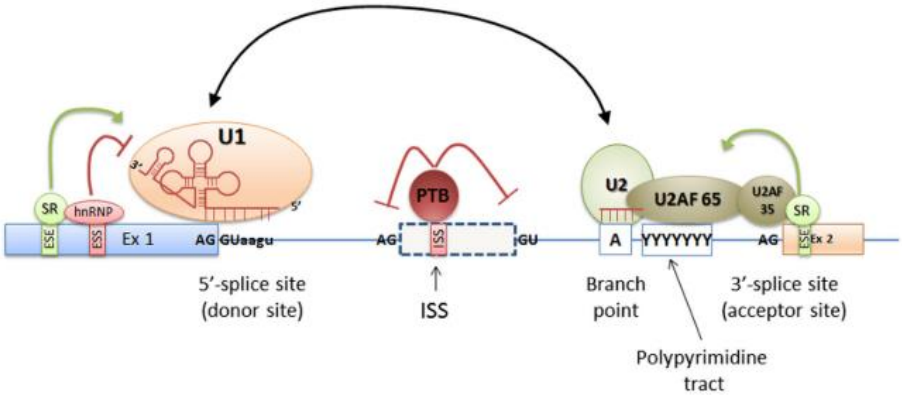


Figure 1.6 – Localization of the core splicing signals and the *cis* and *trans*-splicing regulatory elements, that control the splicing process. Adapted from [18].

Regarding the splicing enhancers mechanism, SR proteins recruit the spliceosome by simultaneously binding to the ESE/ISE in the pre-mRNA and to the U1 snRNP, facilitating the interaction between U1 and the donor splice site. Subsequently, the phosphorylation of the SR proteins is essential for the assembly of the remaining factors of the spliceosome. As opposed to this, the mechanism underlying exon exclusion is not well understood. Recent studies have demonstrated that hnRNP proteins induce exon skipping in the mature mRNA by interfering with the U1 snRNP exchange for U6 snRNP, thus preventing the formation of the catalytic complex [17,20].

Even though SR proteins and hnRNPs traditionally bind to enhancer and silencer sequences, respectively, it has also been shown that these *trans*-regulatory splicing regulators can act both as a splicing repressor or enhancer depending on the genomic context [16,17].

Several of these RNA binding proteins (RBPs) are only expressed in a determined cell-type and modulate splicing in a tissue-specific manner [22]. Since the majority of human genes are alternatively spliced and give rise to protein isoforms with distinct proprieties, abnormal patterns of mRNA splicing are associated with disease [20].

1.2.2 The Role of Alternative Splicing in Cardiac Disease

During cardiac development, the alternative splicing pattern of pre-mRNAs that encode for sarcomeric proteins leads to an isoform switch that results in the expression of slightly different proteins, used to adapt to changes in physiological conditions that cardiomyocytes are subjected to along its maturation. In the case of titin, troponin T and tropomyosin, it has been reported that different protein isoforms are translated from a single pre-mRNA by alternative splicing. Nonetheless, isoforms for proteins such as myosin, actin, troponin I, C and myosin binding protein C are generated from different genes (ex: the *MYH6* and *MYH7* different genes that encode different myosin's heavy chains) [20,22].

Overall, this "programme" of alternative splicing and isoform switch supports the postnatal growth and maturation of the heart. In heart disease, and more specifically in cardiomyopathies, several studies have reported that the fetal cardiac isoform "programme" is reactivated upon stress, with the sarcomeric protein titin being one of the most well described cases [22].

Titin 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 the 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, this large isoform is gradually replaced by a smaller and stiffer one, the N2B. The N2B is the predominant isoform in the normal adult human heart, being expressed in left ventricles in a ratio of 65% as opposed to only 35% of N2BA [16,20,22].

Alterations of the N2B/N2BA expression ratio has been correlated to heart disease, and this shift towards the expression of the N2BA fetal isoform in adult hearts was found in human cardiomyopathies and is mainly associated to Dilated Cardiomyopathy (DCM) [20].

In addition to the shift in the titin alternative splicing "programme", abnormal splicing of *MYH7*, *TNNT2* and *TNNI3* transcripts have also been associated to cardiomyopathies [16].

1.2.3 Splicing Variants

The presence of mutations that alter the splicing regulatory sequences may also lead to aberrant splicing of the pre-mRNA and, consequently, to disease. Splice-altering mutations can occur both in exons and introns, and they often disrupt the recognition of the canonical splice sites [18].

Variants in the consensus 3' (acceptor) and 5' (donor) sequences are denominated splice-site mutations and often have an impact on the recognition of the exon-intron boundary by the spliceosome. Mutations may also lead to the creation or activation of cryptic splice sites that might be recognized in detriment of the canonical ones. Ultimately, the consequence of a mutation in a canonical splice site is dependent on its strength, localization (in neighbourhood introns or exons) and the strength of a generated cryptic splice site, as well as the existence and density of splicing enhancers and/or silencers. Mutations may also alter the *cis*-regulatory sequences, promoting or preventing the assembly of the spliceosome [18].

Single nucleotide substitutions are the type of variants most often found within splice sites or splicing regulatory sequences. As shown in **Figure 1.7**, these mutations can result in the retention of an intron in the mature mRNA, partial of complete exon skipping or usage of alternative or cryptic splice sites [21].

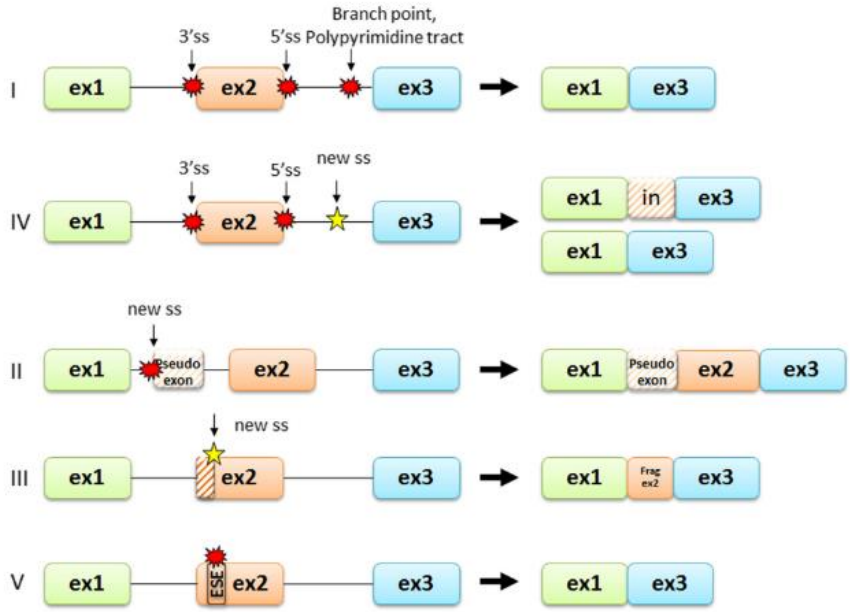


Figure 1.7 – Types of splice-altering mutations and the correspondent outcomes in the mature mRNA. Adapted from [18].

In most cases, splicing mutations lead to the introduction of a PTC in the mature mRNA molecule due to the disruption of its open reading frame (ORF). As a consequence, the mutated mRNA is frequently subjected to degradation by the NMD mechanism. If the deletions/insertions occur in-frame and the ORF is maintained, an altered protein may be translated and have unpredictable consequences to cell function [18,21].

Thus, NMD is a post-transcriptional mechanism consisting of a mRNA surveillance pathway that prevents the production of mutated proteins that might act in a dominant negative or deleterious manner culminating in a pathological situation. Abnormal transcripts are degraded by NMD when the termination codon is located more than 55 nucleotides upstream of the exon-exon junction complex (EJC) [23].

The EJC binds to the mRNA upon splicing and is normally located downstream of the stop codon, being removed by the ribosome during the first round of translation. The location of the stop codon and the EJC determines if the transcript will be degraded or not. Normal stop codons are mostly located within the last exon or are followed by one or two exon-exon junctions 50-55 nucleotides downstream. However, the presence of a PTC in the mRNA results in the displacement of the ribosome prior to reaching the EJC complex, and when this protein complex remains bound to the transcript, the NMD pathway is activated, leading to transcript degradation [23]. Hence, between 5-25% of PTC-containing transcripts were reported to escape the NMD mRNA degradation pathway, resulting in the translation of a mutated protein [24].

The pathogenic effect of a specific variant is mainly predicted based on its location in the genome and its direct effect in the coding sequence. Mutations that occur within exons are most often assumed to have a pathogenic potential because they may affect the encoded aminoacid and, consequently, the protein that is translated. However, variants that alter exonic splicing elements have already been associated to aberrant splicing [18,25]. Thus, disease-causing exonic variants that are thought to cause the pathology due to the alteration of the aminoacid sequence (so-called missense variants) may be unidentified as having an effect on splicing [21]. Such exonic mutations that affect splicing might introduce new donor or acceptor sites or may activate cryptic splice sites that may be recognized by the spliceosome in favour of the canonical splice sites, often leading to the loss of an exonic fragment. In addition, when they are located in the beginning or ending of an exon, they may disrupt the canonical splice sites and prevent the binding of the spliceosome. Moreover, exonic mutations in splicing enhancers or silencers may also lead to abnormal splicing [18].

Another type of variants that might affect splicing and whose pathological potential has been overlooked due to their location within the transcript are deep intronic mutations. These variants are mainly localized within large introns, distant from the canonical splice sites, and frequently result in the inclusion of an intron fragment into the mature transcript, by creating novel acceptor or donor sites or activating intronic cryptic splice sites. Deep intronic variants may also interfere with the recognition of enhancer or inhibitor splicing elements [18,26,27].

In HCM, several variants in the canonical splice sites have already been reported. For instance, in the *TNNT2* gene, a mutation in the donor splice site of intron 15 was shown to result in protein truncation. In addition, splice-site mutations in the *MYBPC3* gene that promote exon skipping have also been described [28,29]. Several deep intronic variants identified in HCM patients are predicted by bioinformatic tools to have an impact on splicing [27].

Approximately 9% of all the mutations reported in the Human Gene Mutation Database (HGMD) are splicing variants. However, this type of analysis only account for mutations within the splice sites, and this number may be largely underestimated due to the misclassification of exonic and deep intronic variants that may have an impact on splicing that has not yet been explored [18].

The majority of splicing mutations identified so far have been mainly reported through the analysis of genomic DNA sequencing. Knowing that variants in *cis* and *trans*-regulatory elements may also have an effect on splicing, RNA sequencing may also be used to pinpoint new splicing variants. To further evaluate the potential of a variant to disrupt splicing, minigene assays can also be performed [18]. In parallel, several *in silico* algorithms have been developed in order to predict if a given variant may affect splicing. Even though these bioinformatic tools were first established for a research purpose, they are already implemented as a diagnosis strategy. These tools have proven to be very useful to assess a possible effect on splicing; However, several publications have reported that often the predictions of these software's fail and do not reflect the functional effect of a mutation. Hence, functional studies are essential to confirm the functional impact of a variant on splicing and to validate the predictions of the *in silico* algorithms [18].

1.3 Allelic Imbalance and Haploinsufficiency in HCM

Two disease mechanisms have so far been described in HCM, as previously mentioned in **Chapter 1.1.1.** The majority of the variants identified in sarcomeric genes are classified as missense mutations that may have severe consequences in protein function and result in the production of a dominant-negative protein. This mutant protein can act as a poison polypeptide and be incorporated in the sarcomere disrupting its normal structure and function [30,31,32].

In contrast, over 90% of HCM-related mutations in *MYBPC3* gene result in the formation of PTCs caused by small insertion/deletions [24,31]. Therefore, haploinsufficiency is proposed as the disease mechanism for the majority *MYBPC3* mutations, in contrast to the poison polypeptide one described for the other sarcomeric genes. In cases of haploinsufficiency, since the PTC-containing transcript produced by the mutated allele is most often degraded by NMD, a functional protein is only translated from the wild-type allele, being insufficient to compensate its low expression levels and achieve a normal phenotype [24,31].

In support of this view, analysis of myocardial tissue isolated from symptomatic HCM patients revealed that the levels of full-length MYBPC3 protein are reduced, when compared to control subjects, with MYBPC3 truncated proteins being undetectable [31]. Furthermore, in heterozygous mouse carriers of *MYBCPC3* truncating mutations, the levels of MYBPC3 vary and are reduced in up to 50% [33]. This data suggests that the mutated protein is either not synthesized and the mRNA is degraded through NMD, or that the truncated protein is rapidly subjected to ubiquitin-mediated proteasomal degradation [30,31]. In HCM patients with *MYBPC3* truncating mutations, the lower expression levels of this protein may be responsible for the disruption of the sarcomere structure, since the stoichiometry of the proteins that comprise the sarcomere is no longer maintained. The difference between a normal heart and a

pathological situation due to insufficient MYBPC3 protein included in the sarcomere structure is shown in **Figure 1.8**.

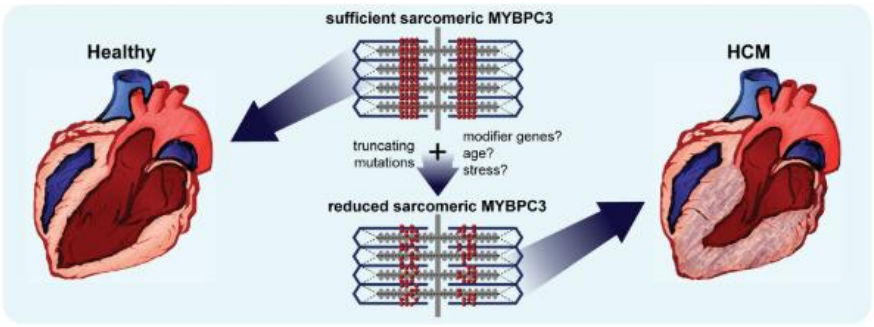


Figure 1.8 – Haploinsufficiency in the *MYBPC3* gene as a causal mechanism of HCM. Adapted from [24].

Allelic imbalance has been reported for many missense HCM-related mutations in *MYH7*, *MYBPC3*, *TNNT2* and *MYL2* genes. By definition, allelic imbalance consists of a differential expression of both alleles, at the mRNA or protein level. Some papers have described equal expression of the normal and mutant alleles, while others described that mutant alleles can have a higher or lower abundance relative to the wild-type [24].

As depicted in **Figure 1.9**, in general, the mechanisms of allelic imbalance can occur at the DNA, RNA or protein levels. Epigenetic modifications such as imprinting or differential methylation along with sequence variants in regulatory elements, such as promotor regions, may influence the availability of the DNA to the transcription machinery. Moreover, heterozygous mutations that originate PTCs may lead to the degradation of the mutated transcripts. Furthermore, mutated proteins that were able to escape the previous degradation pathways often have alterations in their cellular location, folding and three-dimensional structure and protein-protein interactions, among others, and are often less stable and more susceptible to degradation [24].

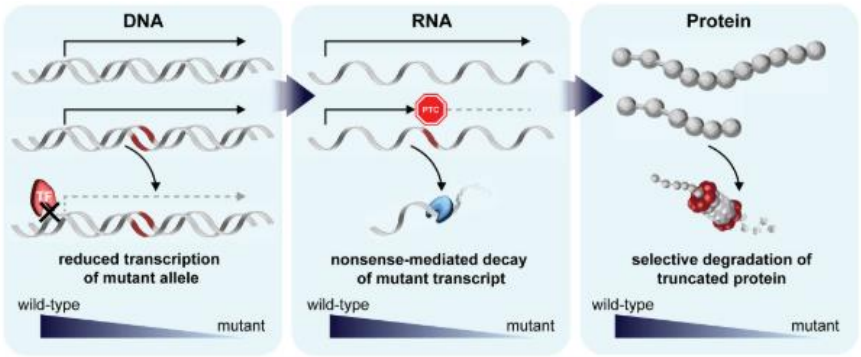


Figure 1.9 – Mechanisms that lead to allelic imbalance, in which the expression ratio of the wild-type and mutated allele differs from the expected 1:1. Adapted from [24].

However, several studies where heterozygous *MYBPC3* truncating mutations were analysed show that wild-type *MYBPC3* is expressed at levels higher than 50% suggesting that there is some sort of compensation [24]. This compensation from the normal allele is thought to be a mechanism involved in the 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, culminating in an haploinsufficiency condition [11].

1.4 Stem Cells in Disease Modelling

Stem cells are defined as undifferentiated cells that have self-renewal and proliferative capacity, along with the ability to differentiate into specialized cell types. Thus, stem cells can be characterized into four distinct types, according to their differentiation potential or so-called potency [34,35].

The zygote is the only **totipotent** cell in the human body, being able to generate all cell types and give rise to an entire organism. On the other hand, stem cells isolated from the inner cell mass of the embryo are **pluripotent** and can differentiate into the three embryonic germ layers - endoderm, mesoderm and ectoderm, but cannot generate cells of the extraembryonic tissue. **Multipotent** cells, such as haematopoietic stem cells, are capable of differentiating into a specific subset of cell lineages. Lastly, spermatogonia stem cells are **unipotent** since they can only originate one specific cell type [34,35,36]. Stem cells can also be classified as either embryonic or adult, depending if they are isolated from the embryo or the adult, post-natal organism. Embryonic stem cells (ESCs) are obtained from the blastocyst and are characterized as pluripotent. Although a large number of evidences pinpoint the existence of adult stem cells in all adult tissues, they are mainly isolated from the bone marrow, where two types of stem cell populations are present: hematopoietic and mesenchymal stem cells [36,37].

1.4.1 Embryonic Stem Cells

Human embryonic stem cells (hESCs) were first isolated in 1998 by James Thomson and presented a similar morphology to those already obtained from mice and non-human primates [38,39,40]. In **Figure 1.10**, an image depicting the standard morphology of an ESC colony is presented. hESCs also displayed an unlimited proliferative capacity *in vitro* while maintaining pluripotency and differentiation ability into cell types of the three germ layers [38,41]. Telomerase activity has been correlated with immortality in human cell lines, since it is known to support ongoing cell division. Hence, the immortality of hESCs is believed to be associated with the high expression level of telomerase in these cells [36,38].

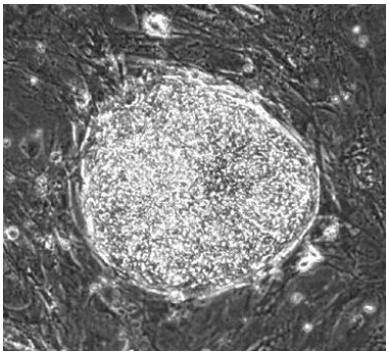


Figure 1.10 – H9 ESC colony in a mouse embryonic fibroblast feeder layer. Adapted from [38].

Following the successful establishment of hESCs, several protocols were developed for the differentiation of hESCs into specialized cell types such as neural cells, cardiomyocytes and pancreatic β-cells, aiming for the creation of disease models and stem cell therapy approaches. However, since the isolation of hESCs involves the manipulation of human embryos in an *in vitro* fertilization context, many ethical issues have been raised. Furthermore, in the case of heterologous stem cell therapy, immune response is also a strong possibility, given that the development of patient-specific ESC lines was not a viable option due to intrinsic challenges and difficulties of the procedure and the ethical problems associated [41].

1.4.2 Induced Pluripotent Stem Cells

Several studies have shown that through nuclear transfer into oocytes or by fusion with ESCs, it is possible to reprogram a somatic nucleus into a pluripotency state [42,43]. Transcription factors such as the octamer-binding transcription factor 3/4 (Oct3/4), sex determining region Y-box 2 (Sox2) and Nanog were shown to be essential for the maintenance of the pluripotency state of hESCs. In addition, genes that are frequently found to be up-regulated in tumours, such as Krüppel-like factor 4 (Klf4) and c-Myc also contribute to the maintenance of the hESCs phenotype and their proliferation capacity whilst in culture [41].

In 2006, Yamanaka and Takahashi were able to reprogram mouse fibroblasts into so-called induced pluripotent stem cells (iPSCs) using a retroviral system to force the expression of genes thought to be involved in pluripotency induction. Out of a pool of 24 candidate genes, they discovered that *Oct3/4*, *Sox2*, *c-Myc*, and *Klf4*, nowadays known as the Yamanaka factors, were necessary and sufficient for the reprogramming of mouse fibroblasts. iPSCs morphology and proliferation *in vitro*, along with the capacity of forming teratomas, resembled those of mouse ESCs, as their data also confirmed the pluripotent nature of these cells [41]. In a similar approach, using the original four reprogramming factors, human iPSCs were first generated from somatic cells in 2007, and alike what was observed in mouse, hiPSCs are comparable to hESCs. For his work, Yamanaka was awarded the Nobel Prize in Physiology or Medicine in 2012 [44,45,46].

iPSCs are mainly characterized based on their morphology, gene expression and capacity of teratoma formation upon injection in immunodeficient mice. Regarding their morphology, hiPSCs typically form compact colonies/clusters of round or oval shape with well-defined and smooth margins. These clusters are comprised of tightly packed cells, with large nucleus, prominent nucleoli and a high nuclear to cytoplasmic ratio. Markers such as transcription factors *Oct4*, *Nanog* and *Sox2* and stage-specific embryonic antigens (*SSEA3/4*) and DNA methyltransferase 3β (*DNMT3β*) can also be used to assess pluripotency [47,48,49,50]. In **Figure 1.11,** images that depict the characteristic morphology of hiPSCs clusters and the expression of pluripotency markers through immunofluorescence staining are shown. Additionally, the expression of the placental alkaline phosphatase has been related to pluripotency and is also widely used as a pluripotency marker [48].

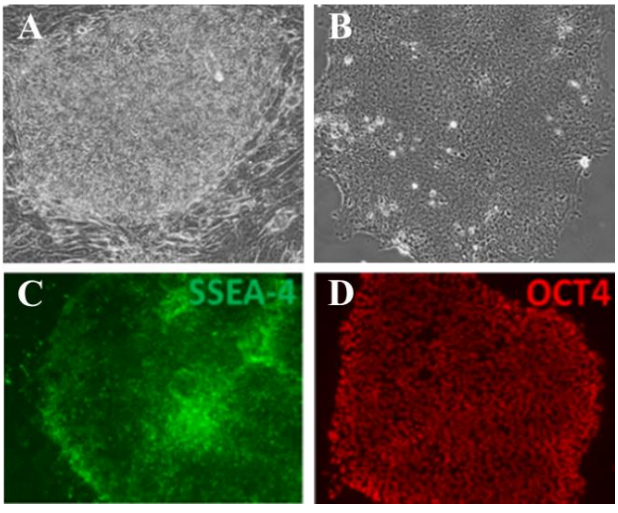


Figure 1.11 – Characteristic morphology of hiPSC colonies growing on **(A)** a feeder layer and **(B)** under feeder-free conditions on a Matrigel substrate. Human iPSC colonies grown on Matrigel showcase a flatter morphology. **(C)** Immunofluorescence analysis of pluripotency markers SSEA-4 and **(D)** OCT4 in an established hiPSC line. Adapted from [49].

Functional tests are also required to ensure that iPSCs have the ability to differentiate into specialized lineages of the three germ layers. This can be achieved either by the formation of embryoid bodies, 3D aggregates of iPSCs that recapitulate the embryo undergoing early gastrulation, or through *in vitro* differentiation protocols. The pluripotency potential may also be evaluated by iPSCs injection into immunodeficient mice by formation of teratomas that comprise cells of the three germ layers [47,48].

A little over a decade after they were first described, human iPSCs (hiPSCs) have emerged as a powerful tool to study development and human genetic disease, as their use circumvents the ethical and moral issues associated with the manipulation of hESCs. In addition, protocols for the reprogramming of several types of somatic cells have been developed, and a skin biopsy or a blood sample is sufficient to generate hiPSCs, facilitating the establishment of patient-specific cell lines. Disease models have evolved from 2D differentiation protocols to 3D organoids that to an extent can recapitulate the complex disease environment. These models are also currently used for drug screening. hiPSCs also comprise an evolving field of patient-specific cell-based therapies, as patient rejection is no longer an issue [51].

1.4.3 Roadblocks in the use of hiPSCs

Even though hiPSCs have revolutionized disease modelling, scientists are still faced with many challenges that need to be solved. The first reprogramming technologies relied on the use of integrating vectors such as retrovirus or lentivirus for the induction of the expression of both transcription factors and tumour-associated genes involved in pluripotency. Despite being associated with higher reprogramming efficiencies, in such delivery systems the possibility of insertional mutagenesis and consequent generation of genomic abnormalities is a strong safety concern. Tumorigenicity is also a drawback, not only due to the forced expression of genes such as *c-myc*, but also because the use of integrative virus may interfere with oncogenes or oncosupressor genes expression when inserted in the particular genomic regions. Thus, non-integrative approaches such as plasmids, adenovirus or Sendai virus were developed and are now preferred, regardless of their lower efficiencies [45,51,52]. The low reprogramming efficiency, reported in Yamanaka's first study in 2006 as 0.01–0.1%, has been circumvented through the use of small molecules such as valproic acid and sodium butyrate, along with miRNAs that were shown to enhance iPSC generation [41,45].

During the reprogramming process, the somatic cell is subjected to a rearrangement of the transcriptome and epigenome, which has been associated with the generation of genomic abnormalities in many hiPSCs lines. Retention of epigenetic memory, a residual DNA methylation pattern that leads to differentially methylated regions and inefficient silencing of genes of the somatic cell, has been reported in iPSCs [45,51]. In addition, throughout time in culture, hiPSCs were also reported to develop chromosomal anomalies such as aneuploidies [52]. Therefore, karyotypic analysis to ensure that the iPSCs lines maintained genetic stability is essential [47].

Nowadays, besides the limitations of the reprogramming process *per se*, one of the main concerns of the use of hiPSCs in disease modelling is the fact that differentiated cells do not reflect the level of maturity and complexity found in human tissues. These hiPSC-derived cells show an embryonic or fetal identity rather than displaying adult characteristics, posing as a hindrance to data analysis. This high level of immaturity has been reported in various 2D and 3D differentiation protocols into many cell types relevant in the context of specific diseases, including the cases of iPSC-derived cardiomyocytes and pancreatic β-cells [53].

Regarding their phenotype and gene expression pattern, a large variability in both iPSCs and cells differentiated from iPSCs has been reported. Given the feasibility with which iPSCs can now be generated from somatic cells of many donors, the different epigenetic profiles and the presence of single nucleotide polymorphisms (SNPs) that alter the genetic background were associated with the existent heterogeneity among iPSCs. These aspects are pointed out as the likely cause of the different behaviours and differentiation capacities described for distinct iPSC lines. Thus, the broad background genetic variation does not allow for the establishment of reliable consistent genotype-phenotype correlations, since comparing data from patient-derived iPSCs and wild-type controls derived from a different

individual may lead to inaccurate conclusions. This heterogeneity amongst hiPSCs lines and the requirement of appropriate controls raises the need to apply gene-editing methodologies in order to obtain isogenic controls and trustworthy disease models [45,53].

1.5 Cardiac Differentiation

Before the breakthrough of hESCs and hiPSCs, *in vitro* cardiac models were based primarily in immortalized cell lines, which can only recapitulate human cardiomyocytes (CMs) characteristics to a certain extent. On the other hand, CMs isolated from other animal models also could not reproduce all of human features. Finally, it is very difficult to obtain cardiac tissue from human patients, as it is limited to extreme clinical situations such as a cardiomyectomy. Thus, the emergence of stem cell technology has allowed for the development of more suitable models [54].

Human CMs were first differentiated from embryonic stem cells cultivated in suspension to form embryoid bodies. In this original experiment, only approximately 8% of cells presented spontaneous contraction and were shown to express cardiac-specific genes and display certain degree of myofibrillar organization. Ever since, several protocols have been established to achieve an higher cardiac differentiation efficiency and better models to study heart disease [55].

For the differentiation of iPSCs into cardiomyocytes, the cells must be exposed to specific stimuli throughout the differentiation process, in order to recapitulate the endogenous signals and consequent shifts in gene expression that drive the development of the human heart. This is achieved through the time-specific control of the major pathways shown to be involved in cardiac development: the activin/nodal/transforming growth factor-β (TGFβ), bone morphogenetic protein 4 (BMP4) and Wnt signalling pathways [56,57]. Reportedly, the Wnt/β-catenin signalling regulates the activin/BMP4 pathway, which is necessary for the initial differentiation into the mesodermal germ layer [58].

Thus, for the generation of hiPSCs derived cardiomyocytes (hiPSC-CMs), the cells must first be subjected to an activation of the BMP4 and Wnt signalling pathways and committed to the mesoderm fate [59]. At this stage, cells express early mesodermal markers such as the transcription factor Brachyury. Following its induction, the Wnt signalling pathway must then be inhibited in order to achieve cardiac commitment, with the formation of cardiogenic mesoderm cells accompanied by the expression of the insulin gene enhancer protein (ISL1). Further induction of the cardiac fate leads to the generation of cardiac progenitors that express early heart development markers such as the homeobox protein Nkx2.5 and, followed by differentiated cardiomyocyte lineage, where sarcomeric-related proteins are expressed and cells present electrical and contractile activity [56,60]. An overview of the cardiac differentiation process, highlighting the role of the Wnt pathway is schematized in **Figure 1.12**.

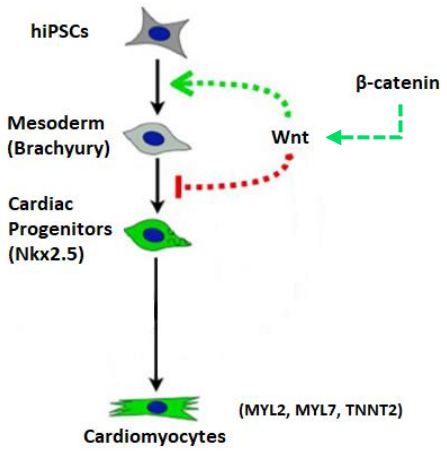


Figure 1.12 – The Wnt signalling pathway plays an essential role during differentiation of iPSCs-CMs. Initially, Wnt is responsible for the differentiation of the iPSCs towards cardiac mesoderm and is afterwards inhibited in order to give rise to cardiac precursors. Adapted from [59].

In 2012, Lian *et al.* developed a protocol that resulted in a high cardiac differentiation efficiency, with 80–98% of the total cell population consisting of functional cardiomyocytes [57].

In this 2D monolayer protocol, glycogen synthase kinase-3β (GSK3β), a transduction factor of the Wnt/β-catenin pathway, is inhibited via the small-molecule inhibitor CHIR-99021. GSK3β is responsible for the phosphorylation of β-catenin and leads to its degradation. Hence, in the absence of GSK3β, the Wnt pathway is activated. Subsequently, the Wnt/β-catenin pathway is inhibited via the small-molecule inhibitors of Wnt ligand production (IWP2/4), inducing the mesoderm progenitor cells to differentiate into cardiac precursors. In addition, the authors stated that the presence of insulin in the initial period of differentiation leads to a decrease in the cardiomyocyte yield, in concordance with previous reports that insulin signalling inhibits cardiac commitment [61]. Thus, B27 minus insulin, a supplement that contains all factors required to support the survival and differentiation of the cells, but lacking insulin, is added to the medium for the first days of differentiation. The latter addition of insulin to the medium is associated with an efficiency increase in cardiac differentiation [57,58,60].

Morphologically, hiPSCs-CMs are small and round when compared to adult cardiomyocytes, resembling the phenotype of embryonic and fetal CMs. Adult CMs are not only larger, but the cells are also more elongated and often bi- or multinucleated, as opposed to hiPSCs-CMs that are mainly mononuclear [56,62]. These immature hiPSCs-CMs also display a reduced sarcomerc structural organization, a lower expression of cardiac-specific genes when compared to the levels in the adult human heart, as well as a fetal isoform expression pattern [45]. In addition, several studies reported that iPSC-CMs present mainly a glycolytic metabolic phenotype, the same as embryonic CMs, as opposed to the primarily oxidative metabolism of adult cardiomyocytes [62].

Nonetheless, differentiation protocols have already been improved in order to achieve more mature cardiomyocytes, essential for disease modelling and possible biomedical applications. The foetal-to-adult isoform switch and gene-expression characteristics closer to those of adult CMs results in highly organized sarcomeric structures, myofibril alignment, more prolonged cell morphology and the improvement of electrophysiological proprieties [45,58]. An immunofluorescence image of an hiPSC-derived cardiomyocyte can be seen in **Figure 1.13**.

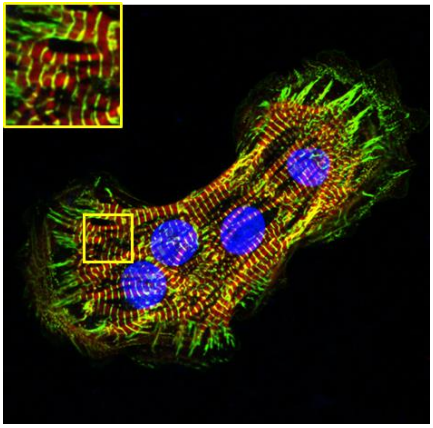


Figure 1.13 – Immunofluorescence staining of hiPSC-CMs for cardiac troponin T (green) and α-sarcomeric actinin (red), depicting the level of maturation and organization of the sarcomere structure that is associated to the expression of these cardiac proteins.

However, the reproducibility and efficiency of the cardiac differentiation protocols continue to be two of the main challenges for the generation of iPSCs-CMs. The establishment of 3D and variations on monolayer protocols, such as the supplementation with growth factors and small molecules that are known to promote cardiac differentiation, glucose deprivation or prolonged time in culture reportedly result in higher differentiation efficiencies. Furthermore, the differentiation process often results in a heterogeneous mix of several cell types, i.e. endothelial cells and smooth muscle cells, and a pure

population of CMs can be achieved through sorting of these cells for specific markers such as the vascular cell adhesion molecule 1 (VCAM-1) [45].

