University of Southampton

Faculty of Medicine

Human Development and Health

Characterisation and interpretation of genetic variants by functional splicing assays in cardiovascular disorders

by

Ilenia Marino

ORCID ID: 0009-0009-8022-7742

Thesis for the degree of Doctor of Philosophy

June 2024

University of Southampton

Abstract

Faculty of Medicine

Human Development and Health

[Doctor of Philosophy]

Characterisation and identification of genetic variants by functional splicing assays in cardiovascular disorders

by

Ilenia Marino

Advancements in sequencing technologies and their application to patients' care have revolutionised human genetics. A huge number of genetic variants and their association with human diseases have been identified, in particular for cardiovascular disorders (CVDs). CVDs represent a major cause of sudden death, due to genetic alterations of genes encoding for proteins mainly involved in sarcomeric and cytoskeletal function. While genetic variants altering protein function clearly cause disease, not all genetic variants identified in patients are necessarily linked to the development of a cardiac disorders, thus a classification following standard criteria has been introduced at the clinical level by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) which includes classifications of Pathogenic, Likely Pathogenic, and Variants of Unknown Significance (VUS) for variants where is unclear whether they will exert a functional effect on gene expression or function.

Many diseases causing variants do not lie within protein coding segments of the genome and cause disease through alteration in RNA transcription and/or processing. For example,38 to 50% of disease causing variants are associated with altered mRNA splicing (Bryen *et al.*, 2022).

The evaluation of the potential pathogenicity of variants that may affect mRNA splicing is not easy due to the complexity of splicing mechanisms, providing a significant motivation for conducting functional studies to elucidate the effects of gene variants on splicing.

In this thesis, *in silico*, *in vitro* and *in vivo* functional RNA splicing analysis of 24 variants identified in potential candidate genes in patients with CVDs enabled the reclassification of58% Variants of Unknown Significance and 25% of likely pathogenic variants in pathogenic. These findings strongly support the introduction of these assays to determine the pathogenicity of splicing variants in clinical settings.

Additionally, these techniques may allow the development of treatments for genetic diseases with defective splicing. For example, the use of small molecule splicing modulators regulating different steps of mRNA splicing. Here, six small molecules were tested on a series of minigene constructs, to determine whether they can rescue effects of pathogenic variants on mRNA splicing of cardiac genes. Of the compounds tested, 6-Benzyladenine and Branaplam were the most efficient in restoring splicing. Altogether, these results demonstrated that small molecule compounds could represent a valid therapeutic approach by correcting aberrant splicing.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Ilenia Marino

Title of thesis:Characterisation and identification of genetic variants by functional splicing assays in cardiovascular disorders

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature: ..... Date:.....

Acknowledgements

Ph.D. is difficult, no matter how well prepared you image yourself to be, and scientific research requires curiosity, a lot of energy, time and exceptionally hard work.

I am extremely grateful to my PhD supervisors Dr. Diana Baralle, Dr. John Hollowey, and Dr. Marco Baralle, for all that you have done to get me here today. Thank you for giving me a chance to expand my knowledge and repertoire of laboratory skills. It has been an honour under your guidance to learn how to be a scientist.

I would also like to recognise the contributions of the entire team of the Cardiologist Unit (Cattinara Hospital, Italy) and Wessex Laboratory (UK) for their support and collaboration, without whom this study would not have been possible. Thanks to all the individuals who were part of the cohort recruited into this project.

To Dr. Dave Bunyan, thank you for your constant support for variant classification that I was not qualified to lead.

To Carolina and Ian, thank you for your patience and guidance for performing RNA-sequencing analysis. I am grateful for providing me the knowledge necessary to start coding.

Thanks to my friends, Antonio, Himanshi, Lubna, and Yasmine, for being part of this crazy adventure. Thanks for always supporting me, commiserating with me, and reminding me that it is always a good time for lightheartedness. You guys made this experience feel more fun, even in the down moments.

To my labmates and colleagues that I met in these years, thank you for sharing your experience and ideas. You made this experience fell more human and exposed me to the beautiful culture of the world.

To my family – Dad, Mom, and big brother – I am so grateful for your support through my life. Thanks for believing in my ambition and always encouraging me to achieve my goal. Your unwavering love is the source of my strength and motivation.

Ilenia

"Science never solves a problem without creating ten more."

(George Bernard Shaw)

Abbreviations

ACMG/AMP ..... American College of Medical genetics and Genomics/Association for Molecular Pathology

ADPKD ..... Autosomal dominant polycystic kidney disease

APA..... Alternative polyadenylation

ARVC ..... Arrhythmogenic right ventricular cardiomyopathy

AS..... Alternative splicing

ASOs ..... Antisense oligonucleotides

ASS ..... Acceptor splice site

BLASTs ..... Basic local alignment search tool

bp..... base pair

BPS..... Branch point site

6-B..... 6-Benzyladenine

BrS ..... Brugada syndrome

cDNA ..... complementary DNA

CF ..... Cystic fibrosis

CFTR..... Cystic fibrosis transmembrane conductance regulator

CIP ..... Alkaline Calf Intestinal Phosphatase

circRNA ..... circular RNA

CK ..... Cytokinin

CNN ..... Convolutional neuronal network

CPVT..... Catecholaminergic polymorphic ventricular tachycardia

CRT..... Control

CVDs ..... Cardiovascular disorders

DCM ..... Dilated cardiomyopathy

DL..... Deep learning

DMD ..... Dystrofin

DNA..... Deoxyribonucleic Acid

DSP ..... Desmoplakin

dsRNA ..... double-stranded RNA

DSS ..... donor splice site

ESE ..... Exonic splicing enhancers

ESS ..... Exonic splicing silencer

ex ..... Exon

ExAC ..... Exome Aggregation Consortium

FGA ..... Feature generation algorithm

FGFR3 ..... Fibroblast growth factor receptor 3

FLNC ..... Filamic C

FMR1 ..... Fragile X messenger ribonucleoprotein 1

FTDP-17 ..... Frontotemporal dementia with parkinsonism-17

FW ..... Forward

gnomeAD ..... Genome Aggregation Database

HCM ..... Hypertrophic cardiomyopathy

HDAC ..... Histone deacetylase

HMM ..... Hidden Markov model

hnRNP ..... Heterogeneous nuclear ribonucleoproteins

IKBKAP ..... Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein

INDEL ..... insertion/deletion

ISE ..... Intronic splicing enhancer

ISS ..... Intronic splicing silencer

LMNA ..... Lamin

LoF ..... Loss of function

LP ..... Likely Pathogenic

LQTS ..... Long QT syndrome

MDD ..... Maximal Dependence Decomposition

MYBPC3 ..... Myosin binding protein C3

ML ..... Machine learning

MM ..... Markov model

M-MLV RT..... Moloney Murine Leukemia Virus Reverse Transcriptase

MPS ..... Massively parallel sequencing

mRNA ..... messenger RNA

NaBu ..... Sodium Butyrate

NEXN ..... Nexillin F-actin binding protein

NGS ..... Next generation sequencing

NHS (UK)..... National Health Service (United Kingdom)

