Universidade de Lisboa
Faculdade de Medicina

U LISBOA | UNIVERSIDADE DE LISBOA

Exploring new biological frontiers in Hypertrophic Cardiomyopathy

Rita Mingot de Almeida Mendes de Almeida

Orientadores:
Prof.ª Doutora Maria do Carmo Salazar Velez Roque da Fonseca

Prof.ª Doutora Maria Teresa Tenório Figueiredo Carvalho Gonçalves

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas
E especialidade de Biologia Celular e Molecular

2018

Universidade de Lisboa

Faculdade de Medicina

U LISBOA | UNIVERSIDADE DE LISBOA

Exploring new biological frontiers in Hypertrophic Cardiomyopathy

Rita Mingot de Almeida Mendes de Almeida

Orientadores:
Prof.ª Doutora Maria do Carmo Salazar Velez Roque da Fonseca

Prof.ª Doutora Maria Teresa Tenório Figueiredo Carvalho Gonçalves

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas
E especialidade de Biologia Celular e Molecular

Jurí:
Presidente:
- Doutor José Augusto Gamito Melo Cristino, Professor Catedrático e Presidente do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa

Vogais:
- Doutora Maria Alexandra Marques Moreira Mourão do Carmo, Investigadora Principal, Group Leader do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.
- Doutor José Alexandre de Gusmão Rueff Tavares, Professor Catedrático da Faculdade de Ciências Médicas da Universidade Nova de Lisboa
- Doutor Carolino José Nunes Monteiro, Professor Associado com Agregação da Faculdade de Farmácia da Universidade de Lisboa
- Doutora Maria do Carmo Salazar Velez Roque da Fonseca, Professora Catedrática da Faculdade de Medicina da Universidade de Lisboa (Orientadora)
- Doutor Luís Afonso Brás Simões do Rosário, Professor Auxiliar Convidado da Faculdade de Medicina da Universidade de Lisboa
- Doutora Ana Berta da Fonseca Vieira Álvares e Sousa Ferrand de Almeida, Professora Auxiliar Convidada da Faculdade de Medicina da Universidade de Lisboa

2018

Parte do trabalho referido nesta tese encontra-se publicado na seguinte referência:

Mendes de Almeida, R., et al., *Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy.* Plos One, 2017. **12**(8): p. e0182946.

URL link:
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0182946

AGRADECIMENTOS

"The real voyage of discovery consists, not in seeking new landscapes, but in having new eyes."

Marcel Proust

Durante esta viagem que foi o doutoramento tenho de agradecer a muitas pessoas pelo apoio prestado, pelos conhecimentos e pelas novas perspectivas que me deram.

Em primeiro lugar tenho que agradecer às minhas orientadoras, Prof.ª Doutora Maria Carmo-Fonseca e Prof.ª Doutora Teresa Carvalho pela oportunidade que tive em trabalhar no seu laboratório e por toda a ajuda prestada para conseguir desenvolver este projeto. Não posso deixar de também de agradecer à Prof. Doutora Sandra Martins, a minha co-orientadora "não oficial". Obrigada por me darem uma nova perspectiva do mundo científico e a possibilidade de desenvolver o meu pensamento crítico, entre tudo o mais que aprendi.

A todos os que me ensinaram e apoiaram durante estes últimos 5 anos no laboratório tenho a agradecer o caminho percorrido. Um grande obrigada à Prof.ª Doutora Evguenia Bekman pelo apoio e amizade. Gostaria de agradecer também à Marisa Cabrita, à Ana Carolina Freitas e à Ana Jesus que foram igualmente de uma grande ajuda quando iniciei este projeto; ao Sérgio Marinho, ao José Albuquerque e à Dinora Levy pelo apoio técnico, ao Prof. Doutor Francisco Enguita pelo apoio informático e notável bom humor e, ainda, à Vanessa Pires e à Rita Drago pela amizade.

Quero deixar um obrigado muito especial a duas pessoas que contribuíram muito para a elaboração desta tese, sendo que sem elas parte deste projeto não teria chegado a bom termo: à Catarina Santos, cujo o trabalho de mestrado contribuiu na elaboração desta tese e à Joana Tavares, pela a sua ajuda na realização do *paper*.

TABLE OF CONTENTS

IV

Summary

Hypertrophic Cardiomyopathy (HCM) is the most common hereditary disease of the heart (1:500 individuals), and a cause of sudden cardiac death in young adults and athletes. The disease is inherited as an autosomal dominant trait caused by mutations in genes of the cardiac sarcomere. It is a disease with a widely variable genotype and phenotype. At present there is no effective treatment for this genetic disorder.

The goal of my work was to explore applications of recent molecular genetics tools to improve patient diagnosis and develop potential new treatment strategies.

Inspired by recent reports demonstrating the feasibility of performing "molecular RNA surgery" by using a double trans-splicing approach that results in the specific substitution of a given mutated exon, I investigated whether trans-splicing could efficiently correct the expression of a mutant *TNNT2* gene in cardiac cells. The *TNNT2* gene codes for cardiac troponin T, one of the first sarcomeric proteins to be linked to HCM with more than 30 mutations identified to date. Because there is a significant mutational clustering on *TNNT2* exon 9 associated with poor prognosis, I designed a strategy to specifically correct this exon. As a model system I used murine HL-1 cardiomyocytes. Given architectural differences between the human *TNNT2* and mouse *Tnnt2* genes, human exon 9 corresponds to mouse exon 8. The human *TNNT2* exon 9 was used to replace the homologous mouse exon 8, which encodes the same amino acid sequence but differs in nucleotide composition, thus creating unique restriction sites and also a unique binding site for a primer that only hybridizes to the human *TNNT2* exon 9. These unique restriction sites or the specific primer were further used to check the efficiency of trans-splicing events. Briefly, double trans-splicing molecules were constructed containing the replacing exon flanked by artificial intronic sequences with strong splice sites and splicing enhancers connected by a spacer linker to antisense sequences designed to anneal the two introns flanking exon 8 in the target murine *Tnnt2* pre-mRNA. Cells were transfected with the exon exchange constructions cloned under the control of different promoters. The efficiency of trans-splicing was determined by RT-PCR followed by restriction analysis or, alternatively, by RT-PCR using the primer that only hybridizes to the human *TNNT2* exon 9 and thus only amplifying trans-spliced transcripts. An RT-PCR assay using a radioactive γ-^{32}P labelled

primer indicated the presence of only residual amounts of the trans-splicing product. In order to improve efficiency, I designed an alternative strategy that involves a single 3' trans-splicing reaction. In brief, the goal was to replace mouse exon 8 and all exons downstream (exons 8 to 15) with a human cDNA containing the nucleotide sequence corresponding to exons 9 to 16. I constructed a 3' trans-splicing vector containing the human cDNA and upstream an artificial intronic sequence with a strong splice site and splicing enhancers connected by a spacer linker to an antisense sequence designed to anneal to intron 7 upstream of the mouse *Tnnt2* exon 8 in the target pre-mRNA. After transfection, no trans-splicing product was detected. All together, these results argue that trans-splicing does not ensure efficient correction of expression of a *TNNT2* gene in cardiac cells. This could be due to inefficiency of trans-splicing reactions in general, or a particular resistance to trans-splicing of the targeted region in the *Tnnt2* pre-mRNA.

High throughput sequencing technologies have revolutionized the identification of mutations responsible for HCM. Detection of pathogenic mutations has important implications for the medical management of patients and their families. However, approximately 50% of individuals with a clinical diagnosis of HCM have no causal mutation identified. In my host lab, we hypothesized that this may be due to the presence of pathogenic mutations located deep within the introns, which are not detected by conventional sequencing analysis restricted to exons and exon-intron boundaries. The aim of my study was to develop a whole-gene sequencing strategy to prioritize deep intronic variants that may play a role in HCM pathogenesis. In collaboration with other members of the host lab, the full genomic DNA sequence of 26 genes previously associated with HCM was analysed in 16 unrelated patients. We identified likely pathogenic deep intronic variants in *VCL, PRKAG2* and *TTN* genes. These variants, which are predicted to act through disruption of either splicing or transcription factor binding sites, were 3-fold more frequent in our cohort of probands than in normal European populations. Moreover, we found a patient that is compound heterozygous for a splice site mutation in *MYBPC3* and the deep intronic *VCL* variant. Analysis of family members revealed that carriers of the *MYBPC3* mutation alone do not manifest the disease, while family members that are compound heterozygous are

clinically affected. In conclusion, we developed a framework for scrutinizing variation along the complete sequence of HCM-associated genes and our results suggest that deep intronic variation contributes to HCM phenotype.

In order to translate the novel genetic information that we found to clinical decision taking requires further functional analysis. To date, mechanistic and functional studies of HCM mutations have been largely restricted to animal models in part due to difficulties in obtaining human tissue from patients. However, the recent emergence of patient-derived induced pluripotent stem cells (iPSCs) that can be differentiated into functional cardiomyocytes that recapitulate HCM-specific characteristics holds great promise as an exciting new approach to study how gene mutations relate to clinical outcomes and might be applied to test our hypothesis-generating data. Thus, I decided to use the CRISPR-Cas9 genome-editing technology to introduce patient mutations in the genome of embryonic stem (ES) cells that were subsequently differentiated in cardiomyocytes. In collaboration with other members of the host lab, I generated sets of isogenic ES cells that differ exclusively by the presence of HCM-causing mutations in the *TNNT2* gene. We used mouse ES cells, which are easier to manipulate and differentiate than human cells.

In conclusion, during my PhD training I explored the feasibility of inducing trans-splicing as an RNA-targeted therapy to correct the expression of mutant sarcomeric genes in cardiomyocytes, I contributed to the development of a bioinformatics pipeline to identify novel mutations located within intronic regions of sarcomeric genes that may contribute to HCM pathogenesis, and I constructed new genome-edited cellular models of HCM.

Keywords: hypertrophic cardiomyopathy, splicing, intronic mutations, whole-gene sequencing, genome editing

Resumo

A Cardiomiopatia Hipertrófica (CMH) é a doença cardíaca hereditária mais comum (1: 500 indivíduos) e uma das principais causas de morte súbita em adultos jovens e atletas de alta competição. A doença é herdada de forma autossómica dominante e causada por mutações em genes que codificam proteínas sarcoméricas. É uma doença com um genótipo e um fenótipo bastante variáveis. Atualmente, só existe tratamento sintomático para esta doença genética e uma terapia génica convencional não é aplicável.

O objetivo do meu trabalho foi explorar a potencial aplicação de ferramentas de genética molecular "de ponta" visando, por um lado, melhorar o diagnóstico de doentes de CMH e, por outro, o desenvolvimento de estratégias inovadoras para o tratamento destes doentes.

Tendo por base estudos prévios que demonstraram a possibilidade da realização de "cirurgia molecular ao nível do RNA ", isto é, a substituição específica de um exão mutado por '*double trans-splicing*', neste estudo foram realizadas experiências de *trans-splicing* com o objectivo de corrigir a expressão do gene *TNNT2* quando mutado em células cardíacas. O gene *TNNT2* codifica a troponina T cardíaca, uma das primeiras proteínas sarcoméricas a ser associada à CMH com mais de 30 mutações identificadas até a data. Uma vez que existe um elevado número de mutações no exão 9 do gene *TNNT2* associadas a um prognóstico pouco favorável, neste trabalho foi desenvolvida uma estratégia especifica para a correção específica deste exão. Para tal, uma linha de cardiomiócitos murinos, HL-1, foi utilizada como modelo celular. Devida as diferenças estruturais entre os genes *TNNT2* humano e *Tnnt2* murino, o exão 9 humano corresponde ao exão 8 murino; assim, nas células HL-1, o exão 8 do gene *Tnnt2* foi subtituído pelo exão 9, seu homólogo humano, o qual codifica a mesma sequência de aminoácidos, mas difere na sua composição nucleótidica, criando assim locais de restrição únicos e um local de ligação também único para um *primer* que apenas hibrida com este último. Estes locais de restrição únicos, assim como o *primer* específico do exão 9 humano, foram usados para distinguir, por RT-PCR e ensaios de restrição, o produto endógeno de *cis-splicing* (murino) do produto de *trans-splicing* contendo o exão 9 humano. As construções de *double trans-splicing*

foram concebidas para conter o exão 9 humano, flanqueado por sequências intrónicas artificiais contendo *splice sites* e splicing *enhancers* ligados a sequências *antisense,* as quais foram desenhadas para emparelhar com os dois intrões que flanqueiam o exão 8 no pre-mRNA endógeno do *Tnnt2* murino. Sob o controlo de diferentes promotores, as várias construções foram usadas para transfectar células HL-1. A eficiência de *trans-splicing* foi determinada por RT-PCR, seguida de análise de restrição dos produtos de PCR, ou alternativamente, apenas por RT-PCR usando um primer que híbrida especificamente com o exão 9 humano e logo amplifica apenas transcritos que sofreram eventos de *trans-splicing*. Um ensaio de RT-PCR usando este primer especifico marcado radioactivamente com γ-^{32}P permitiu aumentar a sensibilidade de detecção destes transcritos, revelando a presença de quantidades residuais de eventos de trans-splicing. Por forma a melhorar a eficiência do *trans-splicing*, uma estratégia alternativa envolvendo uma única reacção de *3' trans-splicing* foi posteriormente desenhada. Resumidamente, um cDNA do Tnnt2 humano contendo os exões 9 a 16 foi utilizado para substituir os exões 8 a 16 murinos. As construções de *3' trans-splicing* foram desenhadas para conter o cDNA humano de substituição dos exões 9 a 16 e, a montante, uma sequência intrónica artificial, contendo em *splice site* e splicing *enhancers* ligados por sequências *antisense,* as quais foram desenhadas para emparelhar com o intrão 7 a montante do exão 8 nos transcritos endógenos murinos de *Tnnt2*. Após transfecção, nenhum produto de 3´*trans-splicing* foi detectado. Em conjunto, estes resultados revelam que o *trans-splicing* não garante a potencial correção da expressão do gene mutado *Tnnt2* em células cardíacas. Este facto poderá ser devido à ineficiência das reacções de *trans-splicing* no geral ou, especificamente, à resistência ao *trans-splicing* da região alvo escolhida no pre-mRNA *Tnnt2*.

As tecnologias de sequenciação de última geração vieram revolucionar a identificação de mutações associadas à CMH. Embora a detecção de mutações patogénicas tenha implicações importantes no tratamento e diagnóstico de doentes com CMH, estudos prévios mostram que em 50% dos doentes não foram detectadas mutações causadoras da doença. No meu laboratório de acolhimento, foi proposta a hipótese que este factor se devesse à presença de mutações patogénicas em regiões intrónicas profundas em genes associados à CMH, as quais não são detectadas por

análises de sequenciação convencionais que se cingem aos exões e as fronteiras exão-intrão. Um outro objectivo do meu trabalho de doutoramento passou por desenvolver uma estratégia de sequenciação total de um painel de genes associados à CMH, priorizando mutações identificadas em regiões intrónicas profundas e que poderão desempenhar um papel na patogénese da CMH. Em colaboração com outros membros do laboratório, a sequência genómica completa de 26 genes previamente associados à CMH foi analisada em 16 doentes não relacionados. Da análise realizada, várias mutações intrónicas nos genes *VCL, PRKAG2* e *TTN* foram identificadas, prevendo-se que possam interferir com eventos de *splicing* ou com locais de ligação de fatores de transcrição, uma vez que são 3x mais frequentes no nosso *cohort* de doentes do que na população europeia normal. Um dos pacientes deste *cohort* revelou-se heterozigótico para uma mutação num *splice site* do gene *MYBPC3* e para uma mutação intrónica no gene *VCL*. A análise de familiares deste paciente indicou que os que eram portadores da mutação no gene *MYBPC3* não manifestavam a doença; no entanto, os que possuíam as duas mutações, em heterozigotia, eram afectados pela patologia. Em conclusão, ao longo desta análise, foi desenvolvida uma *pipeline* que permitiu identificar variantes genéticas ao longo de toda a sequência de um painel de genes associados à CMH, sendo que os resultados sugerem que variantes em regiões intrónicas profundas nestes genes contribuem para o fenótipo da CMH.

A potencial aplicação clínica das variantes genéticas previamente identificadas através da *pipeline* por nós desenvolvida requer a análise funcional das mesmas. Durante as últimas décadas, os estudos mecanísticos e funcionais de mutações causadoras de CMH têm estado confinados a modelos animais em parte devido à dificuldade de obter tecido humano de doentes. No entanto, mais recentemente, a utilização de *induced pluripotent stem cells* (iPSC) derivadas de doentes, as quais podem ser diferenciadas em cardiomiócitos funcionais que recapitulam características da CMH, tem emergido como uma inovadora e promissora abordagem para o estudo de que como mutações genéticas se correlacionam com os *outcomes* clínicos, podendo ser usada para testar as variantes prioritizadas pela nossa *pipeline* bioinformática. Assim, numa terceira fase do meu projecto de doutoramento, utilizou-se a tecnologia de edição de genoma CRISPR-Cas9 para introduzir em células estaminais embrionárias murinas (mESCs) mutações já descritas em doentes como causadoras de CMH, que

foram subsequentemente diferenciadas em cardiomiócitos. Assim, criaram-se *set*s de linhas de mESCs isogénicas que diferem apenas pela presença de mutações causadoras de CMH no gene *Tnnt2*. Optou-se pelo uso de mESCs para gerar modelos celulares de CMH porque a manipulação destas células é mais fácil do que a manipulação de células humanas.

Em conclusão, durante o meu trabalho de doutoramento, foi explorada a hipótese de se utilizar a técnica de *trans-splicing* como uma nova abordagem terapêutica, tendo como alvo moléculas de RNA, por forma a corrigir a expressão de genes sarcoméricos mutados em cardiomiócitos; também contribuí no desenvolvimento de uma *pipeline* bioinformática para a identificação de novas mutações localizadas em regiões profundas intrónicas de genes sarcoméricos que podem contribuir para a patogénese da CMH, e construí novos modelos celulares de CMH geneticamente editados.

Palavras-chave: cardiomiopatia hipertrófica, *splicing*, mutações intrónicas, sequenciação total de genes, edição de genoma

Index of Tables

Index of Figures

Abbreviations and Units

% - Percent

°C - Degree Celsius

$\gamma^{32}P$ ATP - ATP labelled on the gamma phosphate group with ^{32}P

μg - Microgram

μl - microliter

μM - Micromolar

ACMG - American College of Medical Genetics and Genomics

ANF - atrial natriuretic factor

AS - annealing sequence

ASO - antisense oligonucleotides

ATP - adenosine triphosphate

BBP - branchpoint binding protein

BMP - bone morphogenetic protein

BNP – brain natriuretic peptide

BP - branchpoint

bp - base pairs

Ca^{2+} - calcium cation

CADD - Combined Annotation-Dependent Depletion

cDNA - complementary DNA

Ci/mmol - curie/milimol

CMV - cytomegalovirus

CRISPR - Clustered regularly interspaced short palindromic repeats

crRNA - CRISPR RNA

DKO - double knockout

DNA - deoxyribonucleic acid

dPSI - percent of spliced in

DTS - double trans-splicing

DSB - double strand break

EBs - embryoid bodies

ECG - electrocardiogram

EDTA - ethylenediamine tetra-acetic acid

EF - Enrichment factor

END-2 - visceral endoderm-like cells

ES - embryonic stem

ESE - exon splicing enhancer

ESS - exon splicing silencer

FBS - fetal Bovine Serum

FISH - Fluorescence *in situ* hybridization

GFP - green fluorescence protein

GSK-3 - Glycogen synthase kinase 3

GWAVA - Genome-Wide Annotation of Variants

GWAS – genome wide association studies

h - hour

HA - homology arms

HCl - hydrochloric acid

HCM - hypertrophic cardiomyopathy

HDV - homology donor vector/plasmid

hPSCs - human pluripotent stem cells

HR - homologous recombination

IF - immunofluorescence

Indel - Insertion/deletion

iPSCs - induced pluripotent stem-cell

iPSCs-CM - induced pluripotent stem-cell - derived cardiomyocytes

ISE - intron splicing enhancer

ISS - intron splicing silencer

kb - kilobase

LIF - Leukemia Inhibitory Factor

LNA - locked nucleic acid

M - molar

MAF - minor allele frequency

min - minute

mm - millimeter

mM - millimolar

mRNA - messenger RNA

ms - milliseconds

MHC - myosin heavy chain

MSE - muscle specific enhancer

NaCl - sodium chloride

ng - nanogram

NGS - next-generation sequencing

NHEJ - non-homologous end joining

PAM - protospacer adjacent motif

PBS - phosphate buffer saline

PCR - polymerase chain reaction

PFA - paraformaldehyde

PI3K - phosphoinositide 3-kinase

PolyPhen2 - polymorphism phenotyping tool.

PTC - premature termination codon

Py - polypyrimidine

qRT-PCR - quantative retro-transcriptase PCR

RNA - ribonucleic acid

RNAi - RNA interference

RS - region score

RT-PCR - retro-transcriptase PCR

SDS - sodium dodecyl sulfate

sec - second

sgRNA - single guide RNA

shRNA - short hairpin RNA

SIFT - Sorting Tolerant from Tolerant tool

snRNA - small nuclear RNA

snRNP - small nuclear ribonucleoprotein

SNV - single nucleotide variant

SPIDEX - Splicing Index

ss - splice site

SSC - saline-sodium citrate buffer

TALENs - transcription activator-like effector nucleases

TE - Tris-EDTA

TFBS - transcription factor binding sites

TGFβ - Transforming growth factor beta

tracrRNA - trans-acting crRNA

TS - trans-splicing

TSM - trans-splicing molecule

U - unit

UTP - uridine triphosphate

UTR - untranslated region

V - Volt

VRC - ribonucleoside vanadyl complex

WT - wild type

ZFNs - zinc finger nucleases

CHAPTER 1

BACKGROUND AND STATE OF THE ART

1. Background and state of the art

1.1. Hypertrophic Cardiomyopathy - the disease

1.1.1. Pathophysiology, disease manifestations and symptoms

The European Society of Cardiology defined Hypertrophic Cardiomyopathy (HCM) *"by the presence of increased left ventricular (LV) wall thickness that is not solely explained by abnormal loading conditions"* [1]. It is an autosomal dominant disease, which is phenotypically and genetically heterogeneous [2, 3]. It affects 1:500 individuals, but a new study suggests that the prevalence is actually higher (1:200), due to new techniques in genetic screening, diagnostic methods and more family members being screened for the disease [2, 3].

The disease manifestations range from asymmetrical thickening of the left ventricle leading to myocardial disarray as a compensatory mechanism (**Figure 1.1**) to coronary dysfunction expressed as increased wall thickness/lumen ratio and eventually leading to ischemia, remodelling and fibrosis [3, 4].
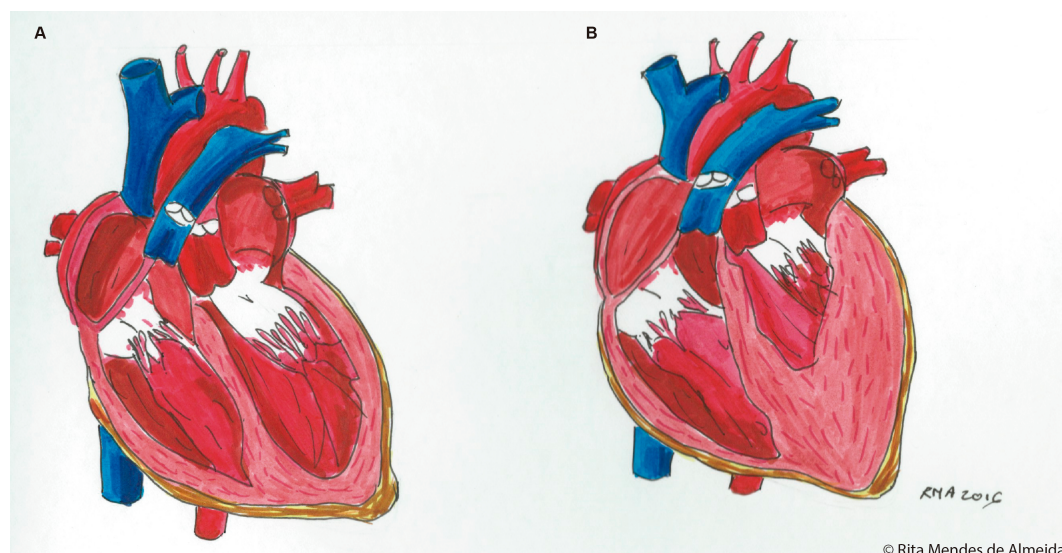
© Rita Mendes de Almeida

Figure 1.1 - Comparison of a healthy heart with a heart with HCM
On the right is the healthy heart (A) and on the left the hypertrophic one (B), the difference in size of the myocardium is clear. (© Rita Mendes de Almeida)

These changes in the cardiac muscle can be detected very early in life or may remain asymptomatic and only be diagnosed *post-mortem* after a sudden death by cardiac arrest. In short, although HCM may be asymptomatic, the main characteristics include angina, shortness of breath, dizziness and stroke and affected individuals may present symptoms ranging from mild to severe heart failure and sudden cardiac death [3-5].

HCM is often referred as a "disease of the sarcomere" [3, 4]. The sarcomere is the functional unit of the cardiomyocytes (**Figure 1.2**). The cardiomyocytes contain thick and thin filaments, and these units repeat themselves and form the myocardial myofibrils [5]. The thin filament includes the following proteins: actin (ACTC1), α-tropomyosin (TPM1), troponin C, T and I (TNNC1, TNNT2, TNNI3). The thick filament is composed of the myosin proteins: myosin binding protein C (MYBPC3), α- and β-myosin heavy chain (MYH6, MYH7) [5]. Around 50% of HCM cases do not have a genetic variation attributed to it [3, 4]. The HCM phenotype is modulated either by a genetic variant or a modifier [5, 6]. To date, more than 1400 mutations have been associated to HCM in at least 13 different genes [2, 4, 5]. HCM mutations can interfere in several processes, such as: cellular calcium signalling, actin ATPase activation and changes in the myosin-actin interactions that interfere in force generation, among others that still remain unclear. Altogether, these mutations lead to different HCM phenotypes [6].
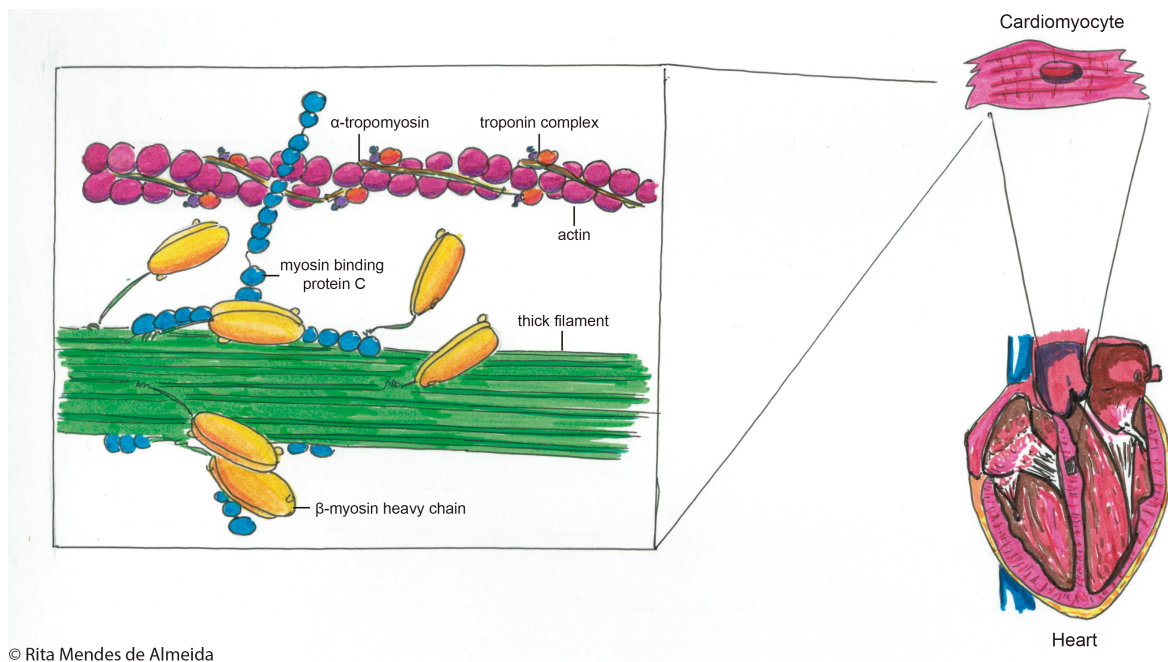
© Rita Mendes de Almeida

Figure 1.2 – The sarcomere: the contractile unit of cardiomyocytes

In the heart's myocardium, cardiomyocytes are responsible for the contractions in the heart. The sarcomere is the contractile unit in the cardiomyocytes. HCM mutations occur usually in proteins of the sarcomere: α-tropomyosin, β-myosin heavy chain, myosin binding protein C, titin, actin and troponin complex (C, T, I). (© Rita Mendes de Almeida)

HCM-associated mutations present in the thick filament are the ones in the myosin genes (*MYBPC3, MYH7, MYH6*) and they account for 75-80% of the sarcomere HCM mutations and 30-40% of all HCM cases [4, 5]. Most of myosin mutations are missense mutations with some exceptions for the *MYBPC3* gene [7]. Mutations found in the *MYBPC3* gene might lead to truncated protein that will be degraded, resulting in insufficient protein for normal functional sarcomeres [5, 8]. For the *MYH7* gene, mutations are aggregated between exons 3 to 23, which code for functionally relevant protein domains of α-myosin heavy chain, mainly the myosin 'head' and 'neck' responsible for the process of contraction by connecting to actin [5]. On the other hand, mutations in the *MYH6* gene showed an increase in actin ATPase dependent activity and disruption of the interaction between actin and myosin, which may lead to alterations of cardiomyocytes and trigger hypertrophy [5]. The HCM mutations in thin filament proteins (actin, α-tropomyosin, troponin C, T and I) account for only 5 to 15% of all HCM cases. These might alter the Ca^{2+} sensitivity of those proteins (essential for contraction) and of the force generation [5]. Despite the lower percentage of HCM-

associated mutations in these genes, they are associated to a more severe phenotype [9].

Even though the major genes associated to HCM are sarcomeric genes, the panel of genes associated to this disease has been growing larger. These new genes code mainly for associated proteins of the Z-disc, of the sarcoplasmic reticulum and of the plasma membrane [6]. Several mutations were identified and associated to a small number of HCM cases, in the following genes (**Table 1.1**): cardiac ankyrin repeat domain 1 (*ANKRD1*), α-actinin 2 (*ACTN2*), myozenin-2 (*MYOZ2*), LIM domain binding 3 (*LDB3*), vinculin (*VCL*), titin (*TTN*), telethonin (*TCAP*), cysteine and glycine rich protein 3, (*CSRP3*), nexilin (*NEXN*), junctophilin-2 (*JPH2*), other proteins that regulate Ca^{2+} (*PLN, CASQ2, CALR3*), among others.

Table 1.1 – List of HCM-associated genes and evidence of their pathogenecity

Genes that have been associated to HCM, with respective evidence and frequency. Adapted from [5].

Disease genes	Evidence for pathogenicity	Locus	Protein	Frequency
Thick filament protein				
MYH7	strong	14q11.2	β-myosin heavy chain	15–25%
MYBPC3	strong	11p11.2	myosin binding protein C	15–25%
MYL2	strong	12q24.1	regulatory myosin light chain	rare
MYL3	strong	3p21.3	essential myosin light chain	rare
MYH6	weak	14q11.2	α-myosin heavy chain	rare
Thin filament protein				
TNNT2	strong	1q32.1	troponin T	≈5%
TNNI3	strong	11p15.5	troponin I	≈5%
TNNC1	strong	3p21.1	troponin C	rare
TPM1	strong	15q22.2	α-tropomyosin	rare
ACTC1	strong	15q14	α-cardiac actin 1	rare
Z-disc protein				
CSRP3	weak	11p15.1	muscle LIM protein	rare
TCAP	weak	17q12	telethonin	rare

VCL	weak	10q22.2	vinculin	rare
LDB3	weak	10q23.2	LIM domain binding 3	rare
ACTN2	strong	1q43	α-actinin 2	rare
MYOZ2	strong	4q26	myozenin 2	rare
ANKRD1	weak	10q23.3	cardiac ankyrin repeat protein	rare
NEXN	weak	1p31.1	nexilin	rare
Ca2+-handling protein				
PLN	weak	6q22.3	phospholamban	rare
CASQ2	weak	1p13.1	calsequestrin 2	rare
CALR3	weak	19p13.1	calreticulin 3	rare
JPH2	weak	20q13.12	junctophilin 2	rare
Other related genes				
TTN	weak	2q31.2	titin	rare
CAV3	weak	3p25.3	caveolin 3	rare

When looking at genotype-phenotype relationships, they remain unclear for most of the mutations that were associated to HCM [6]. In fact, the pathogenesis in HCM can come from a dominant negative effect from an abnormal protein, a problem in the metabolism of the myocardium or haploinsufficiency (insufficient wild-type protein for a normal functional cell) [5]. When it comes to the dominant negative effect of a protein, it means that the protein itself is abnormal and will cause harm to the cell, which can be explained by a possible gain of function of the mutant protein. Since most of the HCM mutations are missense, they will give rise to mutant, abnormal proteins that will be incorporated as if they were wild-type, giving rise to several problems at the cellular level. These can affect the mechanical and the Ca^{2+} signalling in the sarcomeres, leading to potential arrhythmias and cardiac sudden death [5]. Haploinsufficiency maybe the main reason for the pathology of mutations in the *MYBPC3* gene that lead to premature stop codons, meaning that the wild-type allele is not enough to maintain a healthy phenotype (loss of function protein, truncated protein) [5, 6, 8, 10]. Finally, an altered energetic metabolism of the myocardium such as an increased sensitivity and affinity to Ca^2 or increased activity of ATPase activated actin, can indicate a gain of function of the proteins involved and that they are spending more energy than wild-type proteins [5].

1.1.2. Diagnosis and genetic screening

The diagnosis of such phenotypically heterogeneous disease such as HCM involves different techniques. In the 70's, the most common exam for HCM diagnosis was echocardiography. But since the 80's, cardiac magnetic resonance came to be a more enriched tool for diagnosis, which is very useful for characterization of the heart's appearance and function and also to evaluate specific regions of hypertrophy [2, 3]. The imaging techniques are indispensable to make a definitive HCM clinical diagnosis, together with the patient's history, physical exam, symptoms and electrocardiogram, being the ultimate decisive factor the presence of a thickness of the left ventricle wall higher than 15 mm [11, 12].

In both clinically HCM diagnosed patients or first-degree relatives of HCM patients, there is the recommendation for genetic screening, which is useful for patient management [2]. However, and despite the recent evolution of DNA sequencing techniques facilitates the screening of HCM mutations, to date only 50% of patients have potential HCM mutations identified. Moreover, even in these patients where an HCM mutation was detected, the mutation identified might not have a well-known genotype/phenotype correlation. This makes genetic sequencing difficult but still useful, especially for management of relatives of HCM patients that maybe at risk of developing the disease, if they have an HCM associated variant [5]. A genetic screening can also be helpful in distinguishing between an HCM case or other cardiomyopathies, helping physicians to make a proper diagnosis [5]. The impact of NGS and genetic screening in HCM patients is further discussed in section **1.3**.

1.1.3. Impact of HCM mutations at cellular level

There is a complex challenge to develop functional assays to study the effects of a given variant, especially at the cellular level, for a better understanding of the pathogenesis and heterogeneous HCM phenotypes. At the cardiomyocyte level affected by HCM, several issues take place: hypertrophy, fibrosis, altered gene expression and altered signalling pathways (mainly calcium regulation) [13]. In HCM the characteristic hypertrophy may be attributed to a compensatory mechanism, which

independent of the initial cause, will lead to the increased size of the muscle and of the cardiomyocytes [13]. The altered expression of several genes may also help in HCM diagnose and in the understanding of the molecular mechanisms undelaying this disease. For example, the expression levels of atrial natriuretic peptide (*ANF*) and brain natriuretic peptide (*BNP*) are altered in cases of cardiomyopathies; specifically, BNP is secreted in cases of ventricular hypertrophy [13, 14]. In early signs of cardiomyopathy there is also a change in expression of the *MHC* gene, which is tightly regulated, with *β-MHC* levels being increased and *α-MHC* decreased [13, 15]. Two mutations in the *MYH6* gene (R403Q, R719W) lead to the increase of *TGF-β* expression, which contributes to fibrosis [16]. The Akt/phosphoinositide 3-kinase (PI3K) is a signalling pathway involved in cardiac remodelling. If it remains active it may lead to cardiomyopathy. Akt activation may increase cardiomyocyte size, and inactivates glycogen synthase kinase 3 (GSK-3β), which in normal circumstances helps reduce cellular size [13, 17]. Another sign of hypertrophy is the overexpression of calcineurin (which happens with increased Ca^{2+} in the cytoplasm) or nuclear factor of activated T cells (NFAT, which migrates to the nucleus) [13, 18]. These various HCM pathogenic mechanisms were elucidated thanks to animal models and other available models, but at the cell level there is the need for reliable cellular HCM models.

Since obtaining patient's cardiomyocytes is very cumbersome and cardiomyocytes are fully differentiated cells that do not proliferate, studies at cellular level would be complicated [13, 19]. Before the advent of induced pluripotent stem cells (iPSCs), the major disease models used were animal models. Animal models helped understand many of the disease mechanisms, but they are more expensive and time consuming than cellular ones [19]. Some of the HCM animal models available are represented in **Table 1.2**.

Table 1.2 - Examples of HCM mouse models

Example of mouse models from the literature and respective cardiac phenotype that each mutation displays.