Even though pure populations of somewhat mature cardiomyocytes can now be obtained, different hiPSCs lines still respond distinctively to the external cues that drive *in vitro* cardiac differentiation. As previously mention in **Chapter 1.4.3**, this variability is thought to be related to the diverse genetic backgrounds of the donors of the somatic cells that are then reprogrammed into iPSCs and can be circumvented through the use of gene-editing methodologies [45,53].

1.6 CRISPR/Cas9 – A Gene Editing System

In the last decades, advances in the gene editing field have completely revolutionized biological research. Scientists can now insert, delete or modify target DNA sequences in its endogenous context, both *in vitro* and *in vivo*, raising new possibilities for understanding the molecular mechanisms behind diseases, establishing innovative drug development processes and therapeutic approaches, such as gene therapy through the correction of disease-causing mutations [63,64].

Gene-editing technology relies on the use of nucleases that can be directed to specific sites in the genome and cleave the DNA molecule, introducing precise double-strand breaks (DSBs) and thereby promoting the activation of endogenous repair mechanisms [65]. Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are chimeric nucleases that contain a non-specific DNA cleavage domain and a sequence-specific DNA-binding region that can be programmed according to the target site in the genome. They were the most widely used tools until the first description of human and other eukaryotes gene-editing approaches via the CRIPSR/Cas9 system in 2013, by Feng Zang and George M. Church [65,66,67].

The clustered regularly interspaced short palindromic repeats (CRISPR) is a kind of "immune defence system" that has been described both in bacteria and archaea. CRISPR loci is constituted by an array of short, conserved repeated sequences intercalated by nonrepetitive sequences denominated as *spacers*, that consist of foreign genetic sequences. After an infection by a virus or phage, the CRISPR-associated (Cas) nuclease processes the extrinsic DNA into small fragments that are then incorporated into the CRISPR locus of the host genomes as the *spacers*. Thus, in response to a new infection, these *spacer* sequences are used as templates for the production of a long RNA that is cleaved into small CRISPR RNAs (crRNA) and that combined with the transactivating crRNA (tracrRNA) to form a tracrRNA:crRNA complex that guides the Cas nuclease to the target sequence. In the invader genome, the *protospacer* is complementary to the crRNA, and a short protospacer adjacent motif (PAM) is also required for binding of the Cas-RNA complex to the DNA, through Watson-Crick base-pairing. Subsequently, the nuclease domains of the Cas protein are responsible for the cleavage of the target region in the invading virus or phage genome [64,65,68].

So far, three CRISPR/Cas systems with different molecular mechanisms (type I, II, and III) and over 40 distinct Cas protein families have been described. The CRISPR/Cas9 technology is an adaptation of an acquired immunity type II system of the bacteria *Streptococcus pyogenes*. In **Figure 1.14**, a summary of the CRISPR-Cas9 system is presented [64,65,68,69]. For its application in gene-editing, the CRISPR/Cas9 system was engineered in order to be easily modified according to the regions of the genome to be altered (**Figure 1.14B**). Hence, through this technique, instead of a tracrRNA:crRNA complex, only a small guide RNA (sgRNA) that is complementary to the target sequence is needed to guide the Cas9 nuclease to a specific site in the genome. This sgRNA consists of a fusion of a crRNA and a fixed tracrRNA. Twenty nucleotides at the 5′ end of the sgRNA direct the Cas9 to a specific target DNA site using standard RNA-DNA complementarity base-pairing rules. These target sites must lie immediately 5′ of a PAM sequence that matches the canonical form 5′-NGG. Thus, with this system, Cas9 nuclease activity can be directed to any DNA sequence of the form N-NGG simply by altering the first 20-nt of the gRNA to correspond to the target DNA sequence. Once the Cas9 recognizes the PAM that flanks the 3' end of the target DNA, it introduces a blunt double strand break ~3 bp upstream of the PAM sequence, that can be repaired either by the error-prone nonhomologous

end joining (NHEJ) mechanism or through homology-directed repair (HDR), a more reliable repair mechanism [64,65].

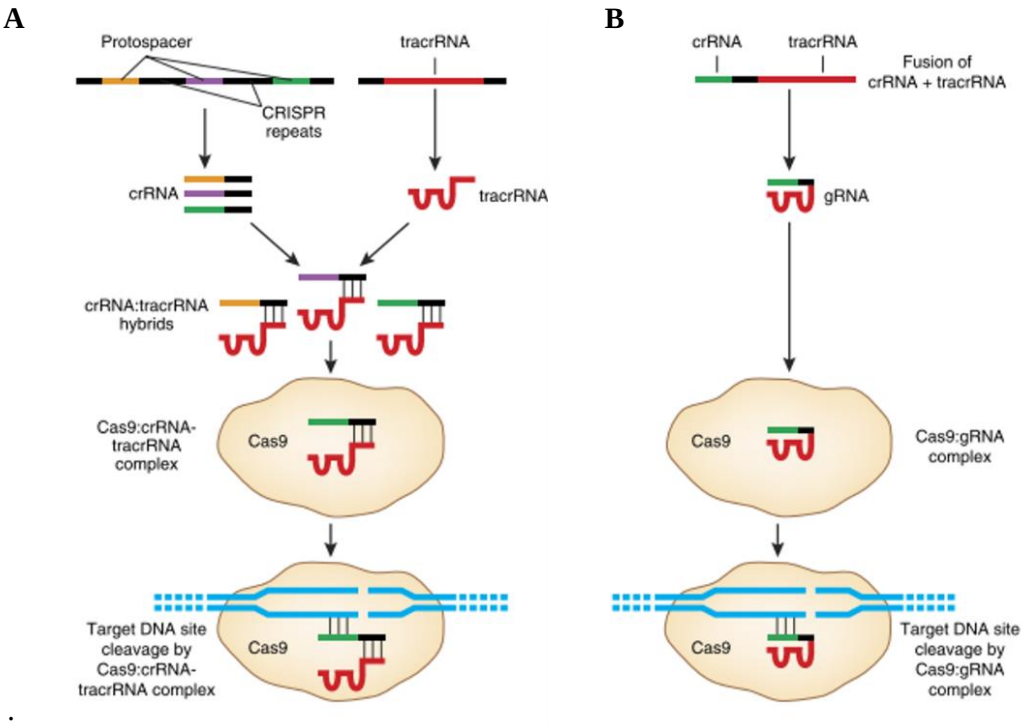


Figure 1.14 – **(A)** The natural occurring adaptive CRISPR/Cas9 immune system. **(B)** The CRISPR/Cas9 system was engineered for an easier application in gene-editing, requiring only a small guide RNA whose sequence can be modified according desired targeted genomic sequence. Adapted from [69].

The HDR mechanism can be used to precisely introduce specific nucleotide modifications or knock-ins in the targeted genome, as it uses a DNA template (either an exogenous template or the already replicated chromosome) to achieve a high-fidelity, error-free repair mechanism. However, the frequency and efficiency of HDR is lower when compared to the NHEJ pathway, that can rapidly ligate the two cleaved ends of the DNA together and repair the damage that was introduced by the Cas9 nuclease. NHEJ often results in the introduction of errors such as insertion or deletion of nucleotides at the target sites, the so-called indels, frequently leading to frameshifts and generation of PTCs (**Figure 1.15**). Hence, the NHEJ pathway can be used in gene-editing to achieve gene knockouts [65,70].

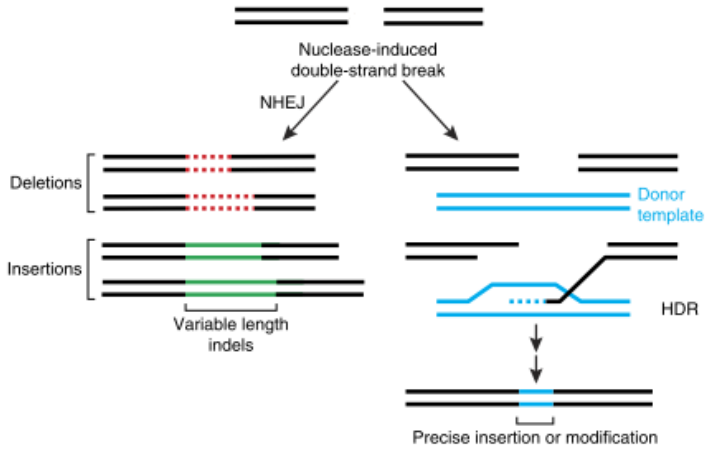


Figure 1.15 – NHEJ or HDR strategies for the repair of double-strand DNA breaks caused by the CRISPR/Cas9 system. Adapted from [69].

Considering that the CRISPR/Cas9 system can be easily adapted because its programmability is only dependant on the sequence specificity of the sgRNAs, it became the most popular approach for gene-editing since the first attempt was performed [64,65].

CRISPR technology has been already used to establish both *in vitro* and *in vivo* animal models of disease, and can even be used to simultaneously introduce genomic alteration in multiple sites [64]. Plus, with the evolution of hiPSCs research, several disease models have also been generated, both through the introducing of a mutation in the genome or its correction to obtain isogenic models [65]. CRISPR/Cas9 gene-editing has also been performed in human embryos and was successful in correcting *MYBPC3* mutations associated with HCM. The first clinical trials using this technology began in 2016, and to this day several trials ranging from the treatment of genetic diseases, such as β-thalassemia and cancer have been approved [71].

1.6.1 Disadvantages of the CRISPR/Cas9 system

Despite its extraordinary potential, there are still several issues that pose as a safety concern for gene therapy studies using this technique and for the establishment of reliable disease models. One of the main drawbacks is possible off-target effects of the CRISPR/Cas9 system, that were reported to be present in high frequencies in human cells. These off-targets are caused due to the similarity of other sequences within the genome to the target region, leading to the unspecific cleavage of the DNA by the Cas9. This unspecific binding of the sgRNA to other sites in the genome can lead to mutations that have unpredictable damaging effects. In addition, after the intended alteration is incorporated in the genome through homologous-direct repair, the CRISPR-Cas9 system can then introduce undesirable on-target indel mutations, also known as re-cutting by the Cas9 nuclease. This occurs because the sgRNA-Cas9 complex can still recognize the now-edited target sequence and introduce a DSB that is in this turn repaired by NHEJ. To prevent this, strategies that resort to inserting additional mutations into the repair template were developed, so that the target sequence is no longer recognized after the first gene-editing. Furthermore, not every location in the genome can be edited, since the recognition/cleavage of the target site is dependent on the presence of the 2-5 nucleotides known as the PAM sequences. Reports have also shown that the efficacy of delivery of the CRISPR/Cas9 machinery and the gene-editing efficiency are highly variable between different cell types [64,65,72].

1.6.2 Gene Editing in hiPSCs

Even though edited hiPSCs lines have already been established, the very low efficiency of gene-editing in these cells remains a serious challenge in the field. In addition to the difficulties of gene-editing that are inherent to the DNA repair mechanisms, particularly the low efficiency of the HDR, it is important to note that hiPSCs are not only very sensitive but also more demanding to work with than other cells, such as immortalized lines. Even though NHEJ has a slightly higher efficacy, different laboratories reported efficiencies of HDR in hiPSCs that range from 0.1 – 1%. This difficulty of gene-editing in hiPSCs is mainly related to the low cell viability after transfection of the CRISPR/Cas9 system components required for gene editing, combined with the difficult manipulation of these cells. Moreover, the transfection efficiencies in hiPSCs are also much lower than those reported for other cell lines. The most effective transfection approach is electroporation; however, it typically leads to major cell death, and this is even more problematic in the case of hiPSCs. Other physical transfection methods, as the lipid-based lipofection approach does not cause such a large decrease in cell viability as that observed in electroporation, but it is also a less efficient method. As a consequence of the low gene-editing efficiency, several hundred clones must be isolated and screened, which translates in a multistep protocol that can last up to 2-3 months, requiring lengthy periods of cells in culture. Since performing gene-editing directly in hiPSCs has proven to have many hindrances, other methods have been developed to circumvent these problems [73,74].

1.6.3 Gene-Editing & Reprogramming of Human Fibroblasts

In a paper published in 2018 by Sara Howden *et al.*, the authors proposed an alternative approach to gene-editing in hiPSCs, that consists of a simultaneous reprogramming and gene-editing protocol of human fibroblasts. They reported that successfully edited hiPSCs, through the HDR mechanism, comprised 2–20% of the total hiPSCs population, as opposed to the <1% that is described when hiPSCs are used as the starting material for gene-editing [75].

For the reprogramming of the human fibroblasts, a non-integrative episomal approach is used, with the forced expression of several factors that were shown to be involved in pluripotency maintenance. Briefly, in addition to the original Yamanaka factors (Oct4, Sox2, c-Myc and Klf4), the expression of Nanog, Lin28 and the SV40 large T antigen was also induced. The SV40 large T antigen is a protein derived from the SV40 virus known to have transforming activity and was reported to increase the reprogramming efficiency of both fetal and adult human fibroblasts up to 70 fold [76]. In turn, the Lin-28 RNA-binding protein is a post transcriptional regulator of genes involved in self-renewal in hESCs [76]. The miR302/367 cluster is able to directly reprogram mouse and human somatic cells to a pluripotent state, and when the authors of the protocol combined its forced expression with that of the other factors, it resulted in a >100-fold increase of the reprogramming efficiency [77]. The mRNA of the Epstein–Barr nuclear antigen-1 (EBNA-1) was also used in this protocol since it enhances the transfection efficiency when co-delivered with EBV-based vectors. The EBNA-1 protein has been shown to be implicated in the nuclear transport of these plasmids. It is proposed that it bind to EBV plasmids in the cytoplasm and subsequently enhances their nuclear uptake by binding to mitotic chromosomes during cell division. A transient expression of this protein using the mRNA transfection instead of the plasmid was shown to be sufficient to enhance the transfection efficiency of the reprogramming plasmids [75,78].

Regarding gene-editing, the mRNA of a variant of the *Streptococcus pyogenes* Cas9 nuclease that is fused to a peptide derived from the human Geminin protein (Cas9-Gem) was further used. Geminin, a DNA replication inhibitor, is absent during the G1 phase and accumulates through S, G2 and M phases of the cell cycle. As demonstrated by the graphs in **Figure 1.16**, the NHEJ pathway is active throughout the whole cell cycle. In contrary, the HDR mechanism is active in S, G2 and M phases and virtually absent in the G1 phase of the cycle. Using this strategy, the Cas9-Gem nuclease is only present in the phases of the cell cycle when the HDR pathway is also active, thus promoting the repair of the DSB through homologous recombination with the template DNA supplied for gene-editing [75,79].

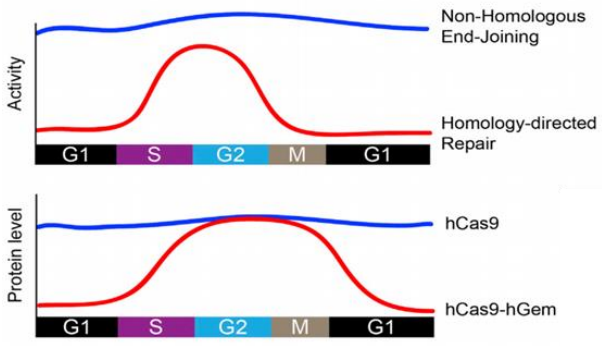


Figure 1.16 – Activity of the NHEJ and the HDR mechanism throughout the cell cycle correlated with the presence of the Cas9-Gem protein. The Cas9-Gem nuclease protein is only present in the stages of the cell cycle in which the HDR mechanism is active, both inhibiting the introduction of indels through the repair of the DSB by NHEJ and inducing HDR. Adapted from [79].

Moreover, by limiting the presence of the Cas9-Gem protein at specific cell cycle phases (both due to the fusion to the Geminin peptide and the transfection of the transient Cas9-Gem mRNA), undesirable secondary mutations within the target locus were minimized [75,79].

2. Objectives and Hypothesis

HCM is the most common genetic heart disease with over 1500 pathogenic mutations associated, with approximately 90% of these variants being classified as missense. Since it is a heterogeneous disease, with patients bearing the same mutation having variable phenotypes and clinical outcomes, a genotype-phenotype correlation has not yet been well established. Therefore, the development of accurate models that reflect the characteristics of human CMs is essential to understand the molecular mechanism that underlies HCM-associated variants. Considering that previous models consisted mainly of immortalized cell lines and CMs obtained from animal models that could not completely reproduce the characteristics of human CMs, hiPSCs are currently the most appealing approach to study how genetic variants can contribute to the complex disease phenotype of HCM.

Exonic mutations that interfere with the splicing pattern and that are related to human disease have already been described and often lead to the retention of intron fragments in the mature mRNA or to exon skipping [80,81]. Therefore, a large number of HCM exonic variants, nowadays classified as missense and synonymous, might in fact be disease causing due to aberrant splicing; thus, these exonic splice-altering variants may comprise an important mechanism underlying human disease and more specifically, HCM.

The main objective of this study is to understand how mutations that have been previously classified as missense may cause a pathogenic effect because they alter the splicing pattern of sarcomeric genes, and not only due to their direct effect in the encoded aminoacid. The proposed hypothesis is that such variants lead to abnormal splicing, resulting in the generation of PTCs that trigger mutant transcript degradation by NMD, culminating in a situation of haploinsufficiency in which the total level of protein produced is reduced up to 50% since it is only translated from the wild-type allele. Haploinsufficiency has been proposed as the major disease mechanism associated to the a large number of *MYBPC3* mutations, and this hypothesis is supported by the fact that several papers report that MYBPC3 truncated proteins are not detected in tissue obtained from human myectomies of such HCM patients [11,14,31]. Additionally, the mutated transcripts that escape NMD pathway may result in the production of a truncated protein that is most likely misfolded and unstable, and consequently targeted for degradation by the ubiquitin-proteasome mechanism; if not degraded, the truncated protein can alternatively be incorporated in the sarcomere and disrupt its structure, having severe consequences in heart function due to a negative dominant effect [30,31,32].

hiPSCs can be differentiated into functional cardiomyocytes, and this hiPSCs-CMs platform is ideal for the study of how exonic HCM genetic variants may interfere with the splicing of a given transcript, considering that these disease models have the potential to better recapitulate the phenotype of healthy and diseased human cardiomyocytes when compared to the pre-existent models. Gene-edited hiPSCs models and the correspondent isogenic controls can be generated resorting to the CRISPR/Cas9 technology, which enables the precise introduction of a specific variant in the genome of a human cell [73,74].

The generation of hiPSCs that harbour HCM missense variants, which by bioinformatic tools were predicted to affect splicing is the first aim of this project. Subsequently, these gene-edited hiPSCs are to be differentiated into CMs using an already established 2D cardiac differentiation protocol, based on the method published by Lian, *et al*., [57]. The expression of sarcomeric genes of the diseased CMs and isogenic controls will be characterized by qRT-PCR and Western Blot analysis in order to identify possible splicing alterations. Morphological analysis of both control and mutated hiPSCs-CMs by immunofluorescence assays will be also performed in order to assess if they are in concordance with the pathological phenotype associated with HCM-CMs. Additionally, a more extensive transcriptomic and proteomic analysis might be required to identify changes in the mRNA and protein levels of sarcomeric genes and how this may affect the overall stoichiometry of the major sarcomeric proteins and, ultimately, sarcomere structure and function.

In parallel, another purpose of this thesis is to explore if in cases of haploinsufficiency, a transcriptional adaption mechanism is activated, in order to compensate for the reduced expression of the affected sarcomeric protein. This compensation might be achieved by the increase of mRNA production from the wild-type allele or from an adaptation of the other sarcomeric transcripts and/or protein levels, to maintain the correct sarcomere stoichiometry.

Therefore, the primary objective of this thesis is to decipher if missense HCM mutations can lead to abnormal splicing and if haploinsufficiency is the disease mechanism underlying these splicing-altering variants. Subsequently, we aim to understand if a mechanism of transcriptional adaptation is activated under these conditions in order to compensate for the lower levels of a given sarcomeric protein.

3. Materials and Methods

3.1 Cloning of CRISPR/Cas9 sgRNAs

3.1.1 Design of sgRNAs

sgRNAs were designed by bioinformatic tools, according to the guidelines previously described in the Nature Protocol *Precise and efficient scarless genome editing in stem cells using CORRECT*, presented in **Table 3.1**, and selected according to their scores and predicted off-target effects [72].

Table 3.1 – Guidelines for sgRNA design. Adapted from [72]

Primer	Sequence (5′-3′)	Purpose
sgRNA-forward	ACACCgNNNNNNNNNNNNNNNNNNNG	To clone sgRNA into MLM3636; required guanine in lowercase
sgRNA-reverse	AAACNNNNNNNNNNNNNNNNNNcG	To clone sgRNA into MLM3636; required cytosine (complementary to guanine) in lowercase

As previously described, to introduce a mutation in homozygosity, the sgRNA should mediate the Cas9 cut at a distance of less than 10 bp of the site to be edited. Furthermore, for an heterozygous HDR event the distance ranges from 2-26 bp, but is optimal between 10-12 bp [72].

3.1.2 MLM3636 Vector Preparation

10 µg of the MLM3636 plasmid were digested with *BsmB*I restriction enzyme in order to linearize the vector. The digestion reaction was incubated for 2 hours at 37ºC. The linearized plasmid was then extracted from an 1% agarose gel and the DNA was subsequently isolated and purified using the Gel Extraction Protocol from the NZYGelpure Kit (NZYTech).

Then, 1 µg of the digested plasmid is dephosphorylated using the Fast AP Enzyme. The removal of the phosphate groups prior to cloning prevents the re-circularization of the plasmid in the presence of the ligase. This reaction was incubated for 10 minutes at 37 ºC and then 5 minutes at 75ºC. After dephosphorylation of the plasmid, we proceeded with the protocol for DNA purification from enzymatic reactions from NZYGelpure Kit (NZYTech).

3.1.3 Phosphorylation and annealing of the sgRNAs

The sense and antisense sgRNAs were annealed and phosphorylated so that they could be cloned in the MLM3636 plasmid, according to the following:

sgRNA sense (100 µM)	1.0 µL
sgRNA antisense (100 µM)	1.0 µL
10x T4 Ligase Buffer	1.0 µL
ATP (NTPs, 10mM)	1.0 µL
PNK	0.5 µL
RNase/DNase Free H_2O	up to 10.0 µL

The reactions were incubated for 30 minutes at 37°C, followed by 5 minutes at 98°C. The temperature was allowed to drop slowly by placing the blocker in the bench at room temperature (RT) for approximately 2h. Each annealed guide was then diluted in a proportion of 1:250.

3.1.4 Ligation of the sgRNAs and cloning vector

50 ng of the dephosphorylated linearized plasmid (deP MLM3636 × BsmBI) were used for the ligation of the phosphorylated annealed sgRNAs. Ligation reaction was set up as following and incubated at 16ºC for 16h.

10x Ligase Buffer	1.0 µL
deP MLM3636 x BsmBI **(**50 ng/µL)	1.9 µL
Insert DNA (1:250 annealed sgRNA)	1.0 µL
ATP (from [NTP's]= 10 mM)	0.5 µL
T4 DNA Ligase (1U)	1.0 µL
RNase/DNase Free H_2O	up to 10.0 µL

The ligation reactions were then used for transformation in DH5α *E.coli* competent bacteria.

3.1.5 Transformation

DH5α competent bacteria (100 µL) were transformed with 5 µL of each ligation reaction.

Briefly, in a sterile 1.5 mL Eppendorf™ tubes, 5 µL of the ligation reaction was added to 100 µL of competent DH5α *E.coli* bacterial suspension and incubated on ice for 10 minutes. Subsequently, the bacteria were heat shocked at 42 °C for 45 seconds and afterwards incubated for 2 minutes on ice.

500 µL of LB medium (liquid lysogeny broth or Luria-Bertani media) was added to each transformation reaction and incubated at 37 °C for 1 hour, under agitation (220 rpm). The bacterial suspensions were briefly spin-downed at 200xg in a microcentrifuge. The majority of the supernatant was discarded, and the cell pellet was resuspended in the remaining volume. The resuspended cells were then transferred to a LB Agar-Ampicillin (100 µg/mL) petri dish. The cells were spread across the dish using a sterile plastic loop to ensure the obtention of isolated colonies and incubated overnight at 37 °C. Considering that the MLM3636 plasmid encodes for the β-lactamase gene that confers ampicillin resistance to the bacteria, this antibiotic was used to select transformed bacteria.

The following day, at least three colonies were picked and inoculated individually in 5 mL of liquid LB medium with ampicillin in a 50 mL Falcon tube. Individual colonies were picked with a sterile toothpick, avoiding contact with other colonies. The toothpick was placed inside the 50 mL Falcon, and the bacterial suspension was incubated overnight at 37ºC, 220 rpm. The next day we proceeded with the Mini Prep Protocol for isolation and purification of plasmid DNA from NZYTech.

3.16 Glycerol Stocks

Glycerol stocks allow for long-term storage of bacterial cultures. In order to prepare stocks of the transformed bacteria with the cloned plasmids, 500 µL of each bacterial culture was added to a cryovial. Afterwards, 500 µL of 30% sterile glycerol (vol/vol) in LB medium was added to the cryovial and the glycerol stocks were immediately stored at -80°C. These bacterial stocks can then be used for

direct inoculation into liquid LB or LB-Agar plates by simply touching the contents of the cryovial with a plastic loop, and then re-storing at -80°C .

3.2 Reprogramming and CRISPR/Cas9 Plasmids Preparation

The necessary plasmids for the reprogramming of human fibroblasts and gene-editing, reported in the paper "Simultaneous reprogramming and gene editing of human fibroblasts" published by S. E. Howden, *et al.*, were ordered from Addgene [75].

All plasmids used in this study are presented in **Figure 8.1**.

Reprogramming

pEP4 E02S ET2K	(Addgene, plasmid ID 20927)
pEP4 E02S EN2L	(Addgene, plasmid ID 20922)
pEP4 E02S EM2K	(Addgene, plasmid ID 20923)
pSimple-miR302/367	(Addgene, plasmid ID 98748)
pSP6-EBNA$^{2A + DBD}$	(Addgene, plasmid ID 98749)

Gene-Editing

pDNR-SpCas9-Gem	(Addgene, plasmid ID 80424)

The transformation protocol was performed as described in **Section 3.1.5.** 1 mL of the bacterial culture was inoculated in 200 mL of LB medium (with 200 µL of ampicillin) in a 250 mL Erlenmeyer and grown overnight at 37ºC, 220 rpm. The following day, bacterial cultures were centrifuged for 10 minutes at 5000 rpm, and isolation of the plasmid DNA was achieved using the Maxi Preps Protocol from NZYTech.

3.3 *In vitro* Transcription

The *in vitro* transcription of the pSP6-EBNA$^{2A + DBD}$ and pDNR-SpCas9-Gem plasmids was performed using the mMessage mMachine SP6 Transcription Kit and the mMessage mMachine T7 Ultra Transcription Kit, respectively.

3.3.1 Linearization of plasmid DNAs

10 µg of the pDNR-SpCas9-Gem and pSP6-EBNA$^{2A + DBD}$ plasmid DNAs (100 ng/µL) were linearized by setting up the following digestion reactions and incubating them at 37°C for 2 hours.

pDNR-SpCas9-Gem	8.0 µL	pSP6-EBNA$^{2A + DBD}$	14.5 µL
10x CutSmart buffer	10.0 µL	10x CutSmart buffer	10.0 µL
PmeI (10 U/µL)	10.0 µL	EcoRI-HF (10 U/ µL)	5.0 µL
RNase/DNase Free H_2O	up to 100.0 µL	RNase/DNase Free H_2O	up to 100.0 µL

The reactions were terminated through the addition of 5 µL 0.5M EDTA (at a final concentration of 6.8 mM), 10 µL of 5M sodium acetate (137 mM) and 250 µL of 100% ethanol. They were subsequently incubated at −20°C for 15 minutes.

The DNA was pelleted through a 15-minute centrifugation at 17000xg at RT and ressuspended in 10 µL of Tris-EDTA (TE) buffer to achieve a final concentration of linearized plasmids of 0.3 – 1 µg/µL.

3.3.2 Transcription Reactions

The transcription reactions were assembled using 3 µg of the linearized plasmids, according to what is described below and incubated at 37ºC for 4 hours.

pDNR-SpCas9-Gem (T7 mMessage mMachine Ultra Kit)

2x NTP/CAP (15 mM)	30.0 µL
GTP (20 mM)	1.5 µL
10x Reaction buffer	6.0 µL
Linearized pDNR-SpCas9-Gem	9.0 µL
Enzyme mix	6.0 µL
RNase/DNase Free H_2O	up to 60.0 µL

pSP6-EBNA$^{2A + DBD}$ (mMessage mMachine SP6 Transcription Kit)

2x NTP/CAP (15 mM)	30.0 µL
GTP (20 mM)	1.5 µL
10x Reaction buffer	6.0 µL
Linearized pSP6-EBNA$^{2A + DBD}$	4.0 µL
Enzyme mix	6.0 µL
RNase/DNase Free H_2O	up to 60.0 µL

After 4h incubation, 3 µl of Turbo DNase was added to each reaction and incubated at 37°C for 15 minutes.

In the case of pSP6-EBNA$^{2A + DBD}$ the reaction was stopped, and the mRNA was precipitated by adding 90 µL of RNase/DNase Free H_2O and 90 µL of lithium chloride (LiCl) precipitation solution. The mRNA was precipitated overnight for 24 hours at − 20°C. The following day, the mRNA was centrifuged for 15 minutes 17x000g at 4°C. The supernatant was removed, the pellet was washed with approximately 1 mL of 70% ethanol (vol/vol) and centrifuged again under the same conditions. Finally, the mRNA pellet was ressuspended in 100 µL of RNase/DNase Free H_2O.

The *in vitro* transcribed mRNA may be stored for over 12 months either at -20 or -80°C. As opposed to the EBNA-1 mRNA, the Cas9-Gem mRNA was subjected to a further polyadenylation reaction in order to allow protein translation upon cell transfection.

3.3.3 Cas9-Gem mRNA polyadenylation

The poly(A) tailing reagents (T7 mMessage mMachine Ultra Kit) were added to the *in vitro* transcription reaction of the SpCas9-Gem.

mRNA	63.0 µL
RNase/DNase Free H_2O	108.0 µL
5x E-PAP buffer	60.0 µL
$MnCl_2$ (25 mM)	30.0 µL
ATP solution (10 mM)	30.0 µL
E-PAP enzyme	12.0 µL
Final volume	303.0 µL

The reactions were incubated at 37°C for 45 minutes, and then stopped by the addition of 300 µL of LiCl. The poly(A)-mRNA was precipitated overnight at − 20°C. The mRNA was centrifuged resorting to the same conditions as mention above for the EBNA-1 mRNA and the pellet was resuspended in 80 µL of RNase/DNase Free H_2O.

3.4 Fibroblasts Cell Culture

Four human cell lines were used in this study: the immortalized human embryonic kidney 293 cell line (HEK293T, from ATCC®), and three primary skin fibroblasts cell lines from donors of different ages, acquired from the Coriell Institute for Medical Research.

	ID
Human Fibroblasts (3 years)	GM05565
Human Fibroblasts (10 years)	GM03348
Human Fibroblasts (11 years)	GM00323

These primary cell lines are hereby designated as HuF3Y, HuF10Y and HuF11Y. Culture mediums were purchased from Thermo Fisher Scientific. For the primary cultures of human fibroblasts (HuF), Minimum Essential Medium (MEM) with Earle's Salts was supplemented with 15% fetal bovine serum (FBS) and 1% L-glutamine. For the plating of HuF cell lines, a coating of gelatin to improve cell adherence is required. Prior to plating the cells, a volume of gelatin sufficient to cover the entire surface of the flasks/plates used was added and incubated for 20-30 minutes at RT.

Regarding HEK293T cells, Dulbecco's Modified Eagle Medium (DMEM 1x) was supplemented with 10% FBS, 1% L-glutamine and 1% Pen-Strep. Cells were maintained in an incubator at 37°C with 5% CO_2.

3.4.1 Thawing of HuF and HEK293T cells

A cryovial of each cell line was removed from liquid nitrogen storage and thawed at 37ºC for 1-2 minutes. Once completely thawed, 1 mL of medium was added to the cell suspension, and the contents of the cryovial were transferred to a sterile 15 mL Falcon tube already containing 5 mL of growth medium. The cells were centrifuged for 3 min at 200xg and most of the supernatant was removed. The pellet was ressuspended in medium and the cells were plated in a T75 Flask. The medium was changed the day after thawing to remove dead cells and every 2 days after that.

3.4.2 Cryopreservation and Expansion of HuF and HEK291T cells

For cryopreservation, cells were washed by rinsing with Phosphate Buffered Saline (PBS) 1x and then a sufficient volume of 0.025% Trypsin-EDTA (1x) was added to cover the cells. Cells were incubated at 37°C for approximately 3-5 minutes to promote enzyme activity. Once the cells were detached from the growing surface, they were ressuspended in medium. The cell suspension was then transferred to a 15 mL Falcon tube and centrifuged for 3 min at 220×g. The cell pellet was ressuspended in 1 mL of cryopreservation medium, composed of 10% DMSO in FBS and transferred to a cryovial. Cryotubes were placed in a controlled-rate freezing cooler, essential for the successful cryopreservation of the cells and the maintenance of their viability, at -80°C for 48h. Afterwards, cryovials were transferred to liquid nitrogen.

For the amplification of HuF and HEK293T cells, the same procedure described above was followed, with the exception that, following centrifugation, the pellet was resuspended in medium and the cell suspension was plated according to the intended dilution. Cells were passed each time they reached 70-80% confluency. Since they are primary cell cultures, for the expansion of HuF cells, the splitting was performed in a proportion of 1:2. Given that HEK293T are immortalized cells that present a much higher growth rate, passages between 1:20 – 1:50 were carried out.