NMD ..... Nonsense Mediated Decay

NN..... Neural network

nt..... nucleotides

OMIM ..... Online Mendelian Inheritance in Man

p. ..... Protein

PAM ..... Protospacer adjacent motif

PCR ..... Polymerase chain reaction

PKD1 ..... Plycystin-1

PKU ..... Phenylketonuria

PPT ..... Polypyrimidine tract

pre-mRNA..... primary transcript

PT ..... Patient

PTC..... Premature termination codon

PTM ..... pre-*trans*-splicing molecule

PWM..... Position weight matrices

QC..... Quality control

RBC ..... Red blood cell

RCM ..... Restrictive cardiomyopathy

RISC ..... RNA-induced silencing complex

RNA ..... Ribonucleic Acid

RNAi ..... RNA interference

RNA-seq ..... RNA sequencing

RRM..... RNA recognition motif

rRNA..... ribosomal RNA

RT-PCR ..... Reverse-transcriptase polymerase chain reaction

RT-qPCR ..... Quantitative reverse-transcriptase polymerase chain reaction

RYR2..... Ryanodine receptors 2

RyRs ..... Ryanodine receptors family

RP..... Retinitis pigmentosa

RV..... Reverse

SD ..... Standard deviation

siRNAs..... Small interfering RNA

SMA..... Spinal muscular atrophy

SMaRT ..... Spliceosome-mediated RNA *trans*-splicing

SMN2..... Survival motor neuron

snRNPs..... Small nuclear ribonucleoprotein

SNVs..... Single Nucleotide Variations

SQTS..... Short QT syndrome

SR..... Serine/Arginine (SR)-rich proteins

SREs ..... Splicing regulatory elements

SVM ..... Support vector machine

TALEN..... Transcription activator-like nucleases

TMEM43..... Transmembrane protein 43

TTN ..... Titin

VCL..... Vinculin

VPA..... Valproic acid

VUS ..... Variant with Uncertain Significance

Abbreviations

WES..... Whole exome sequencing

WGS ..... Whole genome sequencing

wt ..... wild type

5’ ss / 3’ ss ..... 5’ splice site /3’ splice site

ZFPs ..... Zing finger proteins

Chapter 1 Introduction

1.1 Genomic approaches in clinics

Advances in DNA sequencing, reduction in medical costs, and national and international coordinated projects have revolutionised the diagnostic rate of human diseases by enabling the incorporation of next-generation sequencing (NGS) technologies in the procedure.

To diagnose genetic diseases there are several options depending on clinical context (Figure 1). Single-gene testing is the chosen method when the symptom of the patient is unique for the suspected condition and there is established evidence of an association between the disease and a particular gene. It is crucial that clinicians have the clinical expertise to suspect a gene based on clinical findings. For example, this approach is applicable for sequencing *CFTR* for cystic fibrosis (CF) (Rosenstein and Cutting, 1998), *FGFR3* for achondroplasia (Xue *et al.*, 2014), *PAM* for Phenylketonuria (PKU) (Trefz, Lichter-Konecki and Konecki, 1989), and *FMR1* for Fragile-X syndrome. While for heterogeneous diseases in which the patient has symptoms that fit a wide range of conditions, gene panel testing may be used to target genes which can be associated with the disease. The number of genes included in the panel can be from two to >1000 and depends on evidence of disease association and stringency of inclusion criteria determined by testing laboratories (e.g. epilepsy gene panels range from 70 to 377 sequenced genes)(Xue *et al.*, 2015). When there is a multiplicity of phenotypes with indistinct features or genetic cause is unclear, whole exome sequencing (WES) and whole genome sequencing (WGS) are more useful for diagnosis. These approaches may also identify differences throughout the genome not yet associated with disease and clarify disease phenotypes. This is made possible by the ability to sequence millions of small DNA fragments in parallel which allows quick efficient evaluation of several genes or a complete genome (Behjati and Tarpey, 2013). WES and WGS have been utilised for Autism (Ku *et al.*, 2013), Kabuki syndrome (Zaidi *et al.*, 2013), Zellweger syndrome (Majewski *et al.*, 2011), or Oculocutaneous albinism and neutropenia (Cullinane *et al.*, 2011) and are now being adopted as a standard approach in many health systems including the NHS in the UK (NHS, 2021).

WES improves variant classification by only sequencing the protein coding regions of the human genome (~1.5% of genome), which contains ~85% of all identified disease-causing variants (Barbitoff *et al.*, 2020; Taylor, Alloub and Tayoun, 2021). However, the drawback of WES is that classes of variants that are responsible for the disease may be missed due to incomplete sequencing coverage (Meynert *et al.*, 2014). In fact, today we know that the non-coding genome plays an essential role in gene regulation carrying sequence essential for the splicing process

(see section 1.3). Sequence variants at this level may also have an impact on RNA processes such as pre-mRNA splicing and transcription, with additional RNA studies needed to validate a functional impact of the variant to the gene product.

Unlike WES, WGS covers the entire genome improving the diagnostic rate to ~40-60%,in part due to its capacity to detect copy number variants (CNVs), structural rearrangements and repeat expansions (Dolzhenko *et al.*, 2017; Bick *et al.*, 2019).