Organism	Gene	Protein	Mutation	Cardiac phenotype	References
Mouse	*ACTC1*	Cardiac α-actin	E99K	Apical HCM and increased Ca2+ sensitivity	[20]

Mouse	CSRP3	Muscle LIM protein (MLP)	W4R	Late-onset HCM. Blunted response to adrenaline stimulation.	[21]
Mouse	MYH6	α-Myosin heavy chain	R403Q	Left ventricular hypertrophy, myocyte dissaray, myocardial fibrosis	[6, 22]
			R453C	Idem	[6]
			R719W	Idem	[6]
Mouse	MYOZ2	Myozenin-2/calsarcin-1	S48P, I246M	Cardiac hypertrophy with interstitial fibrosis and Z-line abnormalities. Preserved cardiac function.	[23]
Mouse	MYPN	Myopalladin	Y20C	CM with dilatation of both chambers and disrupted intercalated discs.	[24]
Mouse	TNNI3	Troponin I	R145G	Cardiomyocyte disarray, fibrosis and premature death	[25]
Mouse	TNNT2	Troponin T	R92Q	Hypercontractility, diastolic dysfunction and fibrosis	[26]
			I79N	Enhancement of base-line contractility leading to cardiac dysfunction.	[27]
Mouse	TPM1	Tropomyosin	E180G	Severe cardiac hypertrophy, fibrosis	[28, 29]
			N175D	Mild myocyte disorganization and hypertrophy	[30]

But, nowadays, it is already possible to have patient's cells to perform characterization and functional studies without having the need for samples from a patient's heart. This is possible thanks to the ability of reprograming somatic cells into pluripotent stem cells, the so-called induced pluripotent stem cells or iPSCs. This

process was first described by Yamanaka and colleagues, by forcing the expression of certain factors OCT-4, SOX-2, Klf-4 and c-myc ('Yamanaka factors') in differentiated cells [31]. The obtained iPSCs cells can later be differentiated in any cellular type, including cardiomyocytes, taking advantage of the modulation of several signalling pathways such as: Activin/Nodal, transforming growth factor β (TGFβ), glycogen synthase kinase 3 (GSK3), Wnt, bone morphogenic protein (BMP), among others [19, 32]. Using patient's somatic cells and then developing patient specific iPSC - cardiomyocytes (iPSC-CM) allows to generate unique patient specific HCM cellular models [19].

The process used to differentiate iPCSs in cardiomyocytes can be done through three different types of culture: (1) embryoid bodies (EBs), which are round cell agglomerates that recapitulate the three layers of early embryos, (2) co-culture with visceral endoderm-like (END-2) cells and (3) 2D monolayer culture [19, 32, 33].

Regarding the first methodology, in order to obtain a relative pure population of cardiomyocytes, EBs are grown in a monolayer culture in gelatin-coated dishes [33, 34]. At day 8-9 of differentiation, embryoid bodies exhibit contracting areas where cardiomyocytes can be isolated by a complex procedure. But only 1% of the cells in the EBs are actually cardiomyocytes [35, 36].

The co-culture of iPSCs with the endoderm-like cells drives the cardiac differentiation of the iPSCs due to the released factors by the END-2 cells into the medium, with a yield of around 10% of immature cardiomyocytes [32, 33].

The 2D monolayer culture involves the differentiation of the iPSC cells using a set of factors and a specific culture media to induce the differentiation into cardiomyocytes. The cardiomyocytes obtained by this method have been shown to have a more mature phenotype with yields of 85-95%. There are various protocols that involve various and different factors [32]. A recent study proposes three approaches to obtain a high yield of cardiomyocytes (80-98%), using human pluripotent stem cells (hPSCs) in a monolayer based differentiation approach [37]. All three protocols start by pre-treating iPSCs with glycogen synthase kinase 3 (GSK3) inhibitor [37]. It was demonstrated that the use of the GSK3 inhibitor improved cardiac differentiation before EBs are formed, since GSK3 inhibition helps determine cell fate and proliferation in cardiac development [37, 38]. So, after the pre-treatment, the first approach is

based on the use of growth factors of the TGF-β family, specifically activin A and bone morphogenetic protein 4 (BMP4). The second consists in the use of a Wingless (Wnt) inhibitor. The last one, involves the use of inducible expression of a β-catenin shRNA, but requires prior genetic manipulation of the hPSCs [37]. All three protocols have a very successful yield of cardiomyocytes and can be further used to differentiate patient derived IPSCs into cardiomyocytes.

Nevertheless, some disadvantages were found when using these models. Indeed, several factors might contribute to less homogenous iPSC cell populations such as culture conditions, epigenetic markers of the somatic cells used, age of the cells, genetic components, reprograming efficiency might vary from types of somatic cells, etc [19].

Even though, iPSC-CMs have proven themselves as reliable HCM cell models. In **Table 1.3** are represented a few of the HCM iPSC-CM cell model already generated.

Table 1.3 – Examples of iPSC-CMs HCM cell models

Example of iPSC-CMs models from the literature and respective cardiac phenotype that each mutation displays. Partially adapted from [19].

Gene	Protein	Mutation	Cardiac phenotype	References
MYH7	α-Myosin heavy chain	R442G	Enlarged size, disorganized sarcomere structures and arrhythmic beatings. Irregular calcium handling and ion channel functions. Increased expression of HCM related genes.	[39]
MYH7	α-Myosin heavy chain	R663H	Cellular enlargement and contractile arrhythmia at the single-cell level and dysregulation of Ca^{2+} cycling.	[35]
MYBPC3	Myosin binding protein C	G999-Q1004del	Enlarged cells with myofibrillar disarray.	[40]
MYBPC3	Myosin	Q1061X	Larger cardiomyocytes compared	[41]

	binding protein C		to TPM1 mutation (D175N). Arrhythmogenic events, cell elargment and altered gene profile.	
TPM1	Tropomyosin	D175N	More abnormal Ca^{2+} transients compared to MYBPC3 mutation (Q1061X). Arrhythmogenic events, cell elargment and altered gene profile.	[41]

Overall, in all these iPSC-CMs models, there seems to be certain elements in common such as: cardiomyocyte enlargement, altered calcium handling and altered gene expression of the following genes α-myosin heavy chain (*Myh6*), β-myosin heavy chain (*Myh7*), α-actin (*Actc1*), myocyte enhancer factor 2C (*Mef2c*), cardiac troponin T (*Tnnt2*), *GATA4*, atrial natriuretic factor (*Anf*), Connexion 43 (*Gja1*), calcineurin (*Pp3ca*). Indeed, these iPSC-CM HCM cellular models had been very useful as a first screening tool for pathogenicity of a given mutation at the cell-level. So, based on this, establishing of cardiomyocytes cell lines with different HCM mutations appears as a good experimental approach to help in understanding such phenotypically diverse disease. Moreover, the generation of such custom isogenic cell lines, using the existing gene editing tools to induce any HCM mutation and only having a single mutation per cell line (which might not be the case in patient derived iPSCs) will be crucial to understand the individual contribution of each of the identified variants.

1.1.4. Disease management and gene therapy

Disease management for HCM addresses signs and symptoms in order to manage or prevent them, such as: heart failure, ventricular arrhythmias, left ventricular dysfunction, atrial fibrillation and cardioembolisms. Either a pharmacological or a surgical approach can be performed [42]. The approach chosen under the latest guidelines will depend on the symptoms and the degree of severity.

Conventional medical intervention used in HCM solely involves symptomatic treatment either pharmacological or non-pharmacological and those do not focus in the underlying cause of the disease [43]. So in this work, we focused in the much-needed attempt to develop a novel HCM therapy based in a gene therapy approach, as HCM is a genetic disease.

Genetic disorders can be classified as either dominant or recessive [44]. A mutation, associated to a dominant disease, is present in a single allele and such is enough to cause a disorder; on the other hand, a recessive mutation can only lead to a disorder if both alleles are mutated. Disease associated mutations can be classified in two major groups: in-frame mutations (that include missense mutations, in-frame insertions and deletions) or truncating mutations (that include nonsense mutations, frameshift deletions and insertions). An in-frame mutation can eventually produce a protein but with a given defect, whereas a truncating mutation frequently lead to a loss-of-function of the allele, with no protein being produced [44]. Furthermore, these mutations can have various molecular mechanisms that lead to a certain disease. As already mentioned, a recessive mutation usually leads to a loss-of-function; either an in-frame or truncating mutation can be the cause [44]. A dominant mutation can have one of these mechanisms: haploinsufficiency, gain-of-function or a dominant negative effect [44-46]. Haploinsufficiency means that the wild-type allele does not produce enough protein for a normal cell function. An example of this, are missense mutations in the *PDK1* gene leading to polycystic kidney disease; or mutations in the MYBPC3 gene leading to HCM, due to a truncated protein [8, 45]. An example of gain-of-function mutations are mutations in the *RHO* gene that cause autosomal dominant retinitis pigmentosa; and several missense mutations in the *MYH6* and *MYH7* genes were predicted to produce a gain of function of the proteins and lead to an HCM phenotype [10, 45, 47]. A dominant negative effect involves the coexistence of wild-type and altered protein in the same multimeric protein complexes, disrupting the normal function of the complex [44, 45]. An example of the dominant negative effect are mutations in *MYBCP3* gene leading to HCM [45].

 For autosomal dominant monogenic diseases such as HCM, an increase of the ratio of healthy vs. deleterious protein is not always a viable option [43, 48, 49]. HCM is a disease with low penetrance and variable clinical outcomes, ranging from

haploinsufficiency, gain of function or even malfunctioning protein [6, 8, 10]. A normal gene therapy targeting the DNA could be insufficient but a strategy targeting the RNA will be much better as it would also resolve gain of function or even malfunctioning protein situations. [43, 45]

The concept of gene therapy was first introduced in 1972 and was described as the delivery of DNA or RNA to cells in order to treat a disorder using a vector [50, 51]. In order to design a gene therapy, the gene and the mutation that causes the disease must be known, the use of a suitable delivery system is desirable and a disease model for a better understanding of the connection between the mutation and the disease is very helpful [45, 52]. Gene therapies can target two different types of cells: germ cells or somatic cells. When it comes to the germ cells, means that the cells targeted (sperm or egg) with a therapeutic gene would possibly be transmitted to the following generations. This gene therapy in germ cells is not performed in humans due to ethical issues. Somatic cells are all the cells of an organism except the germ cells. When gene therapy is performed on somatic cells, means that the cells are targeted with a therapeutic gene, these would only remain in the target organism and the following generation would not inherit this alteration [52, 53]. In order to correct altered genes with a disease causing-mutation through gene therapy, one of the following approaches can be used: (1) insertion of a wild-type gene somewhere in the genome, (2) replacement of the defective gene by a wild-type one, by homologous recombination (HR), (3) the defective gene could be corrected through reverse selective mutation (returning a gene to a normal functioning) or (4) the controlled expression of a gene (controlled turn on-off of the gene) [52, 54].

In general, gene therapies used can act at two different levels: (1) at the gene level (DNA) and (2) at the transcript level (RNA). Therapeutic molecules can be made from different types of nucleotides such as: double stranded or single stranded DNA or RNA [45, 51, 53]. These must reach their desired target and avoid degradation, and to do so they have to be associated to a delivery system [52, 53]. These can be of two kinds: viral vectors (retrovirus, adenovirus, adeno-associated virus) and non-viral methods (direct injection, electroporation, sonoporation, gene guns, magnetofection, lipoplexes, polyplexes, among others) [52, 53]. Cells to be repaired, can be through two

different ways: ex-vivo (repaired and then delivered to the organism) or in-vivo (vector directly administered to the organism) [51].

At the gene level several tools are available, but lately the most promising ones are genome-editing techniques [51, 55]. Basically, they resort to custom nucleases to perform either a gene inactivation (knockout), a gene insertion (knock in) or a gene repair [55]. This editing by nucleases involves the production of a double strand break (DSB) in a given region of the genome and then the repair of the DSB by the cell. Double strand DNA breaks can be repaired by two different strategies, either by homologous recombination (HR) or nonhomologous end joining (NHEJ) [51, 55]. Briefly, HR is the repair the cell performs after a DSB using the replicated sister chromatid for repair, without loss of genetic information, occurring only in the S/G2 phase of the cell cycle. NHEJ is an error prone repair mechanism, which originates insertions and deletions at the site of repair and occurs throughout the cell's life cycle [55, 56]. There are four types of available custom nucleases: meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the clustered regularly interspaced short palindromic repeats (CRISPR)/ Crispr associated nucleases (CRISPR/Cas system) [51]. The meganucleases are homing nucleases that perform DSB in determined loci of the genome and they only recognize certain sequences. Even being less toxic to cells than ZFNs, there are a limited number of them for each sequence, which limits the genome editing approaches [51]. ZFNs are composed of zinc finger, which are DNA binding domains and each domain only recognizes a set of three specific nucleotides. The ZFNs also have a nuclease domain that is the restriction nuclease FokI. In order for a ZFN to perform a DSB, there must be a set of two ZFN to dimerize in each side of the DNA double helix. Usually, each ZFN recognizes a set of 18 nucleotides and the need for dimerization diminishes the possibility of off-target events, but it is time consuming to develop [51, 55]. The TALENs are also a combination of a DNA binding domain and the nuclease FokI. This binding domain is derived from the TALE protein and each one recognizes one nucleotide. As for the ZFN, two TALEN are necessary to perform the DSB and together they usually recognize 18-20 nucleotides [51, 55]. Finally, the CRISPR/Cas system was adapted from bacteria in order to allow mammalian genome editing. It consists of the Cas (nuclease) together with a guide RNA. The two together are essential to perform the DSB. The guide RNA

recognizes a 20 nucleotides sequence (protospacer) in the genome. Nevertheless to have a successful DSB, this sequence in the genome must be next to a PAM (protospacer adjacent motif) sequence of NGG, where N is any nucleotide. The advantage of this system is that it is very quick to engineer, affordable and highly specific [51, 55].

At the RNA level, there are various techniques available to develop RNA-based gene therapies that go from splicing modulation through anti-sense oligonucleotide (ASOs) to spliceossome-mediated RNA trans-splicing [43]. These RNA-based strategies will be discussed later with further detail.

A viable therapeutic approach for HCM would necessarily include a strategy to impair the expression of un-functional protein in cardiac tissue [43, 48, 49]. The conventional gene therapy does not seem to be the best choice for correcting genetic defects in a complex and highly differentiated tissue such as cardiac muscle. An RNA based approach, aiming at correcting the genetic defect at the mRNA level, may be a better choice, as it does not imply modification of the genomic loci and, in principle, might work efficiently in a wide range of cells.

1.2. RNA therapeutics - an alternative to DNA gene therapy

1.2.1. Mechanics behind the splicing reaction

A possible approach as a gene therapy for HCM would be targeting at the RNA level, to diminish the levels of the mutated transcripts and so of the mutant proteins responsible for the haploinsufficiency, gain-of-function or a dominant negative effect, possible causes of this autosomal dominant disease. Until date, the RNA therapeutic techniques available rely in the interference at the pre-mRNA (messenger RNA) level mainly in a process called splicing. The RNA polymerase II is responsible for the transcription of a gene. During transcription, the pre-mRNA is synthesized and splicing occurs in a co-transcriptional manner for the vast majority of the human transcripts [57].

Splicing can be defined as the removal of introns (non-coding region between exons) from the pre-mRNA and joining of the exons (coding region of the gene) and so producing a mature mRNA, for future translation into a protein [57-59]. The splicing process is represented in **Figure 1.3**.

A complex of several ribonucleoproteins forms the spliceosome, which is responsible for performing the splicing reaction [57, 58]. The exon/intron boundaries, which allow the spliceosome to recognize them and join the exons together after intron excision, are the splice sites (ss). There is the 5'ss (donor site) that is located on the 5' end of the intron and the 3'ss (acceptor site), which is in the on the 3' end of a given intron. These are two of the most conserved sequences in the pre-mRNA, besides the branchpoint (BP, which is an intronic region, with an adenosine involved in the lariat formation) and the polypyrimidine (Py) tract (an intronic region rich in pyrimidine) [57, 58, 60]. Each of these features is in most cases recognized by different spliceosome components [57]. The spliceosome is made from five different small nuclear ribonucleoproteins (snRNPs: U1, U2, U4, U5, U6) and other splicing auxiliary factors, such as U2AF65, U2AF35 and SF1 (branch point binding protein) [57-60]. The splicing reaction starts with identification of the 5'ss by U1 snRNP, of the BP by the branchpoint binding protein (BBP), of the 3'ss by U2AF35 and of the Py tract by U2AF65 [57, 58, 60]. The binding of U1 and the U2AF factors to the respective splice sites leads

to the early formation of the spliceosome, the so called complex E [60]. Then, U2 snRNP binds to the BP, forming complex A of the spliceosome. After this, complex B is formed with the recruitment of U4, U5 and U6 snRNP. These three snRNPs assemble into the active spliceosome and the two splice sites come together, with U1 and U4 snRNPs being released, and leading to the formation of the complex C [59, 60]. Then, the first transesterification reaction occurs between the 2'-OH group of the adenosine that attacks the 5'ss, forming the intron lariat. The second transesterification reaction occurs when the 3'-OH group exposed of the exon upstream attacks the 3'ss and so the two exons are joined together and the intronic lariat is released and degraded [57-60].
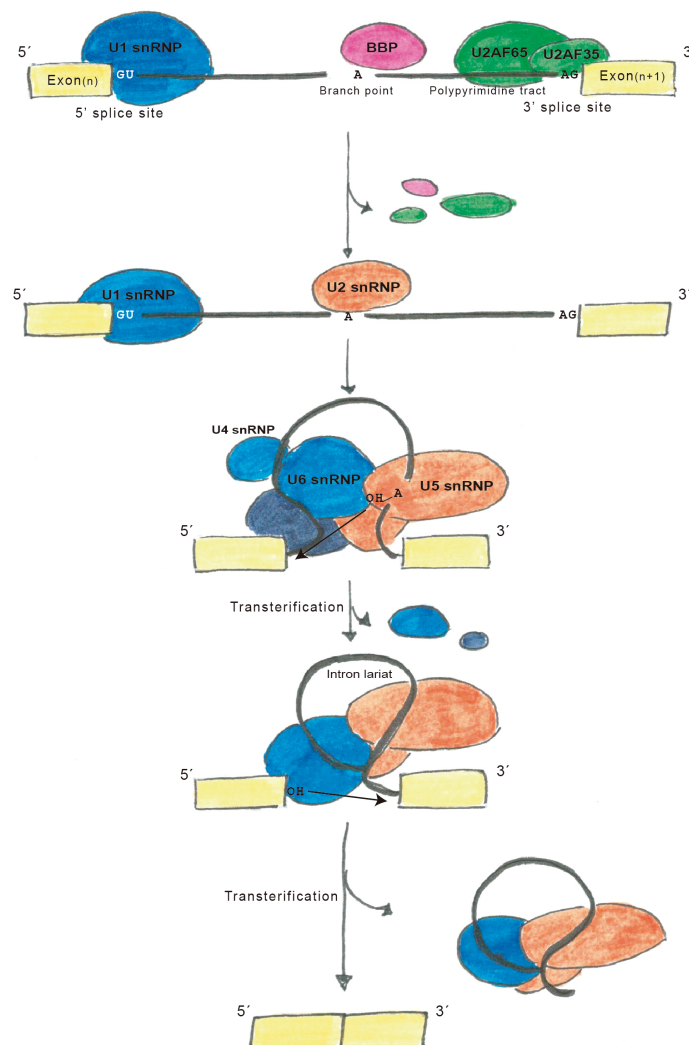
Figure 1.3 - Spliceosome assembly and splicing of the pre-mRNA.
The splicing reaction is the excision of introns and joining of exons, which is performed by the spliceosome. Adapted from [57].

The choice by the spliceosome between splice sites can give rise to different mRNAs and proteins [60]. The splice sites can either be strong or weak (depending on the consensus sequence conservation), and these are the regulatory sequences that make possible alternative splicing. The process of alternative splicing consists of different combinations of exons joined together through splicing and thus, obtaining different mRNAs and protein isoforms from the same gene, based on the choice of a given splice site [61]. If an exon has a strong splice site rather than a weak one, it will more likely be a constitutive exon. The presence of regulatory sequences helps regulate this process and determine the strength of a splice site. Some regulatory sequence can act as an intron or exon splicing silencer or enhancer (ISS, ISE, ESS, ESE) depending of the respective splicing regulatory proteins that bind to these motifs [57, 58]. These regulatory proteins can be from the Ser/Arg-rich and heterogenous nuclear ribonucleoproteins (hnRNPs) families, or be tissue specific and they can promote or not the use of a given splice site [57].

Several of the RNA therapies in study nowadays were designed to interfere in these splicing regulatory steps.

1.2.2. Different RNA therapies

The ability to modulate splicing, with different RNA therapies in order to overturn the possible damage caused by a mutation that could interfere in splicing is becoming a reality, without the need to interfere at the genomic level [58]. Pathogenic mutations that affect splicing are usually located in intronic regions. They usually affect splicing consensus sequences such as 5'ss, 3'ss, BP, Py tract and can result in exon skipping or even intron retention. A splice site mutation can even make the spliceosome use a weaker or cryptic splice site (splice site created by a mutation). Mutations in either exons or introns can also originate splicing regulatory sequences such as enhancers or silencers or even cryptic splice sites. The mutations in these regulatory motifs can interfere in exon inclusion or skipping and so in alternative splicing. Exonic mutations associated to a disease can be classified as missense (a point mutation in which a single nucleotide changes into a codon that codes for a different amino acid), as nonsense (a single nucleotide changes introduces a stop codon) or as

synonymous/silent mutations (resulting in no change of the amino acid) but all can interfere in splicing [58]. The identification of the mutation is key and its impact on the transcript is essential to understand the molecular pathology, mainly, in the case of an autosomal dominant disorder such as HCM, in order to develop an RNA-based approach to target and rescue the mutant transcript, and so the resulting protein.

To date, there are essentially three different RNA therapies described: modified snRNAs, antisense oligunucleotides (ASO) and RNA trans-splicing [43, 58, 62].

The modified snRNAs based approach can be exemplified for example by U1 snRNA that is part of the U1 snRNP and interacts with the 5'ss. If a mutation occurs in the 5'ss it could alter the binding of U1 snRNP and compromise splicesosome assembly. The idea is to create an altered U1 snRNA that could allow binding of the U1 snRNP to the altered 5'ss and resume normal splicing. A disadvantage of this method is the need to develop a vector to express the modified snRNA and arrange for its delivery [58, 63]. However, this approach was successfully performed on several disease models, such as cystic fibrosis, haemophilia B, Fanconi anemia, retinitis pigmentosa, among others [58].

On the other hand, antisense oligonucleotides can work either as splicing modulators or inductors of RNA degradation with the help of enzymes existing in the cell [62]. In a broader sense, antisense oligonucleotides (ASOs) can be defined as oligonucleotides with 15-25 nucleotides, which hybridize with a specific target RNA region. The sequence of the ASO is complementary to its target by Watson-Crick base pairing [58, 62]. ASOs can be used with the objective of promoting degradation of a target RNA and so leading to lower levels of a given protein. There are two sets of enzymes that can be used for this purpose by the cell: RNaseH and Argonaute 2 (Ago2) by RNA interference. Basically, the ASO binds to its target sequence in the RNA, forming a RNA duplex, which is then recognize and cleaved. This strategy was successfully used to perform knockdown of the expression of several genes. An example of this is an ASO against the immediate-early gene (*IE2)* transcript with indication against Cytomegalovirus infection, or even an ASO against the Apolipoprotein B transcript, with an indication in cholesterol management [62]. Moreover, ASOs may be used as splicing modulators; in this situation, they may be designed to target one essential splicing sequence to block or favour splicing in order to

produce a 'healthy' mature mRNA molecule. So, ASOs can either block the consensus sequences (5'ss, 3'ss, BP or Py tract) or a splicing regulatory motif (enhancer or silencer) by binding to it [43, 58, 62]. These ASOs can be specific enough to bind to different isoforms of a pre-mRNA, or specifically to a mutated one. They do not interfere at the genomic level; in the last few years, advances in the ASO design have increased their specificity and stability while decreasing toxicity. The delivery is simpler than other techniques, as cells uptake them easily. These characteristics make them very appealing for RNA therapeutics, being that the ultimate goal is to block the production of a mutant protein or to re-establish normal splicing to produce a normal protein [58]. When, the ASO blocks a splice site and induce exon inclusion or exclusion, it may re-establish the frame-shift of a previous mutated transcript [58]. A good example of this was in dystrophin gene (*DMD)*, responsible for Duchenne muscular dystrophy, where a mutation in the 3'ss was causing exon skipping and so a premature termination codon (PTC). An ASO was used to block this mutated 3'ss and promote the exon skipping in a way to restore the frame-shift [58, 64]. In another case, ASOs proved to be useful in blocking cryptic splice sites and promote the use of the 'real' splice sites. This was applied in several diseases models, like β-thalassemia in β-globin gene (*HBB*) [58, 65]. Furthermore, an ASO can also block a splicing regulatory motif and induce exon inclusion or exclusion and so re-establish the frame-shift. When a mutation creates a new splicing regulatory motif the use of an ASO that can block it may restore normal splicing. A mutation in the fibrinogen β-chain gene (*FGB)* that causes afribrinogenemia, creates a binding site for a splicing factor that caused a pseudo exon inclusion and the use of an ASO to block this new motif restored splicing [58, 66]. In the context of HCM, ASO-based approaches had already been performed, namely for an HCM mutation located in the 5'ss of exon 6 of the *MYBPC3* gene. In this study, the ASOs were designed to block certain exon splicing enhancer existing in exon 5 and 6, in order to promote the use of an alternative splice site versus the mutated 5'ss in exon 6, which led to skipping of the two exons and a viable mRNA transcript [43].

The RNA trans-splicing approach will be discussed further in the next topic.

1.2.3. Trans-splicing as gene therapy strategy at the mRNA level

Since the molecular pathology in HCM is generally attributed to a mutated transcript and in order to target and rescue altered transcripts and the resulting rescued proteins the use of an RNA based therapy such as splicesosome mediated RNA trans-splicing is a very good candidate, namely for not interfering at the genome level [43, 67].

Spliceosome-mediated RNA trans-splicing intend to use the cellular machinery of the spliceosome to promote trans-splicing between an endogenous pre-messenger RNA (pre-mRNA) and an exogenous therapeutic trans-splicing molecule (TSM) instead of cis-splicing within the pre-mRNA. The result is a wild-type repaired transcript. The TSM must include essential features such as annealing sequences (AS) to the target pre-mRNAs, consensus splicing regions (splice sites, branch point and pyrimidine tract), the coding region to be trans-spliced together with the target pre-mRNA and eventually, splicing enhancers to increase the success of trans-splicing [43, 58, 67]. The TSM is only active in cells expressing the pre-mRNA to be targeted and repaired. This approach allows to replace a mutated exon by a wild-type one in any part of the transcript [43]. There are three types of trans-splicing approaches, represented in **Figure 1.4**. The 5' trans-splicing occurs with the 5'end of the transcript and the TSM and so it allows the replacement of any mutated exon in the 5'end (**Figure 1.4A**). In the 3' trans-splicing the only difference from the 5' is that the replacement occurs in the 3' end of the transcript (**Figure 1.4B**). The last one, the double trans-splicing, is the most challenging from the technical point of view due to the need of a double trans-splicing event; it allows the replacement of any mutated exon in the middle of the transcript [67]. When designing a trans-splicing strategy for a TSM, an important feature for an efficient trans-splicing is the choice of the annealing sequences. The position where it binds in the intron, if it blocks any splicing consensus sequence, its length and GC content will impact in its efficiency [48, 67]. The first application of trans-splicing was a 3' approach for the treatment of haemophilia A [68]. Several successful studies using the 3' trans-splicing strategy showed it to be the most efficient one [48, 49, 67, 69, 70]. There was also a study that combined a 3'trans-splicing approach and ASOs against

intron 102 of the *COL7A1* gene (mutations in this gene cause epidermolysis bullosa), which helped increase efficiency of trans-splicing [48].

A

Exon 1 — Exon 2 — Exon 3   pre-mRNA

WT exon 1-2

5' trans-splicing molecule

B

Exon 1 — Exon 2 — Exon 3   pre-mRNA

WT exon 2-3

3' trans-splicing molecule

C

Exon 1 — Exon 2 — Exon 3   pre-mRNA

WT exon 2

Double trans-splicing molecule
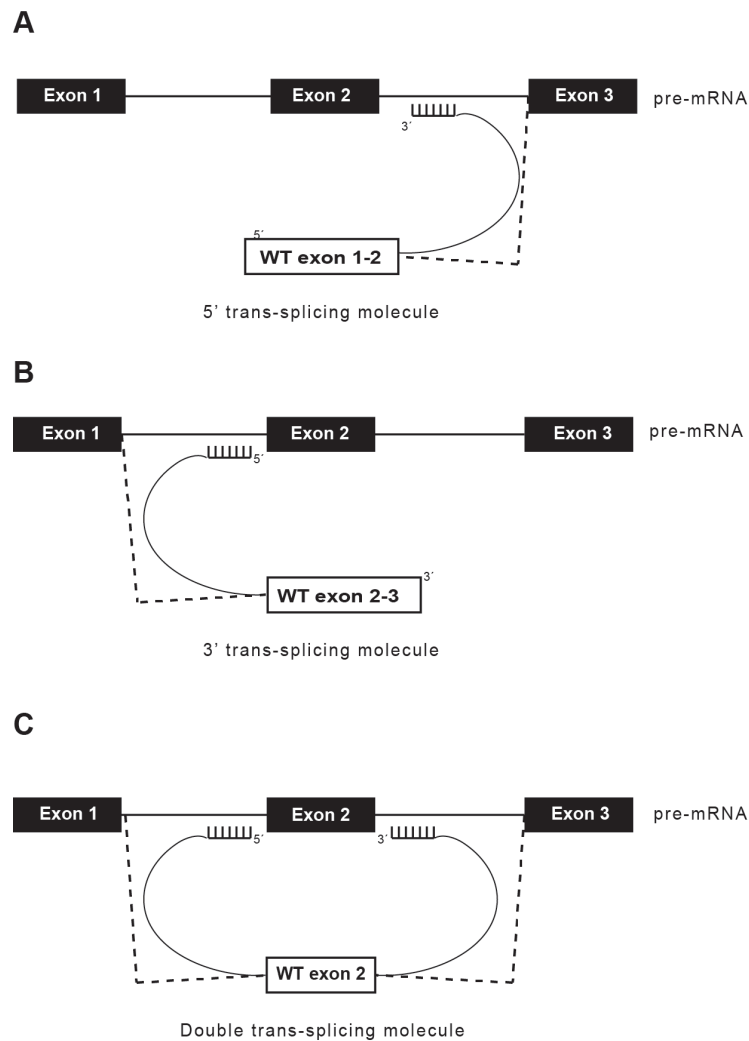
Figure 1.4 – Three different approaches of RNA trans-splicing
In the three approaches, the binding domain of the TSM allows for the pre-mRNA and the TSM to be in close proximity and for trans-splicing occurrence. (A) In the 5' trans-splicing approach, the 5' end of the pre-mRNA would be replaced by the coding part of the TSM. (B) The 3'trans-splicing consists in the replacement of the 3' end of the pre-mRNA by the corresponding present in the TSM. (C) In the double trans-splicing, there is a double trans-splicing reaction, where the TSM binds to both sides of the exon to be replaced.

An RNA trans-splicing approach has three advantages over conventional gene therapy: 1) the normal gene sequence is maintained and the trans-splicing will only work in cells that express the mutated transcript; 2) the delivery is easier than a delivery of a full length cDNA (when using a viral vector due to packaging size

restriction using this type of vector); 3) the mutant transcript is restored, which is why this approach is efficient in an autosomal dominant disorder with the reestablishment of wild-type protein production [67].

Some of the difficult issues found in trans-splicing approaches described were the low efficiency of production of wild-type protein especially in *in vivo* models and also the translation of the TSM, which is undesirable [48, 49, 70, 71]. Nevertheless, a successfully trans-splicing assay was already described *in vitro* and *in vivo* in HCM for the *MYBPC3* gene. The repair efficiency was around 30-60% in the cellular model, but in mouse model the results were less promising with only 7% of the mRNA being repaired and less than 4% of the protein being translated [43, 72]. These reports reveal that there is still room for improvement until trans-splicing can be considered as a viable option for autosomal dominant disease treatment.

HCM-causing mutations in the *TNNT2* gene are associated with the presence of a poorly functioning protein with a dominant negative effect. They are usually associated to myofibrillar disarray and changes in the Ca^{2+} regulation, leading to altered myocardial energetics and contractile capacity [26, 73, 74]. Here, an RNA based-therapy is a very attractive approach to decrease the amount of these deleterious proteins.

In ClinVar (an archive of variants reported in humans), there are 206 mutations identified in the *TNNT2* gene, of which 44 are associated to HCM. Of these 44, 19 were found to be pathogenic, 10 of uncertain significance and 11 had conflicting information [75]. The ones identified as pathogenic are possible candidates for RNA-based therapy.

1.3. Next-Generation Sequencing (NGS) in HCM

1.3.1. NGS nowadays

For more than 30 years, the sequencing reference method was automatic sequencing based on the Sanger method, developed in the 70s. It was an exceptional technique that allowed sequencing of whole genes and eventually whole genomes [76, 77]. The technique evolved over the years allowing, back in 2004, for the first human genome to be sequenced [77, 78]. The only issue was the time, cost and resources needed to sequence a single genome and so the need for a cheaper and quicker method arose [77]. In 2005, the first next-generation sequencing (NGS) technologies came to the market [76, 77]. Several of them were developed: 454, Illumina, SOLiD and Ion Torrent [77, 79, 80].

These new techniques consist in two major steps. The first one consists in the preparation of the library, where the DNA or RNA to be sequenced is fragmented to the desired size, followed by the ligation of adapters to each fragment. The adapters are specific of each system, and are used to attach the fragment to a surface, allowing sequencing to take place [76, 77]. The fragments are then amplified on the surface they are attached to, and this step is essential to get enough signal to be detected with an enormous amount of sequencing reactions taking place at the same time (high-throughput technique) (**Figure 1.6**) [76]. In opposition to Sanger sequencing techniques, in NGS, the sequencing and detection are performed at the same time. Also, the size of the reads is much longer in Sanger sequencing compared to NGS; the read length in NGS is influenced by the signal-to-noise ratio, which varies between NGS platforms and the sequencing reactions. The noise varies from one NGS platform to the other; the biggest issue with this is the read size limitation [76]. Nevertheless, the Sanger sequencing method remains the go-to technique to check for a particular known mutation [81]. The short reads from NGS are aligned to reference genome thanks to newly developed algorithms and the read length and quality of the signal-to-noise ratio has been improved over the last few years [76, 77]. An improvement to the quality of the data was the use of paired sequencing, where the sense and antisense are sequenced [76]. After obtaining the reads, these are aligned to a reference genome

25

and then all variants detected are annotated for further processing using various bioinformatics tools [81, 82].
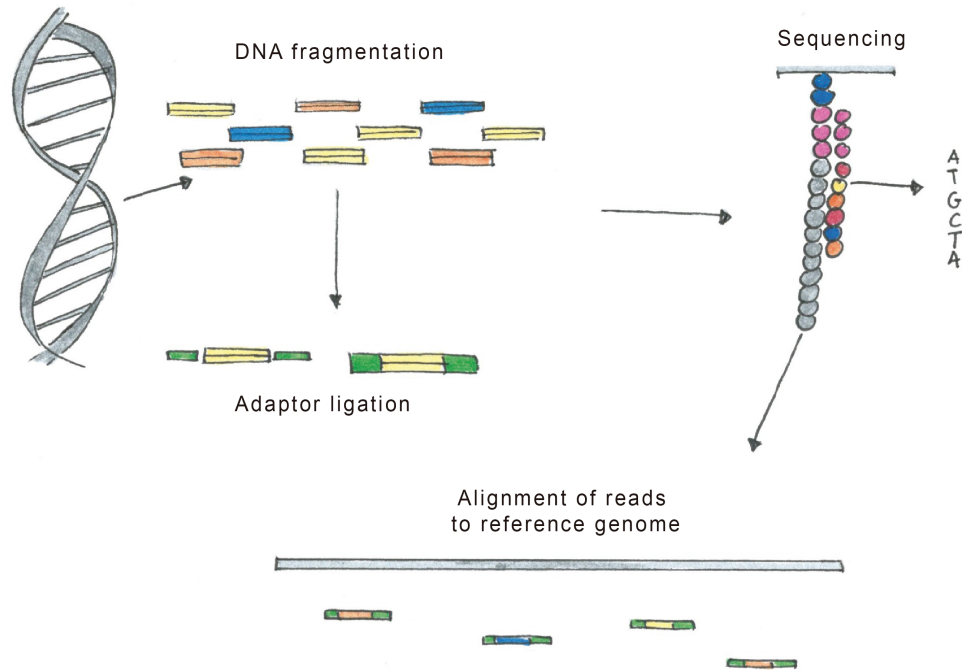
Figure 1.5 – The two major steps of NGS

The NGS techniques involve two steps: library preparation and the product to be sequenced attached to the sequencing platform by the adaptor. After sequencing the reads are aligned to the reference genome. Adapted from [83].

The advances in this field make the cost for genome sequencing very affordable for various clinical and research institutions. Also, whole genome sequencing is more and more present in basic and translational studies and in clinical settings. NGS is also being used for screening in genetic disorders to help better understand of the pathogenesis and then contributing to a better management of a disease [77]. For some applications, there is no need to sequence the whole genome but only the coding region (exome sequencing) or even a panel of genes associated to the disease in study. An advantage for a more selected panel is the higher level of coverage of the reads and no variants going undetected [77, 84].

1.3.2. NGS and the HCM patient screening

The NGS has been evolving over the years making possible the identification of novel pathogenic mutations and contribute to the understanding of pathogenesis of genetic disorders [5, 81]. When it comes to inherited cardiomyopathies there is a wide panel of 30 genes that could be studied easily and fast by NGS, and which are already available in several laboratories [6, 81]. Nevertheless, in 50% of HCM cases a mutation has yet to be identified [5]. In 50-60% of patients who were screened and had an HCM associated mutations identified, one of the following sarcomeric genes were implicated: *ACTC1, MYBPC3, MYH7, MYL2, MYL3, TNNI3, TNNT2, TPM1* [1, 82]. The main reason for genetic testing of patients and their relatives that might be at risk is to help in their diagnosis and disease management [5, 6]. A downfall from the amount of data obtained from NGS is that the number of variants with unknown significance increases, which will influence the genetic diagnosis of HCM [1, 5, 6, 82].