3.5 Testing sgRNA efficiency in HEK293T

3.5.1 Cell counting

HEK293T cells were detached as described in **Section 3.4.2**. After centrifugation, the pellet was ressuspended in a small volume of medium (1-2 mL). Approximately 10 µL of the single-cell suspension was mixed with a 10 µL of Trypan Blue and this mixture was then pipetted into an hemocytometer for cell counting. The unstained live cells were counted in at least 4 quadrants, and an estimative of cell density per millilitre was obtained.

3.5.2 Transfection of HEK293T

Twenty-four hours before transfection, ~130 000 cells were plated per each P24-well plate. On the following day, cells should be at 70-90% of confluency and fresh medium was added to the cells prior to transfection.

HEK293T cells were transfected usingX-tremeGENE9 DNA Transfection Reagent. 50 µL of Opti-MEM was added to a sterile 1.5 mL Eppendorf tube. Subsequently, 1 µL of X-tremeGENE9 Reagent, 400 ng of pCas9-GFP and 150 ng of each sgRNAs to be tested were added to the eppendorf and carefully mixed. This mixture was incubated for 15 minutes at RT (15-25°C). The transfection complex was then added to the cells in a dropwise manner and the plate was swirled to ensure even distribution. The cells were incubated for 48 hours at 37°C with 5% CO_2.

3.5.3 Cleavage Assay

The efficiency of sgRNA was assessed using the GeneArt® Genomic Cleavage Detection Kit. Transfected cells were trypsinized 48 hours after transfection with 0.025% Trypsin-EDTA (1x) and resuspended in PBS 1x. The cell suspension was spin-down for 220xg for 5 minutes at 4°C. The supernatant was carefully discarded, and the cell lysis protocol from the Cleavage Detection Kit was performed. At this stage, as an alternative, the pellet may also be stored at –80°C.

3.5.3.1 Cell Lysis

A mix of 50 µL of Cell Lysis Buffer with 2 µL of Protein Degrader per condition was prepared. 50 µL of this mix was then used to resuspend each cell pellet, and the contents were then transferred to a PCR tube. Tubes were incubated in a thermal cycler at 68°C for 15 minutes and at 95°C for 10 minutes. Next, cell lysates were briefly vortex, and used as template for PCR amplification of the target genomic region. A control reaction to ensure the good performance of the GeneArt® Genomic Cleavage Detection Kit was also prepared. The following reagents were added to a respective PCR tube:

	Sample	**Control**
Cell lysate	2.0 µL	-
10 µM F/R Primer Mix	1.0 µL	-
Control Template & Primers	-	1.0 µL
AmpliTaq Gold® 360 Master Mix	25.0 µL	25.0 µL
RNase/DNase Free H_2O	up to 60.0 µL	

The following program was ran in a thermal cycler:

95°C	10 min	1x
95°C	30 seconds	40x
55°C	30 seconds	
72°C	30 seconds	
72°C	7 minutes	1x
14°C	∞	

3.5.3.2 Denaturation and random re-annealing

This step allows for the random annealing of PCR fragments with and without indels, culminating with the formation of heterogeneous DNA duplexes Thus, 1.0 µL of 10x Detection Reaction Buffer and 5.5 µL of RNase/DNase Free H_2O were added to a PCR tube containing 2.5 µL of a given PCR product. Negative controls for each sample were also prepared.

The subsequent program was set on a thermal cycler:

95°C	5 minutes
95°C–85°C	–2°C/second
85°C–25°C	–0.1°C/second
14	∞

3.5.3.3 Enzyme digestion

The heteroduplex DNAs that contain mismatches were cleaved by the Detection Enzyme. Briefly, 1 µL of Detection Enzyme was added to the re-annealing reaction and 1 µL of RNase/DNase Free H_2O was added to the negative control samples. Reactions were incubated at 37°C for 1 hour and samples were subsequently ran on a 2% agarose gel (0.5x TBE, EtBr) at 120 V for 45 minutes.

3.6 HDR Template Design

Templates for HDR were designed accordingly to the instructions defined in the publication "Precise and efficient scarless genome editing in stem cells using CORRECT" [72]. Templates consist of 100 nucleotides long single-stranded oligonucleotides (ssODN), centred around the Cas9 cut site which have the exact same sequence of the target region, with the exception of the mutation to be introduced in the genome. In addition, the oligonucleotide template is designed for the same DNA strand that the sgRNA anneals to, i.e. it has the same sequence of the sgRNA used, in order to avoid base-pairing between the sgRNA and the repair template.

3.7 One-step Reprogramming and gene-editing of human fibroblasts

An overall timeline of the one-step reprogramming and gene-editing of human fibroblasts process is process is presented in **Figure 3.1**

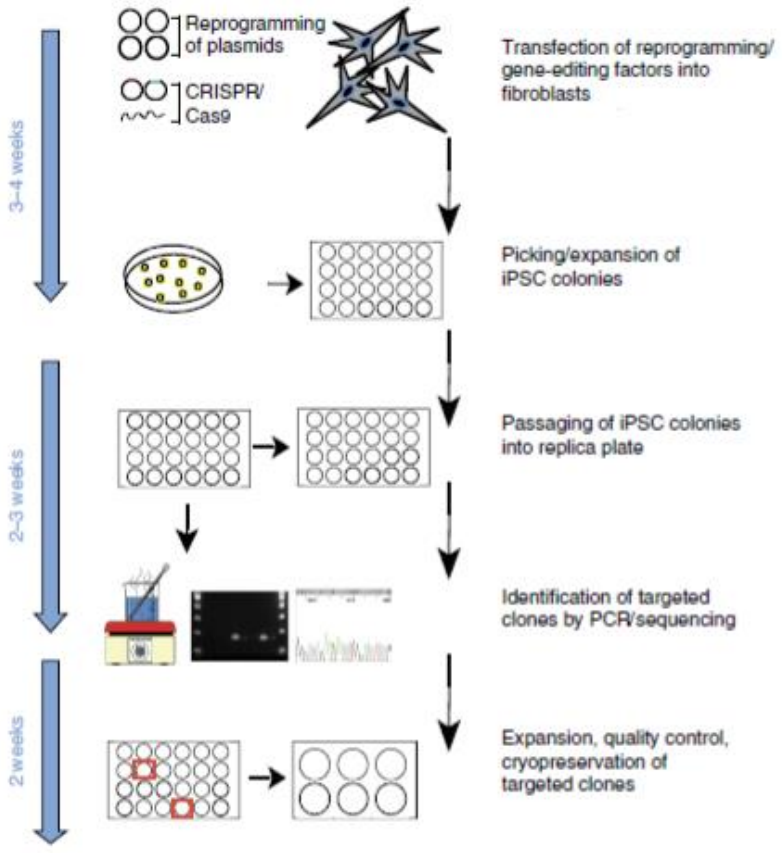


Figure 3.1 – Time line and overview of the processes involved in the one-step reprogramming and gene-editing of human fibroblasts protocol. Adapted from [75].

3.7.1 Transfection of Reprogramming and CRISPR/Cas9 Factors

The reprogramming and gene-editing factors were transfected through electroporation, using the Neon® Transfection System from Invitrogen™.

3.7.1.1 Matrigel Coated Plates Preparation

Matrigel (Corning®) was used as an adhesion substrate for all hiPSCs cell lines. In order to prepare Matrigel coated plates, aliquots were thawed on ice and diluted in a 1:100 proportion in cold DMEM/F12 medium. The diluted Matrigel was then added to the culture plates and left at RT for 2 hours or at 37ºC for 45 minutes to allow polymerization prior to use. Matrigel-coated plates can be stored at 4°C up to 2 weeks. Given that HuF were to be reprogrammed into hiPSCs, the electroporated fibroblasts were plated in Matrigel-coated plates.

3.7.1.2 Electroporation of HuF

Two days after passage, 70% confluent HuF were detached and counted as described in **Section 3.5.1.** HuF 3Y were in passage 13, and HuF 10 and 11Y were in passage 15 at the time of transfection.

Cells were resuspended in Resuspension Buffer R (provided in the Neon Transfection Kit) to achieve a density of $1.0x10^7$ cells per mL. The cells were gently pipetted to obtain a single-cell suspension and kept on ice until electroporation. The following amounts of the DNAs and mRNAs to be transfected were added to 1.5 mL Eppendorf tubes.

pEP4 E02S ET2K	2.5 μg	
pEP4 E02S EN2L	2.5 μg	
pEP4 E02S EM2K	2.5 μg	
pEP4 E02S EM2K	2.5 μg	
pSimple-miR302/367	2.5 μg	
EBNA1 mRNA	5.0 μg	
Cas9-Gem mRNA	5.0 μg	
sgRNA plasmid	2.0 μg	
ssODN (HDR template)	0.3 nmol	→ Gene-editing

mRNAs were added immediately prior to the electroporation to minimize potential degradation by contaminating RNases.

A Neon tube containing 3 mL of Electrolytic Buffer E2 was placed in the Neon pipette station. 100 μL of cell suspension (1.0×10^6 cells) were transferred to the eppendorf containing the DNAs and mRNAs and gently mixed by up-and-down motion. The mixture was then carefully aspirated into the 100-μl Neon tip to avoid formation of air bubbles and placed on the pipette station. Electroporation was performed in the following conditions: 1400 V; 20 milliseconds; 2 pulses.

The electroporated cells were subsequently plated in a Matrigel-coated 6-well with HuF medium and incubated at 37ºC 5% CO_2. Old culture medium was changed on the next day to remove dead cells and debris.

3.7.2 Reprogramming of Human Fibroblasts

On day 4 post-transfection, the fibroblast medium was removed and replaced with reprogramming medium (TesR-E7 supplemented with 100 μM of sodium butyrate). The medium was changed every 2 days until hiPSC colonies began to appear. Once the first hiPSC colonies started to emerge, the reprogramming medium was substituted by Essential 8™ (ThermoFisher), a medium specific for the growth and expansion of human pluripotent stem cells [82]. Daily changes were performed after this point.

3.7.3 Picking of hiPSC colonies

hiPSC colonies should ideally have ~1 mm in diameter when they are picked. Picking iPSCs colonies that are in close proximity to each other should be avoided, given that it could result in the mix of different clones. 24-well Matrigel-plates were prepared for plating of picked hiPSC clones. Prior to picking of the colonies, the medium of the 6-well was changed to Essential 8™ supplemented with 10 μM ROCK inhibitor (Y-27632), an anti-apoptotic molecule that enhances the survival of hiPSCs when they are dissociated at a single-cell stage [83]. hiPSC colonies were scraped using a 200-μL pipette tip and the detached cells were slowly aspirated into a 24-well containing Essential 8™ + ROCK inhibitor. The cells were placed in a 5% CO_2 incubator at 37 °C for 48 hours.

Two days after picking, old medium must be replaced by fresh Essential 8 in order to remove dead cells and the ROCK inhibitor. After that, the Essential 8 medium was changed every day until the iPSC clones have grown enough to be split.

3.7.4 Splitting of hiPSC colonies

hiPSC colonies were often split using the PBS-EDTA dissociation reagent (EDTA 0.5mM and 1.8 g/L of NaCl in PBS). Through this method, the hiPSCs remain in small clusters rather than in a single-cell suspension, increasing their viability and the addition of ROCK inhibitor is not necessary.

Briefly, old medium was removed, and an enough volume of the PBS-EDTA to cover the cells was added. Cells were incubated at room temperature for 3-5 minutes, after that EDTA solution was carefully aspirated, and Essential 8™ medium was added. The cells were mechanically detached using a scraper and resuspended in the fresh medium.

Approximately half of the cell suspension is maintained in culture and plated in another 24-well coated with Matrigel, consisting of a 1:2 passage. The other half is used for DNA isolation, PCR amplification and sequencing of the target region (**Chapters 3.10 - 3.12**).

3.7.5 Cryopreservation of hiPSCs

Cells were detached as mention above in **Section 3.6.4.** The cell suspension was then transferred to a 15 mL Falcon tube and centrifuged for 3 minutes at 200xg. Cell pellet was resuspended in 250 μL of hiPSCs freezing solution, comprised of 10% DMSO in KnockOut Serum Replacement and transferred to a cryotube and stored at – 80 °C for 48 hours.

3.8 Editing of Human Induced Pluripotent Stem Cells

hiPSCs lines were cultured in mTeSR™1 or mTeSR™ Plus mediums from STEMCELLS Technologies.

Three different hiPSC lines are used in the laboratory. The DF6-9-9T.B hiPSC line was acquired from WiCell® and is a vector-free cell line that was reprogrammed from donor's foreskin fibroblasts. The Gibco Human Episomal iPSC Line was obtained from ThermoFisher and is derived from $CD34^+$ cord blood using an EBNA-based episomal system. The TCLab cell line was generated from fibroblasts using a retroviral system and kindly offered by the laboratory Professor Joaquim Cabral Institute for Bioengineering and Biosciences (iBB).

3.8.1 Thawing and Maintenance of hiPSCs

iPSCs were thawed in a similar manner to what was described for human fibroblasts in **Section 3.4.1.** The cryovials were partially thawed for approximately 30 seconds at 37 °C and then fully thawed by adding 1 mL of mTeSR™. Cell suspension was then transferred to a Falcon and centrifuged for 3 min at 200xg. The pellet was resuspended in medium and plated in Matrigel coated dishes.

When using mTeSR™1 to maintain iPSCs in culture, the medium was changed every day. In contrast, mTeSR™ Plus has a newer stabilized formulation that allows for medium changes every other day and skipping of 2 days by double-feeding the cells when they are at a lower confluency.

hiPSCs were split with PBS-EDTA as specified in **Section 3.7.4** in ratios ranging from 1:2 - 1:6 for expansion.

3.8.2 Selection Strategies

In contrary to what happens in the one-step reprogramming and gene-editing protocol, where all hiPSCs clones that are generated from human fibroblasts are picked, when performing gene-editing in hiPSCs a selection strategy is required. In this work, two different conditions were tested as selection strategies of transfected cells.

One of the most commonly used Cas9 variants is the pCas9-GFP (already used for the testing of sgRNAs efficiency, **Chapter 3.5**), that is fused with GFP and consequently allows for the selection of clones by fluorescence-activated cell sorting (FACS) of GFP^+ cells. Thus, GFP-positive cells were known to have been successfully transfected with the Cas9-GFP plasmid. When using the Cas9-Gem mRNA, the same variant used in human fibroblasts, an alternative selection strategy was required. In this case, a pPUR vector that codes for a gene that confers resistance to the antibiotic puromycin was co-transfected. Therefore, puromycin was used to select transfected clones.

3.8.3 Transfection of hiPSCs

hiPSCs were by rinsing with PBS 1x and incubated at 37 °C for 7 minutes with Accutase®, a cell detachment solution composed of proteolytic and collagenolytic enzymes that is a direct alternative to trypsin and suitable for more sensitive cells, and that is used to obtain single-cell suspensions of hiPSCs [84]. Once the cells were detached, two volumes of medium were added to neutralize Accutase® and the suspension was pipetted up and down and transferred to a 15 mL Falcon. The tubes were then centrifuged at 200xg for 3 minutes and pellets were resuspended in mTeSR™1 supplemented with 10 μM ROCK inhibitor.

Cells were counted as described in **Section 3.5.1.** and $200x10^5$ cells were plater per 12 well-plate coated with Matrigel. The following day, medium was changed to remove dead cells and we proceeded with the transfection. The transfection of the gene-editing factors was performed by lipofection, a method that consists on the formation of DNA-lipid complexes to deliver the DNAs/mRNAs to the cells, and that is less aggressive to the cells when compared to electroporation [85]. The Lipofectamine® 3000 Reagent from Invitrogen was used.

The DNA-lipid complexes were prepared in 1.5 mL Eppendorf™ tubes according to the following:

Tube A:	**Tube B:**
100.0 µL Opti-MEM medium	100.0 µL Opti-MEM medium
1.5 µL Lipofectamine® 3000 Reagent	0.1 µg sgRNA
	0.4 µg Cas9-Gem mRNA or pCas9-GFP
	0.6 µg ssODN
	0.1 µg pPUR *
	2 µL/µg DNA P3000 Reagent

* The pPUR plasmid is added for co-transfection with the Cas9-Gem mRNA.

Reactions were incubated at RT for 5 minutes and the DNAs/mRNAs (Tube B) were then added to the tube containing the Lipofectamine® 3000 Reagent (Tube A) in a dropwise manner. The mixture was incubated at RT for 15 minutes to allow formation of the DNA/mRNA-lipid complexes, which were then slowly added to the cells. The plates were swirled to ensure even distributed across the wells. Cells were incubated at 32°C for 48 hours with no medium change.

The incubation of the cells at 32°C (so called cold shocking) after transfection was reported to increase the frequency of HDR mechanism in iPSCs [74]. 48 hours after transfection, the medium was changed and cells were placed at 37 °C for an additional 24 hours to recover before proceeding with a given selection strategy, described in **3.8.4** and **3.8.5**.

3.8.4 FACS of GFP$^+$ hiPSCs

A single-cell suspension is required for sorting of GFP$^+$ cells. Hence, the cells were detached with Accutase® and after centrifugation the pellets where resuspended in 200-1000 µL of mTeSR™ with 10 µM ROCKi. The single-cell suspensions where then filtered through a 70-µm cell strainer to remove any remaining cell clusters. Sorting was performed in the BD FACSAria IIu from BD Biosciences. The GFP$^+$ sorted cells were centrifuged and plated in a 6-well at a density ranging from 3 - 5×10^3, in order to form individualized clones. Cells were incubated at 37 °C, 5% CO_2 for 48 hours. The hiPSC clones were typically ready for picking 1.5 - 2 weeks after FACS.

3.8.5 Puromycin selection of hiPSCs

72 hours after transfection, medium was changed and 2 µg/mL of puromycin was added to each well for an additional 48 hours. After that, puromycin containing the medium was replaced by fresh medium and cells were kept at 37ºC, 5%CO_2. Clones were picked once they had the ideal size.

3.8.6 Picking of hiPSCs

hiPSC clones were picked in a similar manner to what is described in **Section 3.7.3**. In this case, clones were picked by placing a phase contrast microscope inside the flow chamber to facilitate visualization of the clones and minimize the possibility of picking more than one clone at once. Picked clones were plated in 48-well plates with mTeSR™+ROCKi. 48 hours after picking, the ROCKi was

removed and the clones were maintained in culture with daily medium changes until they had enough cells to be split.

3.8.7 Replica Splitting of hiPSC clones

The hiPSC clones were passed with Accutase. A fraction of the cell suspension, depending on the total amount of cells, was maintained in culture in a 48-well plate. The remaining cells were transferred to 1.5 mL Eppendorf tubes and used for DNA isolation and sequencing of the target region to identify correctly gene-edited clones. These methods are described in **Chapters 3.10 - 3.12.**

3.9 Cardiac Differentiation

The protocol used for 2D cardiac differentiation of hiPSCs had already been established in the laboratory [86], was adapted from [57] and is schematically presented in **Figure 3.2**.

Briefly, hiPSCs were plated in 12-well plates at a density of 4×10^5 cells per well. mTeSR was changed daily for 4 days, until cells are at 90-95% confluence, the optimal condition to start the 2D cardiac differentiation assay.

On day 0 of differentiation, 1.5 mL of RPMI/B27-insulin medium supplemented with 6 µM of CHIR99021 was added to each well. After 24 h, CHIR99021 was removed and the medium was changed to RPMI/B27-insulin. At day 3, half of the medium in the plate (750 µL) was transferred to a Falcon and 750 µL of new RPMI/B27-insulin was added to the tube. IWP2 to a final concentration of 5µM was further added. The remaining medium on the plate was aspirated and the conditioned RPMI/B27-insulin+IWP2 was added to the cells. IWP2 was removed on the medium change at day 5 (RPMI/B27-insulin). The medium was substituted by RPMI/B27 on day 7 and changed every 3 days until the end of the experiment. On day 30 of differentiation, cells were collected for RNA extraction; notice that a fraction of hiPSCs-CM population was -plated in cover slips for immunofluorescence protein detection.

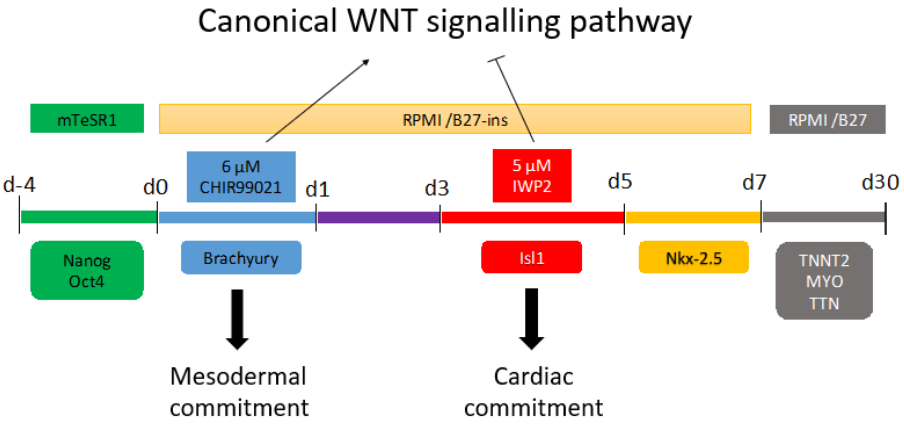


Figure 3.2 – Schematic of the 2D Cardiac Differentiation protocol. Adapted from [86].

3.10 Genomic DNA Isolation

Cells from hiPSC clones were centrifuged at 200xg for 5 minutes and 100 µL of Nuclear Lysis Buffer with Proteinase K, whose composition is described below was added to each cell pellet. The 1.5 mL Eppendorf tubes were briefly vortex and lysis was allowed for at least 4 hours (in most cases overnight) at 65°C. The volume of Nuclear Lysis Buffer added should be adapted according to the

amount of cells per pellet. In the case of hiPSC clones grown on 48-wells, 100 µL of buffer was added. For clones grown in other larger plates, the volume was scaled-up.

Nuclear Lysis Buffer:

100 µL of 1M Tris-HCl

100 µL of 3M NaCl

400 µL of 0.5M EDTA

200 µL of 10% SDS

1200 µL RNase/DNase free H_2O

20 µl of Proteinase K 10mg/mL (Final volume: 2000 µL)

After the lysis is completed, 100 µL of an EtOH/NaCl solution (1.5% 5M NaCl in absolute EtOH) was added to each pellet, in a 1:1 proportion for the amount of Nuclear Lysis Buffer previously used. Tubes were vortex and further incubated at RT for 10 minutes, and then centrifuged for 15 minutes at RT and 2200xg. The supernatant was carefully discarded, and the pellets were washed with EtOH 70%. Upon centrifugation for 10 minutes at RT, 2200xg, all the supernatant was removed, and the pellets were air-dried and resuspended in RNase/DNase free H_2O. The isolated DNAs were then quantified using the NanoDrop™ 2000 (Thermo Scientific™) system.

3.11 PCR Amplification of target region

For PCR amplification, the DNAs were diluted to a final concentration of approximately 3 ng/µL. 20.0 µL of the PCR reaction mix was distributed to each PCR tube, and 5 µL of template DNA (to a final concentration of 15 ng) was added. The NZYLong DNA polymerase from NZYTech was used.

PCR Reaction Mix:

NZYLong Reaction Buffer (10x)	2.5 µL
10 µM dNTPs	0.5 µL
10 µM Primer Fw/Rv	1.0 µL
NZYLong DNA polymerase (5 U/µL)	0.5 µL
RNase/DNase free H_2O	14.5 µL
Final volume	20.0 µL

The following PCR Program was ran in a thermal cycler.

95°C	10'	
95°C	1’	
65°C	1’30"	5×
68°C	2’	
95°C	1'	
65’°C	1’	35×
68°C	2’	
68°C	10’	
14°C	∞	

3 µL of the PCR products were then ran on a 1% agarose gel (0.5x TBE in EtBr) to confirm amplification of the target region. Afterwards, the remaining PCR product was purified using the PCR Clean protocol from the NZYGelpure Kit (NZYTech).

3.12 Sanger Sequencing

The purified PCR products were sequenced to confirm the presence of the intended genomic alteration. For Sanger Sequencing, we resorted to the services of three different companies. A mix of PCR product and primer were sent to either STAB Vida or Eurofins Genomics. Each sample sent for sequencing was prepared according to the conditions defined by each company.

In addition, sequencing of hiPSC clones was also performed in collaboration with GenoMed - Diagnósticos de Medicina Molecular, S.A, a company based in iMM.

3.13 RNA Extraction

A given cell pellet was resuspended in 800 – 1000 µL of NZYol (NZYTech®) and incubated for 5 minutes at RT under rotation. After that, samples may be frozen at -80°C and stored for up to one year.

RNA isolation was performed by adding 200 µL of chloroform to the NZYol extract and shaking the tube vigorously. This was followed by an incubation of 3 minutes at RT under rotation and centrifugation at 11900×g for 15 minutes at 4 °C. Later, the upper aqueous phase (that is mostly constituted by RNA) was carefully transferred to a new 1.5 mL Eppendorf tube. 600 µL of isopropanol was added to the aqueous phase and mixed well by vortex in order to allow RNA precipitation. In this step, 1 µL of GlycoBlue Coprecipitant may also be added. The samples were incubated for 10 minutes at RT and followed by a centrifugation at 11900xg for 10 minutes at 4 °C. The supernatant was carefully discarded, and the RNA pellet was washed with 1 mL of 75 % EtOH and vortex before being centrifuged at 11900xg for 10 minutes at 4 °C. After the supernatant was cautiously discarded once more, the RNA pellet was air-dried at RT and resuspended in 13 µL of RNase/DNase free H_2O and quantified in Nanodrop®.

3.13.1 DNase I Treatment

To remove possible contaminating DNA, a DNase I treatment was performed. 1 - 5 µg of RNA can be treated under these conditions, and the volume of RNA and RNase/DNase free H_2O added is dependent on the concentration of the RNA sample.

1 - 5 µg of RNA	
10x DNase I buffer (Roche®)	5 µL
DNase I (Roche®)	1.0 µL
Ribosafe (RNase Inhibitor, Bioline)	1.0 µL
RNase/DNase free H_2O	up to 50 µL

The reaction was incubated for 30 minutes at 37 °C. To stop the reaction and precipitate the purified RNA, 3V of 100% EtOH (150 µL), 0.1V of AcNa 3M (5 µL) and 1 µL GlycoBlue were added. This mixture was briefly vortex and incubated overnight at -20°C. RNAs were then centrifuged at full speed (~13000xg) for 30 minutes at 4°C. The supernatant was carefully removed, and the RNA pellet was washed with 1 mL of 75% EtOH and re-centrifuged at full speed for 15 minutes at 4°C. After

discarding the supernatant very carefully, the RNA pellet was air dried and resuspended in 13 μL of RNase/DNase free H_2O and quantified in Nanodrop®. Next, we proceeded with the cDNA synthesis protocol. At this point, RNAs may be stored at -80ºC for several months.

3.14 cDNA Synthesis

cDNA synthesis from the isolated total RNA was performed using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche®). To a 200 μL PCR tube, 600 ng of RNA, 2 μL of Random Primers and RNase/DNase free H_2O up to a final volume of 9.4 μL were added. The reaction was mixed by pipetting up and down and briefly centrifuged.

In a thermal cycler, samples were incubated for 10 minutes at 65ºC and immediately placed on ice; then 8.6 μL of the Reverse Transcription mix was added to each cDNA synthesis reaction.

RT mix per cDNA reaction

RT Buffer 5x	4.0 μL
RNase OUT (40U/μL)	0.5 μL
dNTPs 10mM	2.0 μL
Dithiothreitol (DTT) 0.1M	1.0 μL
High Fidelity Reverse Transcriptase	1.1 μL

The tubes were again placed on the thermal cycler and incubated for 10 minutes at 29ºC, 60 minutes at 48ºC and 5 minutes at 85ºC. The cDNA synthetized for RNA can be stored at -20°C.

3.15 Quantitative real-time PCR

The qRT-PCR protocol was performed using the Universal SYBR Green Supermix (Bio-Rad). cDNA samples were diluted in a 1:15 proportion. For each gene to be analysed, a mix with specific primers was prepared.

Primer mix per reaction:

SYBR Green Supermix	6.0 μL
Primer Forward 10 μM	0.5 μL
Primer Reverse 10 μM	0.5 μL

When setting up the qRT-PCR 384 well plate, 7 μL of each primer mix was pipetted into the respective well. Subsequently, 5 μL of each diluted cDNA was added to the corresponding well. The plate was briefly centrifuged for 3 minutes at 700×g and RT. The q-RTPCR reactions were ran in the Applied Biosystems® ViiA™ 7 Real-Time PCR System, using the standard q-RTPCR protocol that used 60ºC as the primer annealing temperature.

3.16 Immunofluorescence

For immunofluorescence, hiPSC-CMs on day 30 of differentiation were re-plated using 0.025% Trypsin-EDTA (1×) and grown on coverslips in 6-wells coated with Matrigel. Day after replatting, RPMI/B27 medium replaced by fresh one, and cells were maintained in culture for an additional 3 days. Then, the medium was removed, and cells were washed three times with 1 mL of PBS 1x.

Cells were then fixed with 1.5 mL of 3.7% paraformaldehyde (PFA) in PBS for 10-12 minutes at RT and then washed twice with 1 mL of PBS 1x.

In order to permeabilize the cells, they were incubated with 0.5% Triton X-100 in PBS 1x for 10 minutes. After permeabilization, cells were washed three times (10 minutes each) with 0.05% TWEEN 20 in PBS 1x.Cells were then covered with blocking solution (5% FBS in PBS 1x) and incubated for 30 minutes at RT.

6 µL of the primary antibody diluted in 1% FBS was placed directly on top of the coverslips and incubated for 1 hour at RT and then overnight at 4ºC in the dark.

Primary Antibodies

α-Actinin (1:75)

Abcam (ab-9465)

α-MYBPC3 (1:50)

Santa Cruz Biotechnology (sc-137181)

α-MYH7 (1:20)

Santa Cruz Biotechnology (sc-53090)

After incubation with the primary antibody, the cells were washed 3× with 1 mL of PBS-TWEEN 0.05% and incubated with 6 µL of the secondary antibody diluted in 1% FBS and incubated for 45 minutes in the dark.

Secondary Antibody

Goat α-mouse Alexa 488 (1:250)

Thermo Fisher (A10684)

6 µL of Phalloidin-TxRed (Thermo Fisher, T7471), diluted in 60 µL of DMSO at a concentration of ~ 165 µM (1:1000 dilution in PBS 1x), a probe that stains filamentous actin, was placed on top of the coverslip and incubated for 1 hour in the dark.

Subsequently, 5 µL of 4',6-diamidino-2-phenylindole (DAPI, Enzo Life Sciences 52404) diluted in sodium bicarbonate (1µg/mL) was added on top to label DNA.

Finally, the cells were washed 3 times for 10 minutes with 1 mL of 0.05% PBS-TWEEN, and then 1 time with PBS 1x.

The coverslip was mounted in 1.6 µL of VECTASHIELD® Mounting Medium and the sides were sealed by painting around the edges with nail polish.

3.17 Materials and Reagents

Table 3.2 – List of the reagents and materials used throughout this project.

Reagents			Reference
Cell Culture			
DMEM 1x		Gibco	41966
MEM 1x		Gibco	10370
OptiMEM 1x		Gibco	31985
TeSR-E7 Basal Medium		STEMCELL	05910
Essential 8 Basal Medium		Gibco	A15469-01
mTeSR 1		STEMCELL	85851
mTeSR Plus		STEMCELL	05826
RPMi 1640 1x no glucose		Gibco	11879-020
TeSR-E7 25x Supplement		STEMCELL	05910
TeSR-E7 500x Supplement		STEMCELL	05910
Essential 8 Supplement (50x)		Gibco	A15171-01
mTeSR1 Supplement		STEMCELL	85852
mTeSR Plus 5x Supplement		STEMCELL	05827
B27 Supplement		Gibco	17504-044
B27 Minus Insulin		Gibco	A18956-01
CHIR 99021		Sigma-Aldrich	SML1046-5MG
IWP2		Sigma-Aldrich	I0536-5MG
Sodium Butyrate		Enzo Life Sciences	ALX-270-301-6001
Fetal Bovine Serum		Gibco	10270-106
KnockOut SR		Gibco	10828-028
Y-27632 ROCK inhibitor		STEMCELL	72302
L-Glutamine 200mM		Gibco	25030-024
Pen Strep		Gibco	15070-063
0,25% Trypsin-EDTA		Gibco	25200-072
Accutase		Sigma-Aldrich	A6964-500ML
DPBS 1x		Gibco	14190
Matrigel Matrix		Corning	354230
Lipofectamine 3000		Invitrogen	L3000-001
X-TremeGene 9 DNA		Roche	06365787001
Puromycin		Sigma-Aldrich	P8833-25MG
DMSO			
Laboratory Reagents and Enzymes			
BsmBI (Esp3I)	10 U/µL	Thermo Scientific	ER0451
FastAP	1 U/µL	Thermo Scientific	EF0654
T4 PNK	10 U/µL	Thermo Scientific	EK0032
T4 DNA ligase	1 U/µL	Thermo Scientific	EL0016
PmeI (MssI)	5 U/µL	Thermo Scientific	ER1341
EcoRI-HF	20 U/µL	New England Biolabs	R3101S
NZYlong DNA Pol	5 U/µL	NZYTech	MB00303

DNase I 10 U/µL	Roche	04716728001
Ribosafe RNase Inhibitor 40U/µL	Bioline	BIO-65028
10x Buffer Tango	Thermo Scientific	BY5
10x CutSmart Buffer	New England Biolabs	B7204S
10x NZYLong Buffer	NZYTech	MB00303
10x T4 DNA ligase 10x buffer	Thermo Scientific	B69
10x T4 ligation buffer	Roche	11243292001
10x DNase I buffer	Roche	04716728001
iTaq Universal SyBr Green Supermix	Bio-Rad	1725120
dNTPs (dATP, dTTP, dCTP, dGTP) 100mM	NZYTech	MB08701
Proteinase K	Sigma-Aldrich	EO0491
NZYol	NZYTech	MB18501
Paraformaldehyde	Sigma-Aldrich	P6148
Kits		
Transcriptor High Fidelity cDNA Synthesis Kit	Roche	05081963001
mMessage mMachine SP6 Transcription Kit	Thermo Scientific	AM1340
mMessage mMachine T7 Ultra Transcription Kit	Thermo Scientific	AM1345
NZYMixiprep	NZYTech	MB051
NZYMiniprep	NZYTech	MB010
NZYGelpure	NZYTech	MB011
Neon Transfection System	Invitrogen	MPK5000
Bacterial Culture		
Glycerol	Sigma-Aldrich	G6279
LB Broth	Sigma-Aldrich	L3022
LB Broth with agar	Sigma-Aldrich	L2897
Ampicillin sodium salt	Sigma-Aldrich	A9518
Materials		
TPP® Tissue Culture Plates	Sigma-Aldrich	Z707910
Nunc™ CryoTube Vials	Thermo Scientific	5000-0020
MicroAmp™ Fast Optical 96-Well Reaction Plate	Applied Biosystems	4346906
MicroAmp™ Optical 384-Well Reaction Plate	Applied Biosystems	4309849

4. Results

4.1 Selecting the Variants

Considering that the main goal of this study is to understand how missense variants can affect the splicing pattern of a pre-mRNA and its potential impact in the disease mechanism of HCM, the first task consisted on the selection of the variants to be tested and further introduced in the human genome through the CRISPR/Cas9 technology. To do so, online databases where variants associated with disease are deposited and bioinformatic tools that predict the possible effect of a specific variant on splicing were used.