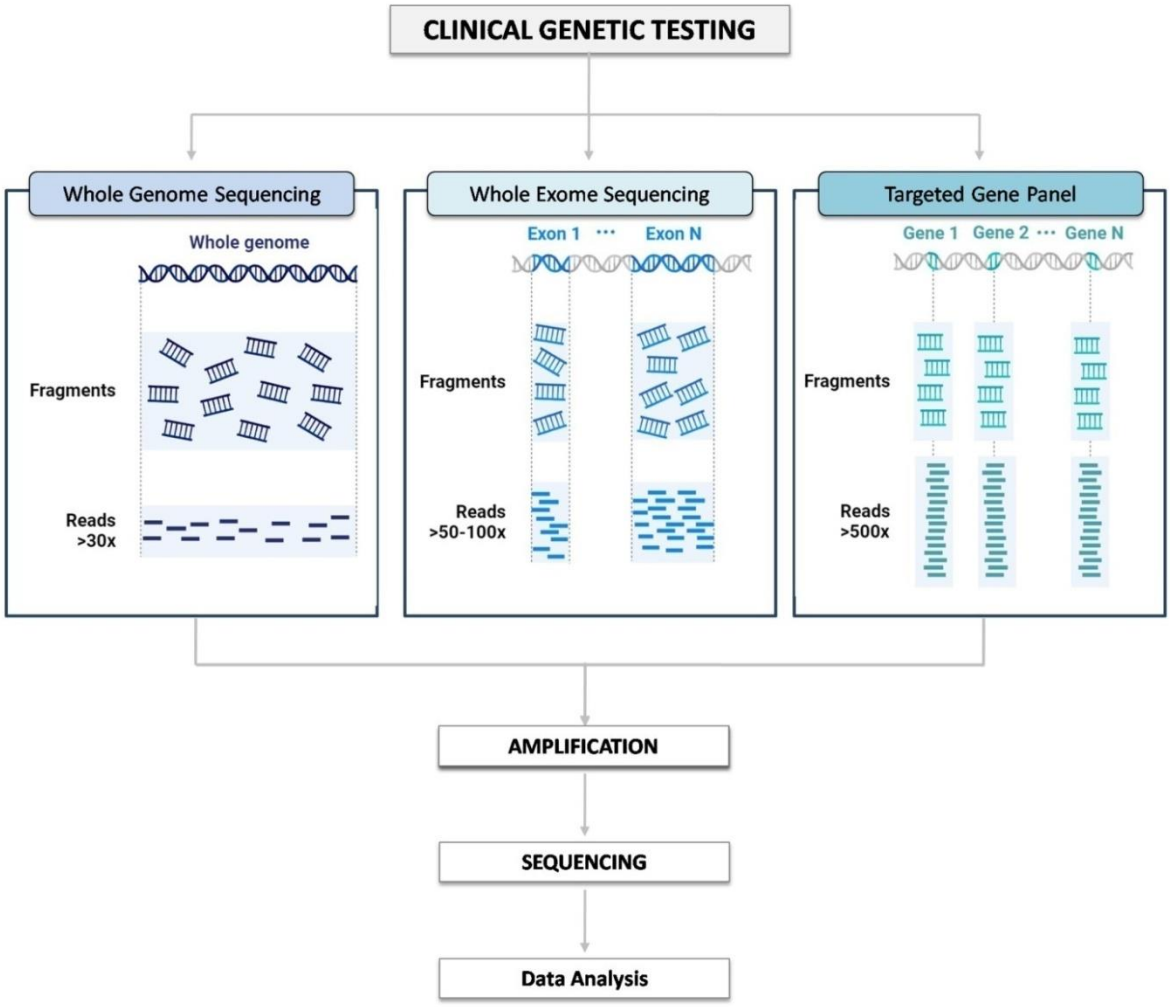


Figure 1. Approaches to genetic testing.

1.2 Clinical variant classification

As described above, technological advances in genome sequencing have led to its application for diagnosis of genetic disorders. This has resulted in the emergence of a new challenge for variant classification (Ackerman, 2015). Using WES and, particularly WGS, results in the identification of many variants in the patient's genome where the potential role for them to affect gene expression and/or function is unclear. These are termed Variants of Unknown Significance (VUSs).

Currently the scale of the VUS problem is colossal; there are 1,279,333 VUSs on the ClinVar database (Updated on 30 December 2023) representing 53 percent of total variants (Henrie *et al.*, 2018). Distinguishing if these are disease-causing from the over 4.5 million single nucleotide variants (SNVs), small insertions and deletions that are located in the non-coding regions of the genome, whose functional impact is unclear, can be challenging. The amount of non-coding variants detected by WGS, combined with a lack of understanding of non-coding sequences, makes effective candidate variant prioritisation extremely difficult (Gloss and Dinger, 2018; French and Edwards, 2020). In fact, approximately half of people with a disease remain without a genetic diagnosis despite the application of WGS (Mattick *et al.*, 2018).

To aid in the interpretation of nucleotide changes, several databases dedicated to the collection and storage of information concerning genomic variants exist. Examples include the Exome Aggregation Consortium (ExAc) (Lek *et al.*, 2016) and Genome Aggregation Database (gnomeAD) where variants are reported and reclassified. Furthermore, a set of criteria for interpreting sequence variants, to provide support for or against the pathogenicity of a variant, has been introduced by the American College of Medical Genetics and Genomics and the Association of Medical Pathologists (ACMG/AMP), which assist clinical genetic testing laboratories in standardisation of the variant classification system and interpretation (Richards *et al.*, 2015).The variant interpretation guideline considers patient phenotype, population data, segregation and allelic evidence, and functional assessment for each variant present in the literature (Richards *et al.*, 2015). The provided criteria consider the level of strength are divided into two sets, 12 Benign or likely benign criteria and 16 Pathogenic and likely pathogenic criteria: benign supporting (BP1-7), benign strong (BS1-4), stand-alone (BA1), or pathogenic supporting (PP1-5), moderate (PM1-6), strong (PS1-4), or very strong (PVS1). The number inside the code does not indicate any strength differences but helps to refer to the different criteria (Figure 2A). These criteria can be combined according to the scoring rules provided in the ACMG/AMP guidelines for classifying genetic variants as pathogenic, likely pathogenic, uncertain significance, likely benign and benign (Figure 2B).

A

		Benign criteria		Pathogenic criteria			
		Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very Strong**
Odds of pathogenicity*		-18.7	-2.08	2.08	4.33	18.7	350.0
ACMG/AMP evidence codes	Population	BA1, BS1, BS2			PM2	PS4	
	Allelic Evidence		BP2		PM3		
	Segregation	BS4		PP1	Increased segregation data		
	De novo data				PM6	PS2	
	Computational and Predictive data		BP1, BP3, BP4, BP7	PP3	PM4, PM5	PS1	PVS1
	Functional assessment	BS3		PP2	PM1	PS3	
	Other databases		BP6	PP5			
	Other data		BP5	PP4			

B

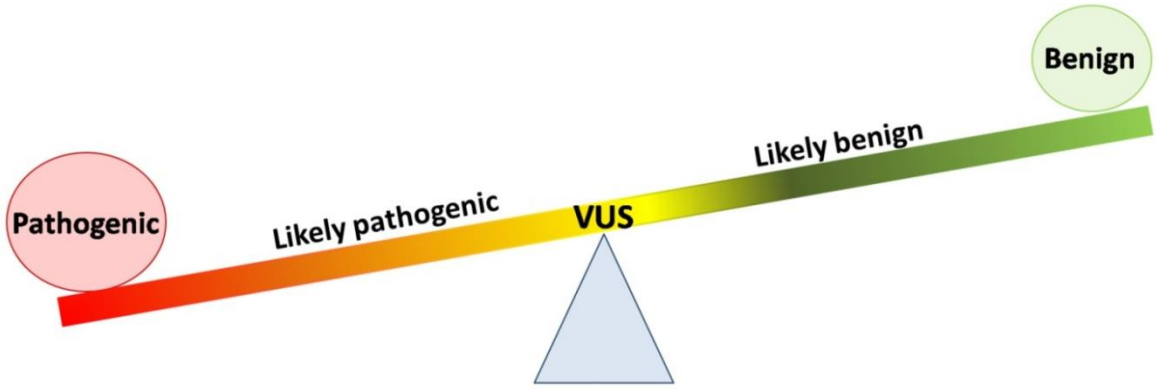


Figure 2. Genetic variant classification. (**A**) Adapted from (Richards *et al.*, 2015; Strande *et al.*, 2018). The table represents the ACMG/AMP criteria organised by type of evidence which supports a benign (b, left side) or pathogenic (P, right site) classification (first letter of code) and given a relative strength (second letter of code): A, stand-alone; P, supporting; M, moderate; S, strong; VS, very strong. For a complete description of the criteria see (Richards *et al.*, 2015).* Odds of pathogenicity corresponding to evidence strength, assuming a prior probability of 0.10 as would be anticipated for single-gene testing. BA1 is not included in the Bayes model used to derive the odds of pathogenicity (Tavtigian *et al.*, 2018). (**B**) Variants are classified into 5 categories based on a range from benign to pathogenic according to the ACMP/AMP criteria.