As mentioned previously, NGS has rapidly evolved over the years and offered a new possibility for genetic screening, clinical management for genetic disorders and contribute in the understanding of these disorders [85, 86]. It offers the possibility to assess a greater number of genes in a shorter time frame and at a lower cost [85]. There are different types of clinical screening possible: a single gene, a gene panel, exome sequencing and whole genome sequencing. Each one is more suited for a given disease [85].

When the association between a disease and a gene is well known, a single gene screening is more reliable. There is less probability of finding variables of unknown significance, which can be a confounding factor and it is more time and cost effective [85]. This type of screening is not applicable in a complex disease like HCM.

When a disorder has such a heterogeneous genotype and phenotype as HCM, the best option is the use of a gene panel to determine the genetic cause of a patient's disorder, the so-called 'molecular diagnosis' in which HCM-associated genes are completely sequenced, allowing the discovery of new or already reported coding and non-coding variants. The screening with a gene panel for a particular disease might

vary among laboratories, due to the regional, institutional guidelines and scientific evidence of a gene being associated to a disease. Even though some laboratories might choose to screen all genes associated, some choose not to and only screen the ones with more clinical significance. In order to include a gene in a gene panel, it should be described extensively and have reported association to the disorder or to genes of diseases with similar phenotypes [85]. The gene panel will also be useful for the detection of variants in introns that might interfere in splicing or even transcription factor binding sites and impact in a disease's phenotype.

In order to find new genes that might be associated to a disease, a promising approach is exome sequencing [85, 87]. Basically, it consists in just sequencing the exons. It allows the screening for new or already characterised coding variants so that they can be associated to a given clinical setting [88]. Besides research purposes, the analysis of the sequencing data from exome sequencing might only focus on a gene pool associated to the disease to allow the molecular diagnosis. Nevertheless, to obtain a genetic profile besides using NGS, other techniques must be used such as Sanger sequencing to confirm NGS results. Another issue that arises with NGS is the increase in the number of variants with unknown significance found. These variants will be considered as potential disease causing mutation candidates if they are expected to affect transcription or RNA processing of a gene reported to be associated with the disorder but they are not represented in any database as disease associated mutations.

Since HCM is such a heterogeneous disease either phenotypically or genetically and the variants of unknown significance are abundant, this makes clinical interpretation and pathogenesis more complex [81, 82]. In order to understand better the impact of newly discovered variants, they should undergo *in vitro* and *in vivo* testing, using cell and animal models [82].

1.4. HCM and genome editing

Hypertrophic Cardiomyopathy (HCM) is an autosomal dominant disorder with low penetrance and variable clinical outcomes, a conventional gene therapy is not applicable [6, 8, 10, 45]. One of the criteria to develop an efficient gene therapy is to better understand phenotype/genotype relationships present in the disease and, for this, cellular and animal models are needed [52]. HCM is a phenotypically heterogeneous and genetically complex disorder, where tractable and accurate disease models are essential for better understanding of its genotype/phenotype relationships. Cellular models are easier to manipulate and to work with when compared with animal ones, and may provide valuable information regarding disease pathogenesis and the development of new therapeutics. However, there is still a grey area when it comes to the understanding of the HCM phenotype at the cellular level, due to the large HCM associated variants and also to the complexity of the affected organ, the heart, where functional impairment may not be reflected in a simplistic cellular model. In order to create reliable HCM cell models, not only a careful characterization of HCM cardiomyocytes is required, but also a simple, time and cost-efficient genome-editing tool is needed. A tool meeting these criteria was developed and used during the past 5 years: the clustered regularly interspaced short palindromic repeats, the CRISPR/Cas9 technology.

1.4.1. The CRISPR/Cas9 genome editing system

Bacteria have developed the equivalent of an adaptive immune system: the CRISPR-Cas9 system [89]. It was first reported in 1987 and found to be present in 50% of all bacteria and most *archae* [89-91].

CRISPR-Cas systems were shown to be an efficient tool against foreign DNAs in prokaryotes, and multiple variants of the system were found that carried different coding sequences for the Cas proteins and different repeat sequences. The different CRISPR/Cas9 systems were further separated in three groups (type I-III) with various sub-groups (A-E) [92]. It took several years to understand the way the system works and to start using it to edit eukaryote genomes [89, 92]. The system preferably used for this purpose belongs to the organism *Streptococcus pyogenes*, which is a type II

CRISPR-Cas system [93-96]. The way this CRISPR-Cas system works in bacteria is represented in **Figure 1.6**. First, the immunization step involves a first contact with a foreign DNA (either a phage or a plasmid DNA). Fragments of the foreign DNA, known as 'spacer', are processed by the Cas complex and then are incorporated in the CRISPR locus between identical sequences called 'repeats', of around 36 nucleotides. The information from different foreign agents is stored for further use, in case the bacteria comes in contact with them again [89, 94-97]. If the latest one occurs, a second phase of immunity takes place with the CRISPR array being transcribed and giving rise to pre-crRNA transcripts that are cleaved into short crRNAs containing the individual spacer sequences and the repeats. In the same CRISPR array, there is also the presence of a sequence that when transcribed codes for another non-coding RNA, the tracrRNA that hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNA/tracrRNA complexes. The crRNA/tracrRNA complexes then directs the Cas9 nuclease (also encode by the CRISPR array) to the target region in the genome by binding by Wastson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM). The PAM, a 5'-NGG sequence, where N is any nucleotide, is an additional requirement for target recognition by the crRNA/tracrRNA complexes. This PAM sequence is on the 3' end of the protospacer. Finally, Cas9 mediates cleavage of target DNA by creating a double-stranded break within the target region in the genome [89, 94-96].
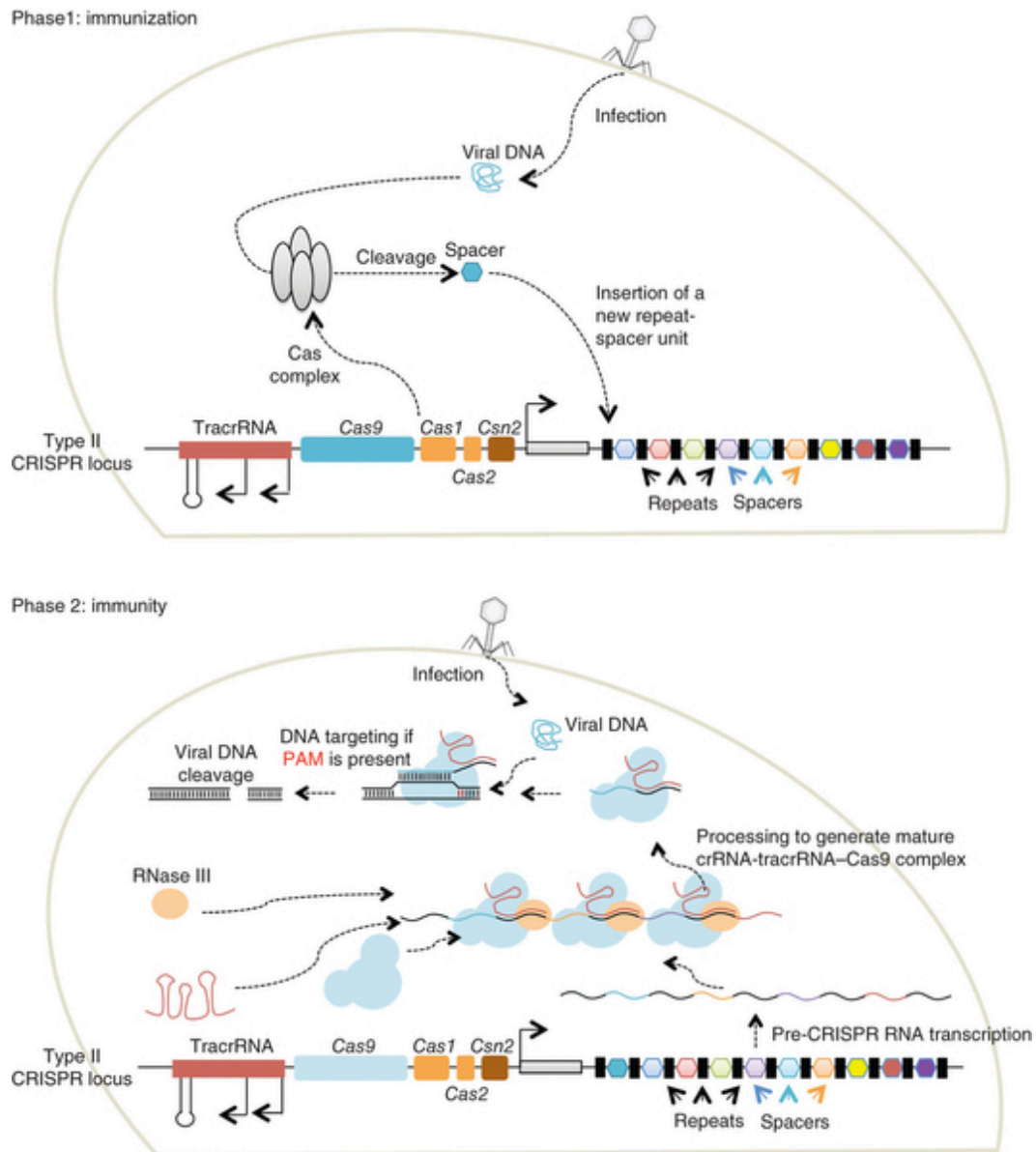
Figure 1.6 – Type II CRISPR-Cas9 system in *Streptococcus pyogenes* bacteria

In the first phase of immunization, bacteria take advantage of the CRISPR locus to store a fragment of DNA from an invading phage or plasmid DNA that is called 'spacer' and will be located between repeat regions in this locus. Upon a secondary contact with the same foreign DNA, the bacteria will express the transcript from the CRISPR locus that includes the spacer and repeats (pre-crRNA). After this, tracrRNA will bind to the repeat regions of the pre-crRNAs and a nuclease Cas9 binds to the each hybrid of crRNA and a tracrRNA. Then, RNaseIII proceeds to the final processing by separating each complex (Cas9-crRNA-tracrRNA). Each complex will bind to the foreign DNA of which the crRNAs is complementary and in the presence of a PAM sequence cleaves it by a DSB (adapted from [89]).

The previously described type II CRISPR-Cas9 system was further adapted to create a simple eukaryote genome editing tool, in which by creating a DSB promotes

either non-homologous end joining (NHEJ) or homologous recombination (HR) through a given donor DNA. Indeed, when a cell suffers a DSB in its genome, it undergoes one of these two pathways. NHEJ is an error filled process, which consists of joining the two ends of the strands where insertions and deletions (indels) can occur. HR consists in the repair using a homologous sequence; usually using the sister chromatid of the gene but in genome editing assay a donor DNA can be used as template [94, 96]. The desired DNA sequence to be inserted in the DSB locus must be flanked by homology arms (HA) with defined size, in order to take advantage of the HR process. For example, to perform a custom homologous recombination the HA can be of the size of 500bp to 1000bp for an insertion of less than 8kb, if the insertion should happen to be bigger than 8kb the HA must be bigger than 1kb [98].

Furthermore, to simplify the use of the CRISPR-Cas9 system a single guide RNA (sgRNA) was created, which is a chimeric RNA that includes the spacer region and the tracrRNA region essential for Cas9 binding [94, 96]. The binding of Cas9 and sgRNA to a target region is represented in **Figure 1.7**, as well the site where the Cas9 performs the DSB: three nucleotides upstream the 3' end of the PAM sequence [96].
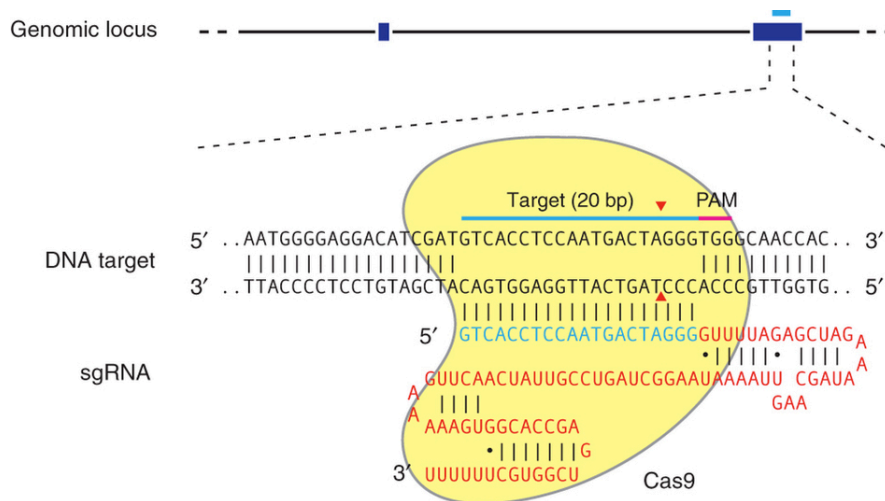
Figure 1.7 – Representation of the CRISPR-Cas9 binding using the chimeric sgRNA

This adaptation of the CRISPR-Cas9 system from *S. pyogenes* uses the Cas9 (in yellow), and instead of the crRNA and tracrRNA complex a single guide RNA (sgRNA) was developed which has the spacer sequence (in blue) and part of the tracrRNA essential for binding of Cas9 (in red). The red arrows represent the site where the DSB is performed. (Adapted from [96])

CHAPTER 2

OBJECTIVES

2. Objectives

For my PhD training, I proposed to explore applications of recent molecular genetics tools to improve the diagnosis of hypertrophic cardiomyopathy patients and develop potential new treatment strategies for this disease. My work focused on three main hypotheses:

Hypothesis I: Trans-splicing may efficiently correct the expression of mutant sarcomeric genes in cardiac cells.

To address this view, my first objective was to investigate the feasibility of inducing trans-splicing in murine cardiomyocytes to replace a specific exon in *TNNT2* pre-mRNA that is frequently mutated in hypertrophic cardiomyopathy patients.

Hypothesis II: Mutations located deep within the introns, which are not detected by conventional sequencing analysis restricted to exons and exon-intron boundaries, may contribute to the pathogenesis of hypertrophic cardiomyopathy.

My second objective was to develop a bioinformatics pipeline for scrutinizing variation along the complete intronic sequence of disease-associated genes and prioritizing candidates for mechanistic and functional analysis.

Hypothesis III: Cardiomyocytes derived from gene-edited embryonic stem cells recapitulate hypertrophic cardiomyopathy characteristics.

To pursue this idea, my last objective was to to use the CRISPR-Cas9 genome-editing technology to produce isogenic embryonic stem cells harboring either a wild-type or mutated *TNNT2* locus.

CHAPTER 3

RESULTS I: Spliceosome-mediated RNA trans-splicing approach for correction of HCM-causing *TNNT2* mutations

3. RESULTS I: Spliceosome-mediated RNA trans-splicing approach for correction of HCM-causing *TNNT2* mutations

3.1. Summary

The objective of this work was to investigate the feasibility of inducing trans-splicing in murine cardiomyocytes to replace a specific exon in *TNNT2* pre-mRNA that is frequently mutated in hypertrophic cardiomyopathy patients.

Two different trans-splicing approaches were chosen as an RNA-based therapy. Briefly, mouse atrial cardiomyocytes (HL-1 cells) were transfected with trans-splicing constructs designed to anneal the two introns flanking the murine *Tnnt2* exon 8 or only the intron upstream of exon 8, which in humans contains a cluster of HCM-associated mutations. These two strategies, double trans-splicing and 3' trans-splicing, respectively, were used with trans-splicing constructs being driven by the endogenous *Tnnt2* promoter or the cytomegalovirus (CMV) promoter. The efficiency of these trans-splicing strategies was evaluated using RT-PCR based techniques, where the replacement of murine exon 8 by the human equivalent exon 9 or the replacement of the murine exons 8-15 by the equivalent human cDNA was assessed. Detection of a residual amount of trans-splicing product was only possible by PCR with a radioactive γ^{32}P-labelled primer, which binds specifically to double trans-splicing molecules. All together, these results argue that trans-splicing does not ensure efficient correction of expression of a mutated *TNNT2* gene in cardiac cells. This could be due to inefficiency of trans-splicing reactions in general, or a particular resistance to trans-splicing of the targeted region in the *TNNT2* pre-mRNA.

3.2. Material and Methods

Cloning of the *Tnnt2* promoters in the pcDNA3 plasmid

The entire *Tnnt2* promoter was amplified by PCR with the following primers ForTNNT2entpro: 5'-TACAGATCTTAGTGAGCAAGCCAGACACAGC-3' and RevTNNT2entpro: 5'-CCAAGCTTCAACTCACTTCCCGTCAAGAAT-3' from mouse genomic DNA [99]. Each primer contains, respectively, the *Bgl*II and *Hind*III restriction sites, for directional cloning into the pcDNA3 plasmid, giving rise to pcDNA3-EntTNNT2 construct. The minimal *Tnnt2* promoter was amplified from mouse genomic DNA with the forward primer (ForTNNT2mpro: 5'-GCTAGATCTCATCTGCTTTATCGGGATTCTCA-3') and reverse primer (RevTNNT2mpro: 5'-CCAAGCTTCACACAGGTCTTGAGGTATCTGTTC-3') [100, 101]. These primers also bear the *Bgl*II and *Hind*III restrictions sites, respectively, for directional cloning into the pcDNA3 plasmid, to originate pcDNA3-S construct. The promoter fragments were amplified using the NZYLong DNA polymerase (Nzytech, Lisboa, Portugal) and purified from agarose gel prior ligation to pcDNA3 by High Pure PCR Product Purification Kit (Roche, Mannheim, Germany) (**Figure 3.1**). The DNA of plasmid constructs (pcDNA3-S and pcDNA3-EntTNNT2) was isolated by GeneJET Plasmid Miniprep Kit (Life Technologies, Paisley, United Kingdom) and sequenced by Sanger sequencing at Stabvida (Stabvida, Caparica, Portugal).
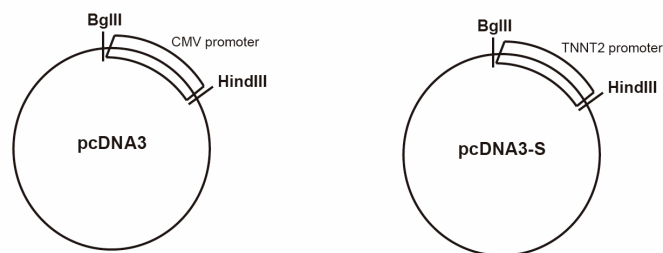
Figure 3.1 - Cloning strategy of the TNNT2 promoter.

The minimal TNNT2 promoter was amplified by PCR with the forward primer For TNNT2 pro and reverse primer Rev TNNT2 pro from mouse genomic DNA. Each primer contains respectively the *Bgl*II and *Hind*III restrictions sites, which were used to clone the mentioned promoter in the pcDNA3 plasmid.

Cloning of GFP reporter downstream of the *TNNT2* promoters

The green fluorescence protein (GFP) sequence was excised with *Kpn*I and *Not*I from the pEGFP-N1 plasmid (Clontech, California, USA) and cloned between *Kpn*I and *Not*I restriction sites downstream of the promoters in pcDNA3-S and pcDNA3-EntTNNT2 constructs.

Construction of the double trans-splicing plasmids

Cloning of an array containing three muscle specific enhancer (MSE$_2$ 3x)

A MSE$_2$ (3x) array was assembled in the pTRE-tight plasmid. Briefly, an oligo was designed with one copy of the MSE$_2$ sequence flanked by *Pvu*II and *Hind*III restriction sites: 5'-CTGTTACTAGTACTGCACCTTTCTTTGTTCCATCTCTCCACCTCTGCTGTGCCTAGGCCA-3' [102], and inserted between these restriction sites of the pTRE vector. In order to insert the second MSE2 copy, pTRE-MSE2(1x) was digested with *Pvu*II and *Avr*II to obtain the MSE2(1x) fragment, which was then cloned in *Pvu*II and *Spe*I (an *Avr*II compatible site) sites of the pTRE-MSE2(1x), generating the pTRE-MSE2(2x) construct. The same approach was performed to insert the third copy of MSE2 in the pTRE-MSE2(2x) to generate pTRE-MSE2(3x) construct (**Figure 3.2**).
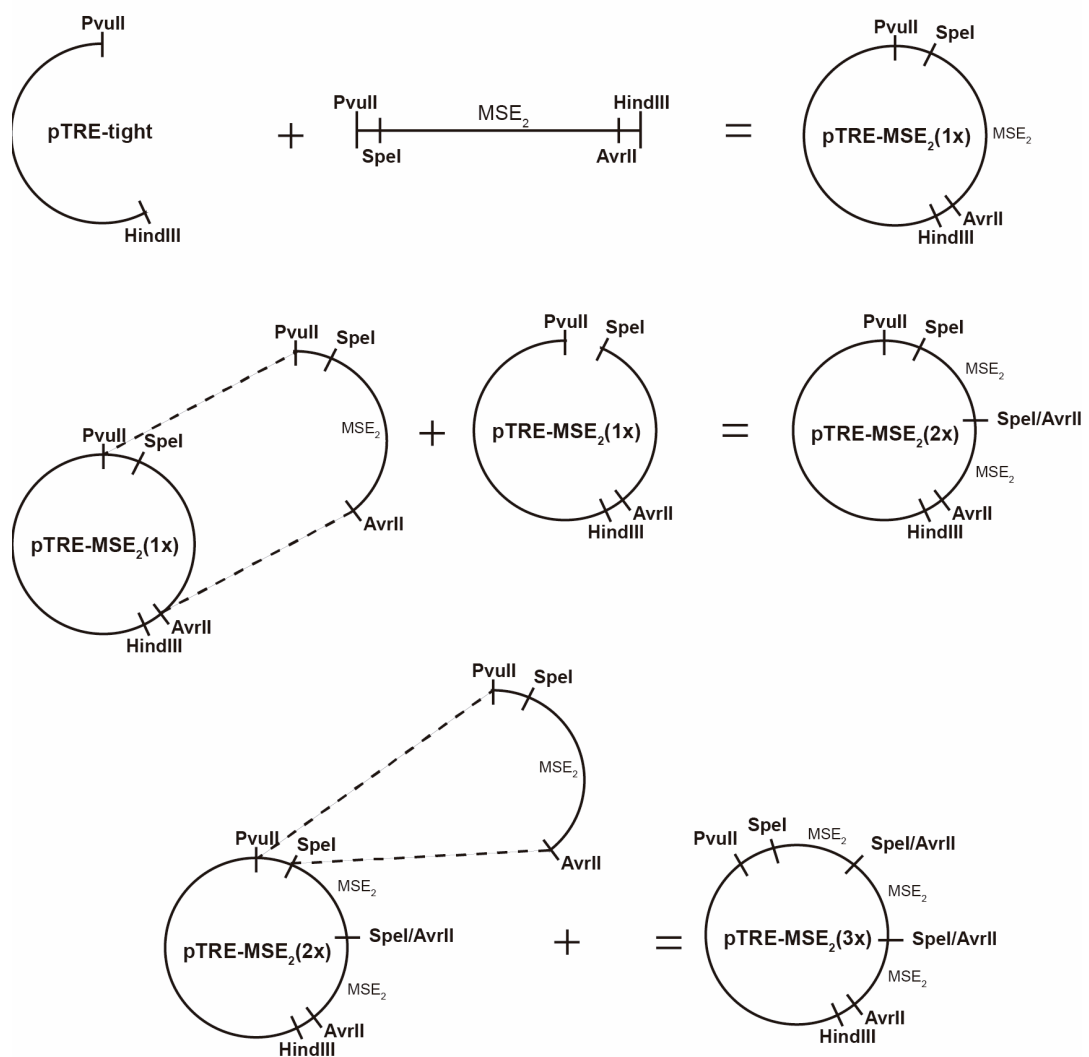
Figure 3.2 - Cloning strategy of the MSE₂ array in intron 8 and 9 of the trans-splicing construct.

The MSE₂(3x) array was assembled in the pTRE-tight plasmid. An oligo with one copy of the MSE₂ sequence was designed flanked by restriction sites. One MSE2 copy was inserted in *Pvu*II and *Hind*III sites of the pTRE vector. In order to insert the second MSE2 copy, pTRE-MSE2(1x) was digested with *Pvu*II and *Avr*II to obtain the MSE2(1x) fragment, which was then cloned in *Pvu*II and *Spe*I sites of the pTRE-MSE2(1x), generating the pTRE-MSE2(2x) construct. The same approach was performed to insert the third copy of MSE2 in the pTRE-MSE2(2x) and generate pTRE-MSE2(3x) construct.

<u>Cloning and mutagenesis of the human *TNNT2* Intron8/Exon9/Intron9 *TNNT2* fragment</u>

The Intron8/Exon9/Intron9 fragment of the human *TNNT2* gene was amplified by PCR from human genomic DNA (ForhuInt8: 5'-GAATTCCTGGGTTCAGGCCACAGTTAC-3', RevhuInt8: 5'-AAGCTTCTGTGCCATATGCTTGGATAGAT-3') and then inserted between the *EcoR*I and *Hind*III restriction sites of the pTRE vector. To insert the MSE2(3x) array

39

in the pTRE Int8/Ex9/Int8 h*TNNT2* construct , two sequential rounds of mutagenesis were performed on the human *TNNT2* intron 8 and 9 sequence in order to generate *Pvu*II and *Avr*II restriction sites in intron 8 and *Nae*I site in intron 9 (**Figure 3.3**). pTRE-MSE2(3x) was then digested with *Pvu*II and *Avr*II and the MSE2 (3x) array was cloned in the intron 8 of the Int8/Ex9/Int9 h*TNNT2* construct. In Intron 9, the same MSE2(3x) array was cloned, in blunt, between the *Nae*I generated ends (**Figure 3.4**). In the end, a Int8-MSE$_2$(3x)/Ex9/Int-9-MSE$_2$(3x) fragment was obtained.
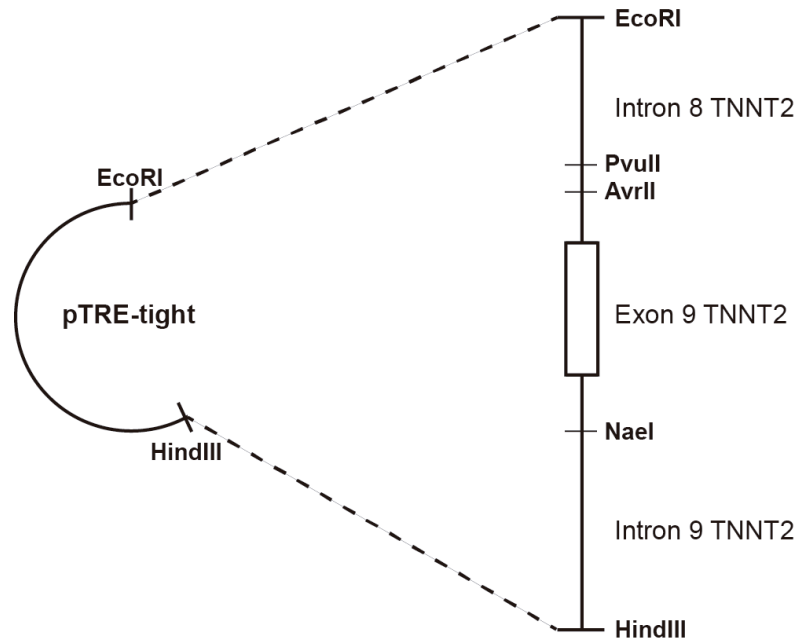
Figure 3.3 - Cloning of the Int8/Ex9/Int9 hTNNT2 in pTRE.

To insert the MSE2(3x) array in the pTRE Int8/Ex9/Int9 hTNNT2 construct , in the human TNNT2 the enzymes *Pvu*II and *Avr*II were used in intron 8 and *Nae*I enzyme in intron 9.
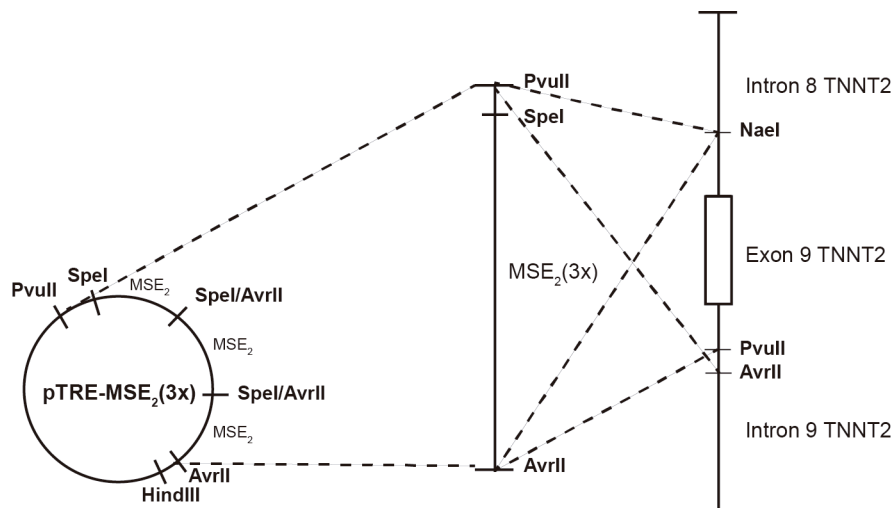
Figure 3.4 - Cloning of the MSE₂ (3x) array.

The pTRE-MSE₂(3x) vector was then digested with *Pvu*II and *Avr*II and the MSE₂ (3x) array was cloned in the intron 8 of the Int8/Ex9/Int9 hTNNT2 construct. In Intron 8, the same MSE₂(3x) array was cloned, in blunt, between the *Nae*I generated ends.

Cloning of two annealing sequences into pcDNA3 constructs

The pcDNA3-S construct containing the minimal TNNT2 promoter and the original pcDNA3 plasmid that bears a CMV promoter were then used to insert the two annealing sequences AS-Int7: 5'-CACTGGCTTGGCTTAGCGCCTGAACCTGAACCATCC TGCCAACAACGGCCTGCTCGCAGCCCACAAGATCAGGAACTCTGTCACTTGTATGAAAGCTCG GTCACCACCCCAGCCTGGCTCGCTCCACCG-3' and AS-Int8: 5'-ACAAGCTGATCCCA TGCTCCCTACGTTAGTGAGGCAGCTCTGACTGATGGGGTCAGGAACAGGGAGAGCACAGACA TTTGACTCGTCTGTGGTCTATAGCCAGTGGTGCCTAGTTAGGCTCCGCGAGAGCCAGAACTTAT CAGGGTCAGTTCCTGTTGGCAAGGAGCACCCACA-3' immediately downstream of each promoter between *Kpn*I and *BamH*I sites for the AS-Int7, and *Not*I and *Apa*I for AS-Int8. The obtained plasmids were named: pcDNA3-AS-Int7/AS-Int8 and pcDNA3-S+AS-Int7/AS-Int8 (**Figure 3.5**).
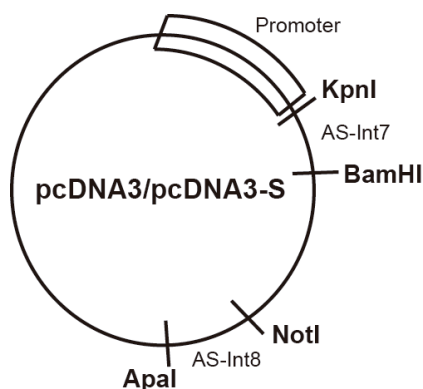
Figure 3.5 - Cloning of the annealing sequences downstream of the promoters.

The pcDNA3 plasmids with the minimal TNNT2 promoter (pcDNA3-S) and the CMV promoter (pcDNA3) were used to insert both annealing sequences AS-Int7 and AS-Int8 immeadiously downstream of these promoters between *Kpn*I and *BamH*I sites for the AS-Int7, and *Not*I and *Apa*I for AS-Int8 and thus obtaining: pcDNA3-S+AS-Int7/AS-Int8 and pcDNA3-AS- Int1/AS-Int8.

41

After the insertion of both annealing sequences, the Int8-MSE$_2$(3x)/Ex9/Int-9-MSE$_2$(3x) fragment was inserted between both of them into pcDNA3-AS-Int7/AS-Int8 and pcDNA3-S+AS-Int7/AS-Int8. The Int8-MSE$_2$(3x)/Ex9/Int-9-MSE$_2$(3x) fragment was cut from the pTRE vector by *EcoR*I and *EcoR*V double digest and was cloned into the pcDNA3-S+AS-Int7/AS-Int8 plasmids using the same sites. The final double trans-splicing constructs were named pcDNA3-ASInt7-Ex9-ASInt8 and pcDNA3-S-ASInt7-Ex9-ASInt8 (**Figure 3.6**).
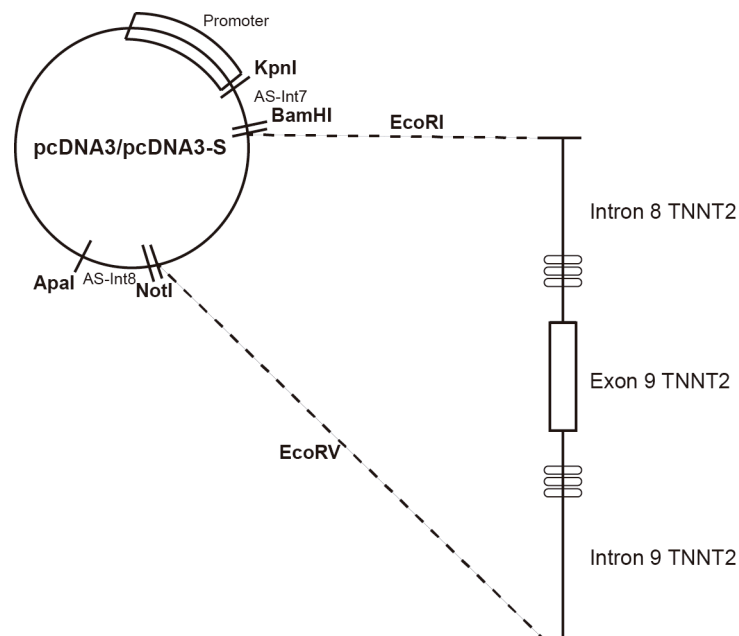
Figure 3.6 - Cloning of the MSE$_2$ (3x) array in the introns flanking the human Exon 9.
The Int8-MSE$_2$(3x)/Ex9/Int-9-MSE$_2$(3x) fragment was obtained from the pTRE vector by *EcoR*I and *EcoR*V double digestion and was cloned into the pcDNA3-S+AS-Int7/AS-Int8 plasmids using the same sites.

Construction of the 3' trans-splicing plasmids

The fragment AS-Int7-Int8-MSE2(3x)-Ex9 from the pcDNA3-ASInt7-Ex9-ASInt8 trans-splicing plasmid was amplified by PCR using ForAS7*KpnI*: 5'-GCGGGTACCACTGGCTTGGCTTAGCG-3' and RevEx9EcoRV: 5'-GCGGATATCAGCGC CTGCAACTCATTC-3' and cloned in pcDNA3 and pcDNA3-S between the *Kpn*I and *EcoR*V restriction sites, obtaining the first fragment for the 3' trans-splicing construct pcDNA3-AS7-Int8-Ex9 and pcDNA3-S-AS7-Int8-Ex9. The human *TNNT2* cDNA from exon 9 to 16

was amplified from human cardiac cDNA using the primers ForcDNAhuEx9: 5'-
CGAGGCTCACTTTGAGAACAGGAAG-3'/ RevcDNAEx16NotI: 5'-TAAGCGGCCGCAT
TACTGGTGTGGAGTGG-3' and cloned between the *EcoR*V and *Not*I sites of pcDNA3-
AS7-Int8-Ex9 and pcDNA3-S- AS7-Int8-Ex9 to generate the final construct pcDNA3- AS7-
Int8-Ex9-16 and pcDNA3-S- AS7-Int8-Ex9-16. These are the final 3'trans-splicing
constructs (**Figure 3.7**).
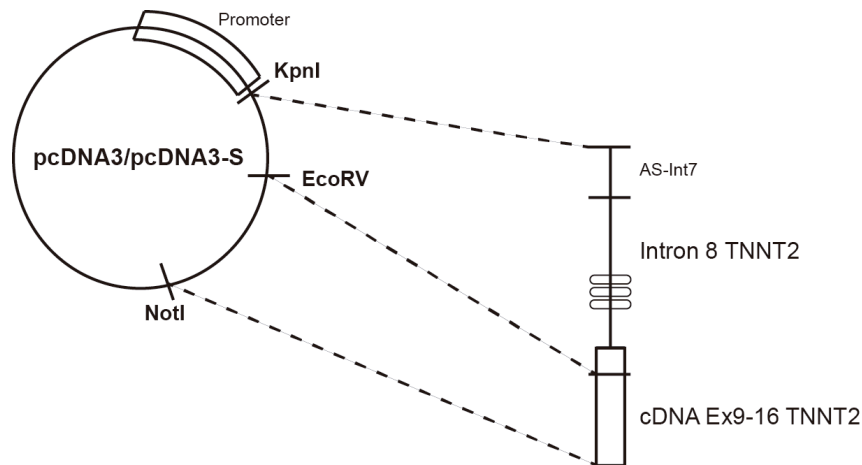
Figure 3.7 - Cloning of the 3' trans-splicing construct.

The fragment AS-Int 7-Int8-MSE2(3x)-Ex9 from the pcDNA3-AS-MSE2(3x)-trans-splicing plasmid was
amplified by PCR using For AS7 *KpnI* and Rev Ex9 EcoRV. It was cloned in pcDNA3 and pcDNA3-S using
the *KpnI* and *EcoRV* restriction sites, obtaining the first fragment for the 3' trans-splicing construct. The
human TNNT2 cDNA from exon 9 to 16 was amplified from human cardiac cDNA using the primers For
cDNA CGA huEx9 / Rev cDNA Ex16 NotI. These were cloned in the *EcoRV* and *NotI* sites.