The ClinVar database, available in http://www.ncbi.nlm.nih.gov/clinvar/ was explored to select the variants to be studied. ClinVar consists of an online repertoire in which human mutations, identified either by clinical testing or through research, and their associated phenotypes are reported [87]. In this database, there are 7277 variants listed as being related to HCM. Over 5000 of these variants are classified as missense, comprising approximately 70% of all variants HCM-variants deposited in this database. However, scanning over 5000 variants for a possible effect on splicing is not feasible and certain filters were applied in ClinVar to restrict the number of variants to be analysed.

Thus, only the mutations that were characterized as being pathogenic and that had no conflicting interpretations of pathogenicity, which correspond to 342 of the initial pool of missense variants were selected. Subsequently, mutations located within the three genes most commonly mutated in HCM (*MYH7*, *MYBCP3* and *TNNT2)* were chosen, further limiting the overall number of variants. Finally, filters that would allow for a higher degree of confidence in the classification of these variants as pathogenic were applied. In the case of *MYH7*, the variants selected were reviewed by an expert panel that approved the classification of these mutations as pathogenic. For *MYBPC3* and *TNNT2*, since an expert panel revision was not available, only variants that were submitted by multiple entities who were all in agreement regarding their classification were chosen. By applying these filters, a final list of twenty variants in *MYH7*, four in *MYBPC3* and one in the *TNNT2* gene was obtained and is presented in **Table 4.1**. These variants were then analysed through a bioinformatic software called Human Splicing Finder (HSF), that is able to predict how a specific variant might have an impact on splicing. This tool employs several algorithms to evaluate the strength of new donor and acceptor splice sites, as well as *cis*-acting elements such as enhancer and silencer motifs, comparing their strength to the consensus sites and issuing the respective score that reflects the probability of splicing alterations due to a specific mutation [88]. According to HSF, out of the twenty-fve previously selected variants, twenty have a potential effect on splicing. The predicted outcomes in splicing caused by these variants are also showcased in **Table 4.1**.

Additionally, these variants were also analysed through the MaxEntScan method. Similar to the approach used in HSF, this splice site modelling tool assigns a score to the 5' and 3' splicing sequences and compares the 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 [89][90]. For the characterization of the splicing effects of these variants, the guidelines that were followed dictated that if the score variation between the wild-type and mutated sequence is higher than 10% for HSF and 30% for MaxEnt, there is a high probability to have created a new splice site. In contrast, when the scores are of less than -10% and -30%, it is correlated with a high probability to have disrupted a splice site.

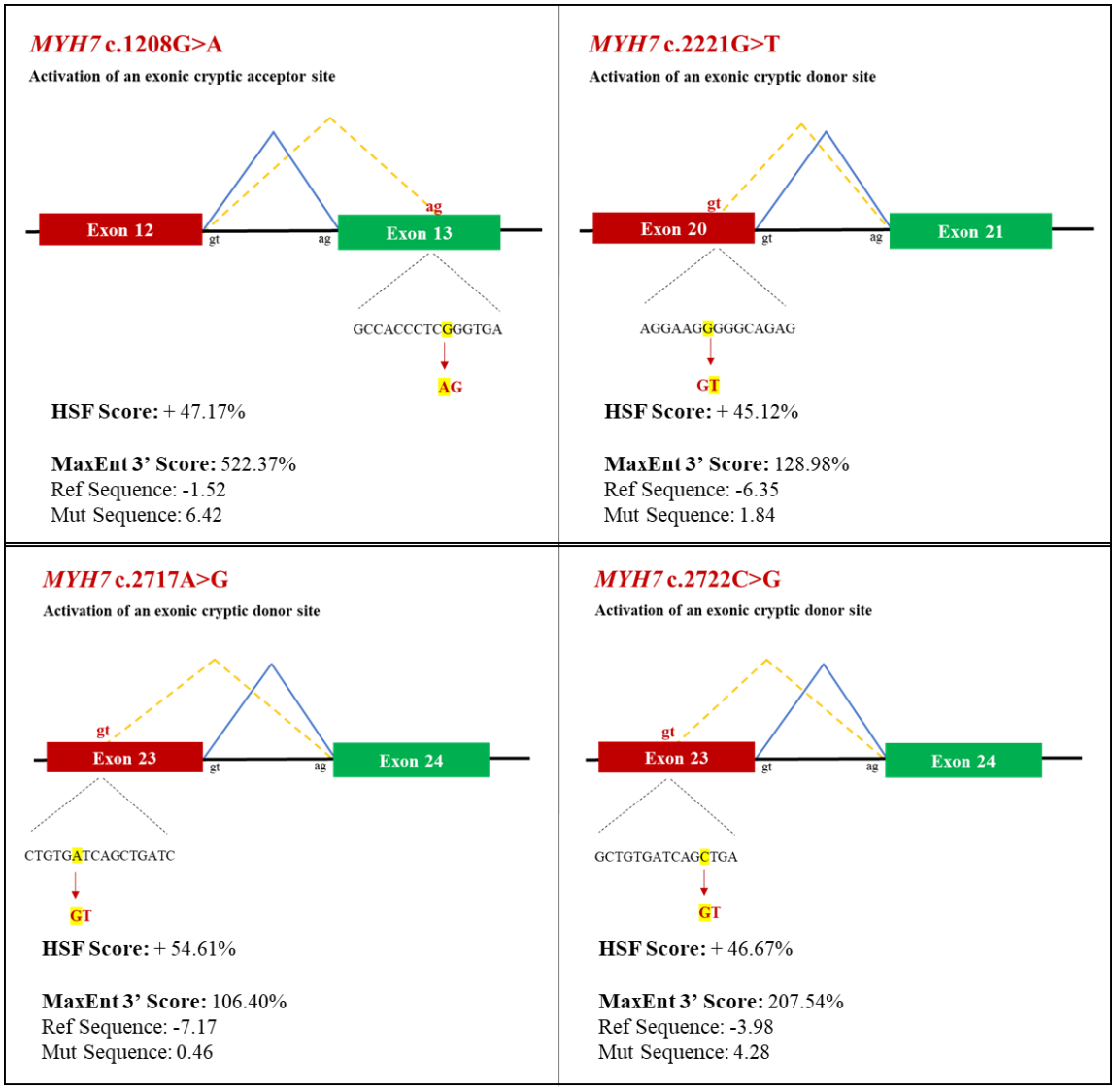
Considering the splicing-altering scores obtained from the bioinformatic tools previously mentioned, the variants that had higher scores and, thus, a superior probability of affecting splicing were selected. Notice that the variants that theoretically lead to the disruption of the wild-type splice sites or to the creation/alteration of cryptic sites presented higher scores than the ones that likely affect exonic splicing enhancers and/or silencers motifs.

Table 4.1 – HCM-related variants in *MYH7, MYBPC3* and *TNNT2* sarcomeric genes reported in ClinVar. Only variants that were classified both as "Missense" and "Pathogenic" were taken into consideration for this study. These variants were then analysed by the Human Splicing Finder software, and the predicted effects on splicing are presented. The following abbreviations are used in this table: Exonic Splicing Enhancer (ESE) and Exonic Splicing Silencer (ESS).

Nucleotide Substitution	Aminoacid Alteration	Variant ID ClinVar	Predicted Effect on Splicing (HSF)
MYH7 **NM_000257.4**			
c.788 T>C	p.Ile263Thr	43106	Alteration of an ESE site
c.1207 C>T	p.Arg403Trp	14102	Creation of an ESS site
c.1208 G>A	p.Arg403Gln	14087	Activation of a cryptic acceptor site Creation of an ESS site
c.1357 C>T	p.Arg453Cys	14089	Creation of an ESS site Alteration of an ESE site
c.1358 G>A	p.Arg453His	42838	Alteration of an ESE site
c.1750 G>C	p.Gly584Arg	14090	Alteration of an ESE site
c.1988 G>A	p.Arg663His	42875	-
c.2146 G>A	p.Gly716Arg	14105	Alteration of an ESE site Creation of an ESS site
c.2155 C>T	p.Arg719Trp	14104	Creation of an ESS site
c.2156 G>A	p.Arg719Gln	14107	-
c.2167 C>T	p.Arg723Cys	14095	Alteration of an ESE site
c.2167 C>G	p.Arg723Gly	42885	Alteration of an ESE site Creation of an ESS site
c.2207 T>C	p.Ile736Thr	164342	-
c.2221 G>C	p.Gly741Arg	14098	-
c.2221 G>T	p.Gly741Trp	177665	Activation of a cryptic donor site
c.2513 C>T	p.Pro838Leu	42910	Creation of an ESS site
c.2609 G>A	p.Arg870His	14120	Alteration of an ESE site
c.2681 A>G	p.Glu894Gly	42922	Alteration of an ESE site
c.2717 A>G	p.Asp906Gly	14125	Activation of a cryptic donor site Creation of an ESS site
c.2722 C>G	p.Leu908Val	14097	Activation of a cryptic donor site Creation of an ESS Alteration of an exonic ESE
MYBPC3 **NM_000256.3**			
c.206 G>A	p.Arg69Gln	42594	Alteration of an ESE site
c.481 C>T	p.Pro161Ser	518242	Activation of a cryptic donor site Alteration of an ESE site
c.1090 G>A	p.Ala364Thr	454301	Alteration of the WT donor site
c.2265 C>A	p.Asn755Lys	407308	-
TNNT2 **NM_001001430.2**			
c.275 G>A	p.Arg92Gln	12409	Activation of a cryptic acceptor site Creation of an ESS site Alteration of an ESE site

Therefore, in result of this bioinformatic analysis, a total of seven variants were selected to be further studied: four mutations in the *MYH7* gene (c.1208G>A, c.2221G>T, c.2717A>G and c. 2722C>G), two in *MYBPC3* (c.481C>T and c.1090G>A) and one in *TNNT2* (c.275G>A).

Schemes of the predicted effect of these variants on the splicing mechanism and their respective HSF and MaxEnt Scores that supported their selection are presented in **Figure 4.1**. Additionally, given that our hypothesis is that missense mutations might cause the disease because they affect splicing and not only due to their direct effect on the encoded aminoacid, a splicing-altering mutation control is required. Hence, the *MYBPC3* splice-site mutation (c.927-2A>G) was selected as it was already confirmed to lead to aberrant splicing in lymphoblasts cells of HCM patients [91]. The mechanism and scores associated to this splice-site mutation are also shown in **Figure 4.1**.



(continues on the next page)

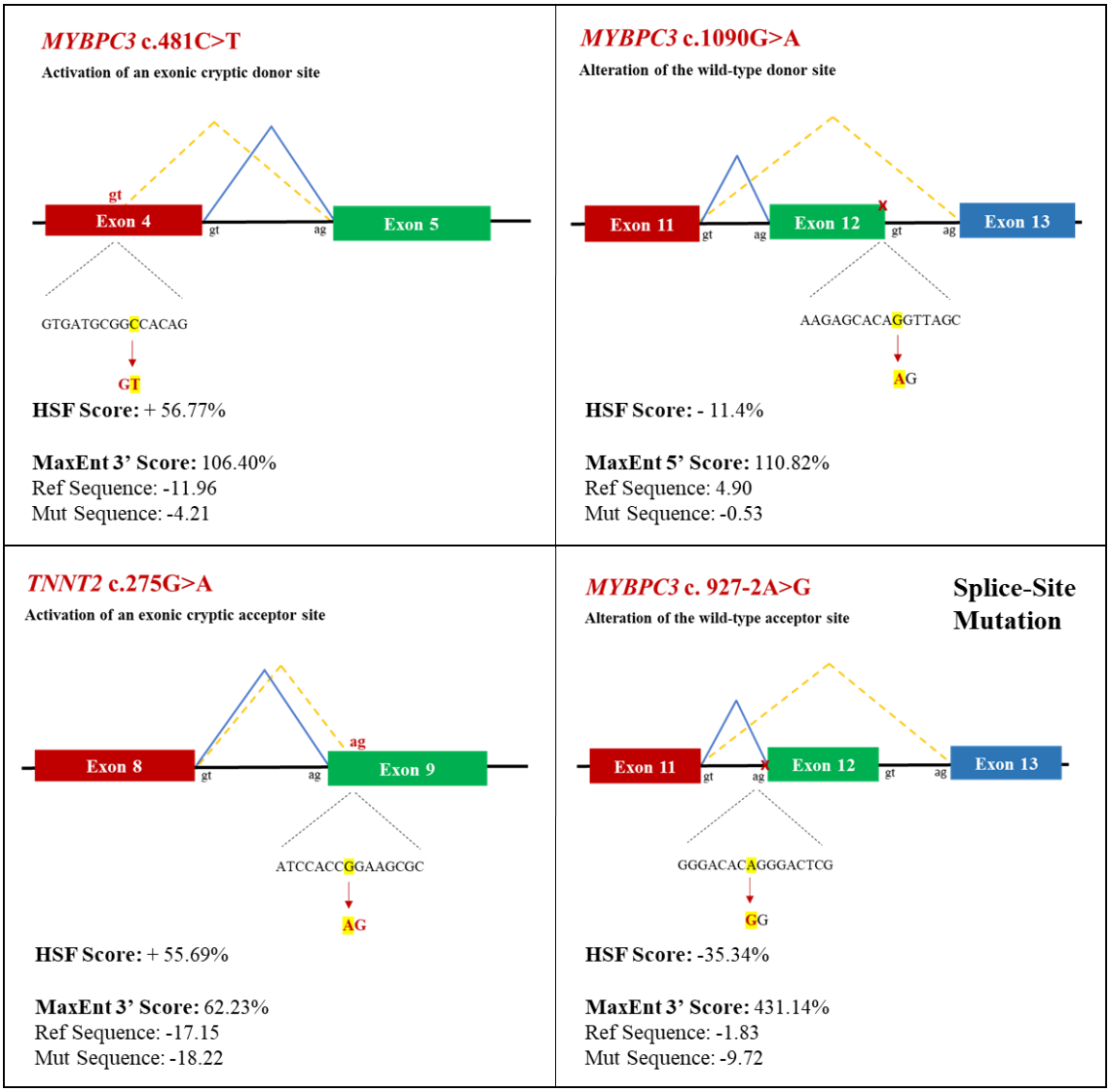


Figure 4.1 – Schemes that depict the normal intron splicing (blue) and the hypothesized prediction of the splicing alteration cause by the selected missense variants (in orange, dashed lines). The HSF and Max Ent Scores for each variants are also presented. In addition to the chosen missense variants, the *MYBPC3* c.927-2A>G mutation, located in the canonical acceptor site and that serves as a positive control for a variant that alters splicing is also presented.

Interestingly, as portrayed in **Figure 4.1**, the *MYBPC3* c.1090G>A mutation potentially disrupts the recognition of the wild-type donor site and is predicted to lead to the skipping of exon 12, the same outcome hypothesized for the *MYBPC3* c.927-2 A>G splice site mutation. Furthermore, two of the selected missense variants hypothetically lead to the activation of exonic cryptic acceptor sites (*MYH7* c.1208G>A and *TNNT2* c.275G>A) and the remaining four may result in the activation of exonic cryptic donors (*MYH7* c.2221G>T, c.2717A>G, c.2722 C>G and *MYBPC3* c.481C>T).

4.2 Evaluating sgRNAs Efficiency

sgRNAs for the previously selected missense and splice-site variants were designed using a bioinformatic tool developed by the Zhang Lab, previously available at crispr.mit.edu, that has since been shut down. For each variant, a minimum of two sgRNAs were selected according to their scores that range from 0-100. The higher the score, the more efficient is, in theory, the sgRNA.

Thus, the choice of sgRNAs was based on a compromise between higher scores, fewer off-target events and the distance from the cut location to the site where the mutation is to be introduced. All sgRNAs chosen had no predicted off-targets within genes. The sequences of the sgRNAs selected are presented in **Table 8.1**. The sgRNAs were then cloned into the MLM3636 plasmid as previously described in **Section 3.1.** and sequenced to confirm the correct insertion of the sgRNA sequence in the vector.

Nevertheless, the results of the bioinformatic tool used for sgRNA design must be validated by an *in vitro* approach. Thus, the cleavage efficiency mediated by the sgRNAs was further evaluated using the GeneArt® Genomic Cleavage Detection Kit, following sgRNAs and pCas9-GFP transfection in HEK293T cells. This procedure was performed for every sgRNA designed for each one of the selected variants. The three sgRNAs that showcased a higher efficiency were selected: sgRNA-1 *MYH7* c.2221G>T, sgRNA-4 *MYBPC3* c.1090 G>A and sgRNA-1 *MYBPC3* c.927-2A>G. These results are presented in **Figure 4.2.**

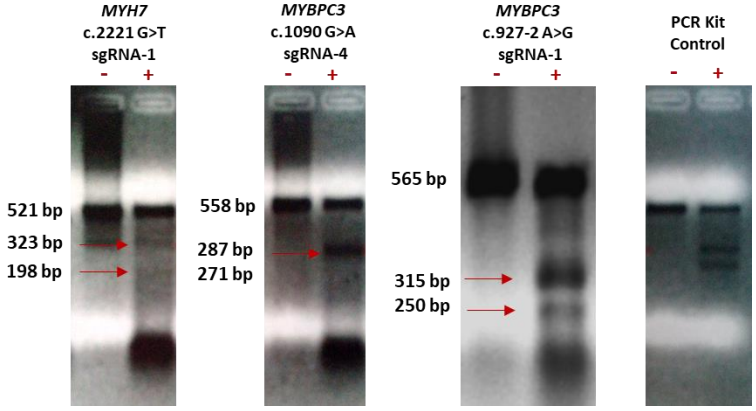


Figure 4.2 – Results of the Cleavage Assay. (-) is the control with H_2O and (+) is the reaction to which the Detection Enzyme was added (GeneArt® Kit). Samples were ran on a 2% agarose gel (0.5x TBE, EtBr) at 120 V for 45 minutes. Two bands corresponding to the cleaved DNA (arrows in red) with sizes consistent with the distance to the cut site are visible for the 3 sgRNAs presented, indicating that they are efficient and can be further used for CRISPR/Cas9 gene-editing assays. The results from the PCR Kit sample are also showed, a kit positive control.

If the sgRNAs are efficient in the guidance of the Cas9 to the intended genomic target site, the nuclease introduces a DSB that, in absence of an HDR template, is repaired by NHEJ resulting in the incorporation of indel mutations. So, after amplification of the targeted genomic region, PCR products are denaturated and randomly reannealed, resulting in the formation of DNA heteroduplexes that contain mismatches, which are subsequently cleaved by the Detection Enzyme. Thus, a band with the size of the PCR product (corresponding to uncleaved DNA) and two smaller bands corresponding to the digested PCR products are obtained when samples are run on an agarose gel. The sizes of these two bands that indicate cleavage by the Cas9 depend on the distance from the region where the primers anneal to the site where the Cas9 introduces the DSB. The list of primers used for the amplification of each target region are presented in **Table 8.2.**

As presented in **Figure 4.2**, for the specific case of the *MYBPC3* c.1090G>A mutation, the cut site is located approximately in the middle of the region amplified by PCR, and the two bands obtained after digestion with the enzyme are very similar in sizes - 271 and 287 bp. Since the resolution of the agarose gel is not sufficient to discriminate between these two bands, they appear as a single one. For *MYH7* c.2221G>T and *MYBPC3* c.927-2A>G, the bands obtained due to cleavage of the amplicons are asymmetric, with 198/323bp and 250/315bp, respectively, appearing as two separate bands on the gel. Thus, from the analysis of these results it is possible to conclude that these three guides are efficient in inducing a DSB, via Cas9, at their target genomic region.

Hence, these were the 3 variants selected to be introduced in the human genome by CRISPR/Cas9: the splice-site mutation that serves as a positive control for the experiment, and two missense variants that may have an impact on the splicing pattern - *MYBPC3* c.1090G>A, that may disrupt the wild-type splice site, and *MYH7* c.2221G>T that is predicted to activate a donor cryptic site. Subsequently, the homologous direct recombination templates were designed for each variant, as defined in **Section 3.6**. The ssODN templates are presented in **Table 8.3**.

4.3 *In Vitro* Transcription pSP6-EBNA$^{2A + DBD}$ and pDNR-SpCas9-Gem

In order to initiate the gene-editing approach in human fibroblasts, all necessary reprogramming and CRISPR/Cas9 plasmids were cloned as previously described. Given that in the case of the Cas9-Gem nuclease and EBNA-1 the mRNAs are transfected instead of the plasmids, the pSP6-EBNA$^{2A + DBD}$ and pDNR-SpCas9-Gem were *in vitro* transcribed. The pSP6-EBNA$^{2A + DBD}$ and pDNR-SpCas9-Gem were linearized prior to being *in vitro* transcribed. The reaction was then ran on an agarose gel to ensure the accurate linearization of the plasmids before following with the *in vitro* transcription, as depicted in **Figure 4.3**.

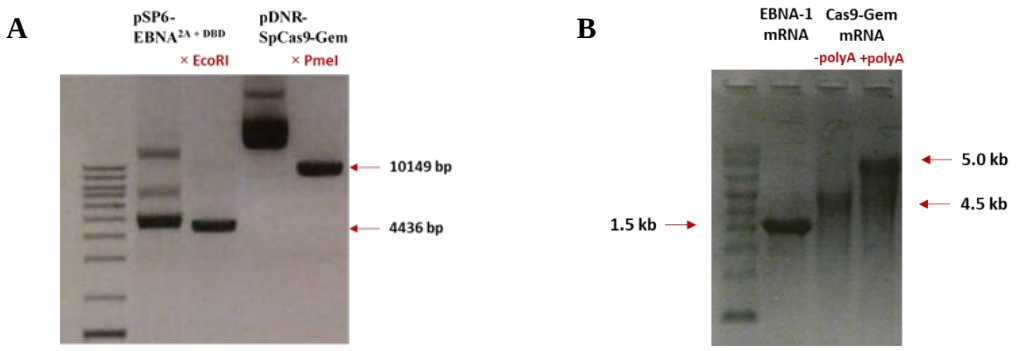


Figure 4.3 – Linearized and in vitro transcribed pSP6-EBNA2A + DBD and pDNR-SpCas9-Gem **(A)** Linearized pSP6-EBNA$^{2A + DBD}$ and pDNR-SpCas9-Gem were ran on a 0.8% agarose gel (0.5× TBE, EtBr) at 100 V for 45 minutes, next to the non-linearized plasmids as a control. After confirming correct linearization of the two plasmids, *in vitro* transcription was performed. **(B)** *In vitro* transcribed EBNA-1 and Cas9-Gem (–) and (+) poly(A) mRNAs were ran on a 1% agarose gel for 1 hour at 100 V.

Subsequently, the linearized plasmid DNA was *in vitro* transcribed using the T7 and SP6 RNA polymerase for Cas9-Gem and EBNA-1, respectively. The Cas9-Gem mRNA was then polyadenylated. The resulting mRNAs were ran on a gel to confirm their quality and to make sure they presented the correct sizes. These results are presented in **Figure 4.3B**, and both EBNA-1 and Cas9-Gem poly(A) mRNAs have the expected sizes, with 1.5kb and 5.0kb, respectively.

4.4 One-step Reprogramming and Gene-Editing of HuF 3, 10 and 11Y

First, the three human fibroblast cell lines were sequenced to ensure that they were wild type for the variants of interest. Sequencing results showed that the HuF 3, 10 and 11Y cell lines are wild-type for the three selected variants that are to be further studied in this project: *MYH7* c.2221G>T, *MYBPC3* c.1090 G>A and *MYBPC3* c.927-2A>G. Since this one-step reprogramming and gene-editing protocol had never been performed in our laboratory, it was initially decided to follow with only two of the selected variants. Thus, the *MYBPC3* c.927-2A>G was selected because it is the splicing-altering positive control variant, whereas *MYBPC3* c.1090 G>A was chosen over the *MYH7* c.2221G>T since it has a higher probability of affecting splicing. In this approach, the obtainment of gene-edited hiPSC

cones is dependent on the reprogramming efficiency and so a control condition in which only the reprogramming factors were transfected was also performed. Thus, a total of 9 electroporations were executed: the two variants of interest and the reprogramming control for each of the 3 cell lines.

On day 7 of reprogramming, three days after the fibroblast medium being replaced to a reprogramming-inducing one, morphological changes started to appear in some of the tested conditions. These changes were much more evident in the HuF3Y *MYBPC3* c.927-2A>G condition. In this case, approximately after 2 weeks into the reprogramming process, cells with a very distinct morphology from that of human fibroblasts started to form colony-like structures, that resembled those of hiPSCs. In contrast to the typical morphology of human fibroblasts, these cells were not as elongated and appeared to be assembling into a more compact, rounded agglomerate of cells. When these clusters started to emerge in some regions of the well, the medium was replaced to Essential 8, which is specific for the growth of hiPSCs. Subsequently, on day 21 of reprogramming, these structures had the desirable size and were picked. In **Figure 4.4A**, a series of images that depict the evolution and changes in cell morphology that occurred during the reprogramming process in the HuF3Y *MYBPC3* c.927-2A>G condition are presented.

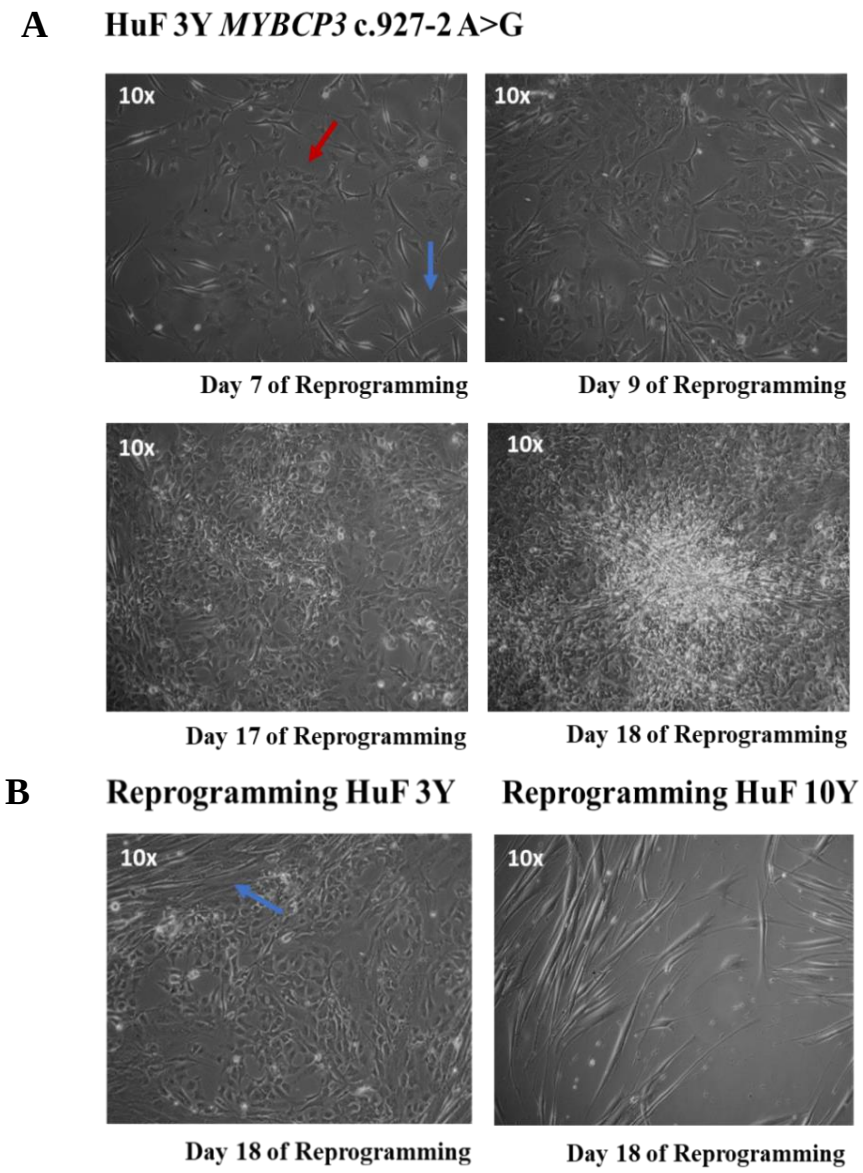


Figure 4.4 – Images that illustrate the changes in cellular morphology that occurred during the reprogramming process of human fibroblasts into hiPSCs. **(A)** In the HuF3Y, *MYBPC3* c.927-2A>G condition, on day 7 of reprogramming, cells with a different morphology to that of fibroblasts (arrow in blue) started to be evident (arrow in red). **(B)** Morphology of the Reprogramming HuF 3Y and 10Y cells.

However, as showed in **Figure 4.4B**, the reprogramming process was only efficient in the HuF3Y *MYBPC3* c.927-2A>G condition. In the others, such as the case of control, non-edited, Reprogramming HuF3Y, cells with a similar morphology to those evident in HuF3Y *MYBPC3* c.927-2A>G condition started to be distinguishable at around day 10 of reprogramming, but by day 18 there were no further changes and the morphology of these cells did not progress into a more hiPSC-like state, as what had happened with the HuF3Y *MYBPC3* c.927-2A>G. Additionally, it is important to note that these small alterations in morphology were restricted to very few regions of the culture plate. In an attempt to understand if these cells could further advance in the reprogramming process, these regions were individually picked so that their growth would not be restricted by fibroblasts. However, after picking these regions, all of the cells died.

Strikingly, on the other seven conditions tested in this experiment no morphological changes were detected, as also shown in **Figure 4.4B**, and all cells kept the typical human fibroblasts morphology. Thus, the reprogramming efficiency on this first assay was lower than expected.

Therefore, further analysis could only be carried out with hiPSCs clones from the HuF3Y *MYBPC3* c.927-2A>G condition. A total of forty-five hiPSC clones were picked. However, only twenty-two of these survived, as several picking strategies were tested and in many cases the clones did not recover after picking. These 22 clones were maintained in culture, DNA was isolated, and the target region was amplified by PCR in order to proceed with sequencing and identification of possible gene-edited clones. However, all clones sequenced were wild-type for the variant of interest, and none presented any alterations in the form of indels that would indicate cleavage by the Cas9-Gem nuclease. Even though none of the clones analysed were successfully gene-edited, they were used for the characterization of the reprogramming method hereby used. Despite the fact that these cells clearly have a very distinct morphology from that of human fibroblasts, the colonies formed did not completely recapitulate the hiPSCs morphology. In the panel presented below, in **Figure 4.5**, images of some of the picked clones can be compared with the HuF3Y cell line from which the cells were reprogramed, and with the control DF6 hiPSC line.

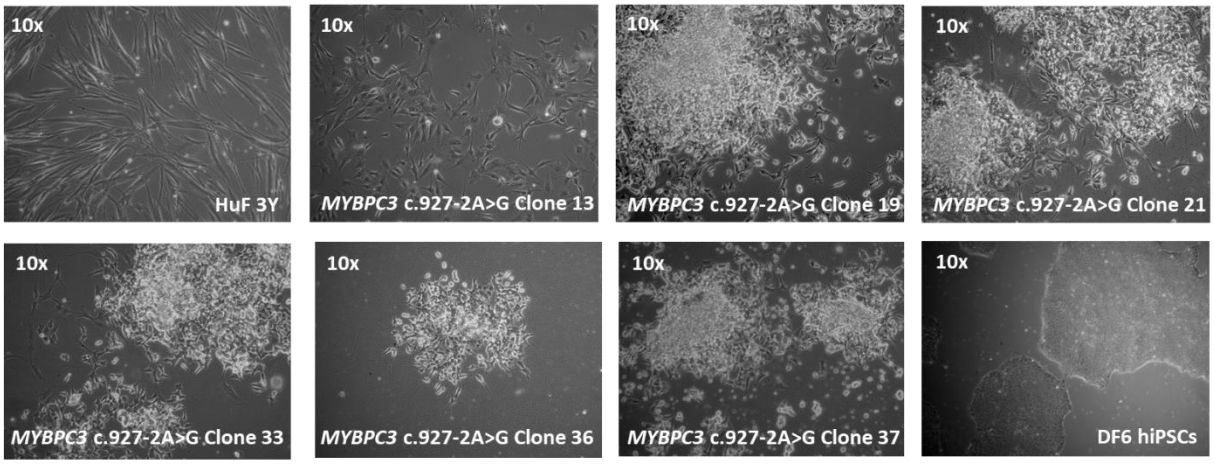


Figure 4.5 – Images of some of the picked clones that demonstrate the evident alteration in cell morphology and their tendency of these cells to form agglomerates and colony-like structures, in a completely distinct matter to what happens in normal fibroblasts.