A pathogenic variant affects gene function with deleterious impact on protein function, resulting in loss of activity and thus causing disease. The definition of a pathogenic variant is supported

by the available evidence which demonstrates that the genome difference impacts on protein production. While a variant is called benign when there is no deleterious influence on the protein. Sometimes there is evidence that supports a pathogenic classification, but not enough to classify the variant as pathogenic. These variants are classified as likely pathogenic. Similarly, a likely benign classification is assigned to variants that have some evidence supporting a benign classification, but not enough to be classified as benign(Karczewski *et al.*, 2019; Wong *et al.*, 2019). Finally, if there is limited clinical and experimental evidence, such as a missense variant or insertion/deletion (INDEL) being either pathogenic or benign, a variant is classed as a VUS (Karczewski *et al.*, 2019). A number of VUS may represent "normal genetic background noise" that exists in the healthy population. However, it is important to distinguish VUS that could be potentially pathogenic from normal genetic background noise (Giudicessi and Ackerman, 2013). This distinction is critical, the results of genetic testing are used for medical care decisions and inform patients and their family members. VUSs are often difficult to interpret for clinical purposes and further analysis is necessary to clarify the role of VUS in disease (Christiaans *et al.*, 2019; Karczewski *et al.*, 2019).

Interestingly, most pathogenic variants tend to be in coding sequences whereas VUSs can occur in both coding and non-coding sequences. This is likely due to the fact that in-depth knowledge of coding sequences, thanks to the fixed amino acid triplet code, makes deciphering pathogenic coding sequence variants relatively easier, and many clinical sequencing strategies concentrate on coding regions. On the other hand, VUSs can also occur in introns, and disrupt *cis* and *trans* regulatory elements, pseudogenes, repeat sequences and telomeres. Current knowledge of these sequences is lacking compared to coding DNA sequence and hinders the efforts to diagnose the VUSs in non-coding areas of the genome.

One type of VUS, often overlooked in clinical setting are VUSs affecting RNA processing, notwithstanding its central role in the classical dogma of biology (Crick, 1970). Indeed, implementing RNA analysis into VUSs analyses has been shown to increase the diagnostic rate by 7.5% to 36% through differential gene expression, splicing and allelic specific expression depending on disease type and tissue source (Murdock, 2020). RNA splicing analysis alone can increase VUS diagnostic rate by 33-35%(Cummings *et al.*, 2017; Wai *et al.*, 2020). Despite these recent studies on RNA splicing by VUSs, specific guidelines for consistent splicing VUS interpretation are yet to be established, although functional assays such as reverse-transcriptase polymerase chain reaction (RT-PCR) (Wai *et al.*, 2020) and minigene analysis (see sections 1.6.2) (Raponi *et al.*, 2014) are recommended as validation matrices for RNA splicing disrupting VUSs (Brnich *et al.*, 2019).

The extent of splicing affecting VUSs in the current VUSs pool is unknown. It was predicted that up to 62% of VUSs can affect splicing (López-Bigas *et al.*, 2005). In contrast, Invitae, reported only 5.4% of their private database VUSs, 4.8% of ClinVar VUS, and 9.4% of gnomAD VUS affect splicing (Truty *et al.*, 2021). These contradictory findings highlight the complexity of the "Splicing Code" (M. Baralle and Baralle 2018) that needs to be considered when defining a splicing disrupting variant. In general, however, the role of VUS affecting splicing is often underestimated in a clinical setting and implementing RNA splicing analysis in routine genetic testing can uplift the diagnostic rate of rare genetic diseases.

The ACMG/AMP guidelines do not give any recommendations on how to interpret splice variants and do not include splicing predictions tools and/or functional studies but only consider alterations in the open reading frame, deletion of one or more exons, or potential to activate a cryptic splice site.

To aid this limitation, several revisions to the ACMG/AMP guidelines have been published in order to improve variant classification and introduce criteria for specific disorders and functional approaches in which *in silico* prediction tools and RNA analysis are included in systematic interpretation.

While the updated revisions provide a strong framework for Mendelian variants, the application of the ACMG/AMP criteria lack of guidance on how to account for mRNA assay data. For example, for variant interpretation relating to potential impact splicing the PVS1 (Pathogenic, Very strong) criterion, defined as "null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss-of-function (LoF) is a known mechanism of disease" (Richards *et al.*, 2015), may be applied when loss of function is the predicted outcome for the essential splice site variants, even without RNA studies. Conversely, when considering functional assay PS3 or BS3 criteria are considered appropriate, but the assays should provide strength evidence of the variant effect on the gene. The "PS3 well-established functional studies" criteria are applied for splicing variants where RNA analysis has confirmed loss of function and the variant classification results are supported by strong evidence. In 2019, the PS3 criteria were updated and take into consideration the use of minigene assays, but these may not represent the proper biological environment. This has led to suggestions that further guidance on how to correctly apply PS3 and PVS1 criteria to splicing assays is required (Brnich *et al.*, 2019). Genetic variants can also have an impact on reading frame, and produce transcripts that can be targeted by Nonsense Mediated Decay (NMD), hence this event should also be considered during variant classification (Abou Tayoun *et al.*, 2018).

Optimising RNA analysis for use in clinics will raise the sensitivity of genomic testing as well as provide conclusive genetic diagnoses.

1.3 Splicing overview

In eukaryotic cells, transcription occurs in the nucleus by RNA polymerase II enzyme (RNA Pol II). Nascent mRNA transcripts (pre-mRNA) undergo modifications to yield mature mRNA transcripts. The main pre-mRNA maturation steps include 5'-capping, intron splicing, and polyadenylation of the 3' end where a poly-A tail is synthesised. The mature mRNA is then transported from the nucleus to the cytoplasm where translation occurs. Sometimes pre-mRNA transcripts are shuttled to other compartments, such as stress-granules or P-bodies for sequestration and/or degradation (Kasprowicz-Maluśki *et al.*, 2016; Nachtergaele and He, 2017).

1.3.1 Pre-mRNA maturation mechanism

A central step in mRNA maturation is splicing, this regulates gene expression in a co- and post-transcriptional way. During transcription, the genomic information is entirely copied into the pre-mRNA which consists of coding sequences (exons) interrupted by non-coding sequences (introns) (Adams, Rudner and Rio, 1996). Intron removal or splicing is characterised by two consecutive transesterification reactions (Figure 3).

In the first reaction, the 5' end of the intron is cleaved at the intron-exon junction resulting in an intermediate composed of an upstream exon that exposes its 3' end and an intron in the form of a lariat. This catalytic step involves the adenosine located at the branch point site (BPS) where a nucleophilic attack is performed by its 2'hydroxyl group (2'-OH) towards the phosphodiester bond of the 5' splice site. The result is a 2'-5' phosphodiester bond in the intron and a free 3'-OH group of the upstream exon. Subsequently, in the second trans-esterification, the 3' splice site is attacked by the 3'-hydroxyl group of the 5' exon, resulting in the joining of the two exons and the release of the intron in lariat form (Brow, 2002).

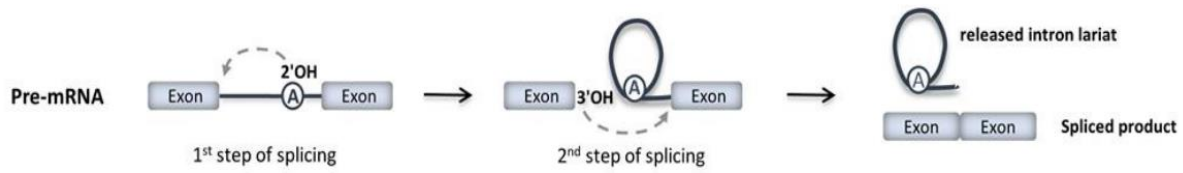


Figure 3.Schematic representation of the two transesterification reactions that result in displaced intron and joined exons from pre-mRNA. Reproduced from (Ribeiro *et al.*, 2020).