HL-1 cell culture

HL-1 cells were a kind gift from Dr. William Claycomb. HL-1 cells are a mouse
atrial cardiomyocyte cell line [103]. They were maintained in Gelatin/Fibronectin
coated flasks in Claycomb Medium (Sigma-Aldrich, St.Louis, USA) supplemented with
10% Fetal Bovine Serum (FBS) (Sigma-Aldrich), 100µg/ml Penicillin/ Streptomycin
(Sigma-Aldrich), 0,1mM Norepinephrine (Sigma-Aldrich) and 2mM L-Glutamine (Sigma-
Aldrich). The medium was changed every 24h. Cells were split at 100% confluency,
when they exhibit a spontaneous contraction phenotype [103]. Cells were dissociated

using 0,05% Trypsin-EDTA (Sigma-Aldrich) inactivated by a Soybean Trypsin Inhibitor (Sigma-Aldrich) and centrifuged 5 min at 500g before plating.

Plasmid transfection of HL-1 cells

For transient transfections, HL-1 cells were seeded in 35mm dishes and transfected with 2,5µg of plasmid DNA using Lipofectamine 2000 (Life Technologies). For stable transfections, HL-1 cells from a confluent T75 flask were electroporated using Gene Pulser Xcell™ Electroporation Systems (Bio-Rad, California, USA) under the following conditions of 250V, 960µF and 100Ω; followed by 2µg/ml neomycin selection started 48h after electroporation.

RNA isolation and cDNA synthesis

RNA was collected 48h after transient transfections and from the stable transfection after 100% confluency was reached in a T75 flask. RNA was isolated using Purezol reagent (Bio-Rad) and cDNA synthesis was prepared according to protocol of the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) using random primers or a specific reverse primer for exon 13 of the mouse *TNNT2*.

RT-PCR and PCR

The expression of the GFP transcript under the *Tnnt2* promoters was assessed using the following primers: ForGFP: 5'-CACATGAAGCAGCACGACTT-3' and RevGFP: 5'-AGTTCACCTTGATGCCGTTC-3'.

In order to check transfection efficiency of the trans-splicing plasmids, PCR was performed with primers against the trans-splicing construct (ForAS7: 5'-CACTGGCTTGGCTTAGCGCCTGAAC-3' and RevInt8: 5'- AAGGGGTAACTGTGGCCTG AACCCA-3'). The same pair of primers was used in RT-PCR reaction to detect trans-splicing transcripts. A pair of primers was also used to detect the endogenous *Tnnt2* transcript levels as loading control (ForEx13: 5'-CGGAAGAGTGGGAAGAGACA-3' and RevEx14/15: 5'-CGCAGAACGTTGATTTCGTA-3'). All PCRs were performed using the NzyLong DNA Polymerase (Nzytech).

Taking advantage of a unique restriction site present in the human exon 9, trans-splicing occurrence was analysed by RT-PCR amplification of m*Tnnt2* Ex7-Ex9

fragment (ForEx7: 5'-AGGCTCTTCATGCCCAACTT-3' and RevEx9: 5'-GGTTCTGCCTTTCCTTCTC-3') followed by *Sac*I digestion. Alternatively, a degenerate and specific reverse primer against the human exon9 (RevhuEx9: 5'-CGAGCTCCTCC-3') was used to generate a PCR product in case of occurrence of trans-splicing events.

Fluorescence *in situ* hybridization (FISH)

Four FISH probes were designed against the endogenous m*Tnnt2* transcripts. For intron 7, one probe of 128bp was generated by PCR amplification using the following pair of primers ForI7: 5´-GGTGGAGCGAGCCAGGCTGG-3'/ RevI7T7: 5'-TAATACGACTCACTATAACTGGCTTGGCTTAGCGCCT-3'. For intron 8, two probes of 174bp and 165bp were made using the following pair of primers respectively ForI8-1: 5'-TGTGGGTGCTCCTTGCCAAC-3'/RevI8-1T7: 5'-TAATACGACTCACTATAACAAGCTGATCCCATGCTCC-3' and ForI8-2: 5'-AACCACCCATTTTTCTCAAC-3'/ RevI8-2T7: 5'-TAATACGACTCACTATACTAGAGGCTGTCTCAACTTT-3'. As a control, a probe for exon 8 was also generated, ForEx8: 5'-GACATCCACAGGAAGCGCGT-3'/ RevEx8T7: 5'-TAATACGACTCACTATAACAAGCTGATCCCATGCTCC-3'. Probes were labelled by *in vitro* transcription with digoxigenin-UTP (Sigma-Aldrich) or directly labelled with Alexa488-UTP (Thermo Scientific, Massachusetts, USA) using T7 RNA polymerase (Promega, Madison, USA). Probes directly labelled with digoxigenin-UTP were detected using an antibody anti-Dig D488 (Vector Laboratories, California USA).

Radioactive PCR using γ^{32}P-ATP labelled oligonucleotide

The primer RevhuEx9 was labelled using γ^{32}P –ATP nucleotide with activity 6000 Ci/mmol (PerkinElmer, Massachusetts, USA). The labelling was performed using 10U of T4 Polynucleotide Kinase (10 U/µL) (Thermo Scientific, Massachusetts, USA) for 30 minutes at 37ºC and another 10U added for another 30 minutes at 37ºC. 500ng of φX174 DNA *BsuR*I/*Hae*III (Thermo Scientific) were also labelled as described above. The reactions were purified by Sephadex G-25 column (GE Healthcare Life Sciences, UK) in Tris-EDTA (TE) buffer. PCR reactions were then performed using m*Tnnt2* ForEx7 and the γ^{32}P -labelled RevhuEx9 primer under the following conditions: 5min at 95ºC, 30 cycles of 30sec at 95, 1min at 57ºC, 1,5min at 68ºC and a final step of 68ºC for 10 min. All reactions were run on 15% acrylamide gel with 6M urea for 45min at 250V. The gels

were then fixed for 5min in a solution of glacial acetic acid/methanol/H_2O (10:20:70 V/V/V), dried for 2h at 50ºC, and exposed to the films (Agfa, Belgium) either at room temperature or at -80ºC for varying amounts of time.

3.3. Results

3.3.1. Trans-splicing strategy

In this study, we established two different trans-splicing strategies, both targeting the murine *Tnnt2* gene, using as a model mouse atrial cardiomyocyte cell line HL-1 [103]. The first strategy consists in a double trans-splicing (DTS) in which the trans-splicing molecule containing the human *TNNT2* exon 9 binds to both flanking introns of the homologous murine exon, and replaces it by trans-splicing (**Figure 3.8A**). The second strategy is a 3' trans-splicing approach, where the trans-splicing molecule contains the human *TNNT2* cDNA from exon 9 to 16 binds to murine *Tnnt2* intron 7 leading to the replacement of all downstream murine *TNNT2* exons (**Figure 3.8B**).
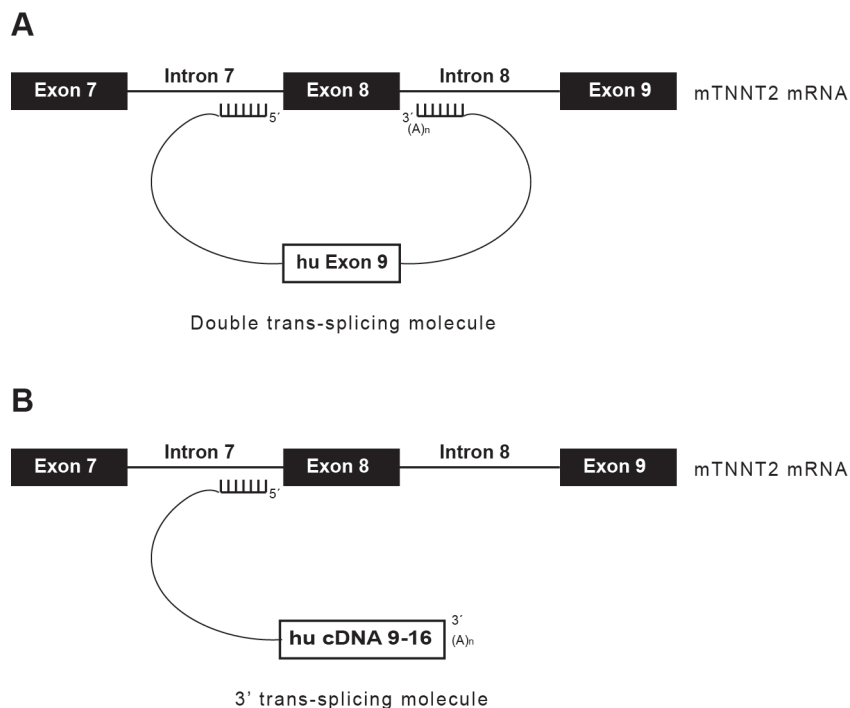
Figure 3.8 - Trans-splicing strategy.

(A) For the double trans-splicing (DTS) approach, the human trans-splicing molecule binds to both flanking introns of murine Tnnt2 exon 8 and, the latter one, is eventually replaced by its human equivalent. **(B)** In the 3' trans-splicing approach, the trans-splicing molecule binds to murine Tnnt2 intron 7 and, upon trans-splicing, all murine exons downstream of exon 8 are replaced by the human TNNT2 cDNA (Ex 9-16).

3.3.2. Selection of *TNNT2* promoters for cardiac expression at physiological levels

A minimal promoter of the cardiac troponin T gene (*Tnnt2*) has been shown to confer cardiac specificity in chicken and rat cardiomyocytes; indeed a fragment of c. around 300 nucleotides of the rat promoter was identified as the minimal sequence that allows cardiac-specific expression [100, 101]. The homologous (orthologous/similar) sequence was identified in the mouse genome. To test the expression of the minimal promoter in HL-1 cells, the entire and minimal *Tnnt2* promoters (**Figure 3.9A**) were cloned into pcDNA3 vector replacing the CMV promoter to drive the expression of a GFP reporter. The two plasmids were transiently transfected in HL-1 cells and RNA samples collected at 24h, 48h and 96h. GFP expression from both promoters had comparable levels and showed similar increase through time (**Figure 3.9B**). Since both promoters exhibited similar behaviour, the minimal promoter was chosen for subsequent experiments.
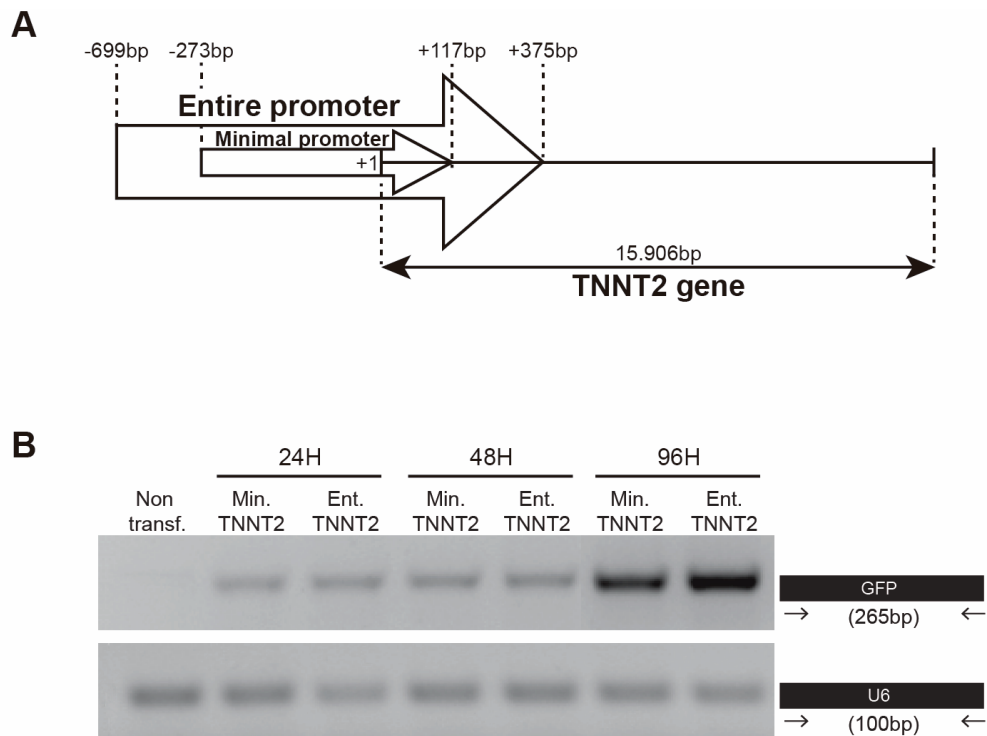
Figure 3.9 - *Tnnt2* promoters expression in HL-1 cells.

(A) The minimal Tnnt2 promoter is essential for cardiac specific transcription; its sequence is in the center of the entire Tnnt2 promoter, which include essential promoter features such as the TATA box. **(B)** Two different plasmids containing a GFP reporter and that are driven by the two versions of the Tnnt2 promoters were transfected in HL-1 cells and samples collected at different time points. The amplification of GFP transcripts was done using specific primers,showing an increased expression through time and efficient expression of the GFP transcript with both promoters. The U6 mRNA was used as a loading control.

Based on these results, both the double and the 3' trans-splicing constructs were cloned under the control of the minimal *Tnnt2* promoter or the CMV promoter. The CMV promoter-bearing constructs were used as a control, to ensure high expression levels of the constructs.

3.3.3. Trans-splicing molecule design: enhancers, annealing sequences and final assembly

The human sequence containing intron 8, exon 9 and intron 9 of the human *TNNT2* was chosen because of the existence of a cluster of pathogenic HCM mutations in exon 9; the few differences between the human exon 9 and the murine exon 8 allow

to determine the occurrence of the trans-splicing event by RT-PCR and restriction analysis.

Instead of amplifying by PCR the whole genomic sequence between the human introns 8 and 9, a study was performed using the Human Splicing Finder online tool, to identify and preserve all possible intronic enhancers alongside with the 3'splice site of intron 8 and 5´splice site of intron 9 [104].

The presence of an array of three muscle specific splicing enhancers, MSE_2 intronic sequences, flanking the exon on both sides, has been described as sufficient for the inclusion of that exon in the processed transcript of the *Tnnt2* gene [102]. Thus, for all constructs, three copies of intronic MSE_2s were inserted in each intronic sequence in order to increase trans-splicing efficiency (**Figure 3.10A, C**) [102].

To determine the best annealing sequences for the trans-splicing constructs, four FISH probes were designed to check the accessibility for annealing of different regions in the endogenous m*Tnnt2* transcripts. Due to the small size of the intron 7, only one probe of 128bp was designed, whereas for intron 8, two probes of 174bp and 165bp were designed. As a control, a probe for exon 8 was also generated. All probes were labelled by *in vitro* transcription with digoxigenin-UTP or directly labelled with Alexa488-UTP nucleotides.

FISH data on HL-1 cells show that both intronic regions flanking m*TNNT2* exon 8 are accessible for annealing with trans-splicing molecules (**Figure 3.10B**). The probe sequence nearest to exon 8 was chosen as the most specific for intron 8. Based on the same results for intron 7, the sequence of the probe was used as the annealing sequence in the trans-splicing constructs. So, the final constructs for the double trans-splicing approach were designed to possess these tested annealing sequences that target intron 7 and intron 8 of the endogenous mouse *Tnnt2* transcripts. Between these two annealing sequences, trans-splicing constructs bear human *TNNT2* exon 9 flanked by its own intronic sequences with the MSEs(3x) array and the essential features to allow the occurrence of trans-splicing (**Figure 3.10C**).

A

-432bp beginning of
promoter

```
mouse   GCTCCCAGCTTGCTCTTCTGGGCCTGGGCTCCTGGCATCTGCTTTATCGGGATTCTCAAG 1231
rat     GCTCCCAGCCTGCTCTTCTGGACCTGGGCTTCTGGCGTCTGCTTTATCGGGATTCTCAAG  782
        ********* ********** ******* ***** *********************

mouse   AGGGACAGCTGGTTTATGTTACAAGCCTGTTCCCTGCATATCTGCTCTGGTTTTAAATAG 1291
rat     AGGGACAGCTGGTTTATGTTACAAGCCTGTTCCCTGCATATCTGCTCTGGTTTTAAATAG  842
        ************************************************************

mouse   CTTTATCTGAGCAGCTGGAGGACCACATGAGCTTATATGGCGTGGGGTACTTGTTCTTTT 1351
rat     CTTTATCTGAGCAGCTGGAGGACCACATGAGCTTATATGGCGTGGGGTACTTGTTCCTTT  902
        ******************************************************** ***

mouse   AGCCCTGTGCCGGGCACCTGCCAAAATAGCAGCCAACACCCCCC--ATTGTGTTGTTCCC 1409
rat     AGCCCTGTGCCGGGCACCTGTCAAAATAGTAGCCAACACCCCCCCCCATTGTGTTGT---- 958
        ******************** ******* ** ************    **********

mouse   CCCCCCCCCCATCTCCTGCTGCACATTCCTCCCTCCGCGGGGCTTGGCTCACAAGGCCCC 1469
rat     CCCCCCCCCCATCTCCTGCTGCACATTCCTCCCACCGTGGGGCTTGGCTCACAAGGCCCC 1018
        ********************************* *** ***********************

mouse   AGCCCACATGCCTGCTTAAAGCTCTCCCCATGCCCTGCCCTCAGCCAGTCCCTGCTGAGG 1529
rat     AGCCCACATGCTTGCTTAAAGCTCTCCCCATGCCCTGCCCTCAGCCAGTCCCTGCTGAGG 1078
        *********** ************************************************
```

302 bp minimal
promoter

300 bp end of
promoter

B

C

Probe-C488/
DAPI

Probe-Dig/
anti-Dig D488/
DAPI

D

Double trans-splicing plasmid

3' trans-splicing plasmid

Figure 3.10 - Trans-splicing construct features.

(A) An array of three muscle- specific splicing enhancers (MSE) was cloned in each of the introns flanking the human exon 9, to favour trans-splicing occurrence. **(B)** Three antisense sequences designed to anneal to the two introns in the target pre-mRNA (flanking endogenous mTnnt2 exon 8) were cloned upstream and downstream of the trans-splicing molecule. Probes were either labelled directly with Dig or Alexa C488. The Dig-probes were then detected by an anti-Dig D488 antibody and nuclei stained with DAPI. **(C)** For the double trans-splicing approach, trans-splicing plasmids containing the human TNNT2 exon 9, flanked by its intronic sequences and enhancers, were constructed. In the 3´trans-splicing approach, trans-splicing constructs contain a human TNNT2 cDNA (Ex 9-Ex16). A single antisense sequence designed to anneal to intron 7 of murine Tnnt2 target pre-mRNA was further cloned upstream of the trans-splicing cDNA molecule.

In the 3´trans-splicing approach, similar constructs were generated but this time containing the human *TNNT2* cDNA from exon 9 to exon 16. Here, the 3´trans-splicing cassette contains a single antisense sequence that anneals to intron 7 of murine *Tnnt2*, cloned upstream of the trans-splicing cDNA molecule. (**Figure 3.10C**).

3.3.4. Detection of trans-splicing transcript and of trans-splicing events

Double (DTS) and 3' trans-splicing (3´TS) constructs were transfected into HL-1 cells in order to test their efficiency in inducing trans-splicing events. The success of trans-splicing was evaluated by the replacement of murine exon 8 in *Tnnt2* pre-mRNA by the human exon 9 via double trans-splicing or by the replacement of all murine exons downstream of exon 8 by the human cDNA from exon 9 to 16 via single 3' trans-splicing. The presence of trans-splicing constructs in HL-1 cells and their transcripts were detected by PCR and RT-PCR using primers ForAS7/RevInt8 (**Figure 3.11A, 3.11B, 3.12A**). All constructs were transiently or stably transfected with success as judged by PCR results (**Figure 3.11A, B** and **Figure 3.12A**). Moreover, regardless of the promoter used (*Tnnt2* or CMV), both DTS and 3'TS constructs were shown to express the therapeutic transcript (**Figure 3.11A, B** and **Figure 3.12A**). Detection of double trans-splicing events was performed by RT-PCR amplification of both endogenous cis-spliced and trans-spliced products using a pair of primers against the murine *Tnnt2* exons 7 and 9, which are flanking the trans-spliced exon 8. Trans-splicing products containing the human exon 9 differ in nucleotide composition, by existence of a unique restriction site for *Sac*I in human exon 9, which can then be used to distinguish the cis-spliced

product from the trans-spliced one. As a positive control, RT-PCR products amplified using the human cardiac cDNA as template were also digested with *Sac*I, to demonstrate the availability of this restriction site in the human *TNNT2* exon 9 (**Figure 3.11C**). However, the digestion was not as efficient as expected. In all transient and stable transfections tested, the restriction approach failed to detect occurrence of trans-splicing (**Figure 3.11C, 3.12C**). These results suggested the use of other detection techniques. In an alternative approach, a small degenerate reverse primer that only hybridizes with the human *TNNT2* exon 9 was used for the detection of double trans-splicing events. However, the trans-splicing-specific 160bp PCR band was only detected on the control human cardiac cDNA (**Figure 3.11D**). The detection of 3´ trans-splicing events was performed by using the same small degenerate primer (**Fig 3.12D**), yielding no positive results.
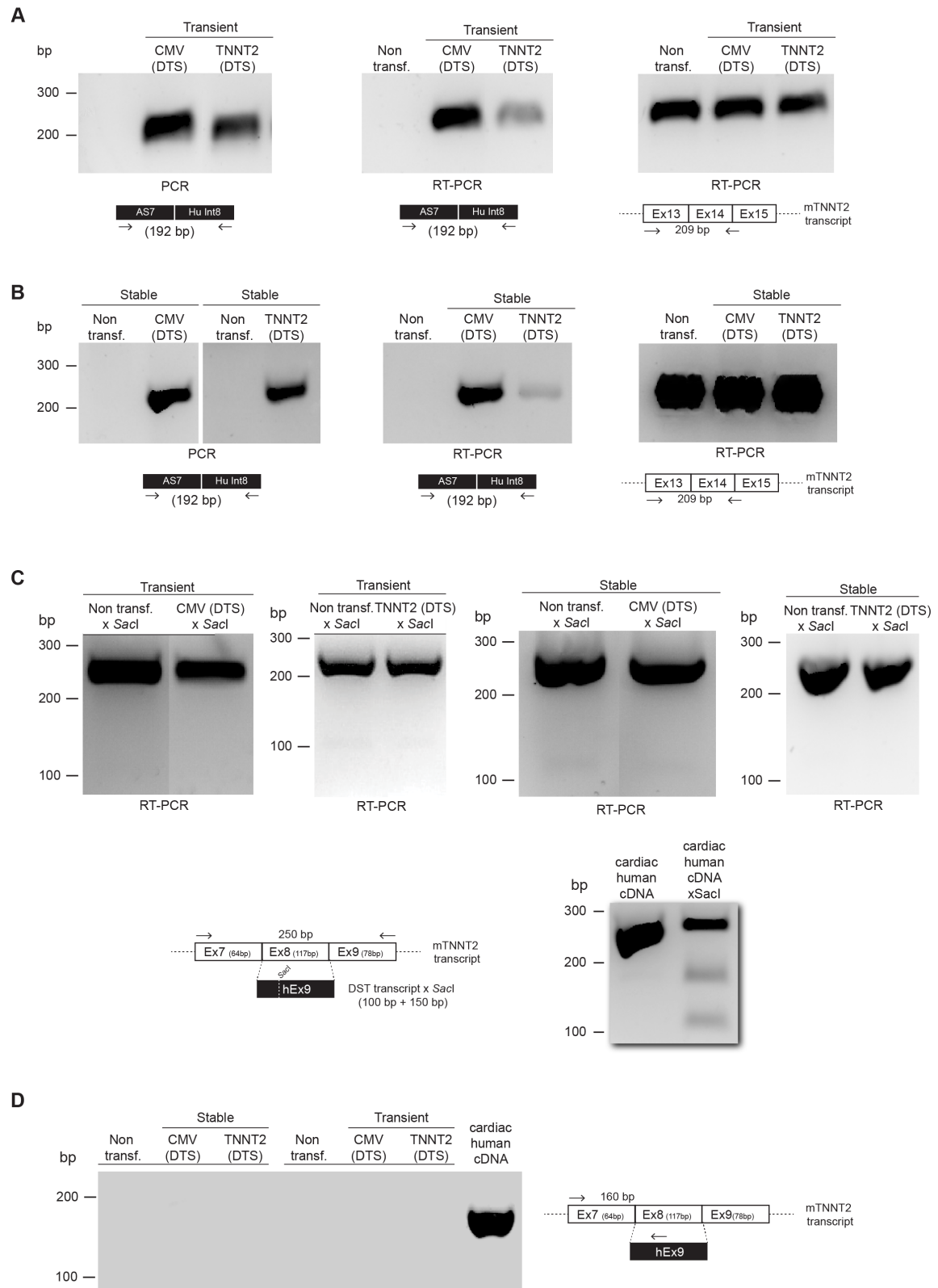
Figure 3.11 - Detection of double trans-splicing events.

(A) The efficiency of transient transfection and of transcription driven by either the CMV promoter or the minimal *Tnnt2* promoter was evaluated by PCR and RT-PCR, respectively, in HL-1 cells, using a pair of primers that allow specific amplification of double trans-splicing molecules. Levels of endogenous *Tnnt2*

transcripts were used as loading control. **(B)** The efficiency of stable transfection and transcription driven by both promoters was also evaluated, as previously described. **(C)** Detection of double trans-splicing events by RT-PCR amplification of both endogenous cis-spliced and trans-spliced products using a pair of primers against the murine *Tnnt2* exons 7 and 9. Amplified PCR products were further digested by the *Sac*I enzyme, with only trans-splicing PCR products being cleaved due to the presence of *Sac*I restriction sites, which only exists in the human TNNT2 exon 9. **(D)** Detection of double trans-splicing events was also performed by using a small degenerate reverse primer that only hybridizes with the human *TNNT2* exon 9. A human cardiac cDNA was used as a positive control (160bp band).
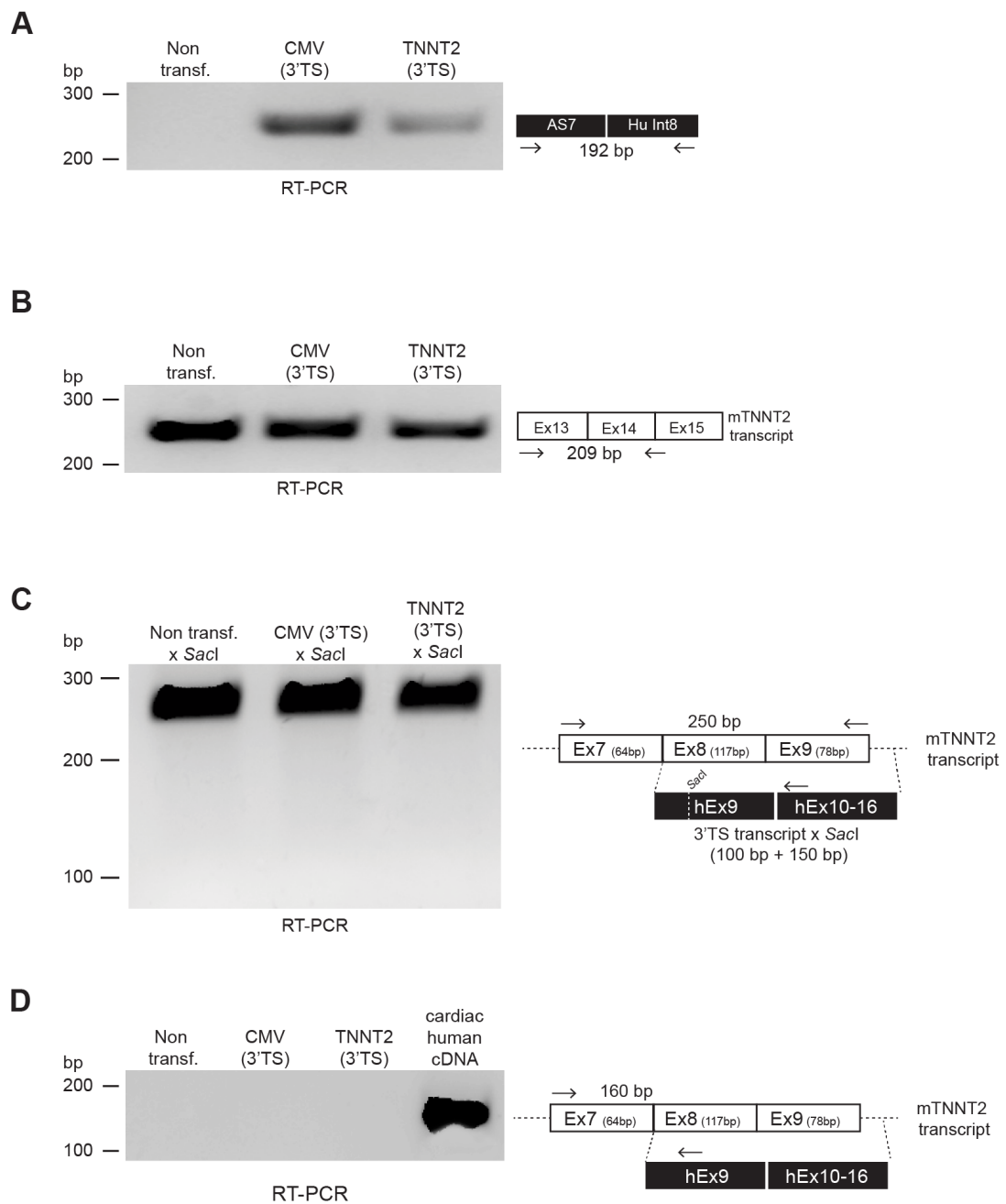
Figure 3.12 - Detection of 3' trans-splicing events.

(A) The efficiency of transcription driven by the CMV and by the endogenous murine *Tnnt2* promoter was evaluated by RT-PCR, in HL-1 cells, using a pair of primers that allow specific amplification of 3' trans-splicing molecules. **(B)** As loading control, levels of endogenous murine *Tnnt2* transcripts were also accessed. **(C)** Detection of 3' trans-splicing events by RT-PCR amplification of both endogenous cis-spliced and trans-spliced products using a pair of primers against the murine *Tnnt2* exons 7 and 9. Amplified PCR products were further digested by the *Sac*I enzyme, with only trans-splicing PCR products being cleaved due to the presence of *Sac*I restriction sites, which only exists in the human TNNT2 exon 9. **(D)** Detection of 3 trans-splicing events was also performed by using a small degenerate reverse primer that only hybridizes with the human *TNNT2* exon 9. A human cardiac cDNA was used as a positive control (160bp band).

To rule out the possibility that the lack of detection of trans-splicing events might be due to insufficient sensitivity of the techniques used, radioactive RT-PCR with a γ^{32}P-ATP radiolabelled primer was performed. This RT-PCR was performed with the same pair of primers ForEx7/RevhuEx9 as above, with γ^{32}P-labelled RevhuEx9. As expected, a 160-bp RT-PCR product was amplified from human cardiac cDNA; additionally, a band was also detected in the stably transfected cells with the double trans-splicing construct under the control of the CMV promoter (**Figure 3.13**), proving that trans-splicing of the mouse *Tnnt2* exon 8 using these constructs does occur.
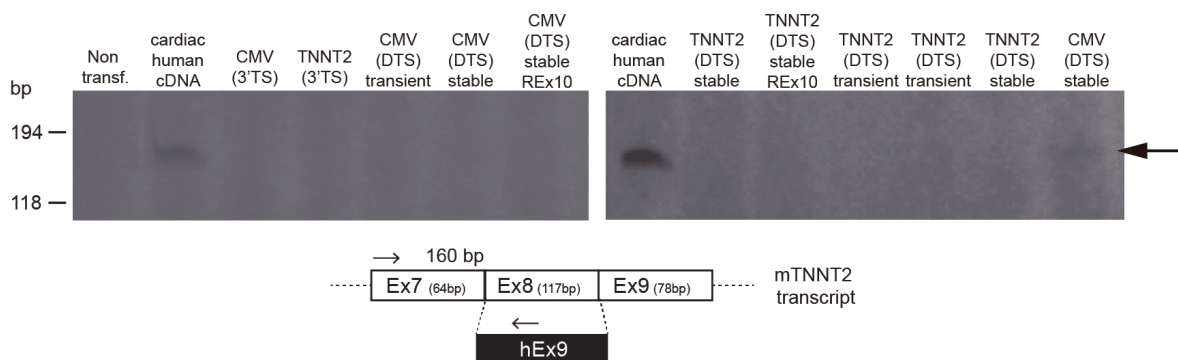
Figure 3.13 - Detection of trans-splicing events by radioactive RT-PCR.
Detection of trans-splicing events by using a small γP_{32}-ATP-labelled reverse primer that only hybridizes in the human exon 9. A human cardiac cDNA was used as a positive control (160bp band).

3.4. Discussion

In several previous studies, RNA trans-splicing was shown to be a promising therapeutic approach for autosomal dominant monogenic diseases, by allowing an increase in the levels of a healthy protein [43, 48, 49]. HCM is an autosomal dominant disease with low penetrance and variable clinical outcomes, ranging from gain of function to haploinsufficiency or malfunctioning protein [6, 8, 10]. In the case of *TNNT2* mutations, the presence of a poorly functioning protein with a dominant negative effect is usually associated to a myofibrillar disarray and changes in the Ca^{2+} regulation, leading to altered myocardial energetics and contractile capacity of the myocardium [26, 73, 74]. Henceforth, an RNA based-therapy is a very attractive approach to decrease the amount of deleterious proteins.

Previous attempts of using trans-splicing strategies in cell culture models with a minigene as target for trans-splicing showed variable efficiencies, ranging 30-60% *in vitro* and 1-10% *in vivo* for the case of adeno-associated viral vectors [49, 71, 72, 105]. The efficiency of trans-splicing also can vary depending on specific gene location. In theory, the 3' trans-splicing appears to be the most efficient strategy. The use of a binding domain closer to the 5' splice site also seems to increase trans-splicing efficiency due to close vicinity of the two splice sites (the endogenous and the construct one) [43, 48, 49, 71, 105]. Moreover, the use of muscle splicing enhancers, carefully chosen annealing sequences, two different promoters, as well as keeping the splicing enhancers of the human *TNNT2* intron 7 and 8 and placing the intron sequences between the annealing sequence and the coding region are all crucial for increasing the chances of trans-splicing. Still, the lack of detection of trans-splicing events with any of the described strategies either by restriction analysis, or by the use of a degenerate primer against the human exon 9 in a conventional RT-PCR clearly shows that there is not enough trans-spliced product to be detected by these techniques. Nevertheless, the detection of a very faint band by RT-PCR when using a radioactively labelled primer clearly demonstrates that trans-splicing events occur even with low efficiency.

One of the reasons for the observed low efficiency of trans-splicing events might be related to the difficulty of choosing the best annealing sequence for introns 7 and 8 of the mouse *Tnnt2* gene because of their small size. In other trans-splicing studies, several annealing regions were tested to choose the best candidate; for example, Koller et al [48] and Berger et al [49] tried to mask different splicing features whenever the intron size allowed to see the effect on trans-splicing [48, 49, 71, 105]. These features can determine the success or not of a given region to be a target for this strategy. In our case, the detection of trans-splicing was more difficult than in these studies because we used the endogenous gene instead of a minigene, to better reproduce physiological conditions.

Another reason for the low efficiency of trans-splicing might be due to endogenous factors that interact with this locus, such as for example, eventual presence of a natural antisense transcript in this particular region, if transcription occurs in the antisense strand it will interfere with sense strand. Additionally, binding of different proteins, such as splicing factors, might explain why the modulation of splicing did not work so efficiently either by trans-splicing or even by the use antisense oligonucleotides (ASO) [43, 48]. Three different ASO LNA (locked nucleic acid) were tested to induce skipping of endogenous mouse exon 8 in order to favour trans-splicing occurrence. Two of the ASOs targeted the 3'splice site of endogenous murine TNNT2 intron 7, whereas the other one target the 5'splice site of intron 8 of the same transcripts. Unfortunately, none of the ASOs was able to promote endogenous exon skipping in the preliminary assays, either alone or in different combinations of the three.

In the future, to increase efficiency of this approach, a new region within the murine *TNNT2* gene could be tested in order to target another intron; alternatively, modulation of different splicing enhancers/silencers may also be address. In fact, the success of trans-splicing is highly variable and has to be evaluated on a case-by-case basis.

In conclusion, these results argue that trans-splicing does not ensure efficient correction of expression of a mutated *TNNT2* gene in cardiac cells. This could be due to inefficiency of trans-splicing reactions in general, or a particular resistance to trans-splicing of the targeted region in the *TNNT2* pre-mRNA.

CHAPTER 4

RESULTS II: Whole gene sequencing identifies novel deep-intronic variants with potential functional impact in patients with HCM

4. RESULTS II: Whole gene sequencing identifies novel deep-intronic variants with potential functional impact in patients with HCM

4.1. Summary

Background High throughput sequencing technologies have revolutionized the identification of mutations responsible for genetic diseases such as hypertrophic cardiomyopathy (HCM). However, approximately 50% of individuals with a clinical diagnosis of HCM have no causal mutation identified. This may be due to the presence of pathogenic mutations located deep within the introns, which are not detected by conventional sequencing analysis restricted to exons and exon-intron boundaries.

Objective The aim of this study was to develop a whole-gene sequencing strategy to prioritize deep intronic variants that may play a role in HCM pathogenesis.

Methods and Results The full genomic DNA sequence of 26 genes previously associated with HCM was analysed in 16 unrelated patients. We identified likely pathogenic deep intronic variants in *VCL, PRKAG2* and *TTN* genes. These variants, which are predicted to act through disruption of either splicing or transcription factor binding sites, are 3-fold more frequent in our cohort of probands than in normal European populations. Moreover, we found a patient that is compound heterozygous for a splice site mutation in *MYBPC3* and the deep intronic *VCL* variant. Analysis of family members revealed that carriers of the *MYBPC3* mutation alone do not manifest the disease, while family members that are compound heterozygous are clinically affected.

Conclusion This study provides a framework for scrutinizing variation along the complete intronic sequence of HCM-associated genes and prioritizing candidates for mechanistic and functional analysis. Our data suggest that deep intronic variation contributes to HCM phenotype.

interpretation of the data (pedigree, transcription binding site assessment, splicing motifs and interpretation of the frequencies in the population) was done by myself. All the results were discussed with all the above responsible for the experimental design.