In the images presented above, the differences between the fibroblast cell morphology and the cells that were obtained through the reprogramming process are evident. When fibroblasts grow to confluency, they tend to align in the same direction, but they do not form colonies. In addition, these cells appear to be smaller and not as elongated as fibroblasts. The morphology and growth rates of the different clones are also variable. Clone 13, shown in **Figure 4.5**, has a morphology that is much more similar to that of HuF cells. This can be explained not only by possible heterogeneity between the reprogramming efficiencies of different clones, but also by technical issues while picking these clones, as some fibroblasts might also be scraped and aspirated. However, it is clear that these cells still have a

very different morphology from a hiPSCs cell line. Even though these clones were maintained in culture and passed for over a month, there were no drastic changes in their morphology that still did not completely resemble that of hiPSCs. In contrast to the DF6 cell line, these clones did not form regular, compact and flat colonies, raising the question that these cells might not have been fully reprogrammed into hiPSCs.

Thus, in order to evaluate the pluripotency levels and characterize these presumed hiPSC clones, RNA was extracted from each one and a qRT-PCR analysis was performed. One of the most widely used markers of pluripotency is Nanog. However, Nanog is one of the transcription factors encoded in the episomal vectors that were transfected to induce reprogramming of the fibroblast cells. Although these episomal vectors are rapidly lost in fast dividing cells like hiPSCs and are usually not detectable after 4-5 passages, assessing the expression levels of a stemness-related gene that was not part of the pool of transfected factors would provide a more reliable result, considering that Nanog could still be extra chromosomally expressed in these cells. Thus, the expression levels of DNA Methyltransferase 3β (*DNMT3B*) which is involved in the *de novo* methylation of DNA during embryogenesis and is also used as an hiPSC marker were analysed [50] [75] [92]. In addition, the levels of collagen type I α-1 (*hCol1a1*), a protein that is highly expressed [93] and used as a fibroblast marker were also evaluated, in order to understand if these clones had any residual fibroblast phenotype. These qRT-PCR results are shown in **Figure 4.6.**

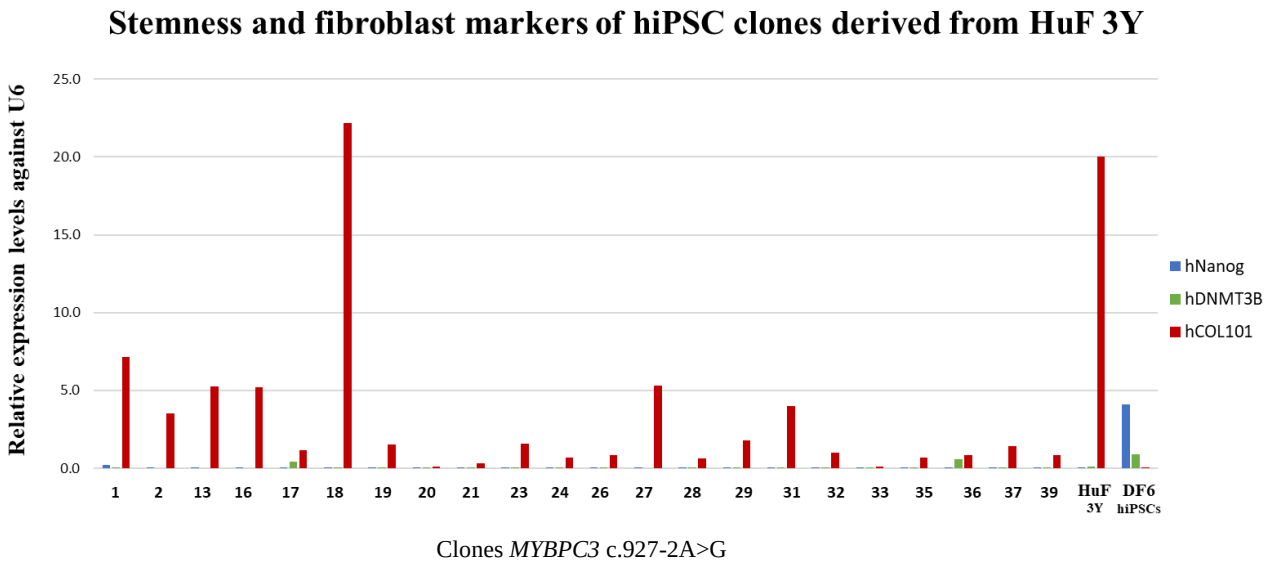


Figure 4.6 – Relative expression levels of pluripotency (*hNanog* and *hDNMT3B*) and fibroblast (*hCol1a1*) markers of the HuF 3Y *MYBPC3* c.927-2A>G clones, normalized against *U6*.

As shown in **Figure 4.6**, most clones express high levels of hCol1a1, and some even express similar levels to the original fibroblast cell line. Additionally, only a few clones express *hDNMT3B* at levels somewhat comparable to the DF6 cell line (clone 17 and 36) and hNanog expression is always very low when compared to the control DF6 hiPSCs cell line. In addition to the morphological differences observed between the picked clones after reprogramming and the DF6 hiPSC line used as a control, the clones did not express pluripotency markers in a similar level to those of hiPSCs. Even in the case of HuF3Y *MYBPC3* c.927-2A>G clones 20, 21 and 33, in which the expression of *hCol1a1* is almost undetectable, the expression levels of pluripotency markers were incomparable to ones obtained in hiPSCs.

All together, these results lead to the conclusion that these cells were not fully reprogrammed into hiPSCs, indicating that they have rather stayed in an intermediate state. Since this problem could be related with a low reprogramming efficiency, the experiment was repeated twice, and the two gene-editing conditions were used in the three cell lines. However, in none of the attempts there were changes

in the morphology of the fibroblast cells, pointing for an unsuccessfully reprogramming efficiency of this approach. Due to the low efficiency that was experienced with this protocol and adding the fact even when morphological changes were observed the cells were neither fully reprogrammed nor gene-edited, the strategy used in order to achieve gene-edited hiPSCs carrying out selected HCM variants had to be switched.

4.5 Gene-editing in hiPSCs

Given that the first task of this study is to generate gene-edited cell lines and the one-step reprogramming and editing technique failed, the CRISPR/Cas9 strategy performed directly in hiPSCs that had already been attempted in the laboratory was re-approached.

Due to the extremely low, 0.5 - 1% editing efficiencies that were previously reported by several publications, it was clear that the main issue was the reduced number of clones previously screened [72]. So, in order to obtain at least one clone with the intended mutation, the experiment needed to be performed in a bigger scale. Thus, the purpose of this first attempt was to screen between 300 - 500 hiPSC clones.

Regarding the Cas9 variant to be used, it was decided to test two different conditions to subsequently select the one that provides a higher gene-editing efficiency: the pCas9-GFP (formerly used in this work to test sgRNAs efficiencies) and the cell-cycle controlled Cas9-Gem mRNA. Furthermore, both Cas9 variant conditions were performed with and without selection of hiPSCs clones. This strategy was employed because previous attempts in the laboratory revealed that hiPSCs are very sensitive to the FACS procedure, and in several experiments, cells died after sorting for GFP positive cells, and there would be either very few hiPSCs clones or none at all to analyse. Since the sorting selection strategy appears to be a limitation, clones from the pCas9-GFP and Cas9-Gem mRNA condition were picked without selection. Due to the large scale of these assays, it was not feasible to work with different variants in simultaneous. Therefore, the splicing control was set aside, and the experiment was followed with the missense *MYBPC3* c.1090G>A variant.

The transfection of the DF6 hiPSCs line was performed as described in **Section 3.8.3**. The selection strategies employed were dependant on the Cas9 used. For the pCas9-GFP, 72h post-transfection cells were sorted and only the cells that were GFP^+, i.e. that had been transfected with the pCas9-GFP, were collected and plated in a 6-well plate. A total of 7394 GFP^+ cells with 98% purity were obtained. Concerning the cells transfected with the Cas9-Gem mRNA, the co-transfection with a plasmid that confers resistance to an antibiotic was required for selection. In this case, a pPUR plasmid was transfected along with the gene-editing factors and puromycin treatment was used for selecting transfected cells. Since after puromycin treatment the hiPSCs clones were in close proximity to each other, a few were picked directly from the 12-well plate and were subsequently diluted into a 10-cm dish. Regarding the conditions without selection, cells were diluted into 10-cm dishes 72h post-transfection and picked afterwards.

A total of >500 clones were picked from the 4 different conditions screened. During the course of the experiment, a total of 472 clones survived and were maintained in culture while the DNA procedures for sequencing of the target region were performed. Considering the two conditions without selection, a total of 144 clones were picked. 171 clones were picked from the Cas9-Gem + puromycin and 157 clones from the GFP^+. Concerning the clones picked without selection, no alterations were detected in the target region upon sequencing, and it was concluded that these settings lowered even further the probabilities of obtaining gene-edited clones. Therefore, the efficiency calculations only take into account the clones picked resorting to the two selection strategies employed.

Thus, 328 clones were picked from the two selection conditions. 46 out of the 328 clones screened contained insertion/deletion mutations indicating cleavage by the Cas9 nuclease and consisting of a 14.0% editing efficiency through the NHEJ repair mechanism. The majority of these mutated clones

(37) are from the Cas9-GFP^{+} condition. In **Figure 4.7**, chromatograms of 2 of these edited clones are presented.

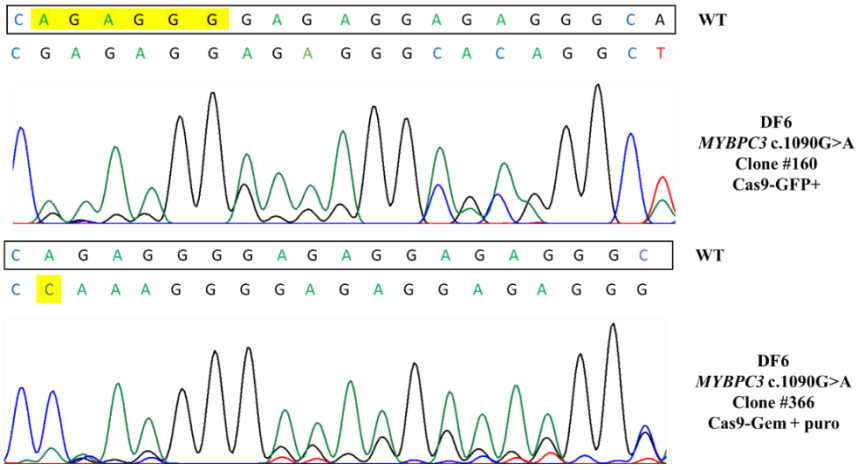


Figure 4.7 – Chromatograms of two gene-edited clones (#160 and #366) that were repaired by NHEJ and harbour indel mutations in heterozygoty.

The clones with indels presented above were characterized as: clone #160 *MYBPC3* c.1090+14_19delAGAGGG and clone #366 *MYBPC3* c.1090+14_15InsC.

Remarkably, 3 successfully gene-edited clones were identified. The DSBs in these clones were repaired with the HDR template and have the intended mutation inserted in the genome, consisting of a 0.9% editing efficiency, in accordance to the numbers described in the literature. The sequencing results of these 3 clones are shown in **Figure 4.8**. The *MYBPC3* c.1090G>A #347 heterozygous clone was isolated from the Cas9-Gem + puromycin condition, along with the homozygous clone #361. Clone #219 is from the pool of cells transfected with the pCas9-GFP that were subsequently sorted for GFP and is heterozygous for the mutation in study. In the region sequenced for all 3 clones (~500 bp), there were no other alterations detected.

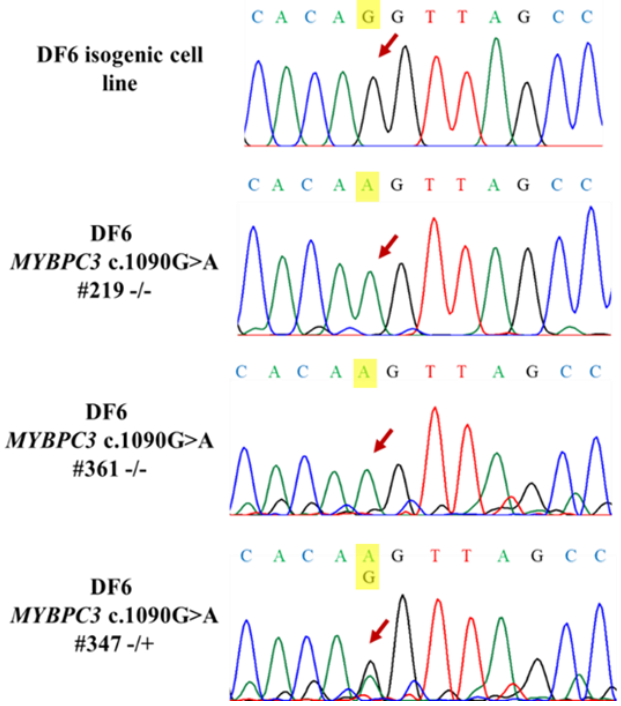


Figure 4.8 – Chromatograms of the three successfully gene-edited clones (DF6 hiPSC line). Clones #219 and #361 have the *MYBPC3* c.1090G>A variant introduced in homozygosity and Clone #347 is heterozygous for the variant in study.

However, these gene-edited clones were only identified in the last plate that was sent to sequencing. At this point of the experiment, all clones in culture had undergone evident morphological changes that indicated spontaneous differentiation, most likely due to suboptimal culture conditions. Cells of the 3 gene-edited clones have a more elongated morphology when compared to other hiPSCs, and in the specific case of clones #219 and #361 they also display long extensions. Strikingly, the cells no longer formed the colonies typical of hiPSCs. Nonetheless, the heterozygous clone #347 has a morphology more similar to hiPSCs, given that in some regions of the plate the cells assemble into small colonies, but that do not completely recapitulate the normal morphology of hiPSCs. Additionally, the division rate of these cells was lower than that of hiPSCs, and it markedly decreased with time in culture.

In accordance to these morphological and growth rate changes, it was hypothesized that these cells had likely undergone spontaneous differentiation. In the images presented in **Figure 4.9**, the contrast between a normal hiPSC cell line (DF6) and the morphology of these gene-edited clones is clear.

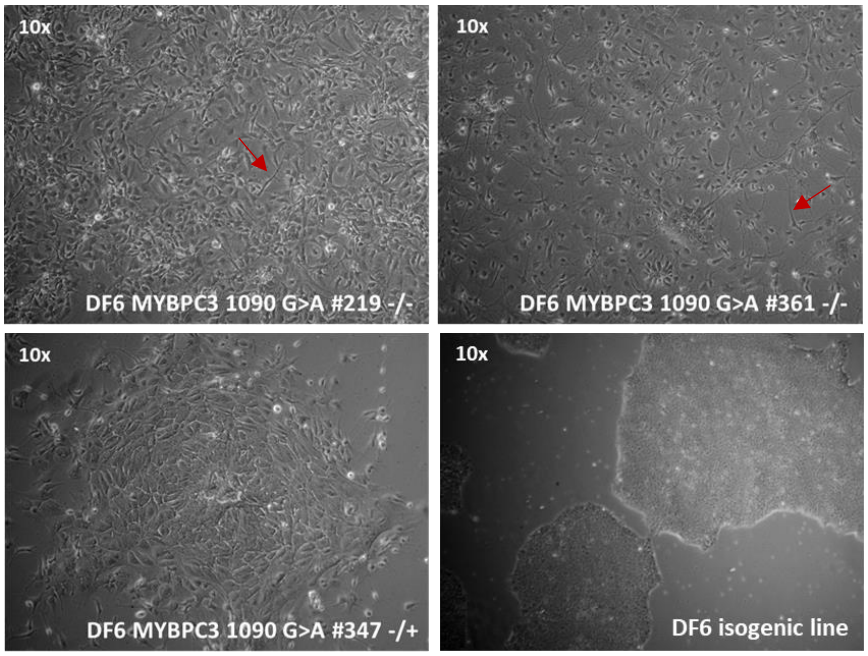


Figure 4.9 – Images depicting the distinct morphology of the gene-edited clones. In clones #219 and #361, some cells present a very elongated morphology (arrow in red).

Hence, even though their morphology presented a strong evidence that these cells lost their pluripotent nature, a qRT-PCR analysis was performed to check the levels of expression of *hNanog* and *hOct4* (stemness markers), along with the expression of markers that would indicate spontaneous differentiation into the three germ layers and specific lineages. Brachyury is a mesoderm marker, whereas Nestin is a neural progenitor that is essential during the early development of the central nervous system [56,60,94]. Pax6 is a key transcription factor that controls neurogenesis and TubIII is a mature neuron marker [95,96].

These qRT-PCR results are shown in **Figure 4.10** and demonstrate that both *MYBPC3* c.1090G>A clones #347 and #361 express the neuron marker *TubIII*, even if at low levels.

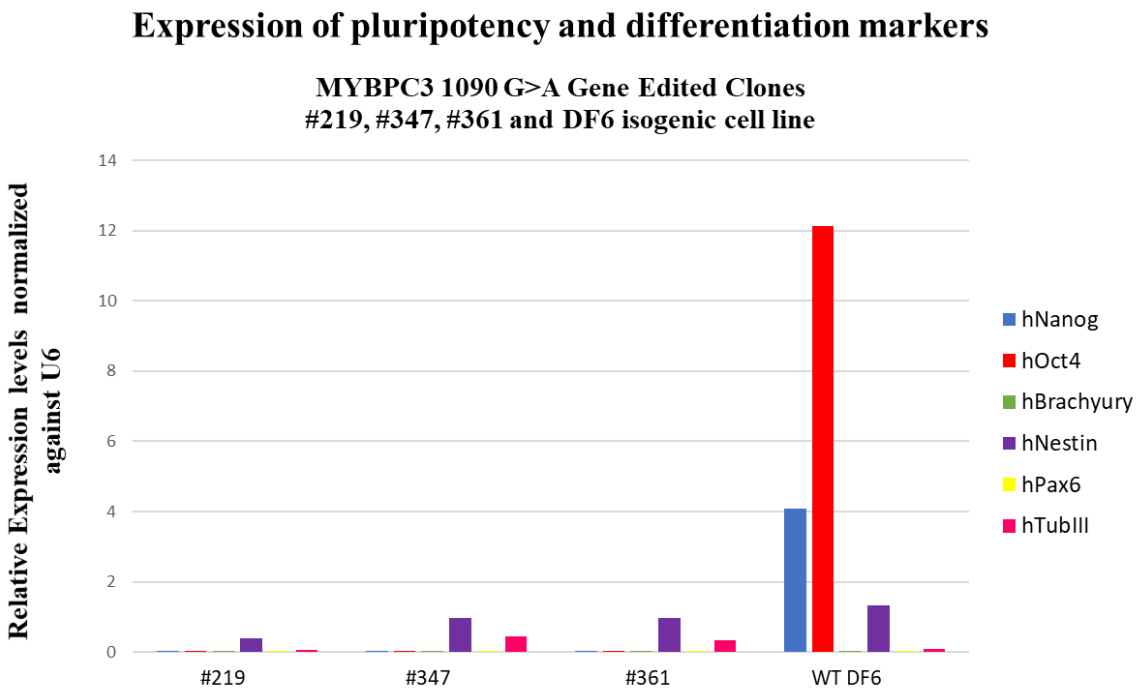


Figure 4.10 – qRT-PCR analysis of pluripotency and differentiation markers for the gene-edited clones, compared to the levels of the DF6 isogenic cell line. The relative expression levels are normalized against *U6*.

In addition, all clones express *hNestin*, a neural progenitor marker that is also expressed by the isogenic DF6 cell line, demonstrating that some degree of spontaneous differentiation in the culture of hiPSCs also occurs. More importantly, all edited clones completely lost the expression of the two pluripotency markers hereby analysed, *hNanog* and *hOct4*.

Thus, considering that the 3 gene-edited clones had undoubtedly differentiated spontaneously, the gene-editing experience was repeated. This time, the Cas9-Gem + puromycin condition was selected, since it was the one that provided the best results; less indels obtained through NHEJ repair and 2 gene-edited clones, one in homozygosity and other in heterozygoty. In order to be able to maintain these hiPSCs clones in the best possible culture conditions, the work scale was lowered to approximately half of the clones. The transfection was performed in the same manner as before, in order to mimic the conditions that supported the obtainment of gene-edited clones as closely as possible.

265 clones were picked from the *MYBPC3* c.1090G>A Cas9-Gem + puromycin condition. All cell culture and DNA procedures had already been optimized to a greater extent when compared to the last experiment. Even though by the end of the protocol all clones presented the normal hiPSCs morphology, no gene-edited clones were identified. Plus, only 4 indels were detected, consisting of a 1.5% editing efficiency by NHEJ, compared to the 5.2% obtained in the previous attempt (9 indels identified in 171 clones).

Considering that the latest approach failed to obtain clones with the selected variant in study, a functional test to attempt a differentiation of the previously obtained gene-edited clones into cardiomyocytes was performed even though the qRT-PCR results indicate that they are no longer pluripotent.

4.6 2D Cardiac Differentiation

The 2D cardiac differentiation was only carried out with the two homozygous clones, given that clone #347 could not be sufficiently expanded to initiate a differentiation assay. In addition to the gene-edited clones, the DF6 isogenic cell line and two other hiPSCs control lines that are currently used in the laboratory, TCLab and Gibco, were also used. Moreover, a patient-derived Xutl hiPSCs line (*MYBPC3* p.I1250fs) carrying a heterozygous *MYBPC3* frameshift mutation that was provided from Marta Ribeiro, a PhD student in the Carmo-Fonseca's Laboratory, was also differentiated into hiPSC-CMs.

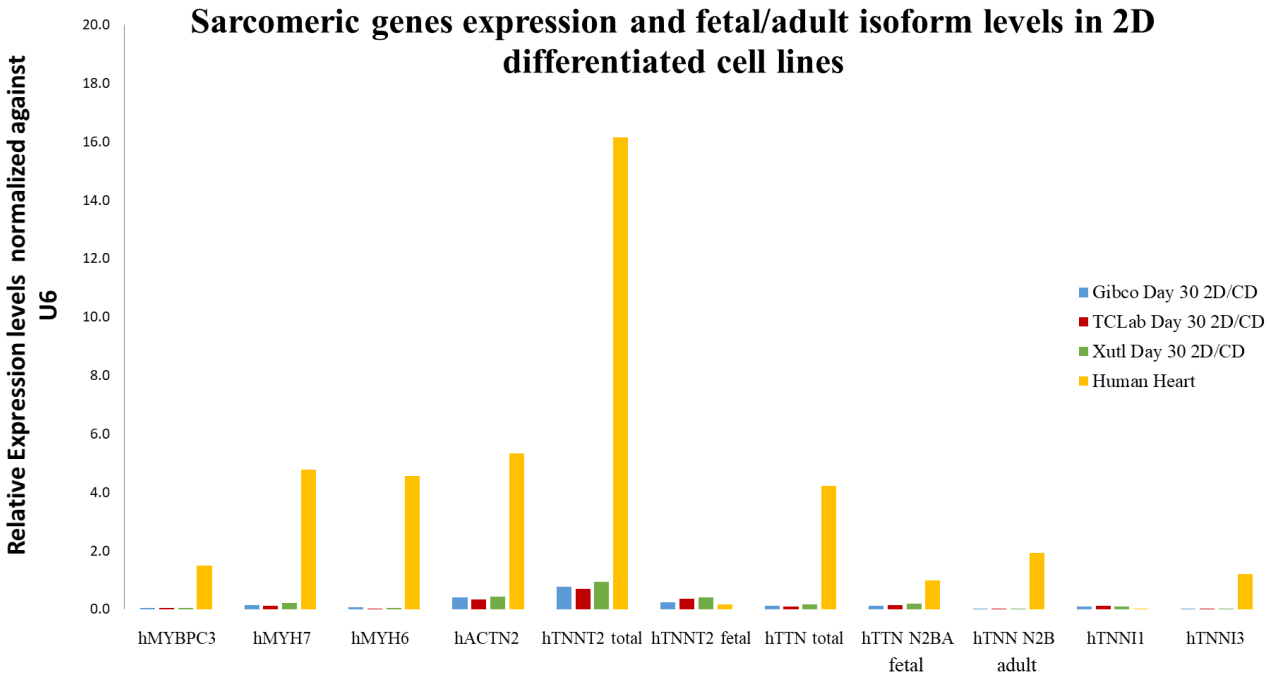
A few days after the beginning of the differentiation process, it was clear that the two gene-edited clones were not undergoing any type of morphological alterations. Importantly, the majority of the cells from clone #361 died upon CHIR99021 addition at day 0 of the differentiation protocol. Nevertheless, these cells were maintained until the end of differentiation process, but the protocol failed providing a definite evidence that the gene-edited clones obtained could not be used for further studies.

Regarding the TCLab, Gibco and Xutl lines, around day 8 of differentiation some cells started to have spontaneous beating; later on, large beating agglomerates were observed throughout the differentiated culture, an exclusive property of CMs. DF6 cardiac differentiation failed, given that only small regions of the cell culture presented spontaneous beating. Thus, the DF6 hiPSCs-CMS were not used for additional experiments.

The TCLab, Gibco and Xutl hiPSCs-CMs were collected at day 30 of differentiation and a qRT-PCR analysis was performed to determine the expression levels of sarcomeric genes and to assess the fetal/adult ratios of the *hTTN* and *hTNNT2* isoforms. These results are presented in **Figure 4.11**.

All 3 cell lines express all sarcomeric genes analysed, with the exception of h*TNNI3* that is only very lowly expressed in the TCLab and Xutl hiPSCs-CMs. When the expression is normalized against U6, reflecting the total population of cells, depicted in **Figure 4.11A**, the expression levels of sarcomeric genes are extremely reduced and cannot be compared to the levels in the human heart.

A



(continues on the next page)

B

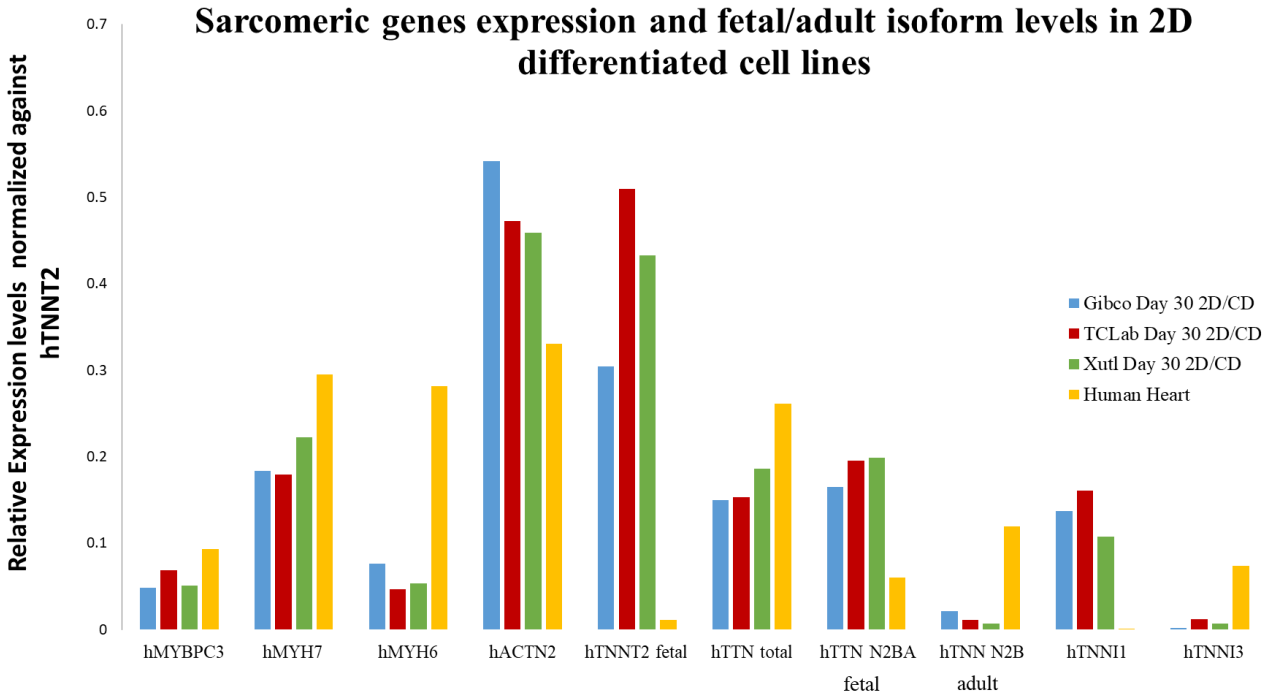


Figure 4.11 – qPCR analysis of the relative expression of sarcomeric genes for the Gipco, TCLab and Xutl hiPSCs-CMs, differentiated using a 2D protocol, compared to the expression levels in the adult human heart. The expression levels of *MYBPC3*, *MYH7*, *MYH6*, *ACTN2*, *TNNT2*, *TTN*, *TNNI1* and *TNNI3* were normalized against **(A)** *U6* and **(B)** *hTNNT2*. Additionally, the expression of fetal and adult isoforms for *TNNT2* and *TTN* was also determined.

However, when the expression is normalized against the levels of h*TNNT2* (**Figure 4.11B**), that is often used as a cardiac marker that reflects the cardiac enriched subpopulation of cells in culture, the expression levels of h*MYBPC3*, h*MYH7*, h*ACTN2* and h*TTN* are rather similar to those found in the human heart. The major difference between the hiPSCs-CMs and the human heart qRT-PCR results are the levels of fetal/adult isoforms. In the human heart sample, the adult *TNNT2* and *TTN* (N2B) isoforms are predominantly expressed. In contrast, the high expression levels of fetal *TNNT2* and *TTN* (N2BA) isoforms versus the reduced levels of adult *TTN* (N2B*)* indicate that the CMs obtained by this protocol are not extensively matured. Additionally, the adult heart only expresses h*TNNI3*, whereas high levels of h*TNNI1* are expressed in the hiPSCs-CMs. The differences in the gene expression levels between the controls and the patient derived Xutl cell line do not appear to be significant. Nevertheless, these results correspond to only one differentiation experiment and since the efficiency of the protocol is variable, no absolute conclusions can be taken from these results and more experiments are required.

In addition to the RNA analysis, cells were also re-plated into coverslips and fixed at day 36 of differentiation. Immunostaining for the β-myosin heavy chain, myosin-binding protein C and α-actinin was performed. For each case, cells are also co-stained with phalloidin, a probe that stains filamentous actin and that enables the visualization of the limits of the cell.

The images are presented in **Figure 4.12**. TCLab, Gibco and Xutl hiPSCs-CMs display an organized sarcomeric structure (which is more evident in **Figure 4.12H,** as the image was taken under higher magnification) and some degree of myofibril alignment. However, the cells are round and do not display the more elongated morphology that is typical of mature cardiomyocytes. Furthermore, in the images presented in **Figure 4.12A**, **C** and **G**, it is evident that in some cells the striation is limited to the centre of the cell and does not spread to the periphery, reinforcing that these CMs are still immature. Moreover, in the fields analysed there were also many cells that do not express sarcomere proteins, indicating that a heterogeneous cell population is obtained through this protocol.

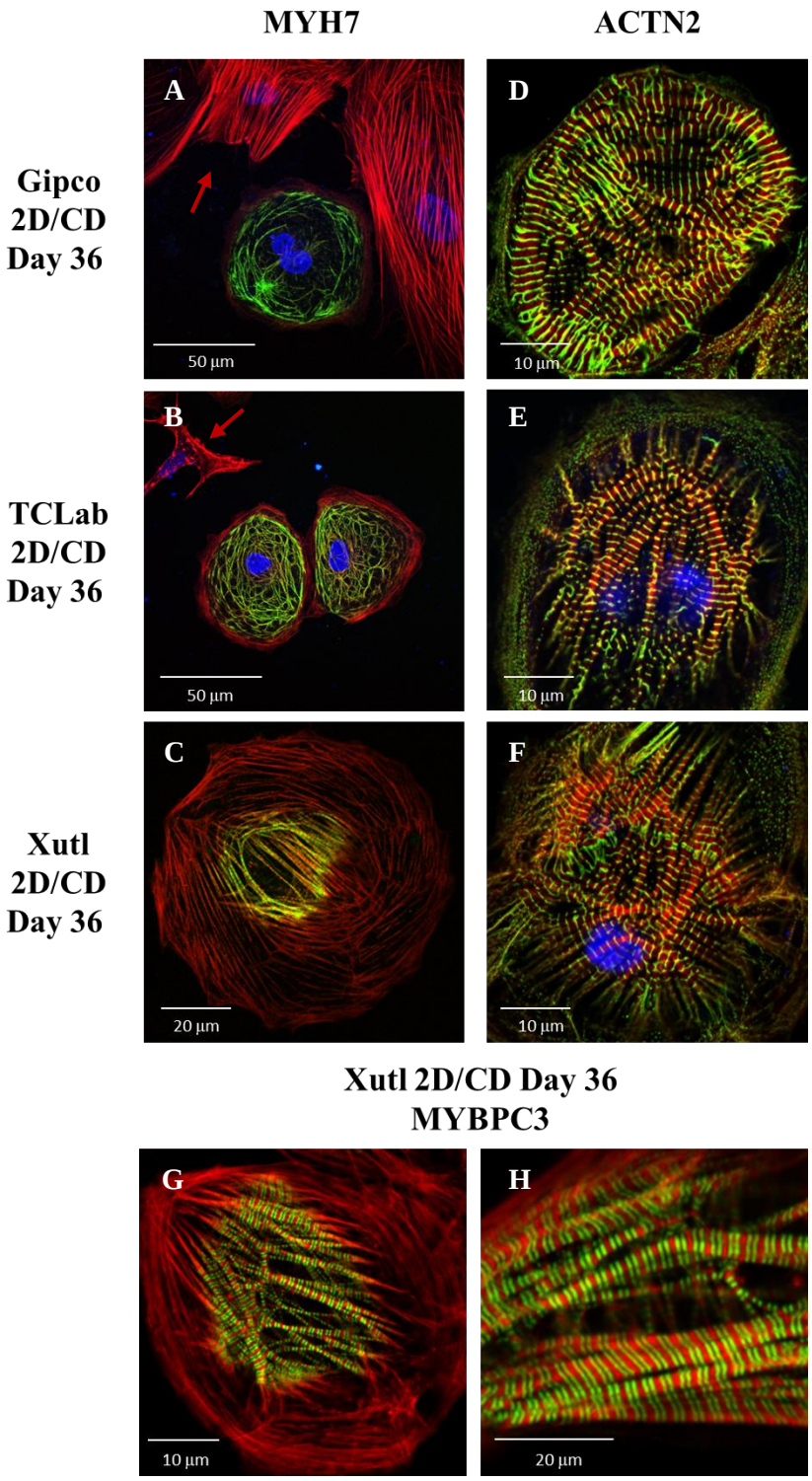


Figure 4.12 – Immunofluorescence analysis of sarcomeric proteins MYH7 **(A,B,C)** ACTN2 **(D,E,F)** and MYBPC3 **(G,H)** in Gibco, TCLab and Xutl hiPSCs-CMs. Actin is co-stained in red due to incubation with phalloidin. Cells that are only stained with phalloidin and that do not express sarcomere proteins are pointed out (arrow in red).