In addition, the length of the mature mRNA is less than the pre-mRNA (Baralle and Buratti, 2017).In mammals, the vast majority of coding sequences are interrupted by long introns (Black, 2003). However, there is a significant difference between unicellular eukaryotes and multicellular eukaryotes in terms of intron density and size distribution (Rogozin *et al.*, 2012). So, the composition of exon-intron into pre-mRNA is key for the splice site recognition efficiency (De Conti, Baralle and Buratti, 2013).

1.3.2 RNA splicing regulation

1.3.2.1 The catalytic step of splicing: spliceosome assembly

The RNA cleavage and ligation processes in splicing are regulated by a huge dynamic machine called the spliceosome. This is formed by the progressive interaction of five small nuclear ribonucleoprotein particles (U1, U2, U4, U5, and U6 snRNPs), and more than 200 proteins (non-snRNPs splicing factors). The interactions across these factors lead to the formation of two possible complexes: minor spliceosome or U12, and the major spliceosome (Wahl, Will and Lührmann, 2009).Biochemical studies have demonstrated that a new spliceosome is repeatedly recruited every time, on each intron of a pre-mRNA through base-pairing interactions between the snRNPs and the conserved sequences that define the exon-intron boundaries(Wahl, Will and Lührmann, 2009; Will and Lührmann, 2011).

The machinery assembly, represented in Figure 4, begins with the identification of the 5’ splice site by base-pairing of U1 snRNP. This interaction originates the initial E complex, which also includes the binding of a protein named splicing factor 1 (SF1) and heterodimer U2AF65/U2AF35to the BPS, polypyrimidine tract, and 3’ acceptor splice site (AG) (Konarska, 1998; Wahl, Will and Lührmann, 2009). Subsequently, the splicing factor U2 snRNP is recruited which leads to the formation of complex A. After this, U4-U6-U5 tri-snRNP joins the pre-spliceosome complex to form the inactive state of complex B (Daguenet, Dujardin and Valcárcel, 2015). At this step, enzymatic reactions occur through a series of conformational changes and compositional rearrangements to form the active complex B. These rearrangements include the release and replacement of the U1 snRNP instead of the U6 which can interact at the 5’ splice site. Simultaneously, an internal stem-loop is created as a critical metal-binding platform and a base-paired complex with U2 snRNP, forming the complex C. Subsequently, the dissociation of the U4 snRNPs permits the U6 and U2 snRNPs interaction with the formation of the catalytically active site of the spliceosome, where the first catalytic transesterification reaction occurs. Following which, the U5 snRNP protein, close to the active core of spliceosome, leads to the second transesterification step with intron excision and exon joining (Konarska, Vilardell and Query, 2006; Wahl, Will and Lührmann, 2009; Will and

Lührmann, 2011; Daguenet, Dujardin and Valcárcel, 2015). Finally, the spliceosome disassembles, and the excised intron is degraded.

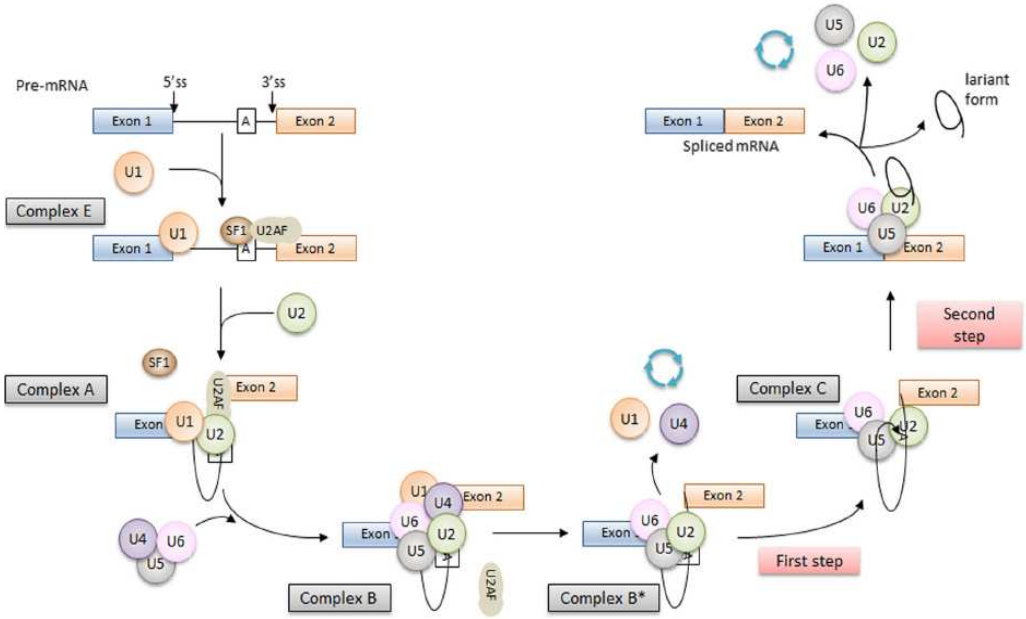


Figure 4. The spliceosome assembly and disassembly cycle. This depicts the stepwise interaction of the spliceosomalsmall nuclear ribonucleoprotein particles (snRNPs) (coloured circles) in the removal of an intron from a pre-mRNA containing two exons (blue and orange); non-snRNP proteins are not represented. The names of the complexes, as well as the first and second catalytic steps are shown. Reproduced from (Anna and Monika, 2018).

The vast majority of introns are recognised and removed by the major spliceosome, while a minority of introns are spliced by the minor spliceosome. In general, the major and minor spliceosome machinery have features in common and the splicing process is verysimilar. One of the major differences is that the four distinctsnRNPs termed U11, U12, U4atac, and U6atatc compose the minor spliceosome, instead of the U1, U2, U4, and U6 respectively (Verma *et al.*, 2018).

1.3.2.2 Splicing regulatory elements and *trans* acting factor

The spliceosome is regulated through the cooperative interaction of auxiliary splicing elements, *cis*-elements (splicing regulatory elements – SREs) on the pre-mRNA and *trans*-binding protein factors. These interactions are due to the size of the human exon which is comprised of between 50-250bp, which is much shorter with respect to the thousands of residues of introns (Berget, 1995; Chen and Manley, 2009).

Interestingly, exon-intron borders are characterised by distinctive conserved sequences which have the ability to determine splicing choices. The core splicing signals are determined by elements localised in every intron in precursor RNAs. These elements include the 5' splice

site(5’ ss), the 3’ splice site (3’ ss), the branch point site (BPS - located upstream of 3’ splice site), and the polypyrimidine tract located between BPS and 3’ ss(Breathnach *et al.*, 1978; Matera and Wang, 2014; Beqqali, 2018) (Figure 5). Studies have identified that, in mammalian genomes, over 98% of splice sites utilise GU-AG as the canonical splice sites and around 1% have GC-AG (Burset, Seledtsov and Solovyev, 2000).

The 5’ ss or splice donor site defines the exon:intron junction at the 5’end of the intron in the pre-mRNA. The consensus sequence consists of a moderately conserved 9 nucleotide motif [C/A)AG/GURAGU (where R is A or G)] and spanning positions -3 to +6, which is the core sequence to bind the U1 snRNP. The GU dinucleotide is conserved phylogenetically and defines the exon:intron junction. In general, variants affecting GU dinucleotides lead to an aberrant splicing event.