4.2. Material and Methods

Patients

The study population comprised 16 unrelated consecutively evaluated patients (8 males, 8 females) referred to the Cardiology Department at University Hospital Santa Maria. For all probands, the personal and family history, physical examination, ECG and echocardiography were consistent with a diagnosis of HCM according to international criteria [107]. Patients were genetically tested at a mean age of 49 years. In addition, family members of two selected probands were clinically and genetically tested. Before blood collection, all patients and relatives provided written informed consent for DNA analysis and received genetic counselling in accordance with guidelines [107]. DNA samples used in this study were residual after conventional diagnostic screening by targeted exome and Sanger sequencing. The project was approved by the Lisbon Academic Medical Center Ethics Committee.

Targeted gene enrichment and sequencing

Blood samples (5-8 mL) were collected into EDTA tubes at routine clinic visits, and DNA was isolated from peripheral blood lymphocytes using standard methods. The study was designed to screen the full genomic DNA sequence of 26 genes indicated in Table 1. These genes are included in many commercially available testing panels. A capture library was designed using SureSelect (Agilent) and target regions were sequenced (paired-end) on an Illumina HiSeq platform with 30-97 base read length. Highly repetitive sequences were excluded. Sample preparation was carried out as recommended by the manufacturer. Relatives were genotyped for selected variants by Sanger sequencing.

Bioinformatic data analysis

Raw sequencing paired-end reads (in *.fastq* format) were aligned using BWA software (version 0.7.12) [108] on the human reference genome (GRCh37) using quality score calibration and Illumina adapter trimming. Following the exclusion of duplicate reads using Picard MarkDuplicates tool (version 1.96) (http://broadinstitute.github.io/picard/), regions around insertion-deletions (indels)

were realigned and each base quality score was recalibrated. For variant calling, we used four distinct tools: GATK-UnifiedGenotyper (version 3.4-46) [109] and SAMtools mpileup (version 1.2) [110], which use alignment-based approaches, and GATK-HaplotypeCaller (version 3.4-46) [109] and FreeBayes (version 0.9.21.26) [111], which use haplotype-based approaches. By comparing the performance of each tool against a standard reference (NA12878, published by Genome in a Bottle consortium [112], we observed a concordance of ~85% (**Figure 4.1**). To take advantage of the strengths of the different tools, we selected variants that were independently called by at least two of them. This strategy showed a better sensitivity (~97%) and precision (~98%) compared to analysis using a single tool (**Figure 4.1**). Variants that were independently selected by at least two tools, and presented a read depth of 20 or more in the targeted genes, were annotated with ANNOVAR [113] (**Figure 4.2**).

For analysis of the clinical impact of coding variants we used the NCBI ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/) [75] and classified the variants according to the American College of Medical Genetics and Genomics (ACMG) guidelines [114]. Prediction of pathogenicity was performed with SIFT [115], PolyPhen2 HVAR [116], Human Splicing Finder (version 3.0) [104], Mutation taster [117], UMD-predictor [118], PROVEAN [119] and FATHMM [120]. Prioritization of noncoding variants was achieved using GWAVA (version 1.0) [121], CADD (version 1.3) [122], SPIDEX [123] and Genomiser [124]. To determine whether a variant may disrupt splicing motifs we used Human Splicing Finder (version 3.0), a tool that predicts potential splice sites, branch points and enhancer/silencer splicing motifs [104]; RegRNA (version 2.0), which searches for enhancer/silencer splicing motifs [125]; and Regulatory Genomics: Branch point analyser, that predicts the presence of branch points and respective polypyrimidine tracts [126].

As deep intronic mutations may result in altered gene expression through either cryptic splicing or disruption of transcription regulatory motifs [127], we investigated whether the identified intronic variants may disrupt transcription factor binding sites (TFBS). We used available tracks in the UCSC genome browser [128-130] (Transcription Factor ChIP-seq Uniform Peaks from ENCODE/Analysis and HMR Conserved Transcription Factor Binding Sites), focusing on TFBS predicted to be targets for

transcription factors that have been implicated in pathways related to cardiac regulation, development or pathophysiology.

Variant frequency was determined using the allele frequency estimates from the 1000 genomes project [131] and gnomAD [132] databases (accessed on June 2017).

Finally, we searched for the potential association of the candidate deep intronic variants with cardiac diseases identified through GWAS (https://www.genome.gov/gwastudies/index.cfm?gene=ESRRG and http://www.ebi.ac.uk/gwas/).
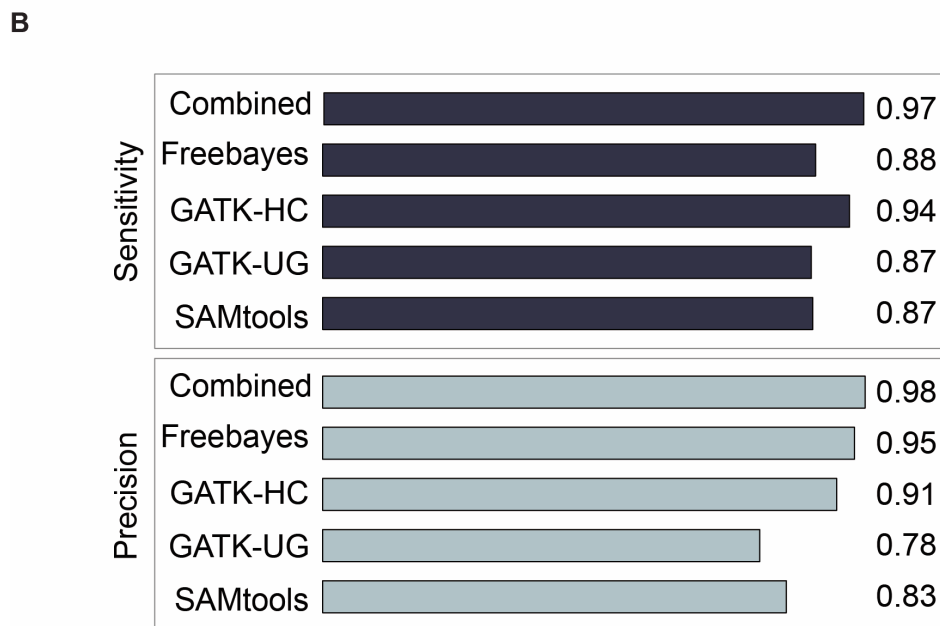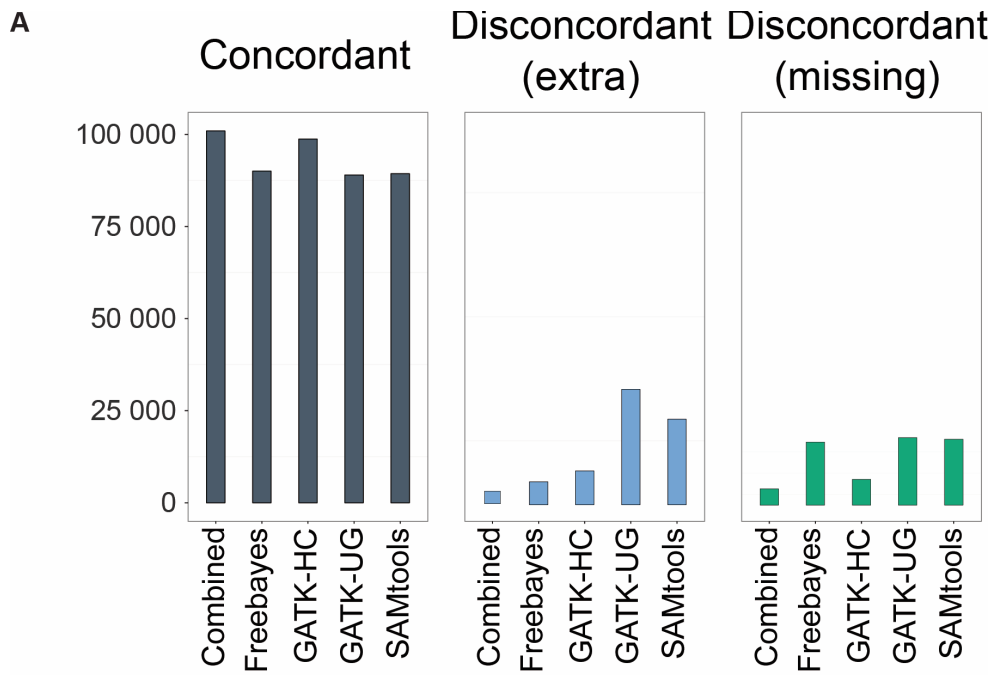
Figure 4.1. Comparison of variant calling strategies.

A) Variants identified by each individual tool and variants that were independently called by at least two tools (combined) were compared to a standard reference (NA12878, [112]). Concordant or true positive (TP) variants are defined as those present in the reference and identified by the indicated calling tool. Discordant extra or false positive (FP) variants are variants not detected in the reference but identified by the calling tool. Discordant missing or false negative (FN) variants are those present in the reference but undetected by the calling tools. **(B)** Sensitivity was assessed by calculating the ratio between TP/(TP+FN). Precision was assessed by calculating the ratio between TP/(TP+FP).
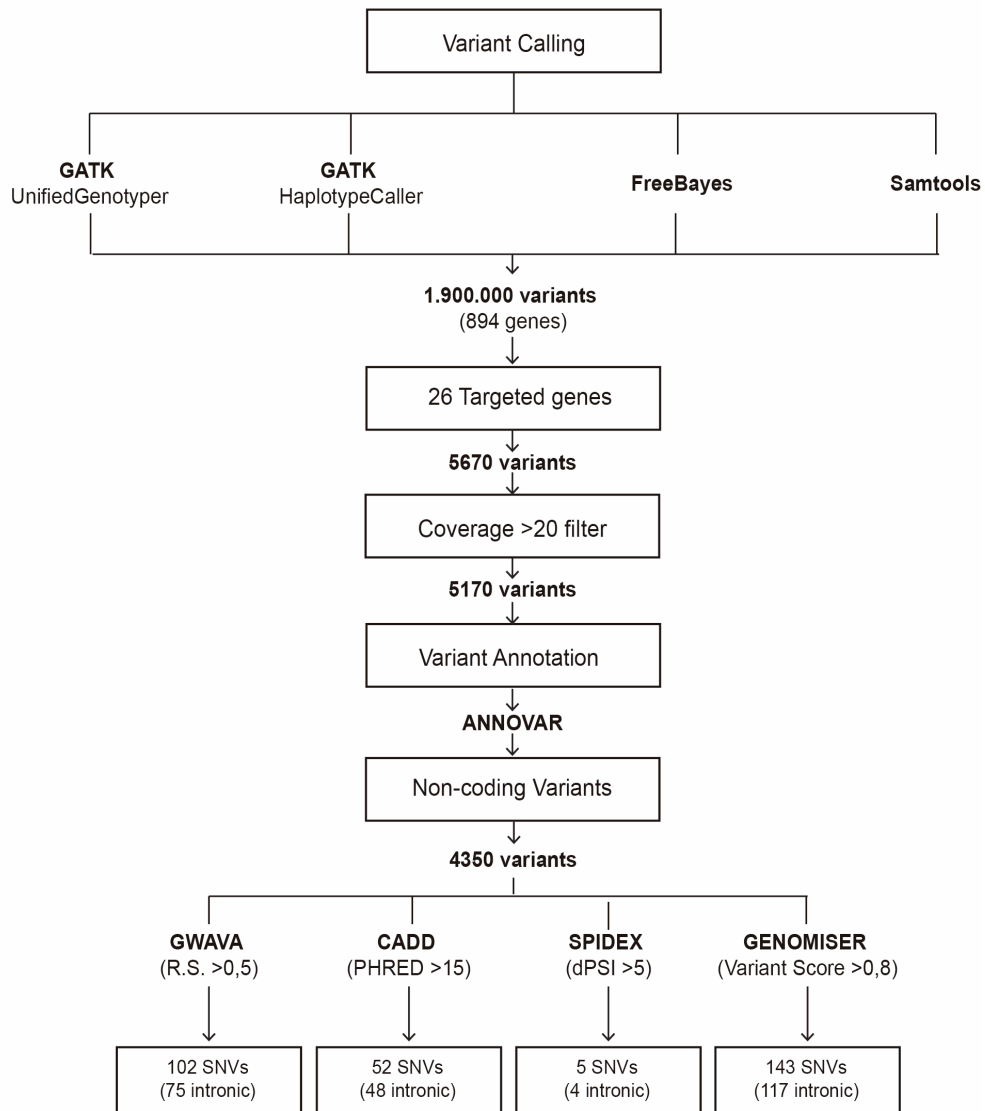
Figure 4.2 - Flowchart of noncoding data analysis.

Variants that were independently selected by at least two tools and presented a read depth of 20 or more in the 26 targeted genes were annotated with ANNOVAR. *In silico* predictions were carried out for noncoding variants that were not classified as either benign/likely benign or pathogenic/likely pathogenic in NCBI ClinVar. All variants with scores above the indicated threshold were single nucleotide substitutions (SNVs). GWAVA: Genome-Wide Annotation of Variants. CADD: Combined Annotation Dependent Depletion. SPIDEX: Splicing Index. R.S.: Region score. PHRED: phred quality score. dPSI: percent of spliced in.

4.3. Results

4.3.1. Quality of sequencing data

Analysis of sequencing data yielded an average of 96.64% confidently mapped reads per gene. For 69% of the targeted genes the average read depth was above 200, and for the remaining genes the average read depth ranged between 130 and 200 (**Figure 4.3A**). The average read depth was slightly lower over noncoding regions (**Figure 4.3B**). The average percentage of covered base pairs was higher than 90 for 85% of the genes, and the lowest coverage was 76% for both coding and noncoding regions (**Figure 4.3A, B**). Following alignment to the reference genome (GRCh37) and variant calling, we removed variants that were off-target, or had an average read depth below 20 (**Figure 4.2**). Single nucleotide substitutions and insertions or deletions of a few bases were identified and considered for further analysis.
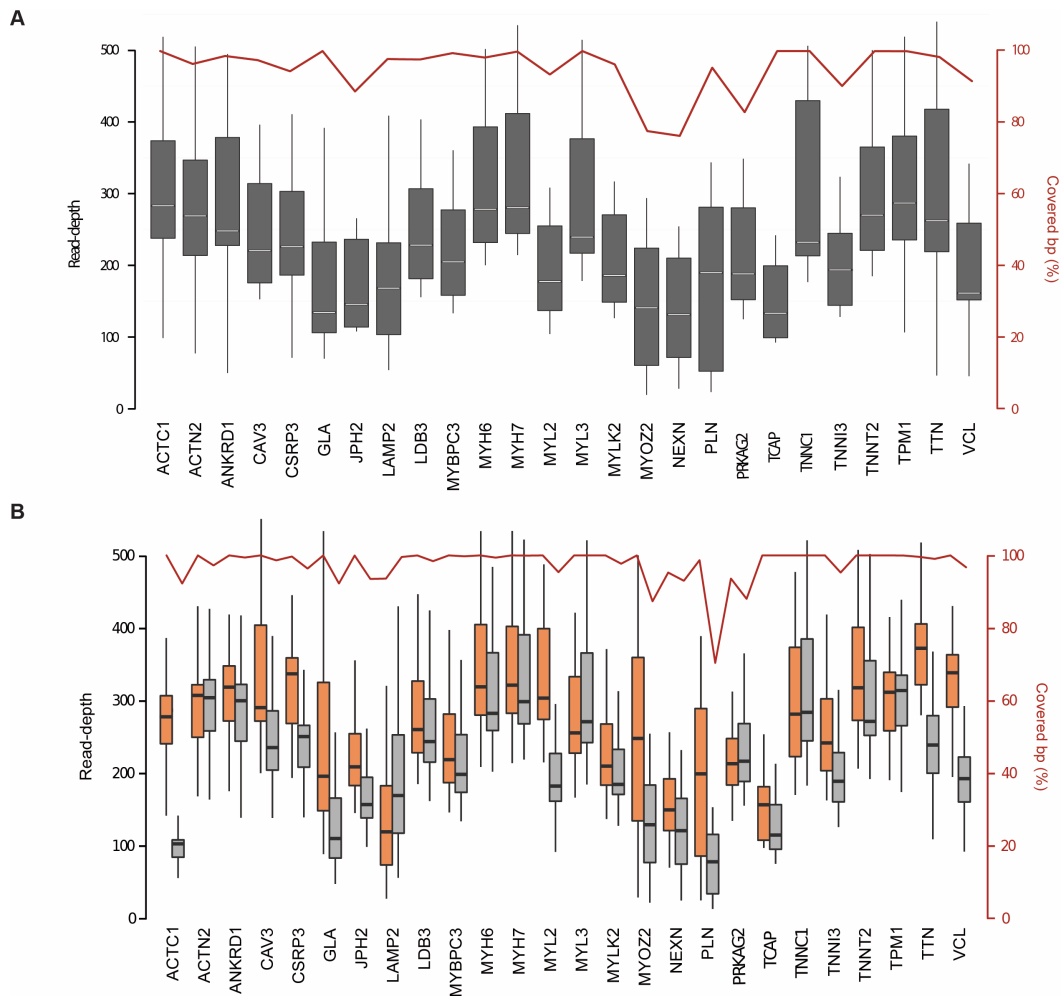
Figure 4.3 - Characterization of sequence data.

(A) Box plots show the read-depths across the targeted genes and the average percentage of covered base pairs per gene is depicted in red. **(B)** Box plots show the read-depths in coding (orange) and noncoding (grey) regions. The average percentage of covered base pairs in each region per gene is depicted in red.

4.3.2. Spectrum of exonic and splice site variants

Previously described disease-causing variants in the *MYBPC3* gene were detected in 3 patients (**Table 4.1**). Rare variants classified in the NCBI ClinVar database and according to the ACMG guidelines as of uncertain significance were additionally detected in the *TNNT2*, *MYBPC3*, *TTN*, *TPM1* and *MYH6* genes; all these scored as likely pathogenic according to multiple *in silico* prediction tools (**Table 4.1**). Noteworthy, one of the patients harboured, in addition to *MYH6* variant rs140596256, a novel variant in

the *GLA* gene that is not listed in online databases but is predicted to be pathogenic by multiple prediction tools (**Table 4.1**).

Table 4.1 - Putative HCM-causing variants located in exons and exon-intron boundaries.

VUS, variant of uncertain significance. ACMG, American College of Medical Genetics and Genomics.

Patient #	Exonic and splice-site variants	ClinVar/Reference	ACMG classification	*In silico* predictions
1	*TNNT2*: c.198G>C (p.Lys76Asn) Het (rs727504869)	ClinVar - VUS	VUS	UMD-predictor: probably pathogenic SIFT: deleterious PROVEAN: damaging PolyPhen2 HVAR: probably damaging FATHMM: damaging Mutation Taster: disease causing
2	*MYBPC3*: c.1224-19G>A Het (rs587776699)	ClinVar - Conflicting interpretations	VUS	Mutation Taster: disease causing Human Splicing Finder: activation of intronic cryptic acceptor site
5	*TTN*: c.57478C>G (p.Leu19160Val) Het (rs781121273)	ClinVar- VUS	VUS	UMD-predictor: probably pathogenic SIFT: deleterious FATHMM: damaging Mutation Taster: disease causing
6	*MYBPC3*: c.1227-13G>A Het (rs397515893)	[133]	Pathogenic (IC)	
7	*TPM1*: c.62G>T (p.Arg21Leu) Het (rs730881151)	ClinVar - Conflicting interpretations	VUS	UMD-predictor: pathogenic SIFT: deleterious FATHMM: damaging

				Mutation Taster: disease causing Human Splicing Finder: potential alteration of splicing
8	*MYBPC3*: c.2827C>T (p.Arg943*) Het (rs387907267)	[134] [135]	Pathogenic (ID)	
12	*TPM1*: c.841A>G (p.Met281Val) Het (rs397516394)	ClinVar - VUS	VUS	UMD-predictor: pathogenic SIFT: deleterious PROVEAN: damaging FATHMM: damaging Human Splicing Finder: potential alteration of splicing
14	*MYH6*: c.292G>A (p.Glu98Lys) Het (rs140596256) *GLA*: c.187T>A (p.Cys63Ser) Het	ClinVar - VUS [136]	VUS	UMD-predictor: pathogenic SIFT: deleterious PROVEAN: damaging PolyPhen2 HVAR: probably damaging FATHMM: damaging Human Splicing Finder: potential alteration of splicing UMD-predictor: pathogenic PolyPhen2 HVAR: probably damaging FATHMM: damaging

				Mutation Taster: disease causing
				Human Splicing Finder: potential alteration of splicing
15	*MYBPC3*: c.1484G>A (p.Arg495Gln) Het (rs200411226)	[134]	Likely pathogenic (III)	
16	*MYH6*: c.292G>A (p.Glu98Lys) Het (rs140596256)	ClinVar - VUS	VUS	UMD-predictor: pathogenic SIFT: deleterious PROVEAN: damaging PolyPhen2 HVAR: probably damaging FATHMM: damaging Human Splicing Finder: potential alteration of splicing

4.3.3. Assessment of deep intronic variants

The noncoding variants were prioritized using GWAVA [121], CADD [122], SPIDEX [123], and Genomiser [124]. The genome-wide annotation of variants (GWAVA) is a computational approach that integrates a wide range of available genomic and epigenomic annotations to predict the functional impact of variants. GWAVA results are in the range 0-1, with higher values indicating variants predicted as more likely to be functional. Variants with a GWAVA score above 0.5 were classified as functional, as in previous studies [121]. The Combined Annotation-Dependent Depletion (CADD) method provides a metric (C score) for deleteriousness, a property that strongly correlates with functionality and pathogenicity [122]. Variants at the top 10% of deleteriousness are assigned a C score of 10, whereas variants at the top 1% are assigned a C score of 20. Variants with C score greater than 15 were selected, as previously described [137]. SPIDEX is a computational model that uses the Percentage of Spliced-In (PSI) metric to evaluate whether a certain splicing isoform is more enriched under the presence/absence of a given variant. SPIDEX scores higher than 5 predict that the variant affects RNA splicing [123]. The Genomiser framework combines a machine learning method and an integrative algorithm for ranking noncoding variants in whole-genome sequence data [124]. Genomiser results are in the range 0-1, with values higher than 0.6-0.9 indicating variants more likely to be pathogenic [124].

We found that all noncoding variants with higher scores for the different prediction metrics correspond to single nucleotide substitutions, the vast majority of which are located within introns (**Figure 4.2**). The position of each variant relative to the nearest canonical splice site ranged between 20 and 2000 nucleotides (**Figure 4.4A**). A comparison of variants prioritized as likely pathogenic by the different tools resulted in the identification of 6 variants that scored with high values using GWAVA, CADD and Genomiser metrics (**Figure 4.4B**). These include two variants in *VCL*, two variants in *TTN,* one variant in *ACTC1* and one variant in *PRKAG2* (**Table 4.2**). Analysis of allele frequency data available for European populations in the 1000 Genomes project [131] and gnomAD [132] databases reveals that two of these variants are more frequent in the patient population than in healthy individuals (**Table 4.2**). Namely, the *VCL* variant c.499+367T>C (rs113195070) was detected in 6 probands, corresponding to

an allele frequency of 19% in the patient population. This contrasts with a frequency of 6-7% in control populations. Similarly, the *PRKAG2* variant c.1234-317T>G was present with an allele frequency of 3% in the patient population contrasting with a frequency of 0.1-0.3% in healthy individuals. Such specific enrichment of certain deep intronic variants in the patient population suggests that these may be contributing to the disease phenotype.
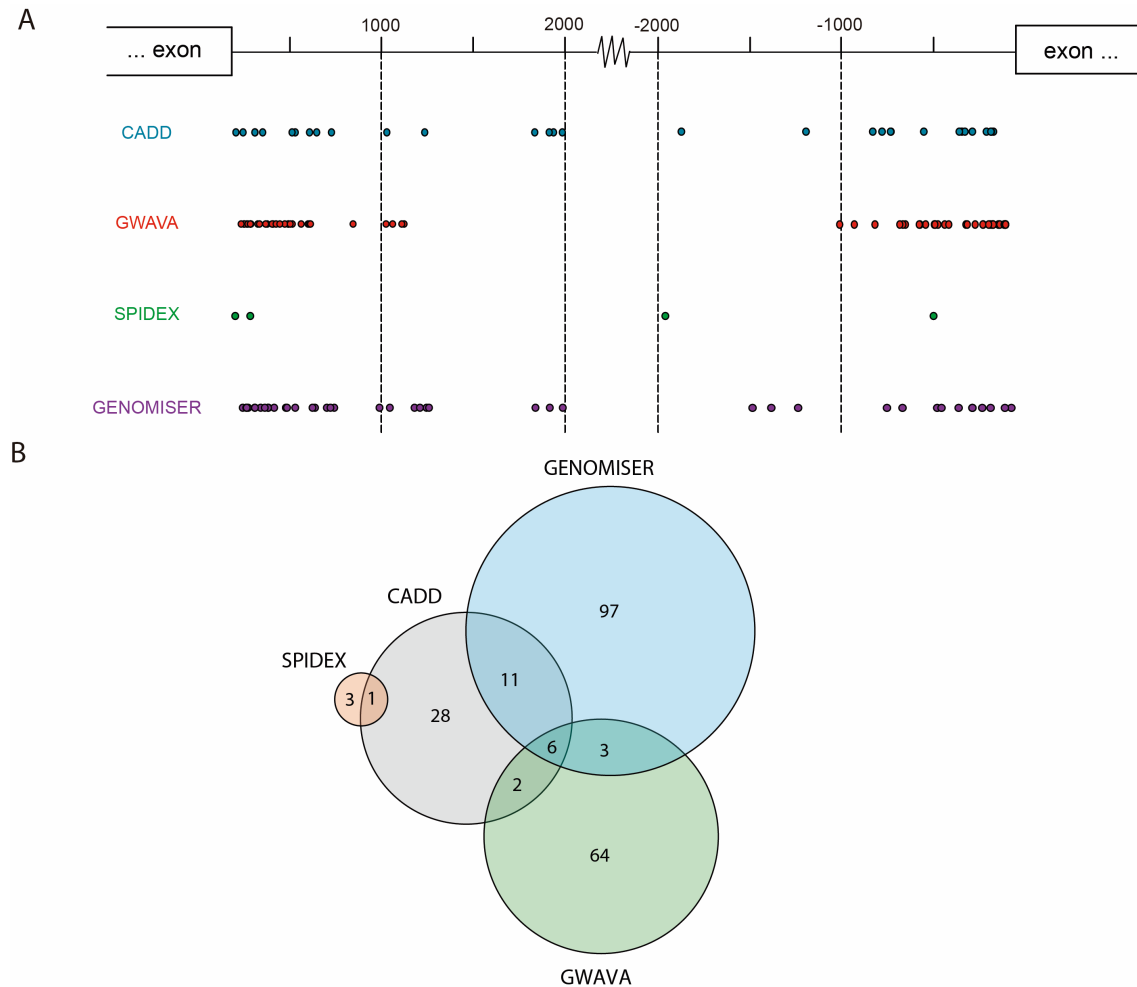
Figure 4.4 - Assessment of intronic variants.

(A) Schematic diagram depicting the position of intronic variants prioritized by each prediction tool. **(B)** Venn diagram illustrating intronic variants that are simultaneously prioritized by multiple prediction tools.

Table 4.2 - Prioritized intronic variants.

For each variant, minor allele frequency (MAF) was determined using European populations in the 1000 genomes project database [131] and gnomAD database [132].

CADD Phred score [122]; GWAVA Region score [121]; Genomiser Variant score [124]; SPIDEX dPSI score [123]. HGVS, Human Genome Variation Society.

Tools	Gene symbol	dbSNP ID	HGVS	CADD score	GWAVA score	GENOMISER score	SPIDEX score	MAF 1000G	MAF gnomAD	Frequency in probands
CADD, GWAVA, GENOMISER	VCL	rs77884406	c.169-2410A>C	17.55	0.59	0.910891	NA	0.019	0.0273	0.03125
	VCL	rs113195070	c.499+367T>C	16.93	0.6	0.963367	NA	0.06	0.0671	0.1875
	ACTC1	rs28595759	c.129+472T>C	21.9	0.53	0.9792082	NA	0.07	0.0604	0.03125
	TTN	rs2243452	c.32929+72T>C	22.2	0.53	0.939604	0.6211	0.029	0.0247	0.03125
	TTN	rs2253324	c.10361-138C>T	18.03	0.51	0.825743	3.3164	0.048	0.0441	0.03125
	PRKAG2	rs141541040	c.1234-317T>G	15.20	0.58	0.872277	NA	0.003	0.0010	0.03125
CADD, GENOMISER	VCL	rs7079796	c.168+1165C>T	15.74	0.35	0.812872	NA	0.2	0.2191	0.15625
	VCL	-	c.169-7572C>T	15.88	NA	0.905941	NA	.	.	0.03125
	VCL	-	c.239+4299C>A	18.78	NA	0.983169	NA	.	0.0001	0.0625
	LDB3	rs12570315	c.93+1827G>A	16.75	0.23	0.858416	NA	0.3	0.3436	0.25
	LDB3	-	c.548+1914C>T	21.4	NA	0.970297	NA	.	.	0.03125
	LDB3	rs779483568	c.548+1993C>T	17.35	NA	0.990099	NA	.	0.0001	0.03125
	MYL2	rs2040571	c.3+604C>T	15.58	0.36	0.89703	NA	0.086	0.0888	0.03125
	PRKAG2	rs62478182	c.467-44847T>G	17.95	0.26	0.880198	NA	0.34	0.3721	0.25
	PRKAG2	rs114394151	c.115-30242C>T	18.38	0.49	0.925743	NA	.	0	0.0625
	LAMP2	rs5956217	c.1094-2886A>G	15.93	0.42	0.881188	NA	0.004	0.0009	0.03125

74

Method	Gene	SNP	Variant							
	LAMP2	rs42887	c.1094-2924C>T	20.4	0.43	0.929703	NA	0.11	0.1711	0.125
CADD, GWAVA	*TTN*	rs12693162	c.37112-700G>A	18.66	0.5	0.190099	NA	0.22	0.2261	0.21875
	LAMP2	rs141348126	c.1094-140A>G	15.68	0.5	0.545545	NA	.	0	0.03125
GWAVA, GENOMISER	*VCL*	rs2131959	c.2132-437G>C	10.13	0.57	0.89604	NA	0.75	0.7445	0.84375
	ANKRD1	rs10509614	c.207+239G>T	13.87	0.52	0.838614	0.7561	0.04	0.0314	0.03125
	TTN	rs80259697	c.10360+317T>C	13.04	0.51	0.821782	NA	.	6.68e-05	0.03125
CADD, SPIDEX	*TTN*	rs142156368	c.31484-286G>T	15.63	0.38	0.425	6.092	0.0089	0.0047	0.0625
SPIDEX	*TTN*	rs2562845	c.32593+111A>G	3.232	0.29	0	9.1111	0.21	0.2015	0.15625
	TTN	rs72650063	c.32077+31C>G	0.713	0.41	0.019802	5.8489	0.021	0.0218	0.0625
	TTN	rs2742353	c.31484+1715A>C	10.79	0.26	0.556436	6.2975	0.029	0.0247	0.03125

We focused on the *VCL* variant c.499+367T>C. We found that two of the probands were compound heterozygous for this variant and a *MYBPC3* mutation previously described as disease-causing. Genotyping of family members of proband #6 showed that the dual presence of the *MYBPC3* splice site mutation (c.1227-13G>A) and the *VCL* variant is associated with the manifestation of the phenotype in the proband (I-2) and his son (II-1), both diagnosed in their 40s (**Figure 4.4**). The other children of the proband (II-2 and II-3), while carrying the *MYBPC3* mutation, did not develop signs of cardiomyopathy when assessed at a similar age. This suggests a possible modifier effect of the *VCL* variant, since the presence of the *MYBPC3* mutation alone is not sufficient for the phenotype to be manifested. Analysis of this family further suggested that the *VCL* variant on its own is not sufficient to cause disease. Genotyping of family members of proband #15 (**Figure 4.5**) indicated that presence of the *MYBPC3* missense mutation (p.Arg495Gln) in the absence of the *VCL* variant appears sufficient to cause the phenotype, even at a pediatric age (III-1). Clinical characteristics for the two families are detailed in **Table 4.3**.
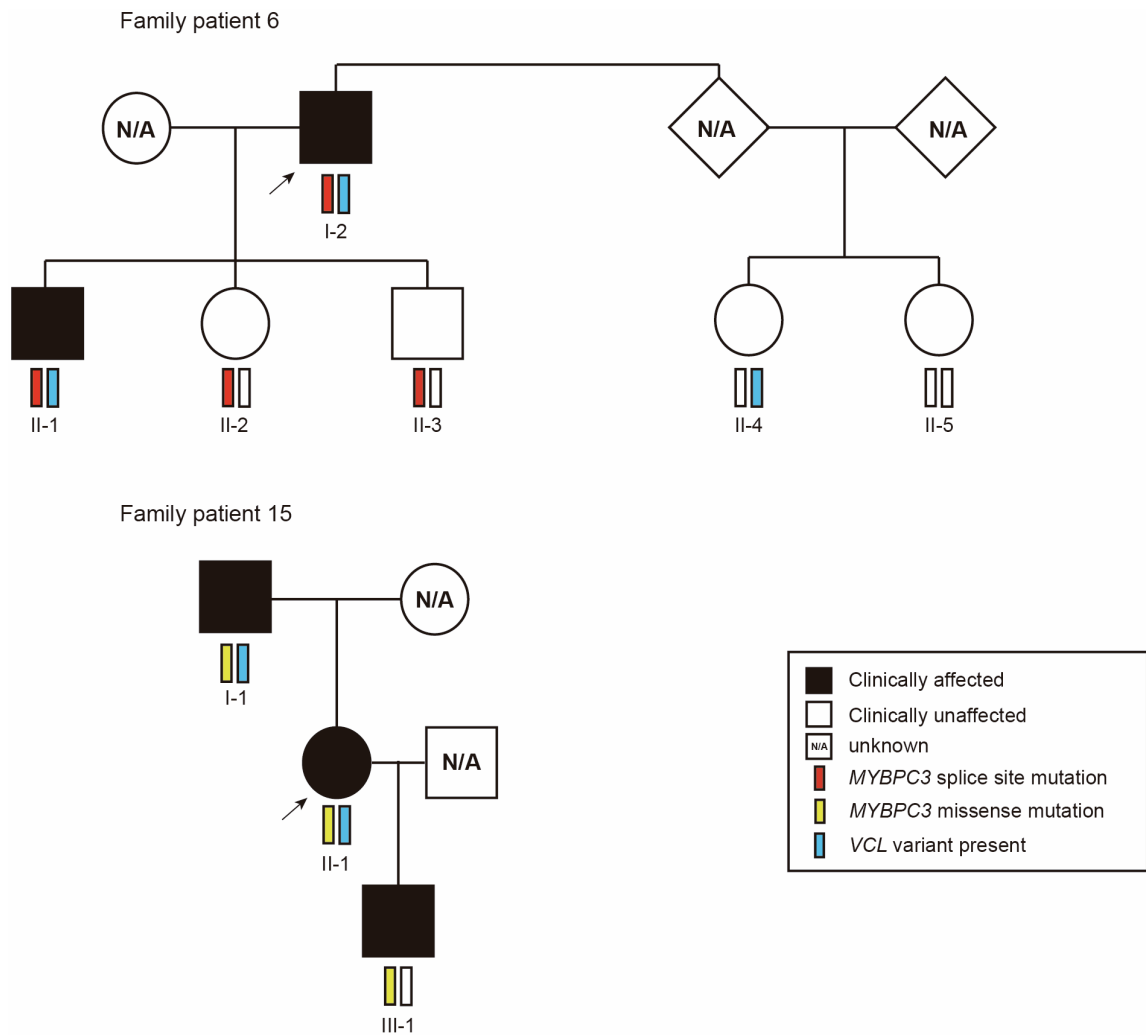
Figure 4.5 - Family pedigrees.

The *MYBPC3* splice site (c.1227-13G>A Het) and missense (p.Arg495Gln) mutations, and the *VCL* variant (c.499+367T>C) identified in probands (arrows) were studied in family members, and their clinical status was ascertained. Circles denote females, squares males, solid symbols clinically affected individuals, open symbols clinically unaffected individuals, and NA unknown clinical status.

Table 4.3 - Clinical, electrocardiographic and echocardiographic data for the proband and relatives of families 6 and 15 (pedigrees are illustrated in figure 3).

Legend: G+= carrier of the causative mutation in the *MYBPC3* gene; G-= not carrier of the causative mutation; Ph+= positive phenotype; Ph-= negative phenotype; Hypertrophic cardiomyopathy (HCM); *diagnosis by familial screening (not symptoms);**symptoms related to HCM; ASH= asymmetrical septal hypertrophy; LVH= left ventricular hypertrophy; & maximal wall thickness (WT) in any left ventricular segment; LA= left atrial dimension (M-mode echocardiography); LVDD= left ventricular diastolic diameter; LVSD= left ventricular systolic dimension; FS= fractional shortening of the left ventricle; LVEF= left ventricular ejection fraction (Simpson method); OB= presence of left ventricular obstruction at rest (left ventricular outflow gradient ≥ 30 mmHg on Doppler evaluation); £ - left ventricular hypertrophy considering pediatric criteria for HCM; Vs= Velocity of the mitral annulus (lateral) by Tissue Doppler imaging (TDI); E/e= ratio of early diastolic velocity of mitral inflow to early diastolic velocity of the mitral annulus (lateral) by TDI.