In conclusion, this 2D differentiation methodology may be used the generation of cardiomyocytes from hiPSCs, although a heterogeneous population of cells is obtained and the CMs display an immature, fetal-like phenotype.

5. Discussion

Mutations that affect the splicing pattern of a pre-mRNA have been implicated in several genetic diseases. These mutations often alter the consensus splicing regions and lead to an incorrect exon/intron boundary recognition, resulting in aberrant splicing. Furthermore, the splicing consequence of a particular variant in an intron or in an exon may depend on their location, either by disrupting the recognition of the canonical splicing sequences, interfering with splicing regulatory elements and/or create or activate cryptic splice sites. On the other hand, exonic variants that are classified as missense and synonymous mutations may also cause splicing alterations. Interestingly, in some of these missense variants, production of two distinct transcripts from the mutated allele, one full-length, with the altered aminoacid, and the other with a different length given that it was abnormally spliced have already been described [18]. Thus, the several types of mutations that generate errors in the splicing process and lead to the production of abnormal transcripts have been reported to play an important role on the pathogenesis of human disease.

In the specific case of HCM, the disease explored in this thesis, there are 212 splice-site variants reported in ClinVar, comprising only 3.7% of the variants deposited in this database. 68 of these variants were identified in *MYBPC3*, 10 in *MYH7* and 7 in *TNNT2*, the 3 genes that are more frequently mutated in HCM. Considering that, according to the literature, there are over 1500 pathogenic variants associated with the disease and reportedly 90% are classified as missense [10], it is logic to presume that a portion of these variants may have an unknown impact on splicing and be misclassified only as missense. Consequently, there may be many other variants that alter the splicing mechanism in HCM than what is recognized today. However, very few studies have delved into this matter so far and assessing the splice-altering potential of exonic and, more precisely, missense variants in HCM was the main objective of this project.

An example of an exonic variant that affects splicing in HCM and that serves as a support for our hypothesis is the *MYBPC3* c.772G>A (p.Glu258Lys) pathogenic mutation. It alters the last base of exon 6 that immediately precedes the consensus donor splice site. Functional studies using lymphocytes samples have revealed that this variant causes an out-of-frame skipping of exon 6 and generates a PTC in exon 9 [80]. Further studies revealed that only wild-type and abnormally spliced transcripts are detected [90]. The *MYBPC3* c.1624G>C (p.Glu542Gln) variant is also located in the last nucleotide of an exon and was recently shown to lead to skipping of exon 17 in minigene assays, lymphoblastoid cells and myocardium samples [81]. In this case, two transcripts were being produced from the mutated allele: one where the exon is skipped, and the full length one containing the altered aminoacid [25,90]. Therefore, some exonic mutations that affect the splicing mechanism have already been characterized resorting to diverse research methodologies.

However, functional assays are both expensive and demanding, and platforms to evaluate the impact of a given variant are not yet optimized to allow for a cost-effective screening of a vast number of variants in an appropriate time window. High-throughput sequencing techniques conferred a substantial advance in clinical diagnostic testing. These technologies have led to the identification of a considerable number of new variants that must be further studied and characterized. In addition, variants that had been previously found in patients with cardiomyopathies and that were thought to have a possible pathogenic effect were revealed to be common in control populations due to the emergence of next-generation sequencing methodologies. Therefore, better strategies are required to evaluate the functional effect of a particular variant [14].

The evaluation of extended gene panels associated with HCM have also led to the identification of an extensive amount of variants of uncertain clinical significance (VUS) [14]. In fact, 4894 of the HCM-related variants deposited in the ClinVar database are of uncertain significance, consisting of 67.2% of the total number of variants identified. Additionally, 1298 have conflicting interpretations of pathogenicity.

When identified, a variant is classified according to its pathogenic potential by considering several guidelines: (a) the frequency of a specific variant in the control population, resorting to databases

such as the Exome Aggregation Consortium (ExAc), (b) the cosegregation of a variant and the associated phenotype amongst patients of the same family, (c) if the variant is located in a highly conserved domain between species, indicating that the region is essential for protein structure and function, (d) other mutations identified in the same or nearby positions (suggesting the existence of hotspots for mutations), (e) if the variant has already been associated with disease and reported in databases such as HGMD and ClinVar, (f) *in silico* algorithms that predict the pathogenicity of a variant and its potential impact on protein structure and/or function, such as MutationTaster, Polyphen2 and HSF and (g) functional studies, which are in most cases inexistent [10,14]. Nonetheless, the results of these analysis are often inconclusive, contributing to the large number of variants deposited in databases that are of uncertain clinical significance.

Consequently, bioinformatic tools are often used for evaluating the potential impact of identified variants, namely the effect of a specific variant on splicing. It is important to note that these algorithms only indicate a possible outcome and several tools should be employed to achieve a higher confidence in the results provided, while also keeping in mind that for a precise classification of a variant in study, these predictions would always need to be verified by functional experiments. This is even more relevant considering the fact that several publications have reported that bioinformatic software's that evaluate the impact on protein structure and function of variants, not only in the specific case of a predictive effect on splicing, are often inaccurate [18].

In the case of a HCM-related variant previously classified as being of uncertain significance in the *MYL3* gene (c.170C>A), an *in silico* analysis tool presented a score of 0.998 confidence (on a scale of 0-1) in classifying the variant as pathogenic. Yet, the authors of the paper performed genetic, morphological and functional assays in hiPSCs-CMs bearing this specific variant in heterozygosity that showed no indications of disease phenotype and revealed that the variant was in fact benign, despite the high *in silico* score [97]. Therefore, the question of low reliability of bioinformatic tools is raised, even though they are widely used and extremely useful for the characterization of variants. Moreover, considering the rapid advances of sequencing technologies and the fact that huge amounts of information are nowadays easily obtained, and many new variants of interest are identified, the need to develop improved and more accurate algorithms is reinforced, as well as complementing these *in silico* tools with functional assays.

For the selection of the exonic missense HCM variants to be studied, two bioinformatic tools – HSF and MaxEnt, were applied for the screening of possible splicing effects of the 20 mutations initially chosen by filtering the ClinVar database according to our criteria. These software's were employed because they are fairly easy-to-use and do not entail extensive bioinformatic training and knowledge. However, since bioinformatic tools have already been shown to not always be precise in predicting the molecular consequences of a variant, more than two software's should be employed. Additionally, these variants should also be analyzed by more complex bioinformatic algorithms, that would provide more accurate splice-altering scores, resulting in a higher confidence in the characterization of the variants.

In order to develop an hiPSC-CMs platform to evaluate the impact of a mutation at the transcript and protein level, a positive control for a splicing alteration is required to ensure that the methodology used is successful in the detection of aberrant splicing and can be used to characterize the disease mechanism underlying HCM variants.

The *MYBPC3* c.927-2A>G (originally described as Int12ASA_2G) is a founder mutation that is the predominant cause of HCM in Iceland, being present in 58% of individuals and that is located in the splice acceptor site of intron 11 [98,80]. This mutation disrupts the canonical splice acceptor site and was shown to lead to aberrant splicing of *MYBPC3* in lymphocytes cells [91]; it is predicted to be a truncating mutation as it leads to skipping of exon 12, resulting in the generation of a PTC in exon 13. [98,80]. Hence, this variant would serve as a positive control, not only because it is a splice-site mutation that has already been proved to lead to abnormal splicing, but since it also results in the production of a PTC that could cause NMD degradation of the mutated transcript, fitting into the haploinsufficiency theory that is proposed.

Similar to the *MYBPC3* c.772G>A and c.1624G>C variants whose mechanism were earlier described in this chapter, the *MYBPC3* c.1090G>A mutation that is classified as missense and is the main focus of this study is also located on the last nucleotide of an exon (in this specific case, of exon 12) and is predicted by bioinformatic algorithms to disrupt the canonical donor splice site and lead to aberrant splicing. The hypothesis proposed is that this variant leads to the skipping of exon 12. Therefore, considering its location in the exon and proximity to the canonical splice site, the similarity to the two variants previously defined and the high predictive score of a 110.82% variation according to the MaxEnt algorithm, this is an exonic variant that has a high probability of interfering with the splicing mechanism. Nonetheless, the skipping of the entire exon 12 is not the only possible outcome, as existent cryptic splice sites might also be used, depending on their strength, given that the canonical site was destroyed. As an example of this, *MYBPC3* c.1928-2A>G, a mutation that disrupts the splice acceptor site, was reported to lead to two aberrant spliced transcripts: one that has a 140 nucleotide deletion that is correspondent to the skipping of the entire exon, and other that consists of a 11-nucleotide deletion that is resultant of the use of a cryptic splice site that is downstream of the canonical one [28].

Additionally, certain regions in sarcomeric genes such as *TNNT2* and *MYBPC3* have already been associated to mutational hotspots, since there is a high frequency of variants located in these sites [99]. In the ClinVar database, 3 donor splice-site region (+1/2) variants in the nucleotides that immediately follows the c.1090 position of the *MYBPC3* gene are described as being pathogenic: *MYBPC3* c.1090+1G>T, *MYBPC3* c.1090+1G>A and *MYBPC3* c.1090+2 T>G, suggesting that this region could consist of a hotspot for mutations. Even though these variants are predicted to result in aberrant splicing, their effect has not yet been confirmed by functional studies.

In January 2019, the work from Singer *et al*., proved that the *MYBPC3* c.1090G>A variant does, indeed, interfere with splicing. The authors of the paper performed both a minigene assay and isolated RNA from venous blood of HCM patients, and their results indicate that the variant causes the skipping of exon 12 and leads to a frameshift, thus confirming our hypothesis [81]. Therefore, this variant possibly generates a PTC with transcripts being subjected to NMD or, alternatively, results in the production of an unstable truncated protein that is likely degraded, supporting an haploinsufficiency mechanism on the basis of HCM disease. Considering that these authors performed no further analysis at the transcriptomic and protein level, it was decided that, since we had already invested in these experiments and despite the fact that the splicing hypothesis had already been proved, we would continue with our work in order to evaluate if the mutated protein is present in hiPSCs-CMs and if, in case of haploinsufficiency, there is a compensatory mechanism from either the wild-type allele and/or the other sarcomeric genes.

In contrast to other sarcomeric genes, in which the vast majority of mutations identified are missense, over 90% of the HCM-associated variants in *MYBPC3* lead to the generation of transcripts containing PTCs that if not degraded by NMD are predicted to result in the translation of a C-terminal truncated protein. Considering that *MYBPC3* has been shown to have a clear predilection for the occurrence of nonsense and frameshift mutations, the disease mechanism most widely proposed is haploinsufficiency, unlike the dominant-negative mechanism associated with the vast majority of HCM mutations in other sarcomeric genes [24]. In agreement with this hypothesis is the fact that most studies of *MYBPC3* truncation mutations reveal that there is no evidence of mutant protein present in cardiac tissue from affected HCM patients, suggesting that their incorporation in the human sarcomere is prevented either because (1) the protein is not expressed, as the PTC-containing transcript lead to the activation of the co-translational NMD pathway or (2) the transcripts that escape the NMD pathway generate premature truncated proteins that are likely unstable and misfolded, and subsequently degraded by the ubiquitin-proteasome system [24,33]. Thus, given that the *MYBPC3* c.1090G>A variant is predicted to generate a PTC, we hypothesize that mutated protein would not be incorporated in the sarcomere due to haploinsufficiency and the MYBPC3 protein embedded in the sarcomere would only be translated from the wild-type allele.

Allelic imbalance which has previously been described in HCM [100], may favour the expression of higher levels of protein from the wild-type MYBPC3 allele, or lower the expression levels of other sarcomeric genes in order to maintain the stoichiometry of the human sarcomere; as such, could

be a mechanism to compensate for haploinsufficiency [24]. A 9-fold upregulation of the mRNA levels of *MYBPC3* was identified in samples containing truncating mutations, and is proposed to be a compensatory mechanism to attempt to maintain the normal levels of MYBPC3 protein in the human heart [101].

In addition to the undetectable expression of truncated MYBPC3, the lower levels of the overall MYBPC3 protein, with a reduction of 30% of the total expression described in some HCM patients also supports the hypothesis of haploinsufficiency. Furthermore, lower levels of mutant transcripts in comparison to the mRNA expression from the wild-type allele have also been identified in cardiac samples, suggesting that such mutated transcripts are partially degraded by NMD. Nonetheless, studies have reported that the reduction in the mutant mRNA levels is rather small, and does not justify the undetectable levels of mutant protein, supporting that the ubiquitin-proteasome pathway plays a major role in the degradation of the truncated proteins when the transcripts escape NMD [102,103].

The reduction of MYBPC3 protein levels in different samples is also variable. Heterogeneity at the transcript levels of mutant and wild-type alleles, which vary from the expected 1:1 ratio has already been identified in HCM, and is proposed to account for the large variability of the clinical phenotype of HCM [11,90]. Also, in samples bearing missense mutations in sarcomeric genes, variations of the 1:1 wild-type/mutant protein ratio were reported, even though the transcript levels are similar, suggesting that mutant proteins are probably unstable and degraded [101].

Despite the attempts performed in this work to generate gene-edited hiPSCs capable of being differentiated into CMs were not successful, there is sufficient data in the literature that supports the further development of this project and that corroborates our hypothesis that haploinsufficiency might be a mechanism involved in the pathogenesis of HCM, regarding the 3 variants of interest: *MYH7* c.2221G>T, *MYBPC3* c.1090G>A and *MYBPC3* c.927-2A>G.

Even that some of the proposed functional analysis can be performed by minigene assays (regarding the splicing alterations) and/or accessing donor cells and reprogramming them into patient-derived hiPSCs that can then be differentiated into cardiomyocytes, the optimization of the gene-editing technology is still indispensable. In fact, since we are trying to evaluate the impact of a single point mutation in the molecular mechanism of a complex genetic disease, it is essential to have isogenic controls in order to prove that the possible genetic, morphological and physiological alterations that are observed in the hiPSCs-CMs are being caused by the variant in study, and that there are no other variables contributing to the disease and interfering with the results of the experiment. Therefore, the influence of a mutation in the pathophysiology of HCM cannot be correctly determined if the controls used in the study derive from cells of another individual. Even though these cells are wild type for the variant of interest, they probably have completely different genetic and epigenetic backgrounds that would influence the results. This issue can be overcomed with the use of the CRISPR/Cas9 editing technology, that allows for the precise generation of isogenic cell lines that differ from each other only in the location of the variant of interest [65].

This topic is even more relevant considering that HCM is an extremely heterogeneous disease and that patients bearing the same mutation sometimes present entirely different phenotypes and clinical outcomes. The phenotypic variability of the disease can be associated not only to factors such as obesity and hypertension, but also to the diverse genetic backgrounds that may contribute to different phenotypes. Additionally, cases of compound heterozygous (~5%) have been identified in patients with HCM, where several sarcomere genes present alterations, and are associated with a more severe phenotype [7]. All together, these evidences suggest that the phenotype of an individual with HCM may often not be a direct result of one single variant but caused by a cumulative effect of several mutations in sarcomeric and sarcomeric-associated genes.

Considering the heterogeneous and complex phenotypes and clinical outcomes of HCM, the genotype-phenotype associations are often inconsistent and the emergence of the CRISPR/Cas9 technique may eliminate these difficulties and lead to the establishment of more accurate correlations. If a specific variant is, in fact, the cause of the disease, in the isogenic genetically corrected control line the reestablishment of a normal, healthy phenotype is expected.

Therefore, the CRISPR/Cas9 technology is important not only for the generation of disease models, but also isogenic controls. Consequently, even if patient-derived cells are available to perform the experiments, the optimization of this gene-editing technique is still required to achieve accurate results. Thus, a carrier-specific hiPSC-CMs platform must be developed in order to decipher the benign or pathogenic phenotype of uncharacterized variants in a patient-specific manner, and to understand the molecular mechanisms underlying the disease.

In the case of the *MYBPC3* c.1090G>A and *MYBPC3* c.927-2A>G variants, after obtaining gene-edited hiPSCs and the correspondent isogenic cell lines, we would be able to establish a disease model, not only to study the impact of the specific variants on splicing but to also demonstrate how the wild-type allele and the other sarcomeric genes possibly adapt and try to overcome this situation, after differentiation of the hiPSCs into cardiomyocytes.

The first approach used for the introduction of the variants of interest in the genome of human cells was based on a recent publication by Sarah E. Howden, *et al*., where a one-step reprogramming and gene-editing protocol from human fibroblasts was described [75].

This strategy presents several advantages when compared to the gene-editing in hiPSCs. Performing CRISPR/Cas9 in hiPSCs requires the picking of several hundred clones, involving not only a considerable amount of cell-culture work for the maintenance of the clones for long periods of time, presenting both a time-consuming and expensive approach that, ultimately, has a very low efficiency of knock-in. The isolation of a correctly targeted hiPSCs population is very demanding, often unsuccessful, principally in the absence of drug and other selection approaches. The isolation of transfected cells by FACS is associated with low cell viability, given that for the cells to be sorted they must be in a single-cell state and the hiPSCs are also sensitive to the sorting process itself. The one-step approach does not require steps such as drug selection or FACS purification, necessary for the initial selection of the hiPSCs clones to be screened and that further complicate the process. Therefore, if the gene-editing is performed directly in fibroblasts, cells that are substantially more robust than hiPSCs, the hindrances associated with the cell death of hiPSCs are possibly overcomed. Furthermore, human fibroblasts also have higher transfection efficiencies associated. More importantly, the efficiency reported for this gene-editing protocol is much higher, allowing for the reduction of the number of clones analysed, as is it stated that the screening of 20-50 clones is enough to identify correctly edited clones by HDR. However, the number of hiPSCs clones available for picking is strongly limited by the reprogramming efficiency of the human fibroblasts [75].

Since human fibroblasts isolated from younger donors are associated with higher reprogramming efficiencies, the cell lines selected for reprogramming are derived from healthy donors of 3, 10 and 11 years old. Interestingly, the only transfected HuF cell line in which morphological alterations that suggested partial reprogramming were detected was the HuF3Y, the one that was derived from the younger donor. Additionally, the number of passages may also influence the reprogramming efficiencies, as these primary cells can quickly become senescent, and low passages are recommended (less than 10). However, the 3 HuF cell lines used were transfected in passages 13-15, which may had also influenced the efficiency the reprogramming process.

The transfection efficiencies were also highly variable between different cell types and lines. Therefore, one of the main problems of our experiment was the possible low transfection efficiency in the HuF lines used. In a subsequent attempt of this protocol, a transfection control should be initially performed, independently in the 3 cell lines used, so that the efficiencies can be optimized. In order to achieve this, the sole transfection of a plasmid encoding GFP can be performed, followed by sorting of GFP^+ fibroblasts. After deciphering the efficiency of transfection in these cells using the parameters described in the Nature Protocol [75], the electroporation settings can be adjusted in order to obtain a higher reprogramming efficiency in our cell lines. In the latest 2 attempts of performing reprogramming and gene-editing in the HuF lines, the main problem faced was the large amount of cell death following electroporation. After plating, the fibroblast cells were extremely individualized, and we hypothesize that this could also have influenced the reprogramming efficiency. Therefore, a transfection strategy such as lipofection could also be tested, as it is associated with a higher cell viability [85].

Furthermore, the picking process was not optimized, as a microscope suitable for picking of hiPSCs clusters was not available at the time. Therefore, the clones were picked when they had grown enough to be visible by the naked eye, a method that lacked precision, and that lead not only to the death of several clones but also to the contamination of some of the clones picked with fibroblasts that were growing in close proximity with the clones. Even though the clones were maintained in culture for several weeks, the fibroblasts continued dividing and contaminating the culture of the proposed reprogrammed cells, suggesting that they could have been immortalized during the reprogramming process.

Overall, marked morphological changes were only detected in 1 out of the total 21 conditions performed in 3 different attempts of this protocol, comprising a discouraging efficiency of less than 5%. In addition, the results obtained suggest that the reprogramming process into hiPSCs was also not fully achieved in the HuF3Y *MYBPC*3 c.927-2A>G condition, the only one where colony-like structures emerged, and clones were picked. Even though there was a clear difference in the morphology of these clones when compared with the original fibroblast line, these cells did not fully recapitulate the morphological characteristics of established hiPSCs lines such as DF6, hereby used as control. The qRT-PCR results presented in **Figure 4.6**, demonstrate that most clones still express high levels of *hCol1a1*, a fibroblast marker, and either do not express the pluripotency-associated genes *hNanog* and *hDNMT3β*, or express them at very low levels. The variability of expression of these genes between the clones also suggests variability in the reprogramming efficiency. In conclusion, this dataset indicates that these clones, primarily the ones that show a considerable decrease in the levels of expression of *hCol1a1*, but present no expression of pluripotency markers, are probably in a some-what intermediate state of reprogramming.

An efficiency of 2–20% gene-editing efficiency is described in this one-step reprogramming and gene-editing protocol [75]. However, 22 clones were sequenced, and none presented genetic alterations, indicating that along with the reprogramming process, the gene-editing was also unsuccessful. These limitations could presumably be overcomed through the optimization of the several steps of this methodology, mainly the electroporation conditions, as obtaining higher transfection levels would likely result in a better reprogramming efficiency, thus allowing for the screening of more hiPSCs clones and the identification of correctly edited cells. Yet, this would require a high degree of optimization and several months of work. Thus, it was decided that for the purpose of this thesis it would be better to change the strategy used, in order to be able to achieve the first goal of the project and generate gene-edited hiPSCs.

Previously in the laboratory, several attempts were performed to generate gene-edited hiPSCs with the *MYBPC3* c.927-2A>G splice-site mutation. Different strategies were employed, the latest being based on the protocol “Precise and efficient scarless genome editing in stem cells using CORRECT” developed by the Marc Tessier-Lavigne group in 2017 [72]. The authors concluded that the majority of genomes where the DSBs had been repaired through HDR also presented additional indel mutations, due to a subsequent re-cutting by the Cas9 of a previously correctly edited locus. Therefore, they developed a two-step gene-editing approach, using CRISPR/Cas-blocking mutations at the PAM site to minimize re-cutting by the Cas9 nuclease. In the second round of CRIPSR/Cas9, the HDR template removes the blocking mutations and maintains the pathogenic mutation [72]. However, this method requires two rounds of gene-editing and in the laboratory not even after the first round there were successful results.

Therefore, an alternative strategy to minimize secondary mutations in the target region is the use of the cell-cycle regulated Cas9-Gem mRNA that was previously employed in the gene-editing and reprogramming of human fibroblasts. Since the HDR pathway is only active in the S and G2 phases of the cell cycle, the degradation of the Cas9 during the G1 phase due to the fusion with the Geminin peptide results in a reduction of the NHEJ activity. Even though this approach has no direct impact in the frequency of HDR, there is a 2 to 3-fold decrease of NHEJ activity. Additionally to the control of the presence of the Cas9 protein in the different cell cycle stages, since the mRNA is transfected instead of the plasmid, the expression of the Cas9 is also more transient, minimizing possible re-cutting activity and the introduction of unwanted mutations in the target region [75].

Therefore, one of the conditions screened for gene-editing in hiPSCs used the Cas9-Gem mRNA and antibiotic selection. Nonetheless, since this strategy had never been tested in the laboratory before, the experiment was also followed with a condition with the pCas9-GFP that allows for single-cell sorting of GFP positive, transfected cells. However, the CORRECT method was not employed. Instead, only one round of gene-editing was performed using the HDR template that only contains the modification of interest. It is possible to identify correctly gene-edited clones that were not re-cut by the Cas9 and, consequently, did not incorporate undesired mutations in the target locus, as it only involves the screening of a higher number of clones. According to the CORRECT paper, the blocking mutation strategy would involve picking a few hundreds of clones, comparing to the possible thousands that would have to be screened if this strategy was not used [72]. However, since gene-editing was achieved in other publications without resorting to this methodology, it was also performed to verify if, even with lower probabilities, it was feasible to identify clones edited by HDR using only one round of gene-editing with the pCas9-GFP.

In hiPSCs, the rate of gene disruption by NHEJ is reportedly between 1–25% without subsequent selection. However, this editing efficiency drops substantially for HDR repair, and selection strategies are usually advised [104]. Yet, considering that in former attempts in the laboratory there were problems regarding high levels of cell death following FACS selection of GFP positive cells, and the puromycin antibiotic treatment had never been tested in hiPSCs, strategies without selection of transfected clones were also performed.

As a method to improve the rate of the HDR mechanism, hiPSCs were exposed to low temperatures following transfection. The so-called "cold shocking" the cells at 32°C is reported to substantially increase the rates of HDR, and to have a more considerable effect when the efficiency of HDR is lower. Even though the mechanism is not understood in detail, it is hypothesized that the growth at 32°C has an influence in the cell cycle, with the cells accumulating at the G2/M stage; since the HDR is active during the G2 phase of the cycle this results in a higher efficiency of the HDR mechanism [74]. Thus, in all conditions tested, cells were incubated at 32°C for 48 hours to enhance HDR.

It was further decided that the different conditions would only be performed with the *MYBPC3* c.1090G>A missense variant, as the main priority of this experiment was to screen different conditions and several hundred hiPSCs clones, in order to identify clones that had been cut by the Cas9 (even if the DSBs were repaired by NHEJ), considering that in earlier attempts in the laboratory no indel mutations had been observed.

In the conditions without selection, no alterations were detected, indicating that the clones picked had not been subjected to the activity of the Cas9 nuclease. This suggests that in order to discard the selection strategies that often pose a hindrance to the experiment, a much higher amount of hiPSCs clones would have to be screened - most likely between hundreds to thousands of clones, in order to identify hiPSCs that incorporated the intended mutation through HDR.

As expected, the majority of the indels were identified in the Cas9-GFP$^+$ condition (80.4%), considering that the main purpose for the use of the Cas9-Gem mRNA is to minimize the frequency of DSB repair by NHEJ, thus presenting a lower frequency of indel mutations when compared to the version of the Cas9 that is not cell-cycle regulated.

Strikingly, one gene-edited clone (#219) with the intended mutation in homozygosity was identified in the Cas9-GFP$^+$ condition, indicating that even though the blocking mutation strategy was not used, this specific hiPSC clone did not harbour re-cutting by the Cas9 nuclease. These results suggest that if the experiment with the pCas9-GFP is performed in a sufficiently large scale, only one round of gene-editing may be required. In this case, the variant was correctly introduced in the genome in 0.6% of the clones analysed. Nonetheless, we can also hypothesize that from the set of all the clones that contain indel mutations, some might have been correctly gene-edited through the HDR mechanism, but were subsequently re-cut by the Cas9, thus disrupting the correctly introduced point mutation.

Using the Cas9-Gem mRNA + puromycin selection, 2 gene-edited clones were identified, as in clone #361 the *MYBPC3* c.1090G>A variant was introduced in homozygosity and in clone #347 in

heterozygoty. Additionally, out of the 171 clones picked from this condition, 9 had indel mutations, and these results support that when compared to the pCas9-GFP strategy, the NHEJ efficiency is reduced. The gene-editing efficiency of this condition through HDR is of 1.2%, consistent with the percentages reported in the literature [75].

When the sequencing results of these clones were obtained, all hiPSCs clones that were still in culture had undergone morphological changes that suggested that they had most likely spontaneous differentiated. These alterations are evident in the images presented in **Figure 4.9**, as the cells from the 2 homozygous clones no longer form the typical hiPSCs colonies and have long extensions. Regarding the heterozygous clone, the colonies that are formed do not completely recapitulate the morphology of hiPSCs colonies. Overall, the morphology of the 3 successfully edited clones was strikingly different from that of hiPSCs. Additionally, the qRT-PCR data that is showcased in **Figure 4.10** reveals that all gene-edited clones lost the expression of pluripotency genes throughout their time in culture. Spontaneous differentiation is reasonably common in hiPSCs cultures, and when it happens, measures are taken to favour the expansion of the hiPSCs and not of the differentiated cells. As an example, spontaneous differentiation often occurs in the periphery of hiPSCs colonies when they become too big, and the centre of the colony is picked to maintain a pure population of hiPSCs in culture. However, considering the large number of clones in culture and the fact that they were grown in 48-well plates, this could not be attempted. Additionally, when the gene-edited clones were expanded, no colonies that resembled that of hiPSCs were detected so that a subcloning event could be carried out.

The primary difficulties of this experience were that, at this point, the cell culture and DNA procedures were not optimized to perform CRISPR/Cas9 in a scale of 500 clones. Maintaining several hundred hiPSCs clones in culture is extremely demanding and difficult to manage. At the time this first gene-editing experiment was performed, the hiPSCs growth medium available was the mTeSR 1, that required daily medium changes. The clones were maintained in 48-well plates throughout the whole experiment, since expanding them to culture plates with higher surface areas would require large amounts of culture medium and the associated costs. Even though a small percentage of cells were plated after splitting (~15-20%), it was unmanageable to split the clones at least 3 times a week, as they should have been, considering that all DNA procedures were also being simultaneously performed. Therefore, when the clones were split, they were often too confluent, and the extended time in culture (approximately 3.5 months for the last clones to be sequenced) in these suboptimal conditions is what we hypothesize caused the loss of pluripotency.

Cryopreservation of the hiPSCs clones would be a cost-effective alternative to maintaining them in culture for long periods of time. However, hiPSCs are very sensitive to the freezing/thawing procedures, and previous attempts in the laboratory to freeze hiPSCs clones in 48-well plates resulted in an extremely low cell viability, leading to the death of the majority of clones. Therefore, freezing the clones would entail a huge risk, as gene-edited clones could be lost after thawing.

Adding to the cell culture difficulties, the sole fact that the PCR amplification was not successful for every DNA sample and that in many cases the sequencing of the PCR products either failed or the sequences were to dirty to decipher if there were alterations or not, also contributed to the prolonged time in culture of the hiPSCs.

When the 2nd gene-editing experiment was performed, a substantially quicker cell-culture system was implemented, and all the DNA procedures, primarily the PCR amplification were already optimized. It was decided that this time working in a smaller scale would be preferable, as screening 500 clones is not feasible if the successfully gene-edited clones are not maintained in the best conditions possible. Therefore, only one condition was performed: *MYBPC3* c.1090G>A Cas9-Gem + puromycin, the one that presented the higher efficiency with 2 correctly gene-edited clones in 171 clones picked. In comparison to the pCas9-GFP method this is also, theoretically, a better approach, since a cell-cycle regulated Cas9 is used.

In the previous experiment, 4 different methods were tested, meaning that the timing for the several procedures of the editing experiment were highly variable depending on the condition in question. For instance, as PCR amplifications were already being performed for the Cas9-Gem mRNA

+ puromycin condition, the clones from the pCas9-GFP + FACS were only starting to be picked. This contributed to the fact that depending on the condition, the hiPSCs clones were at very different stages, which further complicated the process. Thus, performing only one gene-editing condition was essential for the maintenance of the hiPSCs clones in optimal culture conditions. In this attempt, all 265 clones were picked directly from the well where the transfection was performed, removing the variable of diluting the hiPSCs clones that can lead to a mixture of different clones after plating. In addition, the clones were picked in an interval of only 2 days. This meant that, at the time of splitting, all clones were at very similar growth stages which allowed them to be passed simultaneously, making all subsequent procedures more synchronized. Additionally, since all clones from the same plate were in conditions to be split, a 200-µl multichannel pipette was used, making the whole process less time-consuming.

However, even though all conditions were substantially more optimized, no successfully gene-edited clones were identified, regardless of the picking of 265 clones, almost 100 more than the amount screened in the 1st experiment for this condition. This is most likely associated to another technical difficulty experienced, that is the low and variable transfection efficiency of hiPSCs.

In the experiment where the pCas9-GFP was transfected, only ~7400 cells were GFP^+ 72h post-transfection. Adding to the fact that, not only the number of cells transfected is low, but many of the sorted hiPSCs die after plating, even with the addition of an anti-apoptotic molecule, this often results in a very reduced number of clones to pick. Even though with the puromycin selection strategy the quantification of the transfection efficiency is not possible, 265 clones were picked in the 2nd experiment and no gene-edited clones were identified, whereas 171 were picked in the first experiment, resulting in 2 gene-edited clones. Moreover, the NHEJ frequency in this latest experiment was also lower. This reinforces the variability of the transfection efficiency and how it can impact the generation of edited clones. Therefore, the optimization of the transfection efficiency is also essential for the success of the CRISPR/Cas9 methodology. Strategies such as tube electroporation that enhance the transfection efficiency in the hard-to-transfect hiPSCs but present a low cytotoxicity when compared to the conventional electroporation methods present an alternative [104].