The BPS is typically located 18 to 40 nucleotides upstream from the 3’ ss and is followed by the polypyrimidine tract in higher eukaryotes (Will and Lührmann, 2011). BPS is a highly degenerate sequence defined by the YNYURAC consensus (Y: C or T; R: A or G; N: nucleotide), in which the Adenosine (A) provides the nucleophile for 5’ ss cleavage and represents the designed site for lariat formation. Nevertheless, the BPS plays an essential role in the first step of transesterification reactions.

The polypyrimidine tract (poly Y tract or PPT) is composed of a stretch of pyrimidines, uridines in particular, and is located between the branch site and the terminal AG at the intron/exon junction. It has been reported that the PPT has a role in promoting splicing since progressive deletion or elongation of this region can impair or improve splicing efficiency.

The AG dinucleotide defines the 3’ border of the intron and is located just downstream of the PPT. The site is defined by the YAG|G sequence (Y: C or T; |: the exon/intron boundary), where the AG is highly conserved and is fundamental for the second splicing transesterification reaction (Roscigno, Weiner and Garcia-Blanco, 1993).

The core *cis*-acting splicing elements are degenerated, and additional sequences are required for accurate exon recognition in splicing. Additional sequences identified in the exon and neighbouring introns are involved in splicing regulation. The *cis*-elements that assist the inclusion of an exon in the mature mRNA transcripts, termed Exon Splicing Enhancers (ESEs) and Intron Splicing Enhancers (ISEs),Exonic Splicing Silencers (ESSs) and Intronic Splicing Silencers (ISSs) help the exclusion of an exon in the mature mRNA or block the spliceosome assembly. Furthermore, these elements interact with the *trans*-acting factors (Black, 2003).

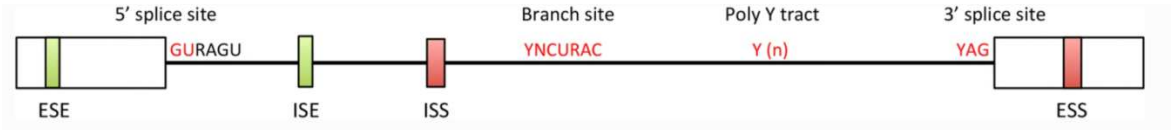


Figure 5. *Cis*-element regulating the splicing of pre-messenger RNA. Major regulatory elements of pre-mRNA splicing, namely 5' splice donor site, 3' splice acceptor site, the branch point sequence, and polypyrimidine tract which are bound by the spliceosome machinery. Exonic and intronic splicing enhancers and silencers (ESE, ISE, ESS, ESI), are shown in green and red rectangle, respectively. Reproduced from (Beqqali, 2018).

1.3.2.3 Trans splicing factors

Auxiliary *cis*-elements can recruit *trans*-regulatory factors that help to carry out the core splicing signal. In general, splicing enhancer elements are characterised by purine-rich sequences which are biding sites for Serine/Arginine-rich (SR) proteins. These proteins are involved in several steps of spliceosome assembly and act as both essential splicing factors and regulatory factors. Although, SR proteins have a common structure which is characterised by the presence of one or two RNA recognition motifs (RRM) at the N-terminal and the serine/arginine-rich domain (RS domain) of at least 50 amino acids with an RS content higher than 40% at the C-terminal (Tacke and Manley, 1999; Manley and Krainer, 2010). The RNA recognition motif domain is required for RNA-binding, whereas the RS domain functions as a protein interaction domain. The predominant location of SR proteins is in the nucleus, in particular in the speckles, which seem to act as sites of storage of these phosphoproteins together with other important splicing factors (Jiménez-García and Spector, 1993; Cáceres, Screaton and Krainer, 1998). However, these factors can be shuttled between the nucleus and the cytoplasm at different dynamic rates (Cáceres, Screaton and Krainer, 1998; Sapra *et al.*, 2009).

Different models have been proposed to explain the mechanisms through which SR proteins facilitate splice site recognition. First, a protein-protein interaction model based on their RS domains and dependent on their phosphorylation state. For example, ASF/SF2 and SC35 can interact with each other, with U1-70K SR related proteins and with U2AF35, suggesting that they facilitate splicing by forming interactions across exons and introns (Fu, 1995; Graveley, 2000). An alternative model for enhancer activity is based on interactions between ESEs and spliceosomal components, through factors such as SRm160 and SRm300. These co-activators, which contain RS domains but lack RRM domains, can generate multiprotein complex with snRNPs and SR proteins bound to the ESE sequence (Blencowe, 2000). In some cases, the binding of the SR proteins can prevent or displace the recruitment of negative regulatory factors to silencers, such as hnRNPA1 (Zhu, Mayeda and Krainer, 2001).

As previously mentioned, splicing enhancers can also be found in introns (ISEs), although the number of positive intronic elements described in the literature is very low in comparison to exonic ones. Contrary to the positive splicing regulators elements of exon inclusion of the splicing, the heterogeneous nuclear ribonucleoproteins (hnRNPs) family bind to exonic and intronic splicing silencers elements and interfere with the spliceosome assembly as a result of exon exclusion (Wang *et al.*, 2006).

The hnRNP protein group includes different proteins with molecular weights between 34 and 120kDa, named alphabetically from hnRNP A to hnRNPU, where their role in splicing regulation has been described for all of them, except for hnRNP U (Dreyfuss *et al.*, 1993; Han, Tang and Smith, 2010).

The diversity of hnRNP proteins is due to the Ribonucleotide-binding motif (RBM) that is responsible for the specific interaction with the pre-mRNA (Geuens, Bouhy and Timmerman, 2016), with some exceptions such as hnRNP K/J and hnRNP E1/E4, which use a KH domain (Dreyfuss, Philipson and Mattaj, 1988; Burd and Dreyfuss, 1994; Chaudhury, Chander and Howe, 2010).

Most hnRNPs also contain an RGG box (a cluster of Arg-Gly-Gly tripeptides) that could be in combination with other RBD (RNA binding domain) or unique, as in hnRNP U (Kiledjian and Dreyfuss, 1992). Moreover, hnRNPs may contain other unstructured auxiliary domains such as glycine-rich or proline-rich domains, that are important in the localisation of the proteins: for example, one glycine-rich nuclear import/export domain (M9) is located at the C-terminal of hnRNPA1 (Weighardt, Biamonti and Riva, 1995).
Regarding the localisation of these *trans*-acting factors, although many of them are nuclear at the steady-state, they can be moved from the nucleus to the cytoplasm (Piñol-Roma and Dreyfuss, 1992), playing an essential role in nucleo-cytoplasmic transport of mRNA(Lee, Henry and Silver, 1996), mRNA localization (Carson, Cui and Barbarese, 2001), translation (Habelhah *et al.*, 2001), and mRNA stability (Xu, Chen and Shyu, 2001). Within the nucleus, hnRNPs take part in various processes such as transcriptional regulation (Miau *et al.*, 1998), RNA splicing(Chan and Black, 1997; Chou *et al.*, 1999; Mourelatos *et al.*, 2001), telomere maintenance (Ford, Wright and Shay, 2002) and 3'-end processing(Kessler *et al.*, 1997).

The regulatory action through which exonic and intronic silencers interfere with splicing may involve the "looping-out" mechanism, where the hnRNP proteins bind both ends of the exon, and the interaction leads to the exon becoming blocked to the spliceosome complex (Chen and Manley, 2009).