	Status	Age at clinical diagnosis or genetic testing (y)	Symptoms** (Yes/No)	Abnormal ECG (Yes/no)	LVH (Yes/No)	Maximal WT (mm)	Type of LVH	OB	LA (mm)	LVDD (mm)	LVSD (mm	FS (%)	LVEF	Vs´ (lateral; cm/s)	E/e' (lateral; cm/s)
Family 6															
I-2	G+/Ph+	49	Yes	Yes	Yes	30	ASH	Yes	48	40	20	50	82	4	6.85
II-1	G+/Ph+	44*	No	Yes	Yes	16	ASH	No	32	44	23	57	87	9	5.9
II-2	G+/Ph-	42*	No	No	No	11	-	-	33	44	23.6	46	76	14	4.5
II-3	G+/Ph-	40*	No	No	No	12	-	-	41	48	25.6	47	78	13	6.18
II-4	G-/Ph-	43	-	-	-	9	-	-	28	43	21.6	52	83	10	6.6
II-5	G-/Ph-	40	-	-	-	11	-	-	32	41	22	46	78	12	5.33

Family 15															
II-1	G+/Ph+	20	Yes	Yes	Yes	19.8	ASH	No	41	39	21.35	44	66	10	7.38
I-1	G+/Ph+	21	Yes	Yes	Yes	15	ASH	No	46	42	22	48	62	9	6.9
III-1	G+/Ph+	3*	No	Yes	Yes	8.8 £	ASH	No	22.34	32	19.7	40	72	7	6.63

We found that, based on Chip-seq experiments [128, 129], the deep intronic *VCL* variant enriched in the patient population (rs113195070) localizes in a region associated with FOS, JUN and EP300 (**Figure 4.6A**). A deep intronic variant in the *PRKAG2* gene (rs114394151) prioritized by CADD and Genomiser is enriched in the patient population (**Table 4.2**) and localizes in a region associated with FOS and JUN (**Figure 4.6B**). FOS and JUN transcription factors are thought to be among the first set of genes to be expressed in the context of pathological cardiac hypertrophy [138], and EP300 has been associated with cardiomyocyte enlargement [139].

Finally, using the SPIDEX tool we identified a variant in the *TTN* gene (rs72650063) that occurs with a frequency of 2% in control European populations and is present in two probands, corresponding to an allele frequency of 6% in the patient population (**Table 4.2**). This variant is predicted by Human Splicing Finder to disrupt binding of splicing factor SRSF1 (**Figure 4.6C**) [140]. Another variant in the *TTN* gene (rs142156368) appears highly enriched in our cohort relative to the general population (**Table 4.2**).

No potential association of the candidate deep intronic variants with cardiac diseases identified through GWAS was found.
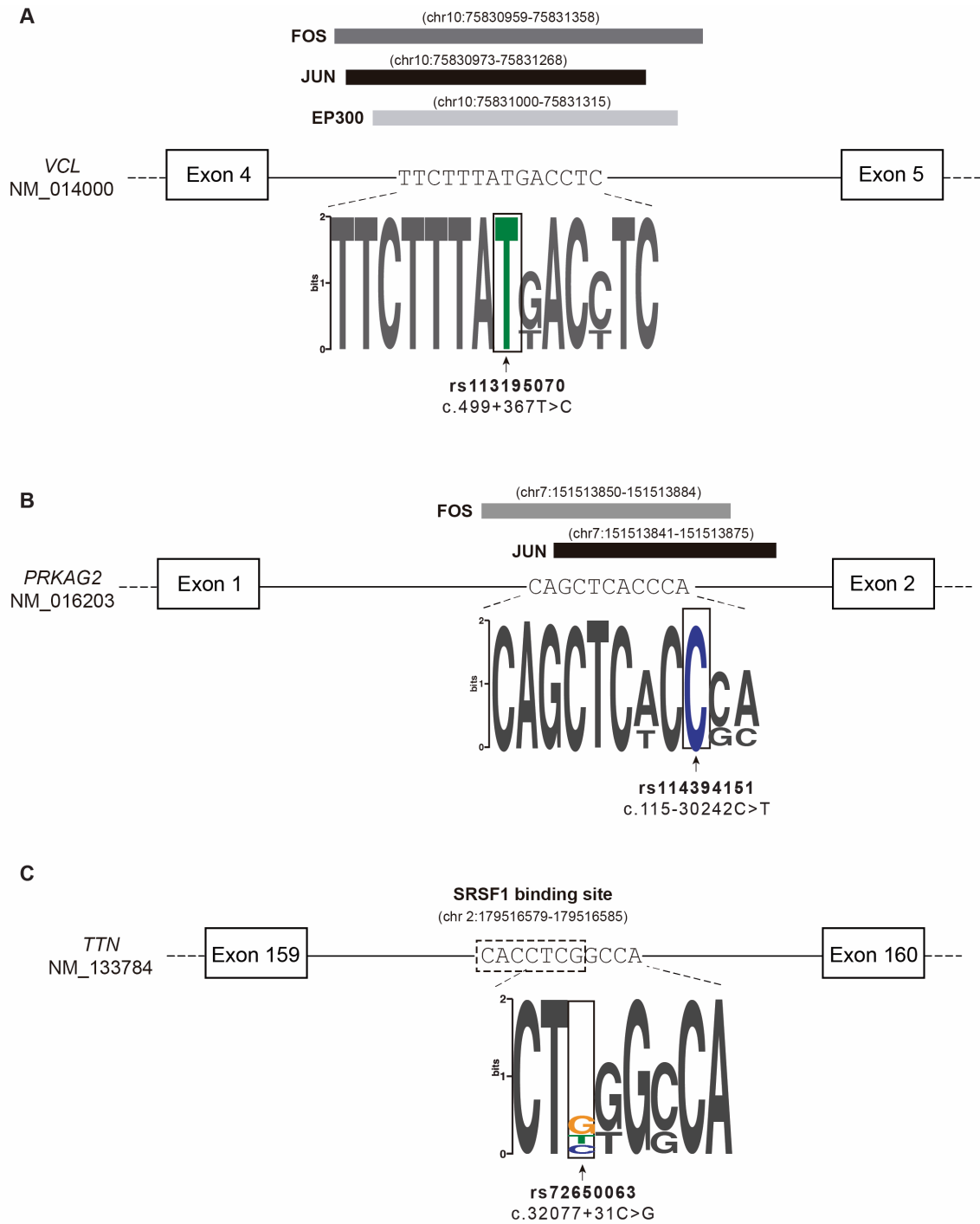
Figure 4.6 - Variants located at binding sites for transcription and splicing factors.

(A) The *VCL* variant c.499+367T>C (rs113195070) is located at a binding site for transcription factors FOS, JUN and EP300. **(B)** The *PRKAG2* variant c.115-30242C>T (rs114394151) is located at a binding site for transcription factors FOS and JUN. **(C)** The *TTN* variant c.32077+31C>G (rs72650063) is predicted to disrupt the binding site of splicing factor SRSF1.

4.4. Discussion

Motivated by the clinical heterogeneity of HCM and the lack of a conclusive genetic diagnosis in approximately 50% of the patients [141-143], we hypothesized that genetic variation within deep intronic regions of sarcomere and sarcomere-related genes contributes to the disease mechanism. Using a targeted high-throughput sequencing strategy, we did a comprehensive screening of 26 genes in a cohort of 16 unrelated HCM patients using recently developed computational models to assess variants.

We identified 3 probands carrying previously described disease-causing variants in the *MYBPC3* gene [133-135, 144], in agreement with the finding that mutations in this gene account for the great majority (30-40%) of identified genetic causes of HCM [145, 146]. We further identified additional probands harbouring rare coding variants of uncertain significance, that are likely pathogenic as assessed by multiple prediction tools, in *TNNT2*, *MYBPC3*, *TPM1*, *TTN* and *MYH6* genes. In one patient we identified a novel variant in the *GLA* gene, associated with Fabry disease [136].

In agreement with previous studies on larger HCM patient cohorts [134, 147], we identified missense and stop-codon mutations, as well as splice site mutations (**Table 4.1**). While missense mutations may code for a pathogenic protein that would cause HCM by a gain-of-function mechanism, stop-codon and splice site mutations are more likely to act through a loss-of-function mechanism due to reduced levels of the normal protein. Demonstrating that particular mutations act through a dominant gain-of-function mechanism and others cause loss-of-function would be critical to understand phenotype-genotype correlations.

In this regard, we analysed the family members of two probands with either a missense or a splice site mutation in *MYBPC3* (**Figure 4.5**) and found that all carriers of the missense mutation (family #15) were clinically affected, as expected for a dominant gain-of-function mechanism. In contrast, in the other family, two individuals are carriers of the splice site mutation (family #6) and do not manifest the disease, consistent with a loss-of-function model. Possibly, the *MYBPC3* splice site mutation present in family members of proband #6 results in mis-splicing by either decreasing the specificity or fidelity of splice site selection or activating cryptic splice sites that are

normally not used. Abnormal splicing often results in a frameshift and consequent introduction of premature termination codons (PTCs), which trigger degradation of the mRNA by nonsense-mediated decay [148]. Thus, this mutation can be functionally equivalent to a null or hypomorphic allele associated with loss-of-function of the protein. In contrast, the *MYBPC3* missense mutation present in family members of proband #15 presumably leads to an abnormal protein containing an amino acid substitution that may cause a gain-of-function phenotype.

We further show that both probands are compound heterozygous for the missense or splice site *MYBPC3* mutation and a deep intronic variant in *VCL*. This variant (rs113195070) is predicted to be pathogenic based on three independent computational tools, GWAVA, CADD and Genomiser (**Table 4.2**). Moreover, it is 3-fold more frequent in our cohort of probands than in normal European populations (**Table 4.2**), further suggesting a direct contribution to the disease phenotype. The variant, which consists of a single nucleotide substitution located at position 367 from the nearest canonical splice site (c.499+367T>C), can potentially disrupt the binding of transcription factors that have been reported as implicated in pathways related to cardiac regulation, development or pathophysiology such as FOS, JUN and EP300 [138, 139]. By interfering with the binding of transcription regulatory factors, the variant is expected to alter the transcription rate of the *VCL* gene. Consistent with this view, sequence elements located within introns of large human genes have been shown to act as transcriptional enhancers [149], and a recent study reported an *IRF4* gene variant located in intron 4 that strongly affects *IRF4* transcription through disruption of an enhancer element [150].

Analysis of family #6 reveals that the presence of the *VCL* variant or the *MYBPC3* mutation in isolation is not sufficient to cause disease phenotype. Indeed, the two clinically affected individuals in this family are compound heterozygous for the *VCL* variant and the *MYBPC3* splice site mutation (**Figure 4.5**), suggesting that the combination of the two mutations triggers the disease. A loss-of-function mechanism for the *MYBPC3* mutation could explain why in family #6 only compound heterozygous members manifest the disease, whereas the presence of the heterozygous gain-of-function mutation in family #15 would be sufficient to cause disease. Complex genotypes, including individuals that carry 2 or more variants in the same or different

sarcomere-related genes, have been reported in 8% of HCM patients [133], and there is evidence indicating that patients with complex genotype and multiple simultaneous mutations may have more severe or early disease expression [151]. However, complex genotype-phenotype correlations focusing specifically on carriers of splice site mutations remain to be investigated.

We further identified two single nucleotide substitutions in the titin gene (rs142156368 and rs72650063) that are 3 to 6-fold more frequent in our cohort of probands than in normal European populations (**Table 4.2**). These variants are located in the PEVK domain that plays a role in extensibility of the sarcomere and contractility of the titin protein [152, 153]. Titin is prone to extensive alternative splicing that can change its size and its elastic/stiffness properties; associations have been established between the ratio of expression levels for the main cardiac isoforms (N2BA and N2B) and genetic and non-genetic forms of cardiac diseases [154, 155]. If these variants do interfere with titin splicing, as predicted by the SPIDEX computational model, they are likely to contribute to HCM phenotype, particularly in combination with other HCM-associated alleles. Supporting this view, titin-truncating splicing isoforms, which are encountered in approximately 1% of the general population are sufficient to induce molecular and physiological effects on the heart [156].

In conclusion, this study provides a framework for scrutinizing variation along the complete sequence of HCM-associated genes and prioritizing candidates for further analysis. Our data suggest that deep intronic variation contributes to HCM phenotype. Translation of genetic information found in an individual to clinical decision taking requires a precise understanding of the molecular mechanisms underlying the disease phenotype. To date, mechanistic and functional studies have been largely restricted to animal models in part due to difficulties in obtaining human tissue from patients. However, the recent emergence of patient-derived induced pluripotent stem cells (iPSCs) that can be differentiated into functional cardiomyocytes recapitulating HCM-specific characteristics [157, 158] holds great promise as an exciting new approach to study how gene mutations relate to clinical outcomes and might be applied to test our hypothesis-generating data.

CHAPTER 5

RESULTS III: Generation gene-edited cellular models of HCM

5. RESULTS III: Generation gene-edited cellular models of HCM

5.1. Summary

The CRISPR-Cas9 genome-editing technology was used to introduce patient mutations in the genome of embryonic stem (ES) cells that were subsequently differentiated in cardiomyocytes. A set of isogenic ES cells that differ exclusively by the presence of HCM-causing mutations in the *Tnnt2* gene were generated.

To target the *Tnnt2* gene we used a plasmid developed by Zhang and colleagues that is represented in **Figure 5.1A**. It includes the sequence that codes for Cas9 and downstream of the U6 promoter, a site to clone the guide RNA wanted or 'spacer', between two *Bbs*I sites. The red sequence represents the tracrRNA portion, in order to generate a working chimeric RNA, the sgRNA, mentioned above [96]. Usually, several sgRNAs are chosen, cloned and tested in order to choose the one that most efficiently induces DSBs in the target sequence [94, 96]. The plasmid represented in **Figure 5.1A** is pX459, which is the available from the laboratory of Zhang and colleagues. The following plasmid pX459 has a puromycin cassette that can be used to select successfully transfected cells.
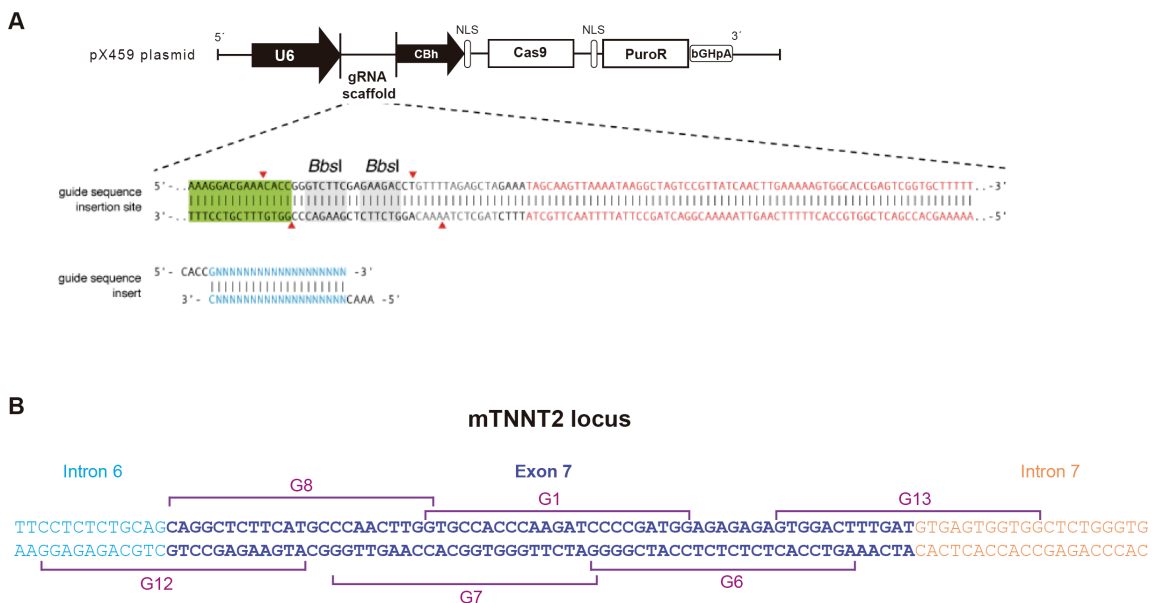
Figure 5.1 - The CRISPR plasmid and the candidate guides cloned

(A) Representation of the plasmid pX459 used for cloning of the guide RNAs in the *Bbs*I sites (adapted from [96]), that also has a puromycin cassette. **(B)** Representation of the mouse *Tnnt2* locus with the several tested guides RNAs.

Several well-characterised mutations of *Tnnt2* were mapped downstream of exon 7 and associated to a severe HCM prognosis. Four of these well-characterised mutations were chosen to develop HCM cellular models: rs121964855 (exon 7, g.135847744 T>A) [159], rs121964856 (exon 8, g. 135847991 G>A) [160], rs121964858 (exon 11, g.135848044 T>A) [161] and rs121964857 (exon 15, g.135852036 C>T) [162]. The variants rs121964855, rs121964856 and rs121964857 were associated to a higher risk of cardiac sudden death, but a lower degree of hypertrophy [159, 160, 162]. The variant rs121964858 presents with different cardiac phenotypes with a more favourable prognosis than the three previous mutations [161]. Each of these 4 mutations was reproduced in two different cell lines using CRISPR genome editing technology. For this purpose, several sgRNAs were tested and cloned in the pX459 plasmid to target the *Tnnt2* locus. The target region in the *Tnnt2* gene was exon 7 because as mentioned above the four mutations chosen are downstream of this exon. The guides were chosen using software available from the Zhang Lab that identifies and scores the potential guides according to the presence of a PAM and off-targets (regions in the genome where the guide might also bind that are not the intended target, available at http://crispr.mit.edu/). The various candidate guides tested are represented in **Figure 5.1B**. The most promising guides were pX459-G1 and pX459-G8; despite G1 showed less off-targets, G8 was the one chosen to follow with most of the further experiments since it was the only one that simultaneously allow the introduction of HCM-associated mutation and of a silence mutation of its PAM sequence in order to prevent further cleavage upon homology recombination events [163]. Exception was the generation of a homozygous knockout of the *Tnnt2* gene by transfection of E14-tg2a cells with pX459-G1, without further mutation of its PAM sequence.

As template for homologous recombination in the *Tnnt2* locus, a homology donor vector (HDV) was constructed containing two homology arms (HA) of around

800bp flanking the *Tnnt2* cDNA from exon 7 to 15 and a selection cassette for puromycin (**Figure 5.2**). The 5'HA corresponds to around 830bp of intron 6 and the 3'HA to intron 7, exon 8 and a part of intron 8; since the DSB is induced on the *Tnnt2* exon 7 everything between these two HAs can be incorporated in the DSB site, as shown on **Figure 5.2A**. The HDV was developed as a plasmid (**Figure 5.2B**) by Santos, C. [163]. This HDV carries the wild-type (WT) version of the *Tnnt2* cDNA (exon7-15) and was further used to introduce by direct point mutagenesis the four different HCM associated mutations giving rise to four different HDV, each carrying a specific HCM-associated mutation.
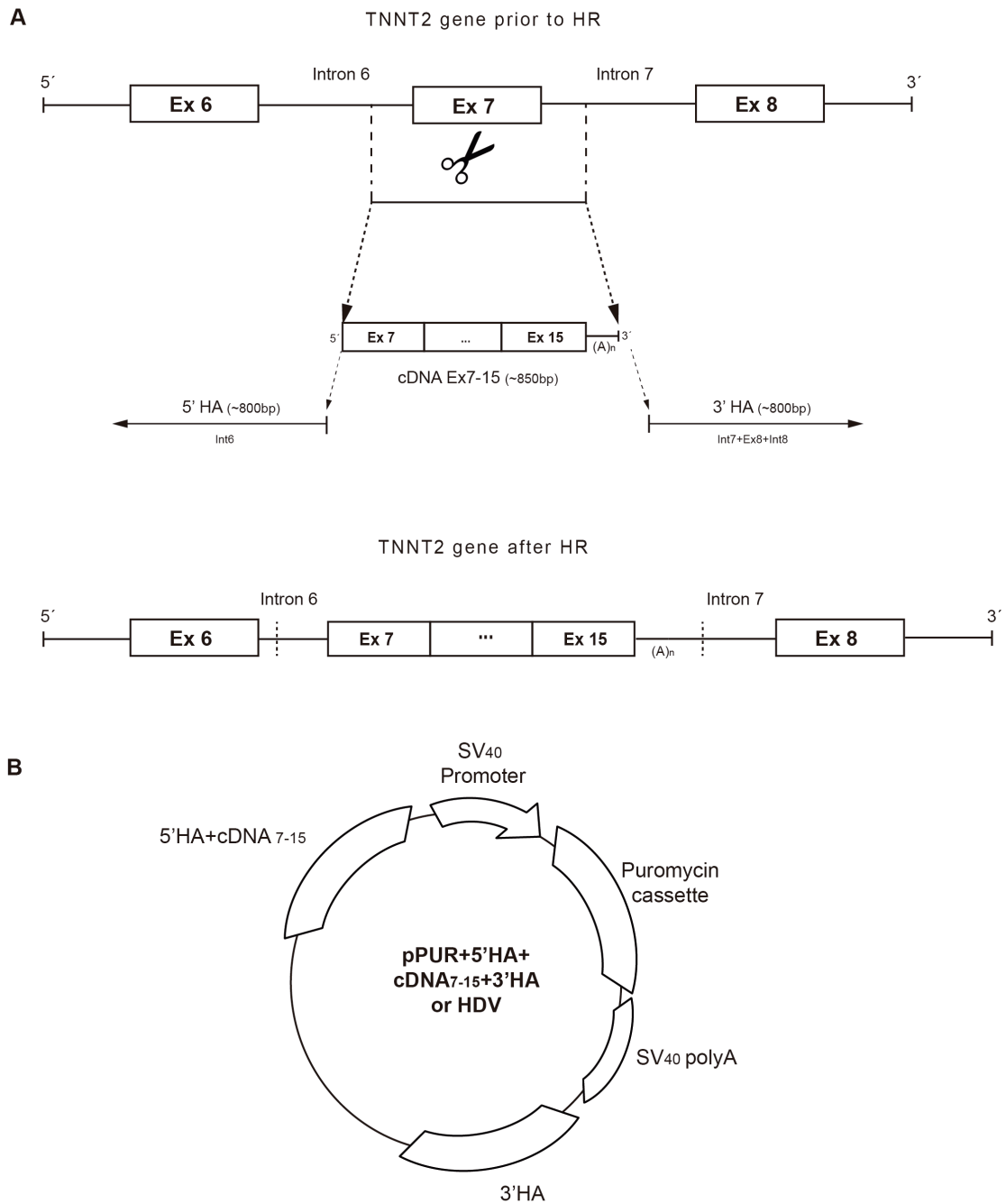
Figure 5.2 - Representation of the *Tnnt2* locus before and after homologous recombination with the HDV

(A) Representation of the *Tnnt2* locus before and after homologous recombination at the DSB performed by Cas9, using the HDV as template. (B) Representation of the HDV, which has the respective homology arms (HA) in either side of the cassette and in the middle the *Tnnt2* cDNA 7-15 and the puromycin cassette that would be inserted at the DSB site. (HA: homology arm)

The experimental design in this chapter was developed by Prof. Doutora Maria do Carmo Fonseca, Prof. Doutora Teresa Carvalho, Prof. Doutora Sandra Martins,

Catarina Santos and myself. The experimental work involving the cloning of the guides, guide cleavage assay and cloning of the original HDV was performed by Catarina Santos. The other experimental work mentioned in this chapter was developed by myself.

5.2. Material and Methods

5.2.1. Donor plasmid directed mutagenesis

The homology donor plasmid, pPUR+5'HA+cDNA$_{7-15}$+3'HA or HDV, was developed previously by Santos, C. [163]. The HDV was used as a template to induce the four chosen TNNT2 mutations. These mutations were introduced in HDV by site directed point mutagenesis. Primer pairs for each mutation are listed in **Table 5.1**:

Table 5.1 - Primers used in the mutagenesis assays to obtain cDNAs carrying the four *TNNT2* mutations

Mutation	Primer	Sequence (5' to 3')
rs121964855	ForEx7T/A	TGGTGCCACCCAAGAACCCCGATGGAGAGAGAGTG
	RevEx7T/A	CACTCTCTCTCCATCGGGGTTCTTGGGTGGCACCA
rs121964856	ForEx8G/A	GGACTTTGATGACATCCACAAGAAGCGCGTGGAGAAGGACC
	RevEx8G/A	GGTCCTTCTCCACGCGCTTCTTGTGGATGTCATCAAAGTCC
rs121964858	ForEx8T/A	GACTCTGATCGAGGCTCACATCGAGAACAGGAAGAAGGAG
	RevEx8T/A	CTCCTTCTTCCTGTTCTCGATGTGAGCCTCGATCAGAGTC
rs121964857	ForEx15C/T	GAAAGTCTCCAAAACTTGTGGGAAGGCCAAAGTC
	RevEx15C/T	GACTTTGGCCTTCCCACAAGTTTTGGAGACTTTC

In summary, we were able to originate the following donor plasmids: HDV-7T/A, HDV-8G/A, HDV-8T/A and HDV-15C/T, each harbouring the indicated HCM-associated mutation.

Then, in all generated donor plasmids, a second mutation was induced in order to create a silent mutation of the PAM sequence of Guide 8 [163]. The following primers were used to perform this site directed mutagenesis, ForPAMG8: 5'-CTCTTCATGCCCAACTTAGTGCCACCCAAGATCC-3' and RevPAMG8: 5'- GGGATCTTG GGTGGCACTAAGTTGGGCATGAAGAG-3'. The positive clones obtained from this second mutagenesis were named HDV-7T/A+mutPAM, HDV-8G/A+mutPAM, HDV-

8T/A+mutPAM and HDV-15C/T+mutPAM. A wild-type HDV with only the PAM mutation, HDV-WT+mutPAM, was used as a control.

5.2.2. E14-tg2a and HL-1 cells cultures

The HL-1 cells were cultured as described in Cap2, session 2.4 of this thesis.

E14-tg2a mouse embryonic stem cells were cultured in Dulbecco's Modified Eagle's medium (Thermo Scientific) supplemented with 15% of FBS embryonic stem cell-qualified (Thermo Scientific), 50mM of β-mercaptoethanol, 2mM L-Glutamine, 1000U/ml of Leukemia Inhibitory Factor (LIF) (Merck Millipore, Darmstadt, Germany) and the 2i inhibitors: 1μM PD0325901 (inhibitor of mitogen-activated protein kinase), 3μM CHIR99021 (inhibitor of glycogen synthase kinase-3) at 37ºC and 5% CO_2 in 0,1% gelatin coated dishes [164-166]. The cells were sub-cultured every 2-3 days using 0,25% trypsin (Thermo Scientific) and split according to the desired density.

In order to obtain embryoid bodies with beating cardiomyocytes (EBs), E14-tg2a cells were plated in hanging drops, at the density of 500 cells per drop, in bacterial grade petri dishes in ES medium without 2i and LIF for two days. Then, dishes were turned in the upright position and filled with 10 ml of ES medium without 2i and LIF. These low adherence conditions were maintained for a total of 6 days since the first plating in drops. The EBs were then collected, replated in a 90mm gelatin coated dish and cultured for a total of 15 days. During this period, cultures were daily observed under light microscopy, checking for spontaneous contractile activity [34, 167].

5.2.3. Transfection of E14-tg2a and HL-1 cells

E14-tg2a and HL-1 cells were transfected using the Neon transfection system (Thermo Scientific), following the manufacturer protocol. Briefly, $3x10^5$ cells of either E14-tg2a or HL-1 were ressuspended in buffer R, containing 5μg of one of the five HDV (HDV-7T/A+mutPAM, HDV-8G/A+mutPAM, HDV-8T/A+mutPAM, HDV-15C/T+mutPAM or HDV-WT+mutPAM) and 5 μg of the pX459-G8. Transfection conditions for the HL-1 cells were the following: 1700V, 20ms, 1 pulse; whereas for the E14-tg2a cells: 1400V,

10ms, 3 pulses. After electroporation, cells were immediately plated onto 35mm dishes with fresh medium. After 48h, puromycin selection was started with 2µg/ml for the E14-tg2a or with 3µg/ml for the HL-1 cells. Individual clones of E14 7TA/mutG8, E14 8TA/mutG8, E14 8GA/mutG8 and E14 15CT/mutG8 were further picked from each dish and transferred into 24well plates for further expansion.

5.2.4. Genomic DNA extraction and PCR genotyping

Cells from all the puromycin selected individual clones growing in 24well plates were trypsinized and collected for genomic DNA extraction and genotyping. Briefly, cell pellets were washed twice in 1x PBS, resuspended in a nuclear lysis buffer (1:1:4:2:12 Tris-HCl 1M pH 8; NaCl 3M;EDTA 0,5M; SDS 10% ; H_2O) and incubated overnight at 55ºC with 80 µg/µl of Proteinase K. After two extractions with phenol/ chloroform/ isoamilic alcohol, and one chloroform extraction, the upper phase was precipitated with 2 volumes of 100% ethanol and 0,1 volume of sodium acetate 3M. After centrifugation, genomic DNA pellets was washed twice with 70% ethanol and resuspended in H_2O.

To genotype the site of the DSB in the *Tnnt2* gene, and address whether or not each allele underwent homologous recombination and whether each clone is an homozygous or heterozygous, a pair of primers was chosen that allow the homozygosity determination of each clone by PCR: ForEx6TNNT2 (5'-TCCAGTAGAGGACACCAAACC-3') and RevI7TNNT2 (5'-CATGAAATGGGTGGCTCAAT-3'). PCR bands of the homozygous or heterozygous clones were purified from the agarose gel using NZYGelpure (Nzytech, Lisboa, Portugal), subsequently cloned in the pBluescript ks+ in the *EcoR*V site (Thermo Scientific) and sequenced with the primers T7 (5'- TAATACGACTCACTATAGG-3') or T3 (5'- CCCTTTAGTGAGGGTTAATT-3'). The sequencing results will help determine the presence or absence of indels in each allele.

5.2.5. Evaluation of cardiac gene expression in selected clones by RT-PCR

In order to evaluate the expression of cardiac markers in the several generated clones, RNA was collected from clones-derived EBs after they start exhibiting spontaneous contractility (day 8 of culture, see above). RNA was isolated using Purezol

reagent (Bio-Rad) and cDNA synthesis was prepared according to protocol of the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) using random primers.

Specific pairs of primers were chosen to evaluate expression of relevant cardiac genes, as depicted in the following table:

Table 5.2 - Primers used for evaluation of cardiac gene expression in selected clones

Gene / Protein	Primer name	Primer sequence (5' to 3')
MYH7 (β- myosin heavy chain)	For MYH7	GGAAGAGCGAGCGGCGCATCAAGG
	Rev MYH7	CTGCTGGAGAGGTTATTCCTCG
ACTC1 (α-actin)	For ACTC1	TGTTACGTCGCCTTCCATTTTGAG
	Rev ACTC1	AAGAGAGAGACATCTCAGAAGC
GJA1 (connexin 43)	For GJA1	GTTCAAGTATGGGATTGAAGAACACGGCAA
	Rev GJA1	TGGTTTTCTCCGTGGGACGTGAGAGGAAGC
ANF (atrial natriuretic factor)	For ANF	CGTGCCCCGACCCACGCCAGCATGGGCTCC
	Rev ANF	GGCTCCGAGGGCCAGCGAGCAGAGCCCTCA
MYH6 (α- myosin heavy chain)	For MYH6	GACAATGCCAATGCGAACAAG
	Rev MYH6	GAAGATCACCCGGGACTTCTC
MYL2 (myosin light chain 2)	For MYL2	CTGAGAGACACCTTTGCTGC
	Rev MYL2	TCCCGAACGTAATCAGCCTT
Ppp3ca (Calcineurin)	For Ppp3ca	CGTTCCATTTCCACCAAGTC
	Rev Ppp3ca	GCGTCGATATCCAGCAAGTT
GATA4	For GATA4	TCTCTTTCCCGGGGACTACT
	Rev GATA4	GGTAGGGGCTGGAGTAGGAG
MEF2C (Myocyte Enhancer Factor 2C)	ForMEF2C	CCGATGCAGACGATTCAGTA
	RevMEF2C	CCAGTGTGCTGACAGGATTG
TNNT2 (cardiac troponin T)	ForEx12/13TNNT2	CGGAAGAGTGGGAAGAGACA
	RevEx14TNNT2	CGCAGAACGTTGATTTCGTA

5.3 Results

5.3.1. Generation and characterization of a knockout *TNNT2* cell line

5.3.1.1. Generation of a *Tnnt2* homozygous knockout cell line

In order to understand the potential impact of a *Tnnt2* knockout at the cardiomyocyte differentiation that occurs in embryoid bodies, a homozygous knockout was performed in E14-tg2a cells [34, 167]. The knockout of the *Tnnt2* gene was obtained by double transfection of E14-tg2a cells with pX459-G1 and WT-HDV without any silence mutation in G1 PAM sequence. After puromycin selection, several clones were tested by PCR and checked by sequencing. One of the clones was successfully characterized as being a homozygous knockout, mainly by the presence of indels in the cleavage site of guide 1 (**Figure 5.3**). This cell line (E14 *Tnnt2* hKO) will be further used as a HCM positive control.

```
                                    DSB site of G1      PAM
A  ref   AACTTGGTGCCACCCAAGATCCC----CGATGGAGAGAGAGTGGACTTTGATGACATCC
   t3    AACTTGgTGCCACCCAAGATCCCATCGCGATGGAGAGAGAGTGGACTTTGATGACATCC
         ************************* *     ** *************************

                                    DSB site of G1      PAM
B  refS  ATGCCCAACTTGGTGCCACCCAAGATCCCCGATGGAGAGAGAGTGGACTTTGATGTGAG
   t3    ATGCCCAACTTGGTGCCACCCAAGA----CGATGGAGAGAGAGTGGACTTTGATGTGAG
         *************************     *****************************
```
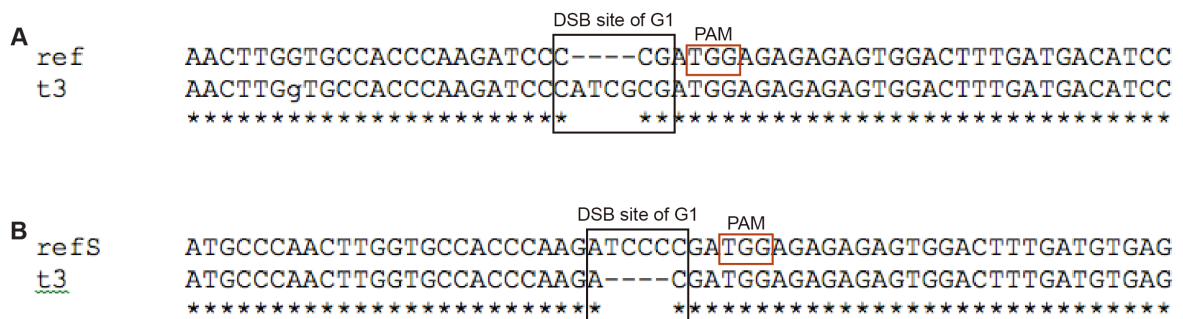
Figure 5.3 – Sequencing of the *Tnnt2* double knockout E14-tg2a clone

Sequencing of both alleles of the Tnnt2 homozygous knockout clone, revealing indels at the cleavage site of G1, due to further activity of Cas9 upon HR events.

5.3.1.2. Embryoid bodies (EB) differentiation of *Tnnt2* homozygous knockout cell line

In two sets of independent experiments, *Tnnt2* homozygous knockout EBs were cultured as floating aggregates until day 5 and showed no difference from non-transfected, wild type ones. However, after being transferred to a gelatin-coated plate at day 5, the homozygous knockout EBs were more irregular and showed no contractile cardiomyocytes on day 8, when compared to the control EBs, which exhibited numerous beating areas and regular edges (**Figure 5.4**).

A   E14 EBs Control

Day 2   Day 3   Day 5

Day 6   Day 7   Day 8

B   E14 EBs TNNT2 DKO

Day 2   Day 3   Day 5

Day 6   Day 7   Day 8
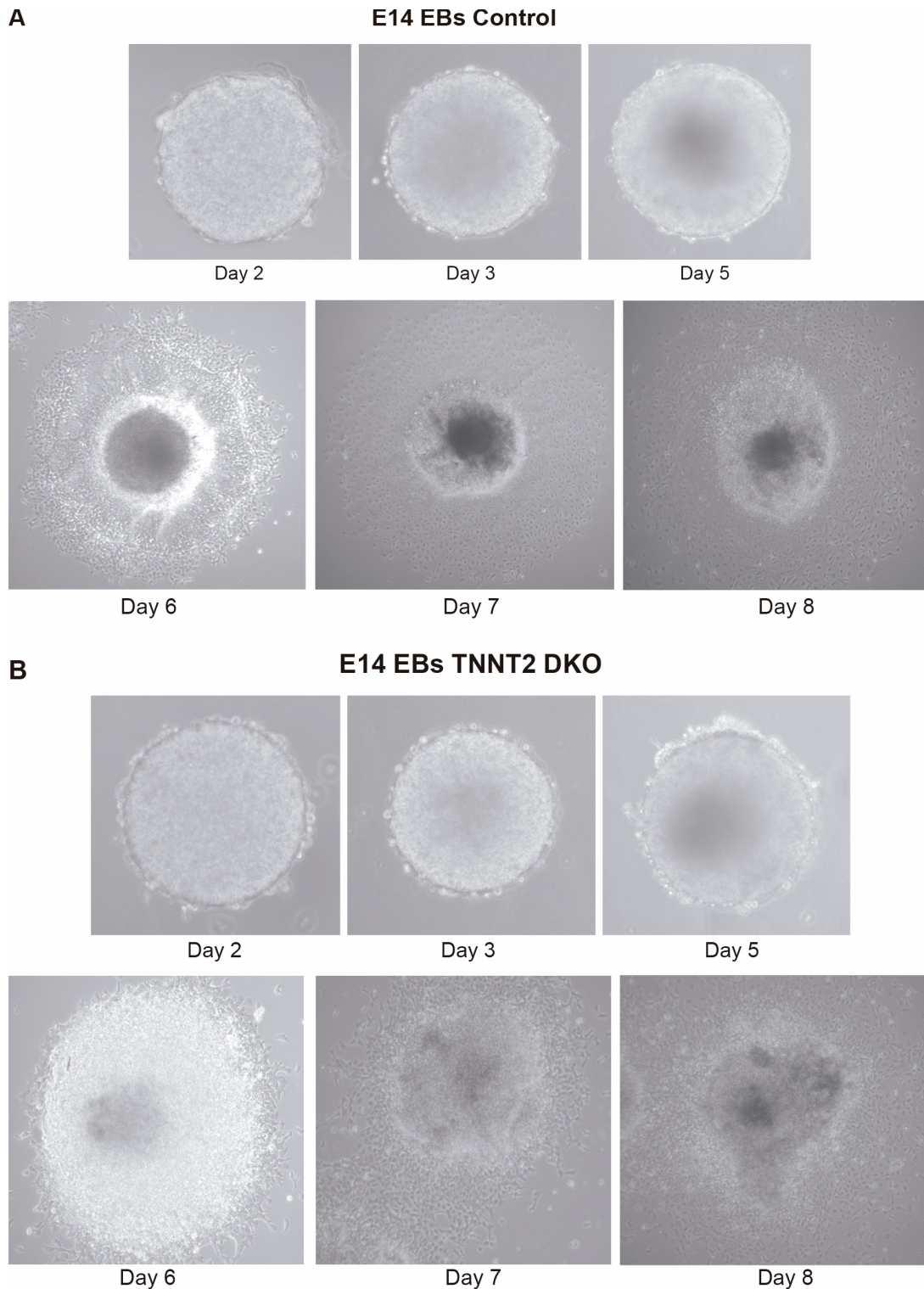
Figure 5.4 - Embryoid bodies of E14-tg2a wild type and *Tnnt2* homozygous knockout

(A) Embryoid bodies (EBs) derived from wild type E14-tg2a cells. From day 0 to day 2, cells were plated in hanging drops-wise manner. On day 3, cells were plated in non-adherent petri dishes, and on day 5 were transferred to gelatin-coated dishes. On day 8, beating EBs were collected and RNA was further extracted to analyse expression of specific cardiac genes. **(B)** EBs derived from *Tnnt2* homozygous knockout. Cells were platted and analysed as described above.