In addition to the transfection optimization, several other approaches have been reported in the literature to improve the editing efficiency through HDR. Considering that the DSB repair strategies NHEJ and HDR compete with each other, the inhibition of the NHEJ pathway would, in theory, favour the HDR mechanism. Cell cycle regulation of the Cas9 to promote HDR (hereby tested using the Cas9-Gem mRNA) and the use of small-molecule inhibitors of the DNA ligase IV to reduce the frequency of NHEJ are some of the approaches proposed. The addition of drugs that interfere with the cell cycle and lead to an accumulation of cells in the G2/M phases are also reported to result in a higher HDR efficiency [105]. In human T cells, the efficiencies of gene KO were optimized to up to 90% through the transfection of Cas9 ribonucleoproteins (RNPs), complexes of recombinant Cas9 with sgRNAs [106].

One of the main factors that determines the efficiency of HDR is the donor template, as the type of template used should be optimized according to the purpose of the experiment. For small changes such as the introduction of point mutations, ssODNs are more efficient than template plasmid DNAs [107].

To facilitate the screening and identification of gene-edited clones, some authors suggest the incorporation of synonymous changes to introduce a restriction site. This would allow an initial screening of the clones by an easy and fast enzyme restriction assay, and only the DNAs that were edited by the Cas9 would then be sequenced. However, this strategy cannot be applied to our cells, since the purpose of this study is to understand the consequences of a single-nucleotide variation in the context of a complex disease, and these synonymous changes may have an influence in the splicing mechanism [72].

It has also been reported that the chromatin status can have an impact on the accessibility of the of the CRISPR/Cas9 complex to the target locus. Therefore, some regions of the genome may be more difficult to edit than others, which can in part explain the variability in editing efficiencies. Additionally, sgRNA efficiency should be assessed in the cell line to be used, as the efficiency can vary for the same genomic location between cell lines [108].

Thus, the main achievement of this thesis is that we were able to introduce the variant of interest in the genome of hiPSCs and isolate 3 gene-edited clones from a pool of >500 clones screened.

Since cardiac differentiation of hiPSCs was also one of the proposed methods to be employed in this thesis, a 2D protocol was performed. Even though the *MYBPC3* c.1090G>A $^{-/-}$ and $^{+/-}$ clones would most likely not differentiate into CMs since they do not express pluripotency markers, a functional test using clones #219 and #361 was performed in a final attempt to use these cells in further studies. Additionally, three control hiPSCs lines currently used in the laboratory – Gibco, TCLab and DF6 (the isogenic cell line for the gene-edited clones), along with a patient-derived hiPSCs line denominated as "Xutl" were also differentiated into cardiomyocytes. This cell line harbours an insertion of a C in exon 13 of the *MYBPC3* gene, that leads to a frameshift mutation - p.I1250fs, and a PTC is predicted to be generated, with the mutated transcripts being predicted to be degraded by the NMD pathway; therefore, this cell line could also be used in our study to test the haploinsufficiency hypothesis and to evaluate possible compensatory mechanisms.

Regarding the 2 gene-edited DF6 clones, no morphological alterations were detected throughout the 30-days of cardiac differentiation, indicating no differentiation into CMs. Therefore, these cells could not be differentiated into cardiomyocytes to perform the intended genetic characterizations. We hypothesize that for the 2 homozygous clones, if the *MYBPC3* c.1090G>A mutation does indeed lead to splicing alterations and the generation of a PTC, this would result in the degradation of the mRNAs transcribed from both alleles, and the cells would likely either not differentiate into CMs or have severe sarcomere disfunction since the MYBPC3 protein, essential for the sarcomere's function, is not produced.

Furthermore, in all gene-edited clones screened, no other mutations in the region of about ~500 bp sequenced were observed. However, if the 3 DF6 *MYBPC3* c.1090G>A gene-edited clones were to be further differentiated into CMs and characterized, at least the sequencing of the top predicted off-target regions or an exome sequencing would be required to verify the presence of possible off-target mutations, that could strongly influence the experiment and the results.

The DF6 hiPSCs line was not successfully differentiated, since only very restricted regions presented beating clusters of cells. Concerning the TCLab, Gibco and Xutl cell lines, the first beating agglomerates of cells were observed on days 8-10 of differentiation, and, throughout time in culture, robust spontaneous contraction became evident and the hiPSCs-CMs started beating in dense masses. According to the results presented in **Figure 4.11A**, in which the expression levels are normalized against U6, the 3 cell lines successfully differentiated express all sarcomeric genes analysed, but in extremely low levels when compared to the expression in the adult human heart. However, in this differentiation protocol an heterogenous population of cells is obtained. Knowing that the genes analysed are sarcomere-specific, by normalizing their expression levels to a cardiac marker, h*TNNT2*, the results would reflect the expression in cardiomyocytes, rather than in the overall population of cells obtained. Therefore, the expression levels were normalized against h*TNNT2* since it starts to be expressed at day 7 of differentiation [57] and previous unpublished laboratory results revealed that the levels of *TNNT2* are very constant between differentiations. Such analysis show that the expression levels of h*MYBPC3*, h*MYH7*, h*TTN* and h*ACTN2* in the hiPSCs-CMs at day 30 of differentiation are actually very similar to those found in human heart (**Figure 4.11B**). This indicates that the cardiomyocytes obtained through this 2D protocol express similar levels of sarcomeric genes to those found in human heart but considering that when normalized against U6 the levels are very low, an heterogenous population of cells is generated where the majority are not cardiomyocytes.

The high expression levels of the fetal *TNNT2* and *TTN* isoforms and the low expression of adult *TTN* are consistent with CMs at an embryonic state. *TNNI3*, the gene that encodes for cardiac troponin I, is expressed in adult hearts, whereas *TNNI1* is expressed in human embryonic cardiomyocytes. All differentiated cell lines have expression of *TNNI1* and only in the TCLab and Xutl hiPSCs-CMs there is a very small expression of h*TNNI3*, further supporting that the hiPSCs-CMs obtained are not extensively maturated. The low expression levels of the adult isoforms of sarcomeric genes and a less-organized sarcomeric structure have been reported in the literature in immature cardiomyocytes [45]. In the adult heart, very small levels of h*TNNT2* fetal isoform is expressed, and the adult h*TTN* N2B isoform

is more expressed than the fetal N2BA. Additionally, h*TNNI1* (fetal) is not expressed. Therefore, one of the main issues with the 2D cardiac differentiation protocol used is the reduced maturity level of the hiPSCs-cardiomyocytes obtained.

In the immunofluorescence images presented in **Figure 4.12**, the expression of MYBPC3, MYH7 and α-ACTN2 protein is verified, and it is also clear that the hiPSCs-CMs differentiated through this method present a relatively high degree of organization of the sarcomere structure, which is especially evident in **Figure 4.12G** and **H**. However, it is also noticeable that the striation appears to be restricted to the centre of the cell and does not spread to the periphery. In addition, the cells are rounder and do not have the elongated morphology that is expected for mature cardiomyocytes, and so the embryonic phenotype characteristics of these cardiomyocytes is reinforced by this analysis.

There were no substantial differences between the qRT-PCR and immunofluorescence results from the 2 control cell lines (Gibco and TCLab) and the patient-derived Xutl cells. HCM-cardiomyocytes are reported to be bigger and rounder than healthy CMs and should present a more fetal-like phenotype. In addition, HCM-CMs are reported to express higher levels of fetal isoforms, consistent with an immature phenotype. However, considering that this is only one differentiation experiment and that a high level of maturity could not be achieved through this protocol, no clear conclusions can be reached. Consequently, additional experiments are required for further improvement of the method itself, and once the protocol is optimized the differences between the control and HCM cell lines can be assessed. Regarding the Xutl line, once a pure population of considerably mature hiPSCs-CMs is achieved, a transcriptomic and proteomic analysis could be performed to understand if the mutated transcripts are degraded by NMD and if there are higher levels of mRNA being produced from the wild-type allele.

The intrinsic variability of the differentiation method should also be addressed, as even though the cells are all exposed to the same stimuli, the protocol often fails. The differentiation efficiency is also dependent on the hiPSCs cell line and the conditions should be optimized in order to achieve a higher cardiomyocyte yield. The starting quality of the hiPSCs may also have an impact in the differentiation, as partially spontaneous differentiated hiPSCs will reduce the efficiency of the protocol [57]. This could explain the fact that the DF6 cell line was not successfully differentiated. The unspecific differentiation into other cell types also poses as a problem. The cells visible in **Figure 4.12** that are only stained with phalloidin are most likely smooth muscle or endothelial cells, and the development of methods that lead to a pure population of cardiomyocytes are necessary.

For disease-modelling purposes, the maturity of the hiPSCs-CMs should be comparable to that of cardiomyocytes in the adult heart. Thus, approaches that increase the maturity and the yield of hiPSCs-CMs have been developed. So, methods that resort to longer periods in culture, mechanical stimulation [45] and 3D differentiation protocols that result in the generation of more mature CMs when compared to a 2D approach were already established [109]. The sorting of cells for specific surface markers such as VCAM-1 in order to purify cardiomyocytes from an heterogeneous population can also be applied to obtain an enriched population of hiPSCs-CMs [110].

6. Conclusions and Future Work

Hypertrophic Cardiomyopathy is a highly heterogenous genetic disease that has a striking prevalence of 0.5% in the general population. Considering that HCM-related mutations have been identified in over 14 genes and are often associated with diverse phenotypes and clinical outcomes, a clear genotype-phenotype association has not yet been established for most causal mutations.

The recent emergence of hiPSCs technology and the development of *in vitro* differentiation protocols into specialized cells of the three germ layers holds great promise for disease modelling, particularly in the field of genetic diseases. Nowadays, hiPSCs can be successfully differentiated into functional cardiomyocytes by a range of diverse protocols that employ different strategies in order to generate models that recapitulate the phenotype of healthy and diseased cardiomyocytes in the adult human heart. The diseased iPSCs-CMs can be characterized at the transcriptomic and proteomic level, and differences in their morphology and physiology, such as calcium handling alterations and differential response to drugs can also be assessed and further compared to that of healthy CMs.

Previous studies have already demonstrated that missense mutations that are hypothesized to cause the disease due to a direct change in the encoded aminoacid, in fact may lead to mis-splicing and are therefore pathogenic due to their impact in this mechanism [25,80].

Out of the 7 missense variants that were chosen following a database search and bioinformatic analysis resorting to tools that predict the possible effect of a specific variant on splicing, 2 missense variants (*MYH7* c.2221G>T, *MYBPC3* c.1090 G>A) were ultimately selected after sgRNA efficiency testing in HEK293T cells. *MYH7* c.2221G>T is predicted to lead to the activation of a cryptic donor site, whereas *MYBPC3* c.1090 G>A variant is thought to disrupt the canonical donor site and is located in the last nucleotide of exon 12, similar to the *MYBPC3* c.772G>A variant that had previously been associated to abnormal splicing. In addition to the missense variants that have a possible effect on splicing, a *MYBPC3* c.927-2 A>G splice-site mutation was to be used as a positive control for a splicing-altering variant. During the course of this thesis, a publication reported that the *MYBPC3* c.1090G>A does indeed lead to abnormal splicing, with the skipping of exon 12, the same mechanism that was hypothesized in our laboratory.

The one-step reprogramming and gene-editing of human fibroblasts was a promising approach, considering that a gene-editing efficiency of 2-20% was reported using this protocol, in contrast to the substantially lower 0.5-1% efficiency of knock-in in hiPSCs. Additionally, this strategy does not require selection strategies that further complicate the process, and the fibroblasts cells are also less sensitive to the overall transfection and culture procedures than hiPSCs. Thus, this method could potentially be used to overcome the previous complications faced in the laboratory with the gene-editing protocols that are performed directly in hiPSCs. However, the efficiency of the CRISPR/Cas9 technology is strongly dependent on the successful reprogramming of the human fibroblast primary cell lines. After several attempts to generate edited hiPSCs using this protocol, none of the clones picked presented any genetic alterations. More importantly, the morphology of those colony-like structures did not completely recapitulate the characteristics of hiPSC clones, and subsequent qRT-PCR analysis revealed that the majority of these clones still expressed high levels of a fibroblast marker and had very low expression levels of the pluripotency genes *hNanog* and *hDNMT3B*. Therefore, in addition to the inefficiency of the CRISPR/Cas9 technology, it is also hypothesized that these cells are most likely in an intermediate state of reprogramming. Consequently, an alternative gene-editing protocol was tested, in order to try to accomplish, within the one-year time period of this thesis, the first task of this project.

Thus, a total of four different conditions of gene-editing in the hiPSCs commercial DF6 cell line were tested, using the cell-cycle regulated Cas9-Gem mRNA and the pCas9-GFP, resorting to the incubation of the cells at 32ºC for 48h post-transfection to promote HDR activity [74]. In this experiment, a NHEJ efficiency of 14.0% was obtained in the two conditions that employed selection strategies. Three correctly gene-edited clones were also identified in both conditions, consisting of an overall knock-in efficiency of 0.9%. However, due to the long period in culture and their maintenance

in suboptimal conditions, the three gene-edited clones that harbour the variant classified as missense *MYBPC3* c.1090G>A lost the expression of pluripotency markers, and therefore could not be differentiated into CMs. Thus, the main achievement of this thesis was the introduction of the variant of interest in the genome of hiPSCs, even though the gene-edited clones could not be used for further studies.

In future experiments, the sgRNAs efficiency test should be performed in the hiPSCs lines, since the editing efficiency of the same locus in different cell lines is variable. Furthermore, considering that the chromatin status can influence the accessibility of the CRISPR/Cas9 complex to the target region, performing the gene-editing experiment in other genes could also be attempted, for example with the *MYH7* c.2221G>T variant, as the editing efficiency in this region could be higher. The overall protocol requires further optimization, and cryopreservation conditions of the hiPSCs clones in 48-well plates are presently being tested, as it would be an alternative that would significatively reduce the costs and the amount of work involved in the experiment.

Despite the remarkable potential of the CRISPR/Cas9 technology, and even though clinical trials are already ongoing, its safety is still a roadblock for clinical translation, mainly due to off-target events of the Cas9 nuclease. These off-targets effects have resulted in large deletions and genomic rearrangements that have been reported both in mouse and human cells. Advances in this area include the development of more accurate Cas9 variants that have reduce off-target activity, therefore improving the precision of gene-editing [71].

Finally, control and patient-derived hiPSCs lines were differentiated into CMs using a 2D approach, resulting in the generation of cardiomyocytes that have similar expression of sarcomeric genes to those found in the human heart. However, these CMs present fetal-like characteristics, such as the predominant expression of the h*TNNI1* gene along with the h*TTN* N2BA and *hTNNT2* fetal isoforms. Therefore, one of the main problems with this protocol is the reduced maturity level of the hiPSCs-cardiomyocytes obtained. Furthermore, a heterogeneous cell population is produced, with the majority of cells not expressing sarcomeric-specific genes. Consequently, the enrichment in CMs during the differentiation process is also required to generate a model that would better recapitulate the phenotype of the adult human heart. No substantial differences were found between the wild-type and the patient derived Xutl (p.I1250fs) CMs and, even though this is only one experiment, it indicates that a deeper analysis could be required to completely understand the mechanism underlying these causal mutations and if a transcriptional compensation system is activated. Thus, despite the great potential of hiPSC-CMs, current cardiac differentiation protocols give rise to cells that differ from the human adult myocardium due to their immaturity and heterogeneity, as the reproducibility and efficiency of the differentiation protocols also continue to be two of the main challenges for the generation of CMs.

The establishment of hiPSCs-based disease models are not only suitable for the understanding of how gene mutations lead to disease, as they could also be used to test different therapeutic approaches that either minor or correct disease-causing mutations. These therapeutically options include, in addition to the CRISPR/Cas9 system that could be used to correct a mutation, technologies such as RNA interference, cDNA overexpression, exon skipping and trans-splicing. Recently, gene replacement and trans-splicing techniques were tested in hiPSCs derived from a patient with HCM bearing a *MYBPC3* truncating mutation. Even though the trans-splicing efficiency was low, *MYBPC3* pre-mRNA replacement with full-length cDNA lead to a 2.5-fold increase in total *MYBPC3* mRNA levels, restoring protein expression in up to 81% [111]. Therefore, once the mechanism by which the mutations studied in this thesis cause the disease are understood, these therapeutic approaches may also be tested in the hiPSCs-CMs models.

In conclusion, the development of cardiac models that are able to recapitulate the phenotype of healthy and diseased human cardiomyocytes is crucial to decipher the molecular mechanisms underlying HCM. In order to achieve this, the CRISPR-Cas9 technology is indispensable, not only to introduce the variant in study in the genome, but also to generate isogenic controls, essential for the accurate determination of genotype-phenotype correlations that would allow for a patient-oriented preventive care and precision medicine.

7. Bibliography

[1] J. R. Eugene Braunwald , Costas T. Lambrew , S. David Rockoff and A. G. Morrow, “Idiopathic Hypertrophic Subaortic Stenosis: I. A Description of the Disease Based Upon an Analysis of 64 Patients,” *Circulation*, vol. IV–3–IV–11, 1964.

[2] B. J. Maron, E. J. Rowin, and M. S. Maron, “Global Burden of Hypertrophic Cardiomyopathy,” *JACC Hear. Fail.*, vol. 6, no. 5, pp. 376–378, 2018.

[3] I. Christiaans and P. M. Elliott, “Hypertrophic cardiomyopathy,” *Lancet*, pp. 61–74, 2013.

[4] S. K. Viswanathan *et al.*, “Hypertrophic cardiomyopathy clinical phenotype is independent of gene mutation and mutation dosage,” *PLoS One*, vol. 12, no. 11, pp. 1–19, 2017.

[5] C. Roma-Rodrigues and A. R. Fernandes, “Genetics of hypertrophic cardiomyopathy: Advances and pitfalls in molecular diagnosis and therapy,” *Appl. Clin. Genet.*, vol. 7, pp. 195–208, 2014.

[6] M. Y. Desai, S. R. Ommen, W. J. McKenna, H. M. Lever, and P. M. Elliott, “Imaging Phenotype Versus Genotype in Hypertrophic Cardiomyopathy,” *Circ. Cardiovasc. Imaging*, vol. 4, no. 2, pp. 156–168, 2011.

[7] J. B. Geske, S. R. Ommen, and B. J. Gersh, “Hypertrophic Cardiomyopathy: Clinical Update,” *JACC Hear. Fail.*, vol. 6, no. 5, pp. 364–375, 2018.

[8] R. M. de A. M. de Almeida, “Exploring new biological frontiers in Hypertrophic Cardiomyopathy,” 2017.

[9] B. J. Maron, “Clinical Course and Management of Hypertrophic Cardiomyopathy,” *N. Engl. J. Med.*, vol. 379, no. 7, pp. 655–668, 2018.

[10] B. J. Maron, M. S. Maron, and C. Semsarian, “Genetics of Hypertrophic Cardiomyopathy After 20 Years,” *J. Am. Coll. Cardiol.*, vol. 60, no. 8, pp. 705–715, 2012.

[11] I. Christiaans and P. M. Elliott, “Hypertrophic cardiomyopathy,” *Clin. Cardiogenetics Second Ed.*, pp. 61–74, 2016.

[12] H. Hartmannova *et al.*, “Isolated X-linked hypertrophic cardiomyopathy caused by a novel mutation of the four-and-a-half LIM domain 1 gene,” *Circ. Cardiovasc. Genet.*, vol. 6, no. 6, pp. 543–551, 2013.

[13] J. L. Zamorano *et al.*, “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),” *Eur. Heart J.*, vol. 35, no. 39, pp. 2733–2779, 2014.

[14] M. Sabater-Molina, I. Pérez-Sánchez, J. P. Hernández del Rincón, and J. R. Gimeno, “Genetics of hypertrophic cardiomyopathy: A review of current state,” *Clin. Genet.*, vol. 93, no. 1, pp. 3–14, 2018.

[15] S. M. Berget, C. Moore, and P. A. Sharp, “Spliced segments at the 5’ terminus of adenovirus 2 late mRNA,” *Proc. Nati. Acad. Sci. USA*, vol. 74, no. 8, pp. 3171–3175, 1977.

[16] A. Beqqali, “Alternative splicing in Heart Disease,” vol. 2700, pp. 1–13, 2006.

[17] E. Lara-Pezzi, J. Gómez-Salinero, A. Gatto, and P. García-Pavía, “The alternative heart: Impact of alternative splicing in heart disease,” *J. Cardiovasc. Transl. Res.*, vol. 6, no. 6, pp. 945–955, 2013.

[18] A. Abramowicz and M. Gos, “Splicing mutations in human genetic disorders: examples, detection, and confirmation,” *J. Appl. Genet.*, vol. 60, no. 2, pp. 231–231, 2019.

[19] S. L. W. and J. A. S. Michael R.Lerner, John A.Boyle, Stephen M.Mount, “Are snRNPs involved in splicing?,” *Nature*, vol. 283, no. 9, pp. 1689–1699, 1980.

[20] C. Zhu, Z. Chen, and W. Guo, “Pre-mRNA mis-splicing of sarcomeric genes in heart failure,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1863, no. 8, pp. 2056–2063, 2017.

[21] N. Andre and T. A. Cooper, “Pre-mRNA splicing and human disease,” vol. 14, no. 4, pp. 568–594, 2016.

[22] C. J. Weeland, M. M. van den Hoogenhof, A. Beqqali, and E. E. Creemers, “Insights into alternative splicing of sarcomeric genes in the heart,” *J. Mol. Cell. Cardiol.*, vol. 81, pp. 107–113, 2015.

[23] L. E. Maquat, “Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics,” *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 2, pp. 89–99, 2004.

[24] A. A. Glazier, A. Thompson, and S. M. Day, “Allelic imbalance and haploinsufficiency in MYBPC3-linked hypertrophic cardiomyopathy,” *Pflugers Arch. Eur. J. Physiol.*, vol. 471, no. 5, pp. 781–793, 2019.

[25] P. S. Andersen *et al.*, “Genetic and phenotypic characterization of mutations in myosin-binding protein C (MYBPC3) in 81 families with familial hypertrophic cardiomyopathy : total or partial haploinsufficiency,” *Eur. J. Hum. Genet.*, vol. 12, pp. 673–677, 2004.

[26] R. Vaz-Drago, N. Custódio, and M. Carmo-Fonseca, “Deep intronic mutations and human disease,” *Hum. Genet.*, vol. 136, no. 9, pp. 1093–1111, 2017.

[27] R. Mendes De Almeida *et al.*, “Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy,” *PLoS One*, vol. 12, no. 8, pp. 1–19, 2017.

[28] G. Bonne *et al.*, “Cardiac myosin binding protein–C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy,” *Nat. Genet.*, vol. 11, no. 4, pp. 438–440, 1995.

[29] L. Thierfelder *et al.*, “α-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: A disease of the sarcomere,” *Cell*, vol. 77, no. 5, pp. 701–712, 1994.

[30] P. E. Mohammed Akhtar*, “The genetics of hypertrophic cardiomyopathy,” *Glob. Cardiol. Sci. Pract.*, vol. 36, pp. 2733–2779, 2018.

[31] J. L. Strande, “Haploinsufficiency MYBPC3 mutations: Another stress induced cardiomyopathy? Let’s take a look!,” *J. Mol. Cell. Cardiol.*, vol. 79, pp. 284–286, 2015.

[32] R. Y. Parbhudayal *et al.*, “Variable cardiac myosin binding protein-C expression in the myofilaments due to MYBPC3 mutations in hypertrophic cardiomyopathy,” *J. Mol. Cell. Cardiol.*, vol. 123, no. August, pp. 59–63, 2018.

[33] M. K. David Barefielda, “Haploinsufficiency of MYBPC3 Exacerbates the Development of Hypertrophic Cardiomyopathy in Heterozygous Mice,” *J Mol Cell Cardiol*, vol. 79, pp. 234–243, 2015.

[34] D. Melton, *“*Stemness”: Definitions, Criteria, and Standards," *Essentials of Stem Cell Biology*, vol. 3, pp. 7-17, 2013.

[35] R. S. Mahla, “Stem cells applications in regenerative medicine and disease therapeutics,” *Int. J. Cell Biol.*, vol. 2016, pp. 1-24, 2016.

[36] L. A. Fortier, “Stem cells: Classifications, controversies, and clinical applications,” *Vet. Surg.*, vol. 34, no. 5, pp. 415–423, 2005.

[37] A. Elbuluk, T. A. Einhorn, and R. Iorio, “A comprehensive review of stem-cell therapy,” *JBJS Rev.*, vol. 5, no. 8, pp. 1–9, 2017.

[38] J. A. Thomson, “Embryonic stem cell lines derived from human blastocysts,” *Science (80-.).*, vol. 282, no. 5391, pp. 1145–1147, 1998.

[39] M. H. Kaufman and M. J. Evans, “Establishment in culture of pluripotential cells from mouse embryos,” *Nature*, vol. 292, no. July, pp. 154–156, 1981.

[40] J. A. Thomson *et al.*, “Isolation of a primate embryonic stem cell line.,” *Proc. Natl. Acad. Sci.*, vol. 92, no. 17, pp. 7844–7848, 1995.

[41] K. Takahashi and S. Yamanaka, “Induction of Pluripotent Stem Cells from Mouse Embryonic

and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.

[42] C. A. Cowan, J. Atienza, D. A. Melton, and K. Eggan, "Developmental Biology: Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells," *Science (80-.).*, vol. 309, no. 5739, pp. 1369–1373, 2005.

[43] T. Wakayama, V. Tabar, I. Rodriguez, A. C. F. Perry, L. Studer, and P. Mombaerts, "Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer," *Science (80-.).*, vol. 292, no. 5517, pp. 740–743, 2001.

[44] I. H. Park *et al.*, "Reprogramming of human somatic cells to pluripotency with defined factors," *Nature*, vol. 451, no. 7175, pp. 141–146, 2008.

[45] Y. Yoshida and S. Yamanaka, "Induced Pluripotent Stem Cells 10 Years Later," *Circ. Res.*, vol. 120, no. 12, pp. 1958–1968, 2017.

[46] K. Takahashi *et al.*, "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors," *Cell*, vol. 131, no. 5, pp. 861–872, 2007.

[47] P. J. Gokhale and P. W. Andrews, "Characterization of human pluripotent stem cells," *Nat. Protoc.*, vol. 24, no. 18, pp. 1031–1034, 2013.

[48] D. A. Robinton and G. Q. Daley, "The promise of induced pluripotent stem cells in research and therapy," *Nature*, vol. 481, no. 7381, pp. 295–305, 2012.

[49] M. Diana Neely, Andrew M. Tidball, Asad A. Aboud, Kevin C. Ess, *Induced Pluripotent Stem Cells (iPSCs): An Emerging Model System for the Study of Human Neurotoxicology*, vol.56, pp. 27-61, 2011.

[50] S. P. Bharathan *et al.*, "Systematic evaluation of markers used for the identification of human induced pluripotent stem cells," *Biol. Open*, vol. 6, no. 1, pp. 100–108, 2017.

[51] A. Prasad, D. T. B. Loong, S. M. Chua, A. H. All, J. Manivannan, and P. M. Gharibani, "A review of induced pluripotent stem cell, direct conversion by trans-differentiation, direct reprogramming and oligodendrocyte differentiation," *Regen. Med.*, vol. 11, no. 2, pp. 181–191, 2016.

[52] S. Assou, J. Bouckenheimer, and J. De Vos, "Concise Review: Assessing the Genome Integrity of Human Induced Pluripotent Stem Cells: What Quality Control Metrics?," *Stem Cells*, vol. 36, no. 6, pp. 814–821, 2018.

[53] R. G. Rowe and G. Q. Daley, "Induced pluripotent stem cells in disease modelling and drug discovery," *Nat. Rev. Genet.*, vol. 20, no. 7, pp. 377–388, 2019.

[54] M. F. Hoes, N. Bomer, and P. van der Meer, "Concise Review: The Current State of Human In Vitro Cardiac Disease Modeling: A Focus on Gene Editing and Tissue Engineering," *Stem Cells Transl. Med.*, vol. 8, no. 1, pp. 66–74, 2019.

[55] I. Kehat *et al.*, "Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes," *J. Clin. Invest.*, vol. 108, no. 3, pp. 407–414, 2001.

[56] I. Karakikes, M. Ameen, V. Termglinchan, and J. C. Wu, "Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes: Insights into Molecular, Cellular, and Functional Phenotypes," *Circ. Res.*, vol. 117, no. 1, pp. 80–88, 2015.

[57] X. Lian *et al.*, "Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signalling under fully defined conditions," *Nat. Protoc.*, vol. 8, no. 1, pp. 162–175, 2013.

[58] I. Batalov and A. W. Feinberg, "Differentiation of Cardiomyocytes from Human Pluripotent Stem Cells Using Monolayer Culture ivan," *Biomark. Insights*, vol. 10, pp. 71–76, 2015.

[59] D. E. Reichman *et al.*, "Wnt inhibition promotes vascular specification of embryonic cardiac progenitors," *Development*, vol. 145, no. 1, pp. 1-12, 2018.

[60] S. Peischard, I. Piccini, N. Strutz-Seebohm, B. Greber, and G. Seebohm, "From iPSC towards cardiac tissue—a road under construction," *Pflugers Arch. Eur. J. Physiol.*, vol. 469, no. 10, pp. 1233–1243, 2017.

[61] C. Freund *et al.*, “Insulin Redirects Differentiation from Cardiogenic Mesoderm and Endoderm to Neuroectoderm in Differentiating Human Embryonic Stem Cells,” *Stem Cells*, vol. 26, no. 3, pp. 724–733, 2008.

[62] C. Robertson, D. D. Tran, and S. C. George, “Concise review: Maturation phases of human pluripotent stem cell-derived cardiomyocytes,” *Stem Cells*, vol. 31, no. 5, pp. 829–837, 2013.

[63] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and applications of CRISPR-Cas9 for genome engineering,” *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[64] F. Zhang, Y. Wen, and X. Guo, “CRISPR/Cas9 for genome editing: Progress, implications and challenges,” *Hum. Mol. Genet.*, vol. 23, no. R1, pp. 40–46, 2014.

[65] B. M. Motta, P. P. Pramstaller, A. A. Hicks, and A. Rossini, “The impact of CRISPR/Cas9 technology on cardiac research: From disease modelling to therapeutic approaches,” *Stem Cells Int.*, pp. 1-13, 2017.

[66] and G. M. C. Prashant Mali, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, “RNA-Guided Human Genome Engineering via Cas9,” *Science*, vol. 339, no. 6121, pp. 823–826, 2013.

[67] F. Z. Le Cong, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, Xuebing Wu, Wenyan Jiang, Luciano A. Marraffini, “Multiplex Genome Engineering Using CRISPR/Cas Systems,” *Science*, vol. 339, pp. 819–823, 2013.

[68] J. A. Doudna and E. Charpentier, “The new frontier of genome engineering with CRISPR-Cas9,” *Science (80-.).*, vol. 346, no. 6213, 2014.

[69] J. D. Sander and J. K. Joung, “CRISPR-Cas systems for editing, regulating and targeting genomes,” *Nat. Biotechnol.*, vol. 32, no. 4, pp. 347–350, 2014.

[70] T. Gaj, C. A. Gersbach, and C. F. Barbas, “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering,” *Trends Biotechnol.*, vol. 31, no. 7, pp. 397–405, 2013.

[71] L. You, R. Tong, M. Li, Y. Liu, J. Xue, and Y. Lu, “Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research,” *Mol. Ther. - Methods Clin. Dev.*, vol. 13, no. June, pp. 359–370, 2019.

[72] D. Kwart, D. Paquet, S. Teo, and M. Tessier-Lavigne, “Precise and efficient scarless genome editing in stem cells using CORRECT,” *Nat. Protoc.*, vol. 12, no. 2, pp. 329–334, 2017.

[73] X. L. Li *et al.*, “Highly efficient genome editing via CRISPR–Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression,” *Nucleic Acids Res.*, vol. 46, no. 19, pp. 10195–10215, 2018.

[74] Q. Guo *et al.*, “‘Cold shock’ increases the frequency of homology directed repair gene editing in induced pluripotent stem cells,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[75] S. E. Howden, J. A. Thomson, and M. H. Little, “Simultaneous reprogramming and gene editing of human fibroblasts,” *Nat. Protoc.*, vol. 13, no. 5, pp. 875–898, 2018.

[76] J. Zhang *et al.*, “LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency,” *Cell Stem Cell*, vol. 19, no. 1, pp. 66–80, 2016.

[77] F. Anokye-Danso *et al.*, “Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency,” *Cell Stem Cell*, vol. 8, no. 4, pp. 376–388, 2011.

[78] S. E. Howden *et al.*, “Chromatin-Binding Regions of EBNA1 Protein Facilitate the Enhanced Transfection of Epstein-Barr Virus-Based Vectors,” *Hum. Gene Ther.*, vol. 17, pp. 1-12, 2006.

[79] T. Gutschner, M. Haemmerle, G. Genovese, G. F. Draetta, and L. Chin, “Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair,” *Cell Rep.*, vol. 14, no. 6, pp. 1555–1566, 2016.

[80] L. Carrier, G. Mearini, K. Stathopoulou, and F. Cuello, “Cardiac myosin-binding protein C (MYBPC3) in cardiac pathophysiology,” *Gene*, vol. 573, no. 2, pp. 188–197, 2015.

[81] E. S. Singer, J. Ingles, C. Semsarian, and R. D. Bagnall, “Key Value of RNA Analysis of MYBPC3 Splice-Site Variants in Hypertrophic Cardiomyopathy,” *Circ. Genomic Precis. Med.*,

vol. 12, no. 1, pp. 32-39, 2019.