Another proposed mechanism suggests the multimerization of the *trans*-acting factors along the RNA prevent the splice site usage. A further mechanism of action of silencers occurs through the formation of protein-protein interactions that loop out the alternative exon (Kolovos *et al.*, 2012).

1.4 Alternative splicing

Sequencing of the human genome has shown much higher number of expressed transcripts(mRNAs) than there are human genes. This phenomenon is explained by alternative splicing (AS) which allows the cell to expand diversity in protein expression. Consequently, the proteins translated from these alternatively spliced mRNAs allow the proteins different stability, regulatory properties, intracellular localization, and turn-over rates (Wang *et al.*, 2008; Barbosa-Morais *et al.*, 2012).

Genome-wide studies estimated the frequency of about 90-95% of annotated human genes undergo alternative splicing, these explain the discrepancy between the 24,000 estimated coding genes and the much greater number of different synthesised proteins (Keren, Lev-Maor and Ast, 2010).

Several different types of alternative splicing patterns exist through which exons can enrich the diversity of the mRNA species (Figure 6). The first consists of a cassette exon that is included or skipped in the mature RNA transcript. Cassette exons can be spliced or skipped in tandem or spliced in a mutually exclusive manner, where one or another exon, but never both, are included in the mRNA. Another form of alternative splicing process is the different usage of the canonical splice sites, 5'ss and 3'ss, as a result, short and long forms of exons are present in the mature RNA transcript and create a new open reading frame, which when translated, generates a protein isoform. Finally, intron retention is a process where part of the introns retained in the mature mRNA which is either translated or ends up in the nonsense-mediated decay (NMD) (Van Den Hoogenhof, Pinto and Creemers, 2016; Vanichkina *et al.*, 2018).Back-splicing is a recently recognised category of alternative splicing which results in the formation of circular RNAs (circRNAs) (Nigro *et al.*, 1991; Tan *et al.*, 2017).

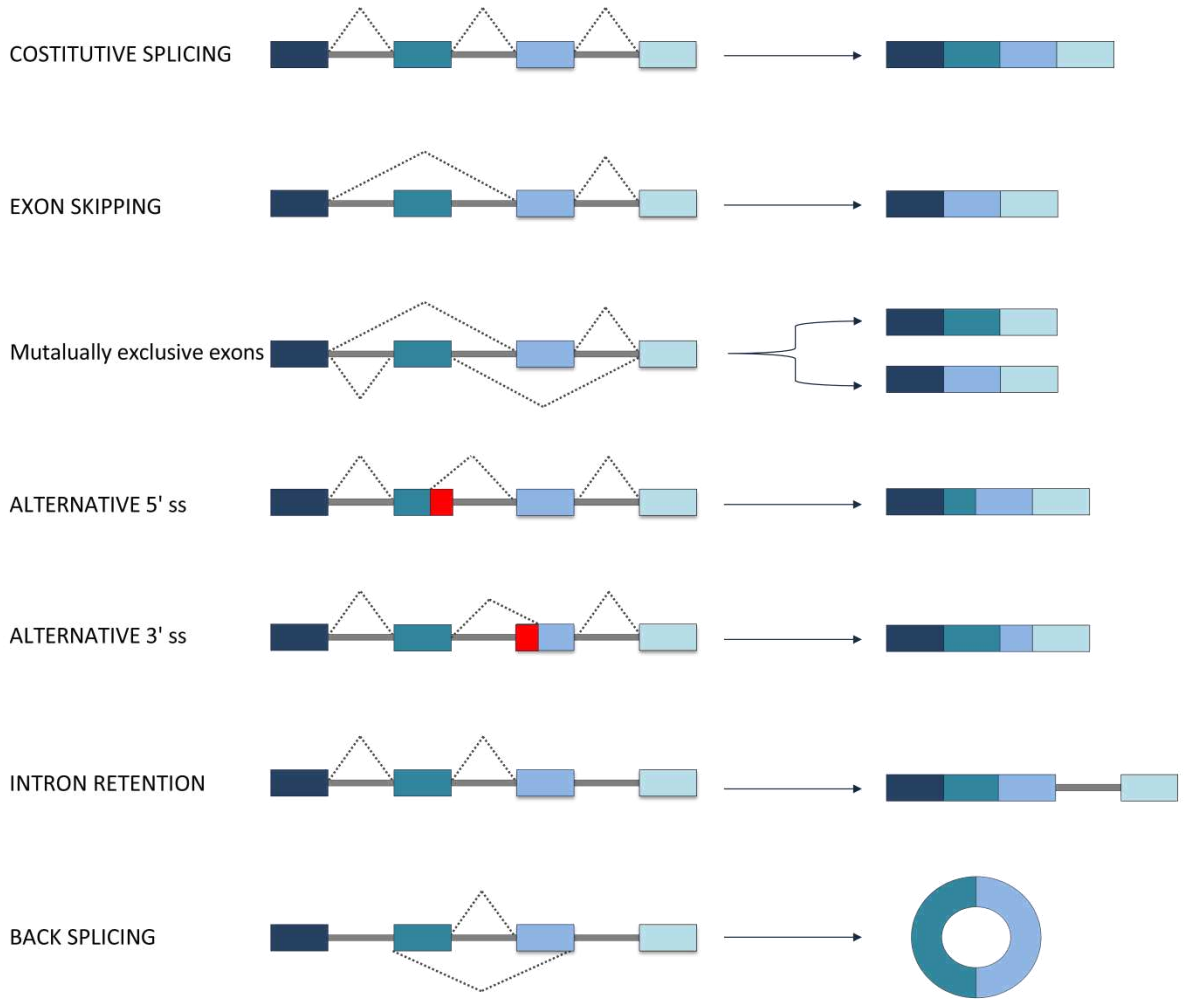


Figure 6. Different types of alternative splicing. Constitutive splicing and the six main types of alternative splicing are reported. The exons are indicated with rectangular boxes with different colours, while introns are in grey lines between exons. The resulted transcripts after intron/exon removal are reported on the right.

This is followed by several functional consequences. For example, if these events occur in the coding region of the RNA transcripts, there is translation of an isoform that may alter or have an opposite functional effect. Alternatively, the mRNA may create a premature termination codon (PTC) that follows the degradation pathway through NMD.

In addition, two other processes can lead to the generation of multiple mRNA transcripts. These include alternative polyadenylation (APA) and alternative first exon usage (or alternative promoter usage).

Alternative splicing is prevalent only in multicellular eukaryotes and, it can be extensive during cardiac development to adapt the heart to physiological conditions. For example, the passive stiffness of mammalian hearts increases during development, an effect that is achieved by adjusting the expression of titin splice variants (Lahmers *et al.*, 2004; Opitz *et al.*, 2004).