As previously mentioned, to further characterize the two different set of EBs in terms of cardiomyocyte differentiation potential, RNA was extracted from the EBs (WT and KO) at day 8. A panel of genes known to be expressed at the early stage of cardiomyocyte development (*Tnnt2*, *Actc1, Mef2c, Gata4, Gja1*) or to be altered in cardiomyopathy (*Anf, Myh6, Myh7, Pp3ca*) were assessed by RT-PCR (**Table 5.2**). Results show that cardiac specific genes such as α-myosin heavy chain (*Myh6*), β-myosin heavy chain (*Myh7*), α-actin (*Actc1*), myocyte enhancer factor 2C (*Mef2c*), cardiac troponin T (*Tnnt2*) are not expressed in none of the EBs at day 1, as expected (**Figure 5.5**). Interestingly, at day 8, only control WT EBs show expression of this set of cardiac specific genes (*Myh6, Myh7, Actc1, Mef2c, Tnnt2*), suggesting that cardiomyocyte differentiation is delayed in EBs derived from *Tnnt2* homozygous KO. On the other hand, GATA4 factor is only expressed in 8 days old EBs and in the HL-1 cells. The calcineurin (*Pp3ca*) is more expressed in 8 days old EBs when compared to EBs of day 1. These last two (GATA4 and *Pp3ca)* show no difference between WT and *Tnnt2* KO- derived cells. As a control, the level of expression of all tested genes was also analysed in primary cardiomyocytes (HL-1). The atrial natriuretic factor (*Anf*) is only expressed in these cells, whereas the connexion 43 (*Gja1*) is expressed in all analysed samples (**Figure 5.5**).
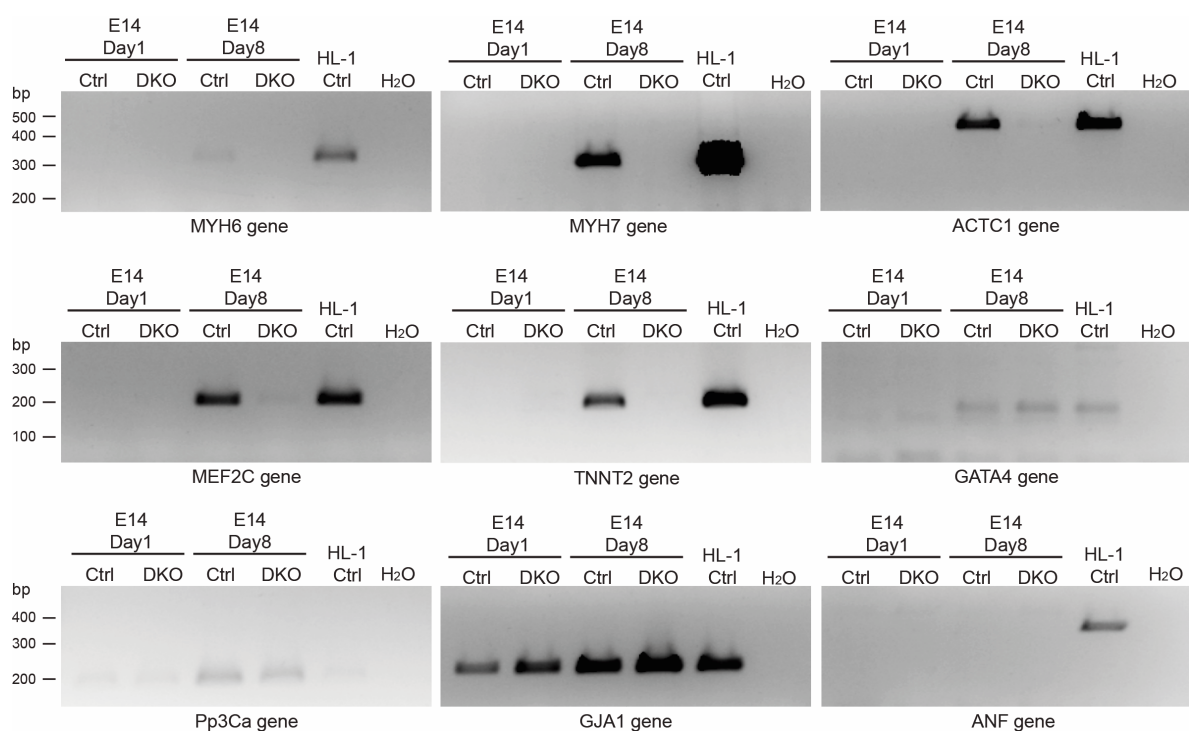
Figure 5.5 - Evaluation of cardiac gene expression in wild-type E14-tg2a and *Tnnt2* homozygous KO-derived EBs (day 1 and day 8), by RT-PCR.

The levels of several cardiac-specific genes, such as α-myosin heavy chain (*Myh6*), β-myosin heavy chain (*Myh7*), α-actin (*Actc1*), myocyte enhancer factor 2C (*Mef2c*) and cardiac troponin T (*Tnnt2*) were analysed by RT-PCR in WT and *Tnnt2* homozygous KO-derived EBs at day 1 and day 8. Levels of *GATA4*, atrial natriuretic factor (*Anf*), Connexion 43 (*Gja1*), calcineurin (*Pp3ca*) were also evaluated. The HL-1 mouse cardiomyocyte cell line was used as a positive control.

5.3.2. Generation of cell lines containing HCM-associated mutations in the *TNNT2* gene using the CRISPR/Cas9 technology

The previously described plasmid pPUR+5'HA+cDNA$_{7-15}$+3'HA (or homology donor vector, HDV) was used for the development of HCM disease cell models. Four previously mentioned *TNNT2* mutations were introduced separately by directed point mutagenesis in the HDV. And so, 4 different HDV were originated: HDV-7T/A, HDV-8G/A, HDV-8T/A and HDV-15C/T. In order to create the HCM cell models, each of the HDV plasmids was co-transfected with the CRISPR-Cas9 plasmid px459-G8, in order to favour HR events at the *Tnnt2* endogenous locus, and so introduction of each specific TNNT2 HCM-associated mutation (**Figure 5.6**). TNNT2 mutations were introduced in two different cell lines: cardiomyocyte cell line HL-1 [168] and mouse embryonic stem cell line E14tg2a.
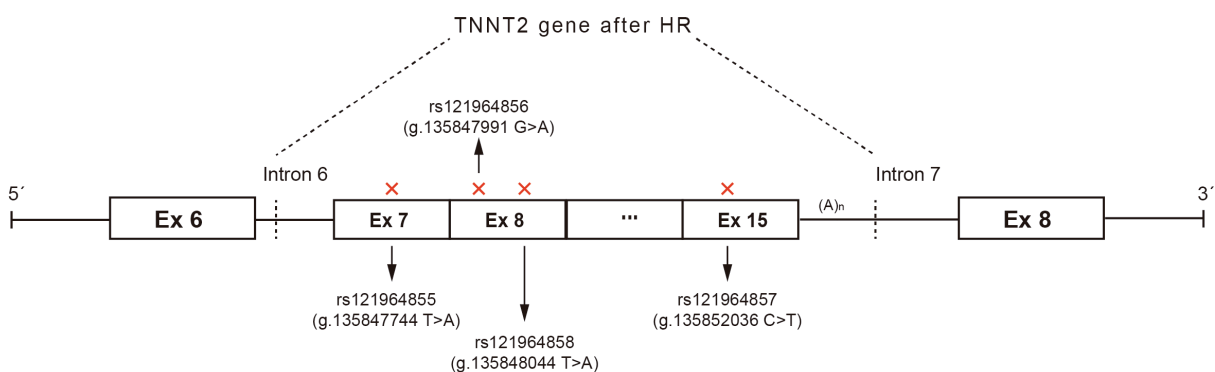
Figure 5.6 - Homologous recombination at the *TNNT2* endogenous locus with representation of the four *TNNT2* HCM-associated mutations.

The red crosses represent the approximate position of the 4 *TNNT2* HCM-associated mutations created by directed point mutagenesis in the HDV, which were incorporated by HR in the *TNNT2* endogenous locus after guide8-mediated cleavage by Cas9.

Notice that in order to avoid further Cas9 cleavage in the *Tnnt2* alleles after recombination, a silent mutation in the PAM sequence recognized by guide 8 was performed in all HDV plasmids (represented in blue, **Figure 5.7**), to avoid the DSB performed by Cas9. To do that site directed point mutagenesis was used to introduce the following silent mutation g.135847730G>A (Leu86Leu) into the HDV-WT, HDV-7T/A, HDV-8G/A, HDV-8T/A and HDV-15C/T (**Figure 5.7**).
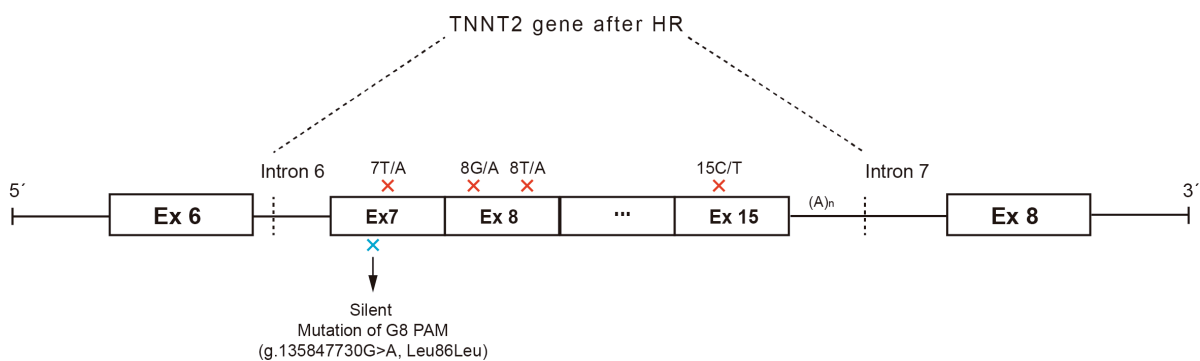
Figure 5.7 - Homologous recombination at the *Tnnt2* locus with representation of the four *TNNT2* HCM-associated mutations and of the silent mutation in the PAM sequence of guide 8.

The red crosses represent the approximate position of the 4 *TNNT2* mutations and the blue one the silent mutation of the PAM sequence created by directed point mutagenesis in the HDV. These were incorporated by HR in the TNNT2 locus thanks to the DSB created by the CRISPR-Cas9 system.

Due to the low frequency of HR in HL-1 cells, a very low number of clones were obtained from this cell line and so, only E14-tg2a derived-clones were used for further characterization. Briefly, E14-tg2a cells were co-transfected with each of the five different linearized HDV plasmids (HDV-WT, HDV-7T/A, HDV-8G/A, HDV-8T/A and HDV-15C/T) together with the pX459-G8. After puromycin selection, PCR screening for the HR-directed incorporation of each HDV was performed by using the pair of primers ForEx6/RevInt6 that allows to distinguish homozygous and heterozygous clones. The appearance of a 3394bp band means that a homologous recombination event had occurred, and a fragment with the cDNA exon10-exon16 was inserted; on the other hand, the presence of a smaller 1220bp band means that no homologous recombination event had occurred. A mutated homozygous clone, where both alleles had suffered a HR event, gives rise to amplification of a single PCR band of 3394bp,

whereas the presence of 3394bp and 1220bp PCR bands simultaneously means a heterozygous clone, with only one of the *Tnnt2* endogenous alleles being target of a HR event (**Figure 5.8A**).
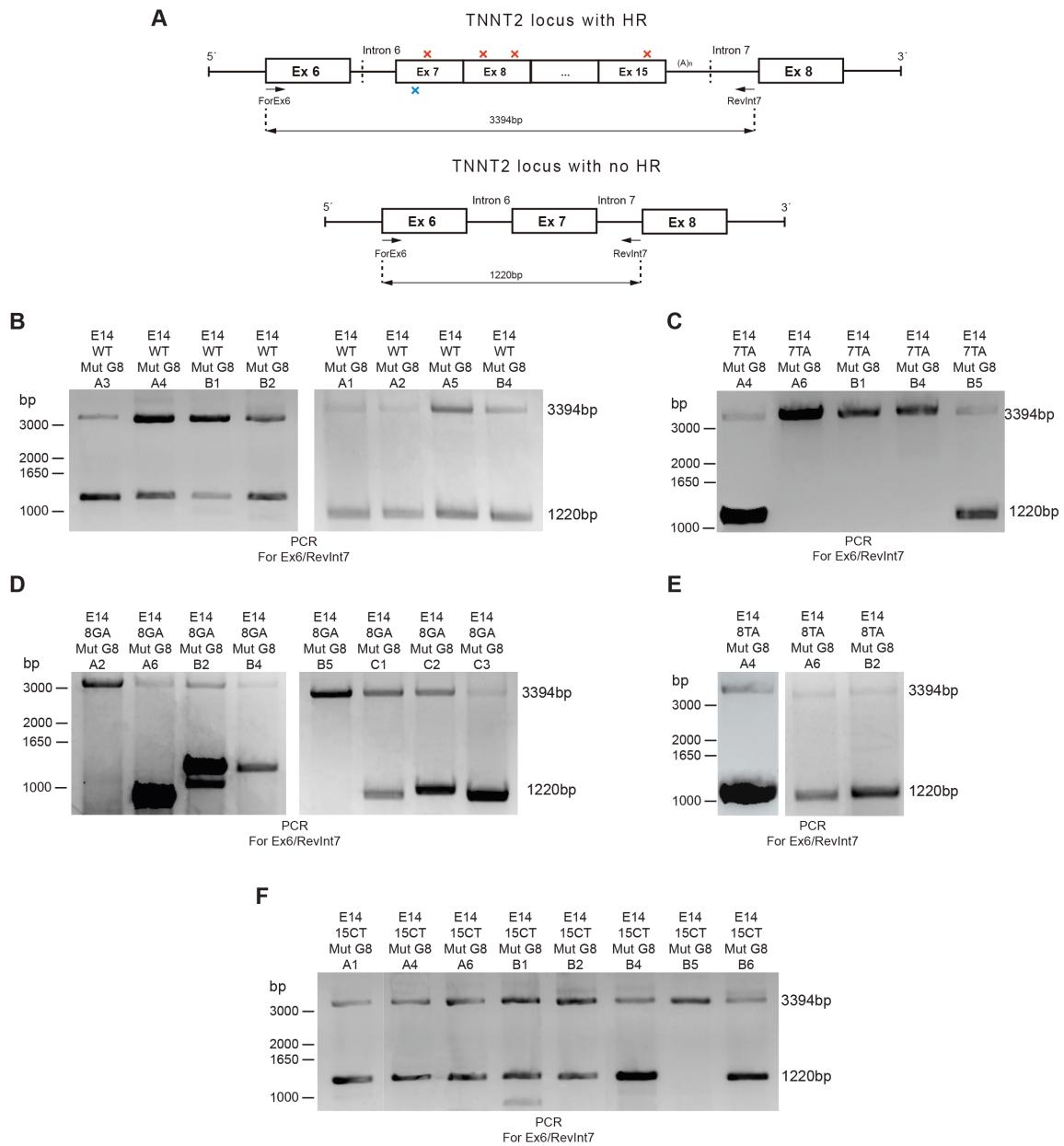
Figure 5.8 - PCR genotyping of selected clones upon co-transfection of E14-tg2a with each of HDV plasmids (HDV-WT, HDV-7T/A, HDV-8T/A, HDV-8G/A and HDV 15C/T) and pX459-G8.

(A) Schematic representation of the position of the pair of primers ForEx6/RevInt7 used for PCR genotyping of selected clones. The size of PCR fragments amplified from TNNT2 alleles that had suffered or not HR events is also represented. **(B-F)** Screening by PCR of the selected clones co-transfected with pX459-G8 and the various HDV: HDV-WT **(B)**, HDV-7T/A **(C)**, HDV-7T/A **(D)**, HDV-8G/A **(E)** and HDV 15C/T **(F)**. The PCR reactions were performed using with the pair of primers ForEx6/RevInt7 to determine the

Tnnt2 cDNA incorporation upon a HR event using genomic DNA extracted from clones transfected with the various HDV, after puromycin selection.

For each of the introduced *Tnnt2* mutations, several homozygous or heterozygous clones were obtained as shown in **Figure 5.8B**. The summary of all the clones obtained is resumed in **Table 5.3**. Both PCR products of one of the heterozygous clones, E14-15CT-A6, were cloned into pBluescript for further sequencing analysis. Sequencing results from the 3394bp PCR product revealed that, as expected, Cas9 did not induce DSB at the mutated PAM sequence (**Figure 5.9A**) and that the 15C/T mutation was successfully introduced into one of the T*nnt2* locus (**Figure 5.9B**).

Table 5.3 - Summary of the genotyping results

HR: homologous recombination

Cell type	Mutation	Clone	Genotype
E14tg2a	8TA	A5	Heterozygous for HR
		A6	Heterozygous for HR
		B2	Heterozygous for HR
	15CT	A1	Heterozygous for HR
		A4	Heterozygous for HR
		A6	Heterozygous for HR
		B1	Heterozygous for HR
		B2	Heterozygous for HR
		B4	Heterozygous for HR
		B5	Homozygous for HR
		B6	Heterozygous for HR
	WT	A1	Heterozygous for HR
		A2	Heterozygous for HR
		A3	Heterozygous for HR
		A4	Heterozygous for HR
		A5	Heterozygous for HR
		B1	Heterozygous for HR
		B2	Heterozygous for HR
		B4	Heterozygous for HR
	8GA	A2	Homozygous for HR
		A6	Heterozygous for HR
		B2	Heterozygous for HR
		B4	Heterozygous for HR
		B5	Homozygous for HR
		C1	Heterozygous for HR

		C2	Heterozygous for HR
		C3	Heterozygous for HR
		A4	Heterozygous for HR
		A6	Homozygous for HR
	7TA	B1	Homozygous for HR
		B4	Homozygous for HR
		B5	Heterozygous for HR

A
```
ref    tgtccccagttgctctggtgacccctgactctccattcctctctgcagCAGGCTCTTCAT
clone  TGTCCCCAGTTGCTCTGGTGACCCCTGACTCTCCATTCCTCTCTGCAGCAGGCTCTTCAT
       ************************************************************
```
DSB site of G8
```
ref    GCCCAACTTGGTGCCACCCAAGATCCCCGATGGAGAGAGAGTGGACTTTGATGACATCCA
clone  GCCCAACTTAGTGCCACCCAAGATCCCCGATGGAGAGAGAGTGGACTTTGATGACATCCA
       *********  *************************************************
```
Silent mutation of G8

HR band

B
15C/T mutation
```
ref    TCCAAAACTCGTGGGAAGGCCAAAGTCACCGGGCGTTGGAAATAGATGAAACTGTTCTCG
clone  TCCAAAACTTGTGGGAAGGCCAAAGTCACCGGGCGTTGGAAATAGATGAAACTGTTCTCG
       ********* ***  ********************************************
```
HR band

Figure 5.9 - Sequencing of the E14-15CT-A6 clone PCR products.

Sequencing results from the 3394bp PCR product revealing that the silent mutation of the PAM sequence of guide 8 is present without indels **(A)** and also the presence of the HCM-associated mutation 15C/T **(B)**.

5.4. Discussion

To date, few studies have been performed in specific HCM cell models to assess the impact of HCM mutations at cell level. Some studies were done for the *MYH7* gene in human patient specific induced pluripotent stem cell derived cardiomyocytes (iPSCs-CM), where the authors were able to reproduce a specific patient/mutation R663H and R442G in the *MYH7* gene [35, 39]. Nevertheless, the use of patient derived cell-lines *per se* does not allow the observation of the phenotype of a single mutation due to patient-to-patient specific genomic context and it also does not allow the study of the phenotype at tissue level, such as fibrosis or even myocyte disarray [35]. So at least, isogenic HCM cell models are required in order to study the phenotypically impact of a single HCM-associated mutation. The use of the CRISPR-Cas9 system allowed during

this work the development of specific mutated HCM cell lines in mouse ES cells. Since, mouse ES cells are of easier manipulation and less time consuming than human ES cells, these were used instead of patient iPSCs cells as a first approach to establish a proof of principle for establishing of HCM cell lines by CRISPR-Cas9 system; hopefully, in the future, and despite the need of generation of human HCM cell models for further comparison with patient iPSC-CMs, these studies may serve as a first cell model to address the morphological and functional consequences of a given HCM-associated at the cell level.

A first step towards this characterization was the development of a homozygous knockout of the *Tnnt2* gene in E14-tg2a cells to assess cardiac specific gene expression in DKO-derived EBs when compared with WT-derived ones, so that the same could later be done with our isogenic *Tnnt2* cell lines. A screening was performed on *Tnnt2* DKO-derived EBs in order to assess the possible impact on the spontaneous beating of the EBs on day 8 of differentiation, showing that these EBs behaved in a similar way until day 5 when compared with WT-derived ones; however, after being seeded in gelatin-coated dishes, DKO-derived EBs begin to be more irregular and show no sign of beating areas when compared to the WT-derived ones on day 8 of differentiation (**Figure 5.4**). Supporting our data, previous studies performed in homozygous KO mice where both alleles of the *Tnnt2* gene were knockdown showed that hearts from mouse embryos around day 8-10 did not beat [169, 170]. Furthermore, these embryo hearts showed sarcomere disassembly and the thin filaments of the sarcomere did not assemble, due to the lack of the troponin T, which is essential for sarcomere assembly [169, 170]. Analysis of RT-PCR results for a set of cardiac specific genes in both WT and *Tnnt2* DKO-derived EBs on day 8 of differentiation show that α-myosin heavy chain (*Myh6*), β-myosin heavy chain (*Myh7*), α-actin (*Actc1*), myocyte enhancer factor 2C (*Mef2c*) and cardiac troponin T (*Tnnt2*) were expressed only in WT-derived EBs and not in the *Tnnt2* DKO-derived ones (**Figure 5.5**). In contrast, the expression of *Gata4* was detected in both *Tnnt2* DKO- and WT-derived EBs. It has been documented both on mouse EBs and in mouse embryos that GATA4 and MEF2C are transcription factors involved in early cardiac development [171-173]; moreover, these factors were reported to be overexpressed in HCM phenotypes [35]. As mentioned above, *Gata4* is expressed in both control and *Tnnt2* DKO-derived EBs, whereas *Mef2c* is present only in

the control WT-derived ones. It has been shown that *GATA4* regulates expression of *MEF2C*, controlling cardiomyocyte proliferation [173, 174]; so, the expression of *GATA4* combined with the lack of *Mef2c* expression in *Tnnt2* DKO-derived EBs might indicate that expression of *Tnnt2* could be also required for sustained expression of *Mef2c*, an essential factor for cardiac muscle development, namely of the left ventricle [175]. On the other hand, calcineurin (*Pp3ca*) is expressed in both WT- and *Tnnt2* DKO-derived EBs, showing increased levels of expression at day 8 of differentiation when compared to day 1. The expression of calcineurin and NFAT pathway is essential for early embryonic development since they are involved in the signalling for cardiovascular development [176, 177]; curiously its overexpression may lead to cardiac hypertrophy [13, 177]. So, the slightly higher calcineurin levels in the *Tnnt2* DKO-derived EBs when compared to the control ones are in agreement with these early observations suggesting a correlation between higher levels of this factor and a hypertrophic clinical condition. Regarding connexin 43 (*Gja1*) levels, they are similar in all analysed samples. This gene was used as a positive control [168], since it encodes a well-known gap junction protein expressed in cardiomyocytes and also in ES cells [178, 179]. Finally, the expression of the ANF factor was also assessed. The expression of ANF is usually observed in embryoid bodies after day 8.5 of differentiation, with this factor being chronically expressed in cases of HCM [13]. In agreement with these findings, ANF expression was only detected in control HL-1 cardiomyocytes, being absent in all analysed EBs at day 8 of differentiation [180, 181]. As previously mentioned, in *Tnnt2* DKO-derived EBs, at day 8 of differentiation, there is no expression of sarcomeric genes (*Myh6, Myh7, Actc1, Tnnt2*). Other studies showed that β-myosin heavy chain (*MYH7*) is expressed earlier during EB development, around day 3-4, and α-myosin heavy chain (*MYH6*) starts to be expressed later, around day 8 [182]. Our data is in agreement with these results, with both *MYH6* and MYH7 genes being expressed only in WT-derived EBs at day 8 of differentiation; moreover, and due to its early expression during EB development, the levels of *MYH7* are higher than the ones of *MYH6* in WT-derived EBs at day 8 of differentiation. In wild-type developing EBs, α-actin and troponin T levels start being expressed as early as day 8 and day 5 respectively [183, 184], in the same direction of our RT-PCR results where both genes only show expression in WT-derived EBs at day 8 of differentiation. All together, our data reveal an important role of *Tnnt2*

gene for proper embryonic cardiac development. Also, these results may be a valuable asset to be used in the further characterization of each of our HCM cell lines carrying specific *Tnnt2* HCM-associated mutations in order to assess their impact on embryonic cardiac development. Quantitative RT-PCR will be essential to determine the relevance of these preliminary results.

In a second step of our work, and to further understand the morphological and functional consequences of a specific HCM-associated at the cell level, HCM cell lines carrying four different *TNNT2* HCM-associated mutations were generated. Briefly, E14-tg2a cells were co-transfected with a plasmid containing a specific mutated *Tnnt2* cDNA pPUR+5'HA+cDNA$_{7-15}$+3'HA (HDV) together with the CRISPR-Cas9 plasmid containing guide 8 (pX459-G8)[163]. Cells that undergo a Cas9-mediated DSB either repair the DNA by non-homologous end joining (NHEJ) or homologous recombination [96, 185, 186]. Taking advantage of this fact, the HDV was developed to promote the insertion of the mutated cDNAs carrying one of the four *Tnnt2* HCM-associated mutations by HR into the endogenous *Tnnt2* locus by the presence of two homology arms that flank the *Tnnt2* cDNA from exon 7 to 15 and a selection cassette for puromycin. Furthermore, and to avoid the occurrence of multiple Cas9-mediated DSB after HR events take place (which could be responsible for indels occurrence in the inserted *Tnnt2* cDNA), a silent mutation of the PAM sequence of guide 8 in the HDV was performed; this has been reported in the literature as being enough to prevent further Cas9 cleavage upon an HR event[187]. After the co-transfection of the various versions of HDV and the pX459-G8 in E14-tg2a cells, several clones were obtained. Notice that HL-1 cells were not used in these assays due to the low rate of HR events observed in this primary cell culture.

Genotyping results of the selected clones show that most of them are heterozygous clones (81%), with only 5 being homozygous (19%). Sequence analysis of one of these clones (E14-15CT-A6) revealed that both mutations, the silent one in the PAM sequence of guide 8 and the *Tnnt2* HCM-associated one, are present without any other alteration in whole sequence of the *Tnnt2* cDNA, proving that mutation of the PAM sequence of guide 8 was sufficient to prevent further Cas9-mediated cleavage.

In order to further carry out this work, all the remaining clones need to be sequenced to confirm that they carry indeed the desired cDNA mutated with each

specific *Tnnt2* HCM-associated mutation. Once this is done, the next step would be to differentiate *Tnnt2* mutated E14-tg2a cell lines into cardiomyocytes and characterised them at the RNA level by RT-PCR and qRT-PCR for the same panel of cardiac specific genes to assess if there are changes in their levels of expression that may be associated to an HCM phenotype. In parallel, immunofluorescence assays will be perform using antibodies to detect specific cardiac proteins, such as troponin t protein and actin, to determine the effect of *TNNT2* mutations on sarcomere assembly, protein distribution and cardiomyocyte morphology. These two proteins were used in another study to verify increased cellular size and multinucleation associated to the HCM phenotype at cell level [35]. Another assay to be performed in order to further characterise the *Tnnt2* mutated cell lines is calcium influx fluorescence imaging. Calcium gradients are an essential part of cardiomyocyte contraction and action potential and this technique is very helpful to detect alterations in the calcium transduction signal in both normal and diseased cardiomyocytes [188]. The ultimate goal of these experiments is to obtain well-characterised HCM cell models in order to use them for further comparison with on patient derived iPSCs-CM.

CHAPTER 6

General Discussion

6. GENERAL DISCUSSION

The work in this thesis was performed in Maria Carmo-Fonseca´s Lab, at Instituto de Medicina Molecular, giving me a unique opportunity to match basic science with translational one. This thesis focuses in a disease that is the most common hereditary disease of the heart: Hypertrophic Cardiomyopathy (HCM). HCM affects 1:500 individuals and, currently, only symptomatic treatment is available [2]. The genetic basis of the disease is largely known with new knowledge being added to it over the years, thanks to genotype/phenotype association analysis and candidate gene approaches. Nevertheless, causal mutations are identified in only 50% of the patients [5]. Even amongst members of the same family, carrying the same HCM-causing mutation, the disease is characterized by a large phenotypic variability [189].

The main goals of this thesis are: 1) to establish a proof-of-principle for a novel gene therapy for HCM at the cell level; 2) the identification of new pathogenic HCM-causing mutations by next-generation sequencing; and 3) the development of new HCM cell models for a better understanding of the HCM genotype/phenotype relationships.

The results obtained during this research project allowed an insight on the future development of RNA based gene therapy for HCM (**Chapter 3**), the identification of candidate HCM causing-mutations that might interfere in splicing or transcription factor binding sites (**Chapter 4**) and the development of cell models of HCM using a cutting-edge technique of genome editing: CRISPR-Cas9 system (**Chapter 5**).

Several types of gene therapy have been designed to target different types of diseases. There are gene therapies that range from gene supplementation, gene silencing, genome editing to spliceosome-mediated mRNA trans-splicing [190, 191]. These are not yet completely efficient to guarantee a reliable therapeutic alternative and so several studies are presently needed. When it comes to recessive disorders, gene supplementation might be used with the delivery of a full cDNA of the respective mutated gene, but this is limited by (1) potential harmful effects due to random

insertion of therapeutic cDNA and (2) the packaging size of current delivery vectors, like the AAV vector[190, 191]. In the latter one, a smaller therapeutic molecule would be the answer like in spliceosome-meditated mRNA trans-splicing. But, when it comes to dominant disorders, there is a need of endogenous regulation of the mutated allele [190]. A dominant disorder can be due to (1) haploinsufficiency, where the levels of wild-type protein produced from the WT allelle are insufficient to maintenance of normal cell function, or to (2) a gain of function of the altered protein produced by the mutated allele, or even both [191]. In these diseases, a promising gene therapy approach would be eliminate the mutated proteins and keep normal levels of the wild-type ones. A gene therapy capable of targeting both issues is spliceosome-meditated mRNA trans-splicing, either for recessive or dominant diseases [190, 191].

In Chapter 2, through a strategy to correct the mutated protein by RNA manipulation, the development of a potential therapeutic approach for HCM is presented [43, 48, 49]. Since HCM is a complex autosomal dominant disease, spliceosome-meditated mRNA trans-splicing was the chosen strategy. In previous studies, an RNA trans-splicing strategy was shown to be a promising gene therapy approach for autosomal dominant disorders, such as retinitis pigmentosa, epidermolysis bullosa, dysferlinopathies, myotonic dystrophy type 1, Huntington's disease and also, hypertrophic cardiomyopathy [43, 48, 49, 190]. One of the most frequently affected genes in HCM is the *TNNT2* gene and some mutations in this gene, mostly downstream of exon 7, show a worst clinical prognosis. The common outcome of these mutations is a poorly functioning protein with a dominant negative effect on myofibrillar organisation and regulation of the intracellular calcium (Ca^{2+}), leading to altered myocardial energetics and reduced contractile capacity of the myocardium [26, 73, 74]. The phenotype resulting from these mutations are well characterized and severe in HCM patients. In order to establish a proof-of-principle to target these mutations at the RNA level, two different RNA trans-splicing strategies were used: double trans-splicing (DTS) and 3'trans-splicing (3'TS). One of the HCM pathogenic mutations in exon 8 of the *TNNT2* gene, the rs121964856 (R92Q) mutation [192], was targeted with a strategy of double trans-splicing approach of exon 8. Other pathogenic HCM mutations downstream of exon 7, such as rs121964858 (exon 8), rs121964857 (exon 15) [162], were targeted using a 3'trans-splicing approach.

When planning a trans-splicing strategy, in the design of the trans-splicing molecule (TSM), there are several features that influence its success, such as: splicing motifs and their strength (splice sites, branch point, polypyrimidine tract), location of the annealing sequences (AS), addition of intronic splicing enhancers and proximity to the target transcript (**Figure 3.8**) [190, 191]. In our assays, all these features were analysed and carefully chosen in order to increase the efficiency of the trans-splicing process. Despite this, is important to take in consideration that the efficiency of trans-splicing can also vary based on specific gene location. So, in the two strategies that were designed in this work, the DTS and 3'TS, all these features were taken in consideration to maximize the success odds of both strategies. Initially, annealing sequences (AS) were carefully chosen and tested by immunofluorescence (IF) to determine if and which endogenous target sequences were available for binding (**Figure 3.10B**). On the other hand, specific muscle splicing enhancers were included in the intronic portion of the TSM, to increase the odds of trans-splicing events (**Figure 3.10A**). Moreover, two different promoters that drove the expression of the TSM were used, an endogenous one (minimal *Tnnt2* mouse promoter) and a viral one (CMV promoter) (**Figure 3.10C**). Finally, carefully selected intronic sequences from the human endogenous intron 8 and 9 that had predicted endogenous enhancers were maintained. Nevertheless, the lack of detection of trans-splicing events with DTS and 3'TS strategies, either by restriction analysis or by the use of a degenerate primer against the human exon 9 in a conventional RT-PCR, showed that there is not enough trans-spliced product to be detected by any of these techniques (**Figure 3.11, 3.12**). The detection of a very faint band by RT-PCR using a radioactively labelled primer clearly demonstrates that the efficiency of the trans-splicing, under these conditions, is very limited (**Figure 3.13**). The small size of introns 7 and 8 of the mouse *Tnnt2* gene may be an obstacle in the choice of the annealing sequences, since there were fewer AS that could be tested for this locus. In some previous studies, it was advantageous to mask certain endogenous splicing features (splice sites, branch point, polypyrimidine tract) to force the spliceosome to choose the splice sites present in the trans-splicing molecule [48, 190]. A good example of this was, the use of antisense nucleotides (ASOs) that successfully promoted trans-splicing by blocking the endogenous splice site forcing the spliceosome to choose the splice site of the TSM, in the *RHO* gene and

COL7A1 gene [48, 190]. According with this point of view, to increase the success rate of double trans-splicing, we had also used different antisense nucleotides (ASOs) that specifically block the endogenous splice sites around murine *Tnnt2* endogenous exon 8. This strategy was tested in WT cultured cells but without success since there was no detection of endogenous exon 8 skipping. Based on this, we decided to not proceed with the use of these ASOs in further trans-splicing assays.

There might be also other important factors, such as an antisense transcription or binding of splicing factors that have a potential to interfere in this given locus of the *Tnnt2* gene. Both may difficult the occurrence trans-splicing reaction and thus interfere with its success. Until date, the efficiency of trans-splicing in several studies ranges from 30-60% *in vitro* and 1-10% *in vivo*, which is still very low for a significant therapeutic effect. However, most of these studies were carried on using artificial systems such as a minigene, and so their efficiency cannot be directly compared with our system, that uses the endogenous gene [49, 71, 72, 105]. Despite this, all together, these assays show that more studies in the design of an optimized TSM are needed [190].

In summary, our data suggest that murine endogenous *Tnnt2* locus around exon 8 is not the best candidate for either 3'TS or DTS. To test the feasibility of trans-splicing in the *Tnnt2* gene, an alternative approach would be the choice of other target introns thus allowing several ASs to be tested; also, other splicing enhancers could be inserted in the TSM in order to favour trans-splicing events. Finally, hiding endogenous splice sites taking advantage of the used ASs could also improve the efficiency of our approaches. Notice that trans-splicing strategies are highly influenced by all the factors discussed above and so their success may vary from case to case. Nevertheless, and despite its low efficiency until date, this RNA-based gene therapy still the most attractive to target autosomal dominant disorders by impairing the production of mature mutated transcripts and so, avoiding the production of a deleterious, gained-of-function protein.

Since, HCM is a complex genetic disorder, with only 50% of patients having a genetic diagnosis [82, 145], there is a need to further screen for HCM pathogenic mutations to better understand the disease and the current knowledge about its

genotype/phenotype relationships [193]. Through the increase number of HCM variants being found by NGS approaches and with the use of cellular or animal disease models more information can be gathered for this end, allowing the genetic sequencing of HCM genes in these patients to help establish a possible genetic diagnosis, with the identification of new pathogenic variants.