[82] S. M. Badenes *et al.*, "Defined essential 8′ medium and vitronectin efficiently support scalable xeno-free expansion of human induced pluripotent stem cells in stirred microcarrier culture systems," *PLoS One*, vol. 11, no. 3, pp. 1–19, 2016.

[83] X. Li, G. Meng, R. Krawetz, S. Liu, and D. E. Rancourt, "The ROCK inhibitor Y-27632 enhances the survival rate of human embryonic stem cells following cryopreservation," *Stem Cells Dev.*, vol. 17, no. 6, pp. 1079–1085, 2008.

[84] F. P. Wachs *et al.*, "High efficacy of clonal growth and expansion of adult neural stem cells," *Lab. Investig.*, vol. 83, no. 7, pp. 949–962, 2003.

[85] P. L. Felgner *et al.*, "Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure," *Proc. Nati. Acad. Sci. USA*, vol. 84, no. November, pp. 7413–7417, 1987.

[86] R. P. Gonçalves "Assessment of hiPSC derived cardiomyocytes maturation in extended culture by analysis of alternative splicing expression patterns," 2017.

[87] M. J. Landrum *et al.*, "ClinVar: Public archive of relationships among sequence variation and human phenotype," *Nucleic Acids Res.*, vol. 42, pp. 980–985, 2014.

[88] F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-Bëroud, M. Claustres, and C. Béroud, "Human Splicing Finder: An online bioinformatics tool to predict splicing signals," *Nucleic Acids Res.*, vol. 37, no. 9, pp. 1–14, 2009.

[89] L. Eng *et al.*, "Nonclassical Splicing Mutations in the Coding and Noncoding Regions of the ATM Gene: Maximum Entropy Estimates of Splice Junction Strengths," *Hum. Mutat.*, vol. 23, no. 1, pp. 67–76, 2004.

[90] R. Leman *et al.*, "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 Res.*, vol. 46, no. 21, pp. 11656–11657, 2018.

[91] M. D. H Ideshi Niimura , M.D., Linda L. B AchinskI , M.D., Somkiat Sangwatanaroj , M.D., Hugh W. Atkins, M.D., P H .D., Albert E. Chudley , M.D., William Mckenna , M.D., Arni Kristinsson , M.D., P H.D., Robert Roberts , M.D., Michael Sole , M., "Mutations in the gene for cardiac myosin-binding protein c and late-onset familial hypertrophic cardiomyopathy," *N. Engl. J. Med.*, vol. 338, no. 18, pp. 1248–1257, 1998.

[92] Y. Zhang *et al.*, "Targets and genomic constraints of ectopic Dnmt3b expression," *Elife*, vol. 7, pp. 1–25, 2018.

[93] S. A. Jimenez and B. Saitta, "Alterations in the regulation of expression of the α1(I) collagen gene (COL1A1) in systemic sclerosis (scleroderma)," *Springer Semin. Immunopathol.*, vol. 21, no. 4, pp. 397–414, 1999.

[94] S. Suzuki, J. Namiki, S. Shibata, Y. Mastuzaki, and H. Okano, "The neural stem/progenitor cell marker nestin is expressed in proliferative endothelial cells, but not in mature vasculature," *J. Histochem. Cytochem.*, vol. 58, no. 8, pp. 721–730, 2010.

[95] K. Sakurai and N. Osumi, "The Neurogenesis-Controlling Factor, Pax6, Inhibits Proliferation and Promotes Maturation in Murine Astrocytes," *J. Neurosci.*, vol. 28, no. 18, pp. 4604–4612, 2008.

[96] J. Behnan *et al.*, "Identification and characterization of a new source of adult human neural progenitors" *Cell Death Dis.*, vol. 8, no. 8, pp. 1-13, 2017.

[97] N. Ma *et al.*, "Determining the pathogenicity of a genomic variant of uncertain significance using CRISPR/Cas9 and human-induced pluripotent stem cells," *Circulation*, vol. 138, no. 23, pp. 2666–2681, 2018.

[98] B. Adalsteinsdottir *et al.*, "Nationwide study on hypertrophic cardiomyopathy in iceland evidence of a MYBPC3 founder mutation," *Circulation*, vol. 130, no. 14, pp. 1158–1167, 2014.

[99] A. Rangaraju, S. M. Lova, N. Calambur, and P. Nallari, "Significance of intronic and synonymous MYBPC3 gene variations in hypertrophic cardiomyopathy," *Gene Reports*, vol. 8,

no. December 2016, pp. 94–99, 2017.

[100] S. Tripathi *et al.*, "Unequal allelic expression of wild-type and mutated β-myosin in familial hypertrophic cardiomyopathy," *Basic Res. Cardiol.*, vol. 106, no. 6, pp. 1041–1055, 2011.

[101] A. S. Helms *et al.*, "Sarcomere mutation-specific expression patterns in human hypertrophic cardiomyopathy," *Circ. Cardiovasc. Genet.*, vol. 7, no. 4, pp. 434–443, 2014.

[102] S. Marston *et al.*, "Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency," *Circ. Res.*, vol. 105, no. 3, pp. 219–222, 2009.

[103] S. J. Van Dijk *et al.*, "Cardiac myosin-binding protein C mutations and hypertrophic ardiomyopathy haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction," *Circulation*, vol. 119, no. 11, pp. 1473–1483, 2009.

[104] X. Xu *et al.*, "Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[105] X. D. Tang, F. Gao, M. J. Liu, Q. L. Fan, D. K. Chen, and W. T. Ma, "Methods for enhancing clustered regularly interspaced short palindromic repeats/Cas9-mediated homology-directed repair efficiency," *Front. Genet.*, vol. 10, no. 551, pp. 1-9, 2019.

[106] A. Seki and S. Rutz, "Optimized RNP transfection for highly efficient CRI SPR/Cas9-mediated gene knockout in primary T cells," *J. Exp. Med.*, vol. 215, no. 3, pp. 985–997, 2018.

[107] F. Song and K. Stieger, "Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks," *Mol. Ther. - Nucleic Acids*, vol. 7, no. June, pp. 53–60, 2017.

[108] S. A. Verkuijl and M. G. Rots, "The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies," *Curr. Opin. Biotechnol.*, vol. 55, pp. 68–73, 2019.

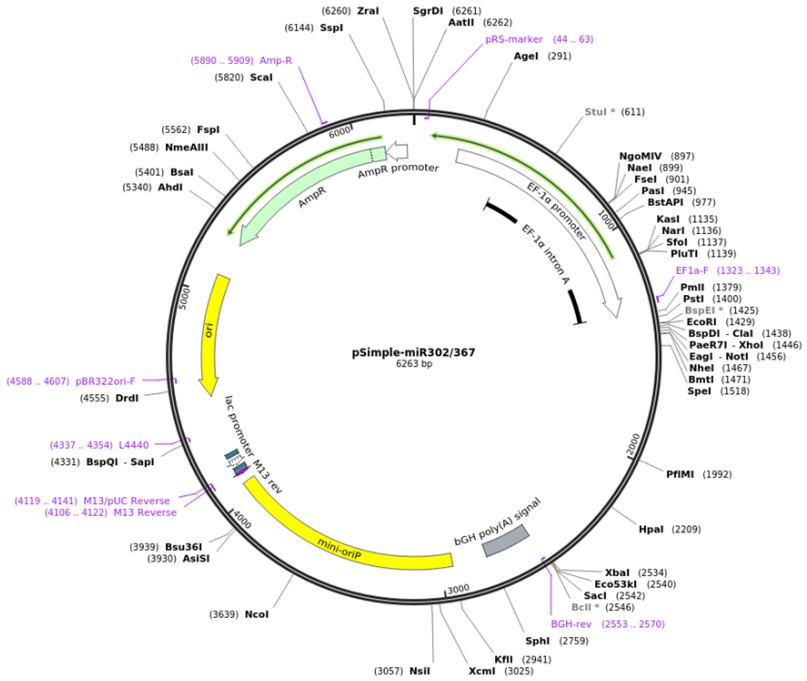
[109] M. A. Branco *et al.*, "Transcriptomic analysis of 3D Cardiac Differentiation of Human Induced Pluripotent Stem Cells Reveals Faster Cardiomyocyte Maturation Compared to 2D Culture," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, 2019.

[110] V. Schwach and R. Passier, "Generation and purification of human stem cell-derived cardiomyocytes," *Differentiation*, vol. 91, no. 4–5, pp. 126–138, 2016.

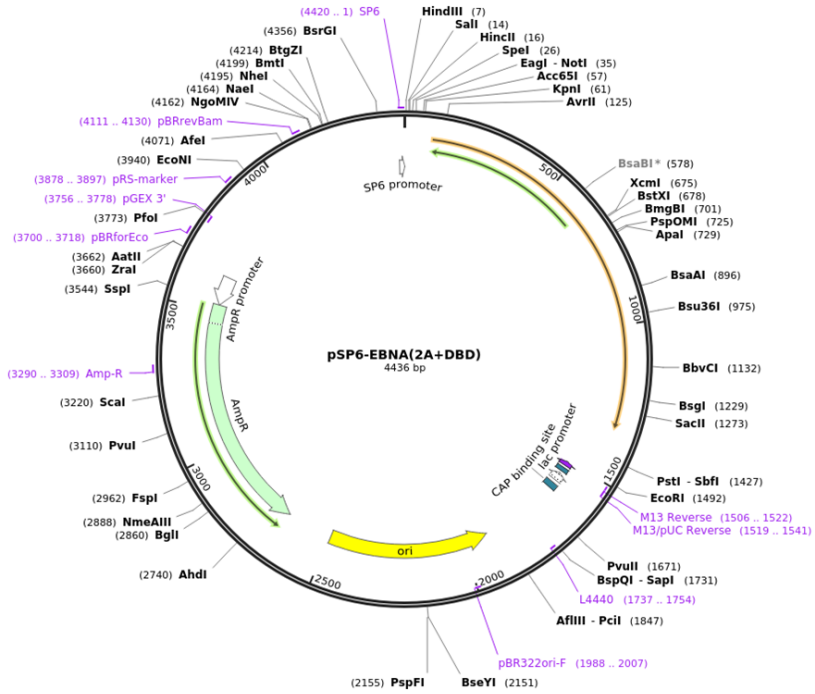
[111] M. Prondzynski *et al.*, "Evaluation of MYBPC3 trans-Splicing and Gene Replacement as Therapeutic Options in Human iPSC-Derived Cardiomyocytes," *Mol. Ther. - Nucleic Acids*, vol. 7, no. June, pp. 475–486, 2017.

8. Attachments

pSimple-miR302/367 **Addgene Plasmid #98748**
Reprogramming vector

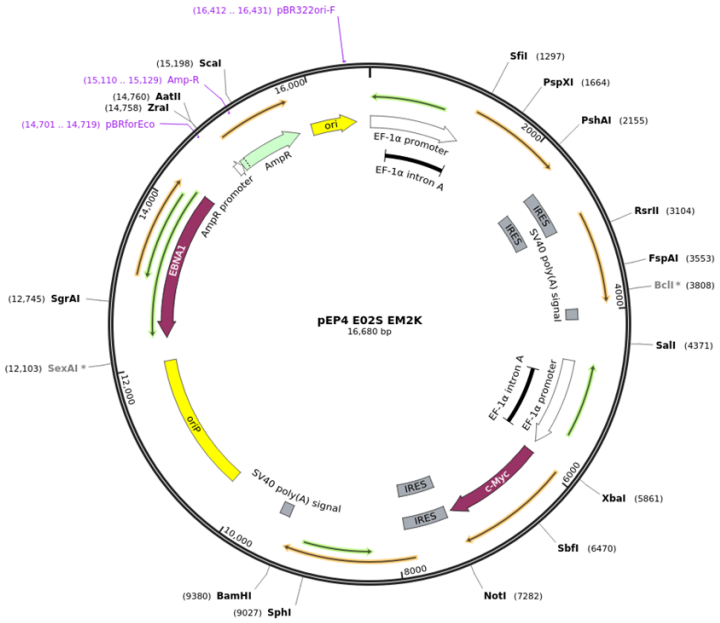


pSP6-EBNA(2A+DBD) **Addgene Plasmid #98749**
Reprogramming vector

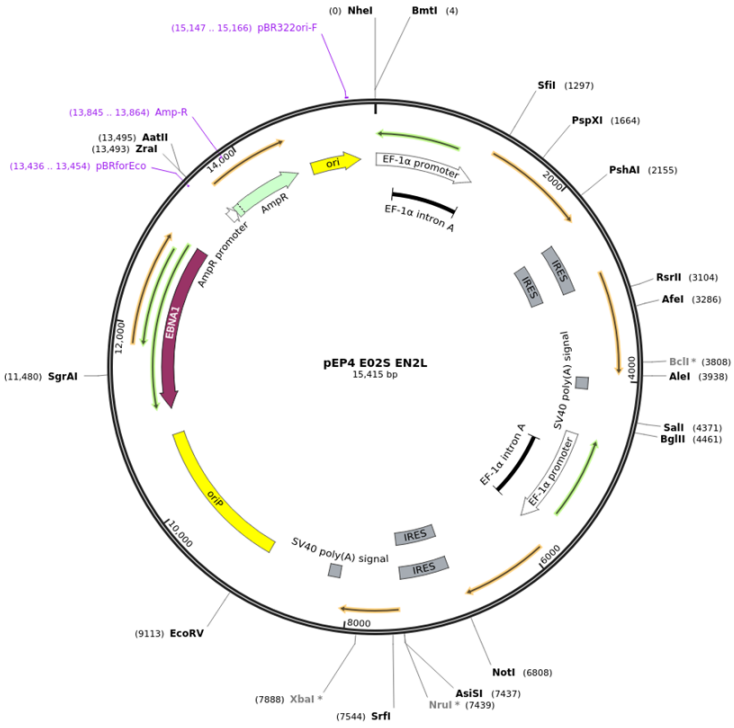


(continues on the next page)

pEP4 E02S EM2K **Addgene Plasmid #2092**
Reprogramming vector

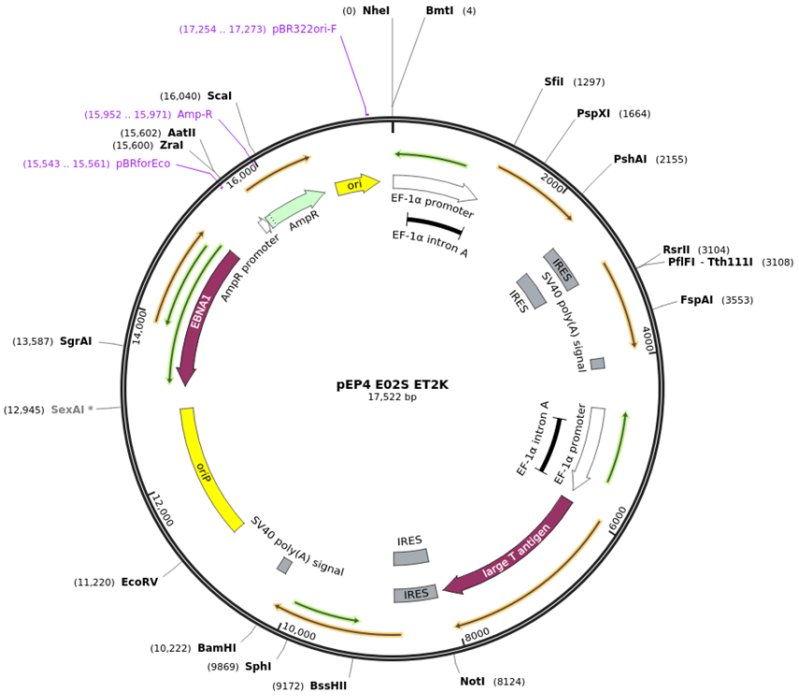


pEP4 E02S EN2L **Addgene Plasmid #20922**
Reprogramming vector

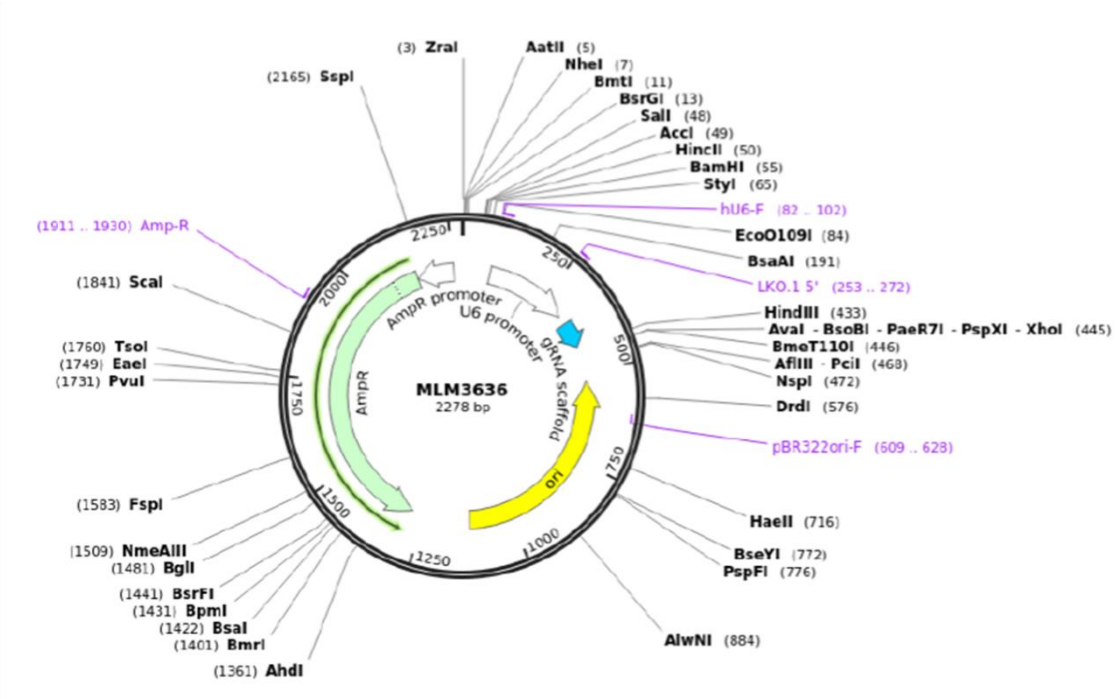


(continues on the next page)

pEP4 E02S ET2K **Addgene Plasmid #20927**
Reprogramming vector

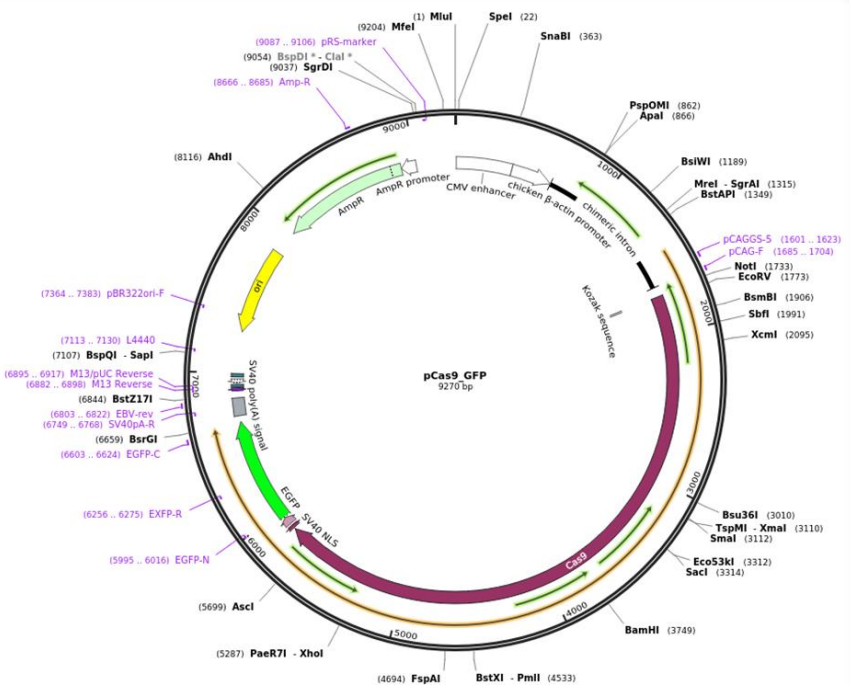


MLM3636 **Addgene Plasmid #43860**
sgRNA expression vector

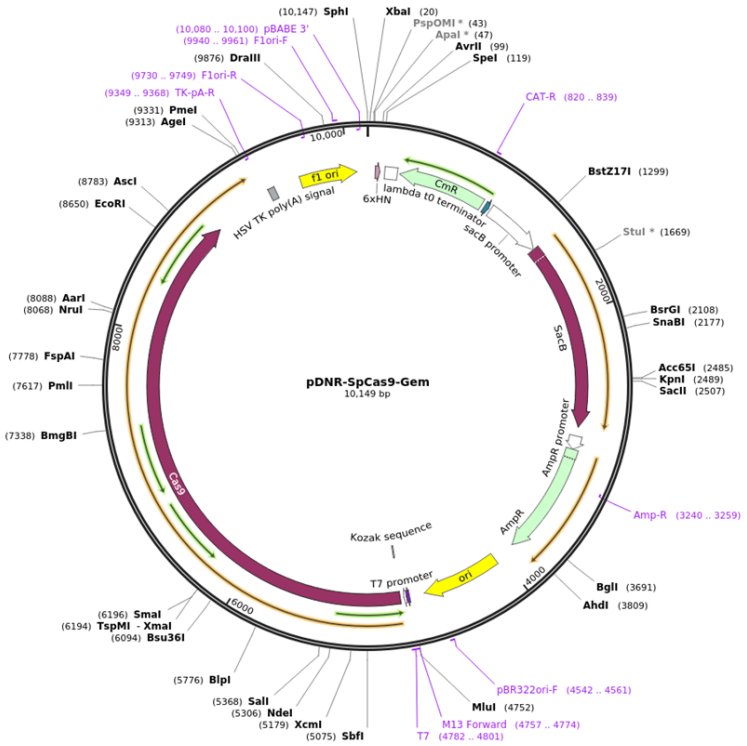


(continues on the next page)

pCas9_GFP **Addgene Plasmid #44719**
CRISPR/Cas9



pDNR-SpCas9-Gem **Addgene Plasmid #80424**
CRISPR/Cas9



(continues on the next page)

pPUR-hU6-sgRNA-Sirius-8XMS2 **Addgene Plasmid #121942**
hiPSCs antibiotic selection

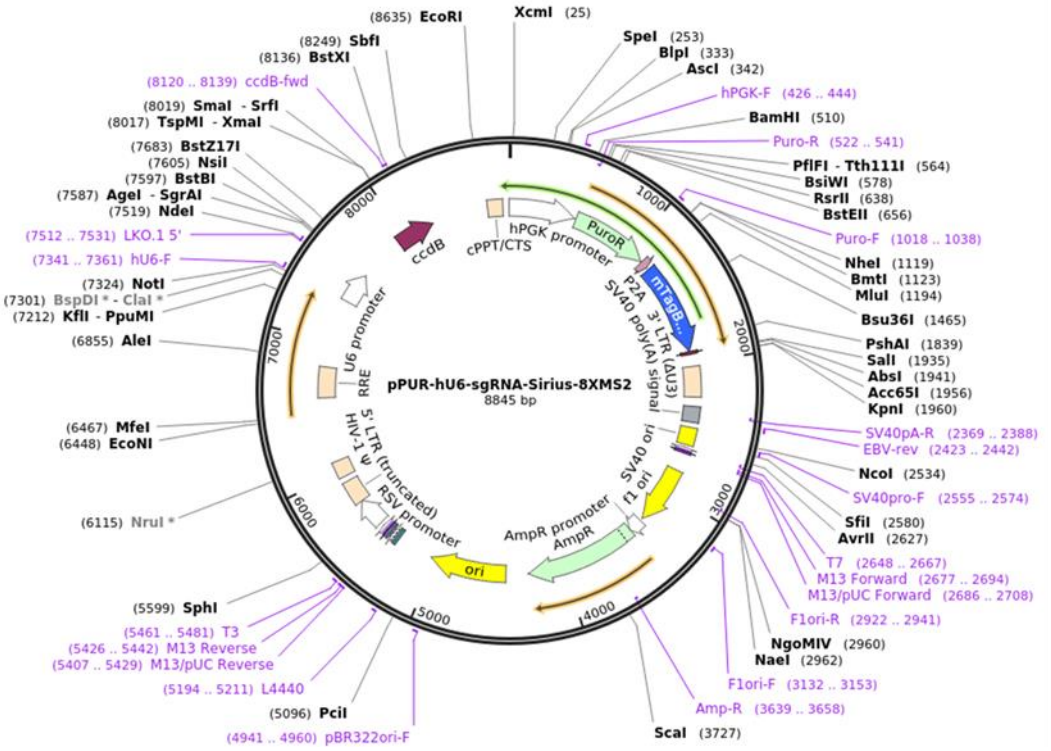


Figure 8.1 – Plasmids used for the reprogramming of human fibroblasts and the gene-editing experiments.

Table 8.1 – sgRNA sequences designed for each variant to be studied. The *MYBPC3* c.927-2A>G sgRNA had been previously selected in the laboratory.

Variants	**Guides**	
MYH7 c.1208G>A	Guide 4	**sgRNA sense** ACACCgCATTGCCCACTTTCACCCGAG **sgRNA antisense** AAACTCGGGTGAAAGTGGGCAATGCG
	Guide 5	**sgRNA sense** ACACCgTCAAGGGGCTGTGCCACCCTG **sgRNA antisense** AAACAGGGTGGCACAGCCCCTTGACG
MYH7 c.2146G>A	Guide 4	**sgRNA sense** ACACCgCTACGGGGACTTCCGGCAGAG **sgRNA antisense** AAACTCTGCCGGAAGTCCCCGTAGcG
	Guide 6	**sgRNA sense** ACACCgAACCGCATCCTCTACGGGGAG **sgRNA antisense** AAACTCCCCGTAGAGGATGCGGTTcG
	Guide 7	**sgRNA sense** ACACCgGCATCCTCTACGGGGACTTCG **sgRNA antisense** AAAC**G**AAGTCCCCGTAGAGGATGCcG
MYH7 c.2221G>T	Guide 1	**sgRNA sense** ACACCgGTTCATTGATAGCAGGAAGGG **sgRNA antisense** AAACCCTTCCTGCTATCAATGAACcG
	Guide 2	**sgRNA sense** ACACCgAGAGAAGCTGCTCAGCTCCCG **sgRNA antisense** AAACGGGAGCTGAGCAGCTTCTCTcG

(continues on the next page)

MYH7 c.2717A>G c.2722C>G	Guide 3	**sgRNA sense** ACACCgAGACAACCTGGCAGATGCTGG **sgRNA antisense** AAACCAGCATCTGCCAGGTTGTCTcG
	Guide 4	**sgRNA sense** ACACCgCTACGGGGACTTCCGGCAGAG **sgRNA antisense** AAACTCTGCCGGAAGTCCCCGTAGcG
MYBPC3 c.481C>T	Guide 2	**sgRNA sense** ACACCgCTTCGTGATGCGGCCACAGGG **sgRNA antisense** AAACCCTGTGGCCGCATCACGAAGcG
	Guide 4	**sgRNA sense** ACACCgCCATTGGCCTCTTCGTGATGG **sgRNA antisense** AAACCATCACGAAGAGGCCAATGGcG
MYBPC3 c.927-2A>G	Guide 1	**sgRNA sense** CACCGCCGGCCACAGCCTAGACTGC **sgRNA antisense** AAACGCAGTCTAGGCTGTGGCCGGC
MYBPC3 c.1090G>A	Guide 2	**sgRNA sense** ACACCgCGCGATGAGAAGAAGAGCACG **sgRNA antisense** AAACGTGCTCTTCTTCTCATCGCGcG
	Guide 4	**sgRNA sense** ACACCgAGGTTAGCCCTTCCTCAGAGG **sgRNA antisense** AAACCTCTGAGGAAGGGCTAACCTcG
	Guide 6	**sgRNA sense** ACACCgACAGGTTAGCCCTTCCTCAGG **sgRNA antisense** AAACCTGAGGAAGGGCTAACCTGTcG
TNNT2 c.275G>A	Guide 3	**sgRNA sense** ACACCgGAAGCGCATGGAGAAGGACCG **sgRNA antisense** AAACGGTCCTTCTCCATGCGCTTCcG
	Guide 7	**sgRNA sense** ACACCgACAGGACATCCACCGGAAGCG **sgRNA antisense** AAACGCTTCCGGTGGATGTCCTGTcG

Table 8.2 – Primers used for the amplification of the genomic target regions of each variant selected.

Variants	Primers
MYH7 c.1208G>A	Forward: CTTGCCTTTTCCTTCCAGAG Reverse: GGGAGCGAGTGAGTGATTGT
MYH7 c.2146G>A	Forward: CTGTTCTGGCTACCACCTATGA Reverse: CAGTGATCTGCTCTGCCCAT
MYH7 c.2221C>T	Forward: GGTATGAGGGTGCACCAGAG Reverse: GCCATGACTAGGGAGGGGTA
MYH7 c.2722C>G *MYH7* c.2717A>G	Forward: GCTCTTGCCAAGTGGGGATA Reverse: GGGTCAAAGGTCACCAGCTT
MYBPC3 c.481C>T	Forward: GGGCAGGTCTCAAAAGCCTT Reverse: GGCTTGGGGAGTGTCCTG
MYBPC3 c.927-2A>G	Forward: AGGTGAGTGCCCAACATAGC Reverse: CCATGTATGTGGACGAGGTG
MYBPC3 c.1090G>A	Forward: GGTGGCTACAGCTCCTTGGT Reverse: GTCCTTTTCCTTGTGGCTGT
TNNT2 c.275G>A	Forward: GGGGTAGAGCATAGGTGGGA Reverse: CCAGTAGGATGGGGAGGGAA

Table 8.3 - 100 nucleotide ssODN templates for homologous directed repair of double-strand breaks, designed for the 3 variants in study: *MYH7* c.2221G>T, *MYBPC3* c.1090G>A and *MYBPC3* c.927-2A>G.

Variants	ssODN Template
MYH7 c.2221G>T	CATCCTGAACCCAGCGGCCATCCCTGAGGGACAGTTCATTGATAGCAGGAAG**T**G GGCAGAGAAGCTGCTCAGCTCCCTGGACATTGATCACAACCAGTAC
MYBPC3 c.1090G>A	CAAGGGCATGAGGCGCGATGAGAAGAAGAGCACA**A**GTTAGCCCTTCCTCAGAG GGGAGAGGAGAGGGCACAGGCTAGCCCTTCCCTACACAGGAGAGGAG
MYBPC3 c.927-2A>G	GGAAGCCCCGCCCAGGGGGCTGCAGTCTTGCCCCCGGCCACAGCCTAGACTGCG GGACAC**G**GGGACTCGAAGCTGGAGGCACCAGCAGAGGAGGACGTGT

Table 8.4 - List of qPCR primers.

Gene	Primers
hCol1a1	Forward: GAGGGCCAAGACGAAGACATC Reverse: CAGATCACGTCATCGCACAAC
hBrachyury	Forward: TCAGCAAAGTCAAGCTCACCA Reverse: CCCCAACTCTCACTATGTGGATT
hNestin	Forward: GAAACAGCCATAGAGGGCAAA Reverse: TGGTTTTCCAGAGTCTTCAGTGA
hPax6	Forward: GAATCAGAGAAGACAGGCCA Reverse: GTGTAGGTATCATAACTCCG
hTubIII	Forward: CTCAGGGGCCTTTGGACATC Reverse: CAGGCAGTCGCAGTTTTCAC
hNanog	Forward: ACCAGAACTGTGTTCTCTTCCACC Reverse: CCATTGCTATTCTTCGGCCAGTTG
hOct4	Forward: GTGGAGGAAGCTGACAACAA Reverse: ATTCTCCAGGTTGCCTCTCA
hDNMT3B	Forward: ATAAGTCGAAGGTGCGTCGT Reverse: GGCAACATCTGAAGCCATTT
U6	Forward: CGTTCGGCAGCACATATAC Reverse: AAATATGGAACGCTTCACGA
hMYBPC3	Forward: TAGCAAGAAGCCACGGTCAG Reverse: CCAGCAATGACTGCGTAAGA
hMYH7	Forward: TCGTGCCTGATGACAAACAGGAGT Reverse: ATACTCGGTCTCGGCAGTGACTTT
hMYH6	Forward: GATAGAGAGACTCCTGCGGC Reverse: CCGTCTTCCCATTCTCGGTT
hACTN2	Forward: ACTCCCGAGAAGACCATGC Reverse: CAGGACGGTTGCTGATCC
hTNNT2 total	Forward: AGAGCGGAAAAGTGGGAAGA Reverse: GCTGATCTTCATTCAGGTGGT
hTNNT2 fetal	Forward: AGAGGAGGACTGGAGAGAGG Reverse: CTGGGCTTTGGTTTGGACTC
hTTN total	Forward: TTCAGAAGCAACCTTGGGCT Reverse: GTAACGGCTGCGTAAACGTC
hTTN N2BA	Forward: GCCTGGAATGAGCCTCACAT Reverse: ATGTTGCATGACTCCCCAGG
hTTN N2B	Forward: GGTTGACTGCGGCGAGTATA Reverse: ACAACTTCTTCCTTTGGTTCAGG

hTTN N2B	Forward: GGTTGACTGCGGCGAGTATA Reverse: ACAACTTCTTCCTTTGGTTCAGG
hTNNI1	Forward: CAGCTCCACGAGGACTGAAC Reverse: CTCTTCAGCAAGAGTTTGCG
hTNNI3	Forward: CCTCAAGCAGGTGAAGAAGG Reverse: CAGTAGGCAGGAAGGCTCAG