So, in Chapter 3, we presented a method that may expand the number and type of new candidate HCM causing-mutations. It hypothesizes that the expression of certain HCM related genes could also be affected by changes in non-coding regulatory regions of the genome. These may be putative targets for transcription factors or splicing regulation and thus influence the HCM phenotype. To test this hypothesis, using targeted high throughput sequencing and computational approaches to identify the noncoding variants, we had identify new HCM candidate mutations that can interfere in those regions.

In a cohort of 16 HCM-diagnosed patients, we had three patients with disease-causing coding mutations in the *MYBC3* gene, and additional patients with variants of uncertain significance in the *TNNT2, MYBPC3, TTN, TPM1* and *MYH6* genes; all these scored as likely pathogenic according to multiple in silico prediction tools.

In this set of patients, 6 variants were identified that scored with high values using GWAVA, CADD and Genomiser metrics (**Figure 4.4B**). These included two variants in *VCL*, two variants in *TTN*, one variant in *ACTC1* and one variant in *PRKAG2* (**Table 4.2**). The *VC*L variant c.499+367T>C (rs113195070) was detected in 6 patients. One of these patients was a compound heterozygous for a splice site mutation in *MYBPC3* and the *VCL* variant. Analysis of family members revealed that carriers of the *MYBPC3* mutation did not manifest the disease, but family members that are compound heterozygous are clinically affected (**Figure 4.5**). Furthermore, this variant was enriched in the patient population (rs113195070) localized in a region associated with FOS, JUN and EP300 (**Figure 4.6A**).

Also, two single nucleotide substitutions in the titin gene (rs142156368 and rs72650063) were identified that are 3 to 6-fold more frequent in our cohort of probands than in normal European populations. If these two variants do interfere with titin splicing, as predicted by the SPIDEX computational model, they are likely to

contribute to HCM phenotype. One of these variants (rs72650063) was even predicted by Human Splicing Finder to disrupt binding of splicing factor SRSF1 (**Figure 4.6C**).

In this part of the current work, we developed and described a pipeline analysis that can be used to prioritize the variants found within the non-coding regions and identify those with a potential impact on HCM. One of the main challenges with the NGS gene panels still remains the increasing number of variants, and the ability to distinguish a polymorphism from a potential pathogenic mutation [194]. These variants will require experimental and functional studies to determine the final impact of these potential pathogenic noncoding variants in cell and animal models. Based on this, in the last Chapter of this work, HCM cell models were developed in order to determine the functional impact of a given HCM-associated variant.

In fact, nowadays, there is a clear need for HCM cell models in order to expand the knowledge about the genotype/phenotype relationships at the cellular level. Previous studies of several HCM patient specific iPSCs cardiomyocytes with specific mutations in the genes *MYH7, MYBPC3, TPM1* were developed and proved to be very helpful in that endeavour [35, 39-41]. However, patient specific cell lines might have other mutations contributing to the HCM phenotype and so the use of isogenic cell lines is of great value, to specifically study the effect of a given variant. This can be accomplished thanks to genome editing techniques, like the CRISPR-Cas9 system, either by creating the desired mutation in a wild-type ES cell line or by correcting any other mutations in patient specific cell lines [195].

In Chapter 4, we present custom engineered E14-tg2a-based cell lines, by using the powerful CRISPR/Cas9 genome-editing tool to induce a desired HCM-associated mutation. We had chosen 4 well-known *TNNT2* HCM-causing mutations and mouse ES cells as cell model due to technical constraints (easier manipulation and less time consuming than human ES cells). Our results establish a proof of principle that by using the CRISPR-Cas9 gene-editing tool we are able to developed mouse ES cell lines carrying specific HCM-causing mutations, which in the future may serve as control to morphological and functional characterization of patient specific iPSC-CMs at the cellular level.

A first step towards this characterization was the production of EBs derived from a double knockout (DKO) of the *Tnnt2* gene to assess its potential impact on stem cell differentiation into cardiomyocytes. Briefly, both WT and *Tnnt2* DKO murine ES cells were differentiated into EBs and, at day 8 of differentiation. Morphologically, the most significant differences were observed at day 8 of differentiation, where the *Tnnt2* DKO derived-EBs were more irregular and had no beating areas (**Figure 5.4**). Previous studies performed in *Tnnt2* DKO mice documented that hearts from mouse embryos around day 8-10 also did not beat [169, 170]. When analysing the results of a set of cardiac specific genes in both WT and *Tnnt2* DKO derived-EBs by RT-PCR (**Figure 5.5**), at day 8 of differentiation, a pattern became apparent with *Myh6*, *Myh7*, *Actc1*, *Mef2c* and *Tnnt2* genes being expressed only in WT EBs and not in the *Tnnt2* DKO ones. In contrary, expression of *Gata4* was detected in both WT and *Tnnt2* DKO EBs. It has been shown that *Gata4* regulates expression of *Mef2C* controlling cardiomyocyte proliferation [173, 174]. So, the lack of *Mef2c* expression in *Tnnt2* DKO derived-EBs might indicate that expression of *Tnnt2* could be also required for sustained expression of *Mef2c*. As previously mentioned, at day 8 of differentiation, sarcomeric genes expression (*Myh6, Myh7, Actc1, Tnnt2)* was absent in the *Tnnt2* DKO derived-EBs. In other studies using *Tnnt2* DKO mice, the embryo hearts showed sarcomere disassembly and the thin filaments of the sarcomere did not assemble, due to the lack of the Troponin T, which is essential protein for sarcomere assembly [169, 170]. Our results from the DKO of the *Tnnt2* gene show a possible role of this gene in embryonic cardiac development.

Next, and in order to create isogenic HCM cell models, puromycin-resistant homology donor vectors, (HDV) containing cDNAs carrying one of four well-known HCM-causing mutations were generated and used to edit the murine endogenous *Tnnt2* locus taking advantage of a CRISPR/Cas9 genome editing approach. Also, a silent mutation was induced by site directed point mutagenesis in the PAM sequence of guide 8 in all HDV, to prevent Cas9 from performing a DSB in the allele where HR occurred. Upon co-transfection of E14-tg2a cells with each of the linearized HDV plasmids together with the pX459-G8 that encodes the Cas9 nuclease and the gRNA8, several clones were isolated by puromycin selection. PCR genotyping of these clones revealed that they were mostly heterozygous (81%), with only 5 being homozygous

(19%). Sequencing analysis of one of the clones (E14-15CT-A6) revealed that both the silent PAM mutation of guide 8, as well the specific HCM causing-mutation was present in the murine endogenous *Tnnt2* locus. Once the rest of the clones are sequenced, they can be differentiated into cardiomyocytes and further characterized. In the time frame of this project, these were the reachable results. But, we were able to establish a proof-of-principle of the development of mouse HCM isogenic ES cell models with the use of the genome-engineering tool, CRISPR-Cas9 system.

The ultimate goal of these experiments is to obtain well-characterised HCM cell models that can be use in the future to compare with patient specific ones, allowing a better understanding of genotype/phenotype relationships in HCM.

The findings reported in this thesis meet the goals establish when it was first defined as a project. They will contribute, either at cell level or in a clinical setting, to the understanding and management of HCM.

First, we attempted to develop an RNA based gene therapy for HCM in the *TNNT2* gene. Despite the low efficiency, the double trans-splicing approach for exon 11 leaves some room for improvement. Second, a pipeline was developed to prioritize non-coding variants that might interfere in splicing or in transcription factor binding sites and have an impact in HCM. Third, HCM isogenic cell models were developed with the CRISPR-Cas9 system for further studies of the HCM genotype/phenotype relationships.

In the near future, RNA based therapy for HCM should be further investigated. The pipeline developed can be further used for the identification of novel variants in new cohorts of HCM patients and even contribute for their 'molecular diagnosis'. Finally, the developed HCM cell lines carrying the 4 well-known *TNNT2* HCM-causing mutations need to be fully characterized so they can be further used as controls for characterizing patient derived HCM cell lines.

Overall, the work presented here contributes with novel knowledge for the HCM field at three different levels: mutation screening for HCM diagnosis/prognosis, HCM cell models for genotype/phenotype relationships studies, and RNA trans-splicing assays as a potential HCM RNA based gene therapy.

7. REFERENCES

1. Elliott, P.M., et al., *[2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy].* Kardiol Pol, 2014. **72**(11): p. 1054-126.
2. Fraiche, A. and A. Wang, *Hypertrophic Cardiomyopathy: New Evidence Since the 2011 American Cardiology of Cardiology Foundation and American Heart Association Guideline.* Curr Cardiol Rep, 2016. **18**(8): p. 70.
3. Semsarian, C., et al., *New perspectives on the prevalence of hypertrophic cardiomyopathy.* J Am Coll Cardiol, 2015. **65**(12): p. 1249-54.
4. Hensley, N., et al., *Hypertrophic cardiomyopathy: a review.* Anesth Analg, 2015. **120**(3): p. 554-69.
5. Tian, T., et al., *Progress in the molecular genetics of hypertrophic cardiomyopathy: a mini-review.* Gerontology, 2013. **59**(3): p. 199-205.
6. Teekakirikul, P., et al., *Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era.* J Mol Diagn, 2013. **15**(2): p. 158-70.
7. Teekakirikul, P., et al., *Hypertrophic cardiomyopathy: translating cellular cross talk into therapeutics.* J Cell Biol, 2012. **199**(3): p. 417-21.
8. Barefield, D., et al., *Haploinsufficiency of MYBPC3 exacerbates the development of hypertrophic cardiomyopathy in heterozygous mice.* J Mol Cell Cardiol, 2015. **79**: p. 234-43.
9. Coppini, R., et al., *Clinical Phenotype and Outcome of Hypertrophic Cardiomyopathy Associated With Thin-Filament Gene Mutations.* Journal of the American College of Cardiology, 2014. **64**(24): p. 2589-2600.
10. Moore, J.R., L. Leinwand, and D.M. Warshaw, *Understanding cardiomyopathy phenotypes based on the functional impact of mutations in the myosin motor.* Circ Res, 2012. **111**(3): p. 375-85.
11. Houston, B.A. and G.R. Stevens, *Hypertrophic cardiomyopathy: a review.* Clin Med Insights Cardiol, 2014. **8**(Suppl 1): p. 53-65.
12. Pantazis, A., et al., *Diagnosis and management of hypertrophic cardiomyopathy.* Echo Res Pract, 2015. **2**(1): p. R45-53.
13. Harvey, P.A. and L.A. Leinwand, *Cellular mechanisms of cardiomyopathy.* The Journal of Cell Biology, 2011. **194**(3): p. 355-365.

14. Kinnunen, P., O. Vuolteenaho, and H. Ruskoaho, *Mechanisms of atrial and brain natriuretic peptide release from rat ventricular myocardium: effect of stretching.* Endocrinology, 1993. **132**(5): p. 1961-70.

15. Miyata, S., et al., *Myosin heavy chain isoform expression in the failing and nonfailing human heart.* Circ Res, 2000. **86**(4): p. 386-90.

16. Teekakirikul, P., et al., *Cardiac fibrosis in mice with hypertrophic cardiomyopathy is mediated by non-myocyte proliferation and requires Tgf-beta.* J Clin Invest, 2010. **120**(10): p. 3520-9.

17. Matsui, T., et al., *Phenotypic spectrum caused by transgenic overexpression of activated Akt in the heart.* J Biol Chem, 2002. **277**(25): p. 22896-901.

18. Molkentin, J.D., et al., *A calcineurin-dependent transcriptional pathway for cardiac hypertrophy.* Cell, 1998. **93**(2): p. 215-28.

19. Ross, S.B., S.T. Fraser, and C. Semsarian, *Induced pluripotent stem cells in the inherited cardiomyopathies: From disease mechanisms to novel therapies.* Trends in Cardiovascular Medicine, 2016.

20. Song, W., et al., *Molecular mechanism of the E99K mutation in cardiac actin (ACTC Gene) that causes apical hypertrophy in man and mouse.* J Biol Chem, 2011. **286**(31): p. 27582-93.

21. Knoll, R., et al., *A common MLP (muscle LIM protein) variant is associated with cardiomyopathy.* Circ Res, 2010. **106**(4): p. 695-704.

22. Luczak, E.D., et al., *Remodeling the cardiac transcriptional landscape with diet.* Physiological Genomics, 2011. **43**(12): p. 772-780.

23. Ruggiero, A., et al., *Pathogenesis of hypertrophic cardiomyopathy caused by myozenin 2 mutations is independent of calcineurin activity.* Cardiovasc Res, 2013. **97**(1): p. 44-54.

24. Purevjav, E., et al., *Molecular basis for clinical heterogeneity in inherited cardiomyopathies due to myopalladin mutations.* Hum Mol Genet, 2012. **21**(9): p. 2039-53.

25. James, J., et al., *Transgenic modeling of a cardiac troponin I mutation linked to familial hypertrophic cardiomyopathy.* Circ Res, 2000. **87**(9): p. 805-11.

26. Tardiff, J.C., et al., *Cardiac troponin T mutations result in allele-specific phenotypes in a mouse model for hypertrophic cardiomyopathy.* J Clin Invest, 1999. **104**(4): p. 469-81.

27. Knollmann, B.C., et al., *Inotropic Stimulation Induces Cardiac Dysfunction in Transgenic Mice Expressing a Troponin T (I79N) Mutation Linked to Familial Hypertrophic Cardiomyopathy.* Journal of Biological Chemistry, 2001. **276**(13): p. 10039-10048.

28. Prabhakar, R., et al., *A familial hypertrophic cardiomyopathy alpha-tropomyosin mutation causes severe cardiac hypertrophy and death in mice.* J Mol Cell Cardiol, 2001. **33**(10): p. 1815-28.

29. Rajan, S., et al., *Microarray analysis of active cardiac remodeling genes in a familial hypertrophic cardiomyopathy mouse model rescued by a phospholamban knockout.* Physiol Genomics, 2013. **45**(17): p. 764-73.

30. Muthuchamy, M., et al., *Mouse Model of a Familial Hypertrophic Cardiomyopathy Mutation in α-Tropomyosin Manifests Cardiac Dysfunction.* Circulation Research, 1999. **85**(1): p. 47-56.

31. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.

32. Batalov, I. and A.W. Feinberg, *Differentiation of Cardiomyocytes from Human Pluripotent Stem Cells Using Monolayer Culture.* Biomarker Insights, 2015. **10**(Suppl 1): p. 71-76.

33. Lee, Y.K., K.M. Ng, and H.F. Tse, *Modeling of human cardiomyopathy with induced pluripotent stem cells.* J Biomed Nanotechnol, 2014. **10**(10): p. 2562-85.

34. Fuegemann, C.J., et al., *Differentiation of mouse embryonic stem cells into cardiomyocytes via the hanging-drop and mass culture methods.* Curr Protoc Stem Cell Biol, 2010. **Chapter 1**: p. Unit 1F.11.

35. Lan, F., et al., *Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells.* Cell Stem Cell, 2013. **12**(1): p. 101-13.

36. Yang, L., et al., *Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population.* Nature, 2008. **453**(7194): p. 524-8.

37. Lian, X., et al., *Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions.* Nat Protoc, 2013. **8**(1): p. 162-75.

38. Lian, X., et al., *Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling.* Proceedings of the National Academy of Sciences, 2012. **109**(27): p. E1848–E1857.

39. Han, L., et al., *Study familial hypertrophic cardiomyopathy using patient-specific induced pluripotent stem cells.* Cardiovasc Res, 2014. **104**(2): p. 258-69.

40. Tanaka, A., et al., *Endothelin-1 induces myofibrillar disarray and contractile vector variability in hypertrophic cardiomyopathy-*

induced pluripotent stem cell-derived cardiomyocytes. J Am Heart Assoc, 2014. **3**(6): p. e001263.

41. Ojala, M., et al., *Mutation-Specific Phenotypes in hiPSC-Derived Cardiomyocytes Carrying Either Myosin-Binding Protein C Or alpha-Tropomyosin Mutation for Hypertrophic Cardiomyopathy.* Stem Cells Int, 2016. **2016**: p. 1684792.

42. Ammirati, E., et al., *Pharmacological treatment of hypertrophic cardiomyopathy: current practice and novel perspectives.* European Journal of Heart Failure, 2016: p. n/a-n/a.

43. Behrens-Gawlik, V., et al., *MYBPC3 in hypertrophic cardiomyopathy: from mutation identification to RNA-based correction.* Pflugers Arch, 2014. **466**(2): p. 215-23.

44. Guo, Y., et al., *Dissecting Disease Inheritance Modes in a Three-Dimensional Protein Network Challenges the "Guilt-by-Association" Principle.* American Journal of Human Genetics, 2013. **93**(1): p. 78-89.

45. Lewin, A.S., P.M. Glazer, and L.M. Milstone, *Gene therapy for autosomal dominant disorders of keratin.* J Investig Dermatol Symp Proc, 2005. **10**(1): p. 47-61.

46. Lodish H, B.A., Zipursky SL, et al., *Molecular Cell Biology*, in *Molecular Cell Biology*. 2000: New York: W. H. Freeman.

47. Comitato, A., et al., *Dominant and recessive mutations in rhodopsin activate different cell death pathways.* Human Molecular Genetics, 2016.

48. Koller, U., et al., *Trans-splicing improvement by the combined application of antisense strategies.* Int J Mol Sci, 2015. **16**(1): p. 1179-91.

49. Berger, A., et al., *Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa.* Mol Ther, 2015. **23**(5): p. 918-30.

50. Friedmann, T. and R. Roblin, *Gene therapy for human genetic disease?* Science, 1972. **175**(4025): p. 949-55.

51. Saada, Y.B., et al., *Genome- and Cell-Based Strategies in Therapy of Muscular Dystrophies.* Biochemistry (Mosc), 2016. **81**(7): p. 678-90.

52. Misra, S., *Human gene therapy: a brief overview of the genetic revolution.* J Assoc Physicians India, 2013. **61**(2): p. 127-33.

53. Ibraheem, D., A. Elaissari, and H. Fessi, *Gene therapy and DNA delivery systems.* International Journal of Pharmaceutics, 2014. **459**(1–2): p. 70-83.

54. Chatterjee, A., N. Singh, and M. Saluja, *Gene therapy in periodontics.* Journal of Indian Society of Periodontology, 2013. **17**(2): p. 156-161.

55. Chugunova, A.A., O.A. Dontsova, and P.V. Sergiev, *Methods of Genome Engineering: a New Era of Molecular Biology.* Biochemistry (Mosc), 2016. **81**(7): p. 662-77.

56. Katsuki, Y. and M. Takata, *Defects in homologous recombination repair behind the human diseases: FA and HBOC.* Endocr Relat Cancer, 2016. **23**(10): p. T19-37.

57. Kornblihtt, A.R., et al., *Alternative splicing: a pivotal step between eukaryotic transcription and translation.* Nat Rev Mol Cell Biol, 2013. **14**(3): p. 153-65.

58. Havens, M.A., D.M. Duelli, and M.L. Hastings, *Targeting RNA splicing for disease therapy.* Wiley Interdiscip Rev RNA, 2013. **4**(3): p. 247-66.

59. Effenberger, K.A., V.K. Urabe, and M.S. Jurica, *Modulating splicing with small molecular inhibitors of the spliceosome.* Wiley Interdisciplinary Reviews: RNA, 2016: p. n/a-n/a.

60. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.

61. Modrek, B. and C. Lee, *A genomic view of alternative splicing.* Nat Genet, 2002. **30**(1): p. 13-19.

62. Bennett, C.F. and E.E. Swayze, *RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform.* Annu Rev Pharmacol Toxicol, 2010. **50**: p. 259-93.

63. Schmid, F., et al., *A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs.* Hum Gene Ther, 2013. **24**(1): p. 97-104.

64. Yokota, T., E. Hoffman, and S. Takeda, *Antisense oligo-mediated multiple exon skipping in a dog model of duchenne muscular dystrophy.* Methods Mol Biol, 2011. **709**: p. 299-312.

65. El-Beshlawy, A., et al., *Correction of aberrant pre-mRNA splicing by antisense oligonucleotides in beta-thalassemia Egyptian patients with IVSI-110 mutation.* J Pediatr Hematol Oncol, 2008. **30**(4): p. 281-4.

66. Davis, R.L., et al., *A deep intronic mutation in FGB creates a consensus exonic splicing enhancer motif that results in afibrinogenemia caused by aberrant mRNA splicing, which can be corrected in vitro with antisense oligonucleotide treatment.* Hum Mutat, 2009. **30**(2): p. 221-7.

67. Wally, V., E.M. Murauer, and J.W. Bauer, *Spliceosome-mediated trans-splicing: the therapeutic cut and paste.* J Invest Dermatol, 2012. **132**(8): p. 1959-66.

68. Chao, H., et al., *Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing.* Nat Med, 2003. **9**(8): p. 1015-9.

69. Gruber, C., et al., *The design and optimization of RNA trans-splicing molecules for skin cancer therapy.* Mol Oncol, 2013. **7**(6): p. 1056-68.

70. Lorain, S., et al., *Dystrophin rescue by trans-splicing: a strategy for DMD genotypes not eligible for exon skipping approaches.* Nucleic Acids Res, 2013. **41**(17): p. 8391-402.

71. Monjaret, F., et al., *Cis-splicing and translation of the pre-trans-splicing molecule combine with efficiency in spliceosome-mediated RNA trans-splicing.* Mol Ther, 2014. **22**(6): p. 1176-87.

72. Mearini, G., et al., *Repair of Mybpc3 mRNA by 5'-trans-splicing in a Mouse Model of Hypertrophic Cardiomyopathy.* Mol Ther Nucleic Acids, 2013. **2**: p. e102.

73. Sweeney, H.L., et al., *Functional analyses of troponin T mutations that cause hypertrophic cardiomyopathy: insights into disease pathogenesis and troponin function.* Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14406-10.

74. Tardiff, J.C., et al., *A truncated cardiac troponin T molecule in transgenic mice suggests multiple cellular mechanisms for familial hypertrophic cardiomyopathy.* J Clin Invest, 1998. **101**(12): p. 2800-11.

75. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-8.

76. Mardis, E.R., *Next-generation sequencing platforms.* Annu Rev Anal Chem (Palo Alto Calif), 2013. **6**: p. 287-303.

77. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology.* Trends Genet, 2014. **30**(9): p. 418-26.

78. Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-945.

79. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-380.

80. Valouev, A., et al., *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning.* Genome Research, 2008. **18**(7): p. 1051-1063.

81. Meder, B., et al., *Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies.* Circ Cardiovasc Genet, 2011. **4**(2): p. 110-22.

82. Lopes, L.R., et al., *Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing.* J Med Genet, 2013. **50**(4): p. 228-39.

83. Faita, F., et al., *Next generation sequencing in cardiovascular diseases.* World J Cardiol, 2012. **4**(10): p. 288-95.

84. Koboldt, Daniel C., et al., *The Next-Generation Sequencing Revolution and Its Impact on Genomics.* Cell, 2013. **155**(1): p. 27-38.

85. Xue, Y., et al., *Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing.* Genet Med, 2015. **17**(6): p. 444-51.

86. Yang, Y., et al., *Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders.* New England Journal of Medicine, 2013. **369**(16): p. 1502-1511.

87. Kiezun, A., et al., *Exome sequencing and the genetic basis of complex traits.* Nat Genet, 2012. **44**(6): p. 623-30.

88. Kiezun, A., et al., *Exome sequencing and the genetic basis of complex traits.* Nat Genet, 2012. **44**(6): p. 623-630.

89. Mali, P. and K.M. Esvelt, *Cas9 as a versatile tool for engineering biology.* 2013. **10**(10): p. 957-63.

90. Grissa, I., G. Vergnaud, and C. Pourcel, *The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats.* BMC Bioinformatics, 2007. **8**(1): p. 1-10.

91. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.

92. Bondy-Denomy, J. and A.R. Davidson, *To acquire or resist: the complex biological effects of CRISPR-Cas systems.* Trends Microbiol, 2014. **22**(4): p. 218-25.

93. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-607.

94. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

95. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

96. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat. Protocols, 2013. **8**(11): p. 2281-2308.

97. Chylinski, K., A. Le Rhun, and E. Charpentier, *The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems.* RNA Biology, 2013. **10**(5): p. 726-737.

98. Wu, S., et al., *A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond.* Nat. Protocols, 2008. **3**(6): p. 1056-1076.

99. Zhao, F., et al., *TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies.* Nucleic Acids Res, 2005. **33**(Database issue): p. D103-7.

100. Ma, H., et al., *Cell-specific expression of SERCA, the exogenous Ca2+ transport ATPase, in cardiac myocytes.* Am J Physiol Cell Physiol, 2004. **286**(3): p. C556-64.

101. Wang, G., H.I. Yeh, and J.J. Lin, *Characterization of cis-regulating elements and trans-activating factors of the rat cardiac troponin T gene.* J Biol Chem, 1994. **269**(48): p. 30595-603.

102. Cooper, T.A., *Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene.* Mol Cell Biol, 1998. **18**(8): p. 4519-25.

103. White, S.M., P.E. Constantin, and W.C. Claycomb, *Cardiac physiology at the cellular level: use of cultured HL-1 cardiomyocytes for studies of cardiac muscle cell structure and function.* Am J Physiol Heart Circ Physiol, 2004. **286**(3): p. H823-9.

104. Desmet, F.O., et al., *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic Acids Res, 2009. **37**(9): p. e67.

105. Philippi, S., et al., *Dysferlin rescue by spliceosome-mediated pre-mRNA trans-splicing targeting introns harbouring weakly defined 3' splice sites.* Hum Mol Genet, 2015. **24**(14): p. 4049-60.

106. Mendes de Almeida, R., et al., *Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy.* 2017. **12**(8): p. e0182946.

107. Authors/Task Force, m., 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, 2014. **35**(39): p. 2733-79.

108. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

109. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.

110. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.

111. Garrison, E.M., G, *Haplotype-based variant detection from short-read sequencing.* Genomics (q-bio.GN); Quantitative Methods (q-bio.QM), 2012.

112. Zook, J.M., et al., *Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls.* Nat Biotech, 2014. **32**(3): p. 246-251.

113. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

114. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* 2015. **17**(5): p. 405-24.

115. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.

116. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

117. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.

118. Salgado, D., et al., *UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution.* Hum Mutat, 2016. **37**(5): p. 439-46.

119. Choi, Y. and A.P. Chan, *PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels.* Bioinformatics, 2015. **31**(16): p. 2745-7.

120. Shihab, H.A., et al., *Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models.* Hum Mutat, 2013. **34**(1): p. 57-65.

121. Ritchie, G.R., et al., *Functional annotation of noncoding sequence variants.* Nat Methods, 2014. **11**(3): p. 294-6.

122. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.

123. Xiong, H.Y., et al., *RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease.* Science, 2015. **347**(6218): p. 1254806.

124. Smedley, D., et al., *A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease.* Am J Hum Genet, 2016. **99**(3): p. 595-606.

125. Chang, T.H., et al., *An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs.* BMC Bioinformatics, 2013. **14 Suppl 2**: p. S4.

126. Corvelo, A., et al., *Genome-wide association between branch point properties and alternative splicing.* PLoS Comput Biol, 2010. **6**(11): p. e1001016.

127. Vaz-Drago, R., N. Custodio, and M. Carmo-Fonseca, *Deep intronic mutations and human disease.* Hum Genet, 2017.

128. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

129. Rosenbloom, K.R., et al., *The UCSC Genome Browser database: 2015 update.* Nucleic Acids Res, 2015. **43**(Database issue): p. D670-81.

130. Rosenbloom, K.R., et al., *ENCODE data in the UCSC Genome Browser: year 5 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D56-63.

131. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

132. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

133. Alfares, A.A. and M.A. Kelly, *Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity.* 2015. **17**(11): p. 880-8.

134. Van Driest, S.L., et al., *Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy.* J Am Coll Cardiol, 2004. **44**(9): p. 1903-10.

135. Tajsharghi, H., et al., *Unexpected myopathy associated with a mutation in MYBPC3 and misplacement of the cardiac myosin binding protein C.* J Med Genet, 2010. **47**(8): p. 575-7.

136. Brito, D., et al., *Cardiac Anderson-Fabry disease: lessons from a 25-year-follow up.* Rev Port Cardiol, 2014. **33**(4): p. 247 e1-7.

137. Mather, C.A., et al., *CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel.* Genet Med, 2016. **18**(12): p. 1269-1275.

138. Carreno, J.E., et al., *[Cardiac hypertrophy: molecular and cellular events].* Rev Esp Cardiol, 2006. **59**(5): p. 473-86.

139. Yanazume, T., et al., *Cardiac p300 is involved in myocyte growth with decompensated heart failure.* Mol Cell Biol, 2003. **23**(10): p. 3593-606.

140. Cartegni, L., et al., *ESEfinder: a web resource to identify exonic splicing enhancers.* Nucleic Acids Research, 2003. **31**(13): p. 3568-3571.

141. Lopes, L.R., M.S. Rahman, and P.M. Elliott, *A systematic review and meta-analysis of genotype-phenotype associations in patients with hypertrophic cardiomyopathy caused by sarcomeric protein mutations.* Heart, 2013. **99**(24): p. 1800-11.

142. Sen-Chowdhry, S., et al., *Update on hypertrophic cardiomyopathy and a guide to the guidelines.* Nat Rev Cardiol, 2016. **13**(11): p. 651-675.

143. Ingles, J., et al., *Application of Genetic Testing in Hypertrophic Cardiomyopathy for Preclinical Disease Detection.* Circ Cardiovasc Genet, 2015. **8**(6): p. 852-9.

144. Morita, H., et al., *Shared genetic causes of cardiac hypertrophy in children and adults.* N Engl J Med, 2008. **358**(18): p. 1899-908.

145. Ho, C.Y., et al., *Genetic advances in sarcomeric cardiomyopathies: state of the art.* Cardiovasc Res, 2015. **105**(4): p. 397-408.

146. Authors/Task Force, m., 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, 2014. **35**(39): p. 2733-79.

147. Millat, G., et al., *Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy.* Eur J Med Genet, 2010. **53**(5): p. 261-7.

148. Popp, M.W. and L.E. Maquat, *Organizing principles of mammalian nonsense-mediated mRNA decay.* Annu Rev Genet, 2013. **47**: p. 139-65.

149. Ott, C.J., et al., *Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus.* Proceedings of the National Academy of Sciences, 2009. **106**(47): p. 19934-19939.

150. Visser, M., R.J. Palstra, and M. Kayser, *Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter.* Hum Mol Genet, 2015. **24**(9): p. 2649-61.

151. Biagini, E., et al., *Significance of sarcomere gene mutations analysis in the end-stage phase of hypertrophic cardiomyopathy.* Am J Cardiol, 2014. **114**(5): p. 769-76.

152. Granzier, H.L. and S. Labeit, *The giant protein titin: a major player in myocardial mechanics, signaling, and disease.* Circ Res, 2004. **94**(3): p. 284-95.

153. Linke, W.A., et al., *PEVK domain of titin: an entropic spring with actin-binding properties.* J Struct Biol, 2002. **137**(1-2): p. 194-205.

154. Weeland, C.J., et al., *Insights into alternative splicing of sarcomeric genes in the heart.* J Mol Cell Cardiol, 2015. **81**: p. 107-13.

155. Yin, Z., J. Ren, and W. Guo, *Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure.* Biochim Biophys Acta, 2015. **1852**(1): p. 47-52.

156. Schafer, S. and A. de Marvao, *Titin-truncating variants affect heart function in disease cohorts and the general population.* 2017. **49**(1): p. 46-53.

157. Buikema, J.W. and S.M. Wu, *Untangling the Biology of Genetic Cardiomyopathies with Pluripotent Stem Cell Disease Models.* Curr Cardiol Rep, 2017. **19**(4): p. 30.

158. Ross, S.B., S.T. Fraser, and C. Semsarian, *Induced pluripotent stem cells in the inherited cardiomyopathies: From disease mechanisms to novel therapies.* Trends Cardiovasc Med, 2016. **26**(8): p. 663-672.

159. Menon, S.C., et al., *Cardiac troponin T mutation in familial cardiomyopathy with variable remodeling and restrictive physiology.* Clin Genet, 2008. **74**(5): p. 445-54.

160. Chandra, M., et al., *Ca(2+) activation of myofilaments from transgenic mouse hearts expressing R92Q mutant cardiac troponin T.* Am J Physiol Heart Circ Physiol, 2001. **280**(2): p. H705-13.

161. Anan, R., et al., *Patients with familial hypertrophic cardiomyopathy caused by a Phe110Ile missense mutation in the cardiac troponin T gene have variable cardiac morphologies and a favorable prognosis.* Circulation, 1998. **98**(5): p. 391-7.

162. Watkins, H., et al., *Mutations in the genes for cardiac troponin T and alpha-tropomyosin in hypertrophic cardiomyopathy.* N Engl J Med, 1995. **332**(16): p. 1058-64.

163. Santos, C., *Exploring new therapeutic approaches for Hypertrophic cardiomyopathy*, in *Biologia (Biologia Humana e Ambiente)*. 2014, Universidade de Lisboa, Faculdade de Ciências.

164. Marks, H., et al., *The Transcriptional and Epigenomic Foundations of Ground State Pluripotency.* Cell. **149**(3): p. 590-604.

165. Ying, Q.-L., et al., *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-523.

166. Tsuji, Y., et al., *Maintenance of undifferentiated mouse embryonic stem cells in suspension by the serum- and feeder-free defined culture condition.* Dev Dyn, 2008. **237**(8): p. 2129-38.

167. Boheler, K.R., et al., *Differentiation of pluripotent embryonic stem cells into cardiomyocytes.* Circ Res, 2002. **91**(3): p. 189-201.

168. Claycomb, W.C., et al., *HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2979-84.

169. Ahmad, F., et al., *The role of cardiac troponin T quantity and function in cardiac development and dilated cardiomyopathy.* PLoS One, 2008. **3**(7): p. e2642.

170. Nishii, K., et al., *Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo.* Dev Biol, 2008. **322**(1): p. 65-73.

171. Franco, D., et al., *[Regulation of myocardial gene expression during heart development].* Rev Esp Cardiol, 2002. **55**(2): p. 167-84.

172. Terami, H., et al., *Efficient capture of cardiogenesis-associated genes expressed in ES cells.* Biochem Biophys Res Commun, 2007. **355**(1): p. 47-53.

173. Xin, M., E.N. Olson, and R. Bassel-Duby, *Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair.* Nat Rev Mol Cell Biol, 2013. **14**(8): p. 529-41.

174. Rojas, A., et al., *GATA4 is a direct transcriptional activator of cyclin D2 and Cdk4 and is required for cardiomyocyte proliferation in anterior heart field-derived myocardium.* Mol Cell Biol, 2008. **28**(17): p. 5420-31.

175. Lin, Q., et al., *Control of Mouse Cardiac Morphogenesis and Myogenesis by Transcription Factor MEF2C.* Science, 1997. **276**(5317): p. 1404-1407.

176. Li, X., et al., *Calcineurin-NFAT signaling critically regulates early lineage specification in mouse embryonic stem cells and embryos.* Cell Stem Cell, 2011. **8**(1): p. 46-58.

177. Schulz, R.A. and K.E. Yutzey, *Calcineurin signaling and NFAT activation in cardiovascular and skeletal muscle development.* Developmental Biology, 2004. **266**(1): p. 1-16.

178. Worsdorfer, P., et al., *Connexin expression and functional analysis of gap junctional communication in mouse embryonic stem cells.* Stem Cells, 2008. **26**(2): p. 431-9.

179. Young Lee, M., et al., *High density cultures of embryoid bodies enhanced cardiac differentiation of murine embryonic stem cells.* Biochemical and biophysical research communications, 2011. **416**(1-2): p. 51-57.

180. Doi, K., et al., *Expression of natriuretic peptide system during embryonic stem cell vasculogenesis.* Heart Vessels, 1997. **Suppl 12**: p. 18-22.

181. Miao, L., et al., *Atrial Natriuretic Peptide Regulates Ca2+ Channel in Early Developmental Cardiomyocytes.* PLoS ONE, 2010. **5**(1): p. e8847.

182. Sanchez, A., et al., *Myosin heavy chain gene expression in mouse embryoid bodies. An in vitro developmental study.* J Biol Chem, 1991. **266**(33): p. 22419-26.

183. Ban, K., et al., *Purification of cardiomyocytes from differentiating pluripotent stem cells using molecular beacons that target cardiomyocyte-specific mRNA.* Circulation, 2013. **128**(17): p. 1897-909.

184. Mummery, C., et al., *Cardiomyocyte differentiation of mouse and human embryonic stem cells.* Journal of Anatomy, 2002. **200**(3): p. 233-242.

185. Li, K., et al., *Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System.* PLoS ONE, 2014. **9**(8): p. e105779.

186. Maruyama, T., et al., *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.* Nat Biotech, 2015. **33**(5): p. 538-542.

187. Farboud, B. and B.J. Meyer, *Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design.* Genetics, 2015. **199**(4): p. 959-71.

188. Guatimosim, S., C. Guatimosim, and L.-S. Song, *Imaging Calcium Sparks in Cardiac Myocytes.* Methods in molecular biology (Clifton, N.J.), 2011. **689**: p. 205-214.

189. Elliott, P.M., 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), 2014.

190. Berger, A., et al., *mRNA trans-splicing in gene therapy for genetic diseases.* Wiley Interdiscip Rev RNA, 2016. **7**(4): p. 487-98.

191. D'Allard, D. and J. Liu, *Towards RNA Repair of Diamond-Blackfan Anemia Hematopoietic Stem Cells.* Hum Gene Ther, 2016.

192. Thierfelder, L., et al., *Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere.* Cell, 1994. **77**(5): p. 701-12.

193. Rubattu, S., et al., *A Next-Generation Sequencing Approach to Identify Gene Mutations in Early- and Late-Onset Hypertrophic Cardiomyopathy Patients of an Italian Cohort.* Int J Mol Sci, 2016. **17**(8).

194. Millat, G., et al., *Functional characterization of putative novel splicing mutations in the cardiomyopathy-causing genes.* DNA Cell Biol, 2015. **34**(7): p. 489-96.

195. Marisa, O. and A.-S. Katriina, *Modeling Hypertrophic Cardiomyopathy with Human Induced Pluripotent Stem Cells*. Pluripotent Stem Cells - From the Bench to the Clinic. 2016.

8. Supplementary Information