Examining within the context of the cardiovascular system (see section 1.8) alternative splicing is normally used by cardiomyocytes at different pre-and post-natal stages and gives rise to different protein isoforms. Kalsotra*et al*. performed one of the first studies of transcriptome-wide changes in alternative splicing during heart development of mice using exon arrays (Kalsotra *et al.*, 2008). This study demonstrated that in the cardiac cells, 63 alternative splicing events occur. Subsequently, large-scale RNA-seq analysis, it was demonstrated that AS works predominantly during late embryonic and postnatal stages, and protein isoforms play a role as regulatory components of postnatal cardiac development. Further studies, conducted on alternative splicing profiles on the human heart by RNA-seq data, revealed that intron retention occurred more frequently in fetal hearts than in adult hearts, and suggested that intron retention is the main form of alternative splicing involved in human heart development (Wang *et al.*, 2016). Finally, critical genes for cardiac development have also been shown to be regulated by isoforms that have distinct functions obtained by AS process. An example of genes orchestrated by AS are the ion channel genes and the sarcomeric genes, such as myomesin (*MYOM*), titin (*TTN*), and LIM domain-binding 3 (*Ldb3*).

1.5 Genetic variants associated with mis- splicing disease

Sequence changes that cause errors in splicing have been observed as a frequent disease-causing phenomenon in human genetic disorders, either monogenic or complex, and splicing profiles can impact the pathogenic classification (Scotti and Swanson, 2016). Perturbation of mRNA splicing is also linked to incorrect protein synthesis, and can lead to a decrease of the specific messenger RNA and thus of the protein, that can result in aberrant cellular metabolism and/or function (Chabot and Shkreta, 2016). In addition, alteration of mRNA splicing complexity increases due to the influence of external factors, and both are dysregulated in the diseased state. It has now been established that an expanding list of diseases are attributed to splicing mis-regulation, including cardiovascular diseases (CVDs) (Wang and Cooper, 2007; Solis, Shariat and Patton, 2008).

The aberrant splicing of premature RNA transcripts can result from nucleotide changes that occur in specific positions on the pre-mRNA. These mutations affect pre-mRNA maturation which influence splicing, leading to consequences such as disrupting splicing regulatory elements (SREs); altering mRNA secondary structure, or even introducing a new cryptic splice site that compete with canonical splicing signals (Baralle and Baralle, 2005).Therefore, when substitutions affect the canonical 5’ and 3’ splice sites (typically GT and AG, respectively)in haplo-insufficient autosomal dominant genes, they are classified as pathogenic variants. Whereas variants in the region near the splice site are often diagnostically classified as VUS (Rivas *et al.*, 2015; Lek *et al.*, 2016; Zhang *et al.*, 2018). The nucleotide variants may lead to exon

skipping, cryptic splice site activation, intron retention, or the generation of 'fake' exons. The nucleotide variants associated with splicing dysfunction in human disease principally affect the activity of the major spliceosome, with less than 0.5% in the minor spliceosome (Padgett, 2012; Buratti, 2016).

Some studies have shown that in cardiac disorders (section 1.8), such as cardiac hypertrophy, there is re-activation of the foetal cardiac gene program. *Kong* and colleagues (Kong *et al.*, 2010a)were the first group to demonstrate that the diseased human heart has predominant usage of alternative splicing. They detected a notable decrease in the efficiency of the splicing process in the damaged heart using exon arrays. In particular, they focused on splicing of key sarcomere genes, which encode for proteins comprising the basic contractile unit of muscle, like troponin T, β-myosin heavy chain, filamin C, and titin which present significant aberrant splicing (Kong *et al.*, 2010a). In 2009, the first mutation in a cardiac splicing factor, RNA binding motif protein 20 (*RBM20*) was identified, which was associated with human heart failure, and the underlying mechanism involved altered splicing of the giant sarcomeric protein titin (Brauch *et al.*, 2009; Guo *et al.*, 2012).

Detection of a disease-predisposing pathogenic variant and its influence on RNA processing, in particular for cardiac genes, would be clinically beneficial, for diagnostic, prognostic and therapeutic purposes (Baralle, Lucassen and Buratti, 2009; Singh and Cooper, 2012).

1.6 RNA analysis approach of potential splicing variants in clinic

Considering the vast number of *cis*- acting regulatory elements necessary for correct mRNA splicing, mutations in these may often be deemed a VUS and consequently the analysis of the mRNA splicing pattern is important. Nucleotide variations can occur in the coding sequences (exons) and non-coding sequences (introns). Variants in the coding sequences are named missense, nonsense, and translationally silent variants, while variants located in introns can be called splicing or silent variants.

The interpretation of the consequences of most reported genetic variants depends on their location and this can lead to misclassification, especially for variants located in exons. This is in part due to the presence of other exonic splicing regulatory regions, such as ESE and ESS, which are involved in splicing and can be disrupted by mutations (Cartegni, Chew and Krainer, 2002). Therefore, interpretation of VUSs can be supported by RNA analysis and avoid underestimation of splicing mutations (Cartegni, Chew and Krainer, 2002; Baralle and Baralle, 2005). Each sequence variant should be considered as a potential candidate for splicing anomalies independent of its location (Zatkova *et al.*, 2004; Dehainault *et al.*, 2007). It is important that classified variants can be regularly evaluated as new data becomes available with regards to their significance in clinical settings (Baralle, Lucassen and Buratti, 2009). Therefore, to

determine how genetic variants affect splicing, further studies may be necessary for a complete comprehension of their functional implications. Several bioinformatic tools have been developed to investigate the possible effects of splicing mutations (see section 1.6.1) but cannot be solely relied upon functional analysis. The advantage of the *in silico* step should be to determine the variants that are likely to have a strong impact on RNA splicing (Desmet and Béroud, 2012).Obviously, the most reliable method to study splicing events is the direct analysis of RNA from tissues relevant to the disease, e.g. the heart for cardiac disease. However, this is often not available or invasive to obtain, therefore blood RNA can be used as a surrogate tissue for genes associated with cardiac disease but that are also expressed in blood. If no biological samples are available, *in vitro* methods can be used, such as a minigene assay (see section 1.6.2).The minigene assay, which is based on the patient's genomic DNA, represents a valid and powerful approach to study splicing patterns (Cooper, 2005).

1.6.1 *In silico* prediction: SpliceAI tool

Numerous bioinformatics programs now exist to assess the effect of a putative pathogenic variant or a VUS on splicing. Although nucleotide substitutions in the natural AG and GT splice sites are certain to disrupt splicing, predicting the impact of changes in sequence elements required for correct splicing or in deep intronic regions is significantly more difficult. Although the splicing analysis tools were originally developed for research purposes, nowadays they are introduced in diagnostics as a component for variant characterisation. It has been reported that bioinformatic tools have a higher sensitivity rate (~90–100%) compared to specificity (~60–80%) (Riolo, Cantara and Ricci, 2021).The algorithms applied by the bioinformatic tools generated to determine the possible consequences on splicing are based on the core splicing signals responsible for the splicing process (Pacini and Koziol, 2018).

Some of the bioinformatic tools used to test the possible impact of putative pathogenic variants are listed in Table 1 below.

Tool	Model	Algorithm	Charateristics	Reference
CADD	First version based on SVM, ubsequently L2-regularized logistic regression	ML	Predict the pathogenicity of a variant	(Rentzsch *et al.*, 2019)
DSSP	CNN with random forest, and linear regression models	DL	Prediction of the variant impact on 5' and 3'ss	(Naito, 2019)
ESE Finder	PWM	MB	Identification of ESE	(Cartegni *et al.*, 2003; Smith *et al.*, 2006)

GENSCAN	MDD	-	Predicts 5' and 3' ss	(Burge and Karlin, 1997)
GeneSplicer	MDD, MM	ML	Predicts 5'and 3'ss	(Pertea, Lin and Salzberg, 2001)
IntSplice	SVM	-	Prediction of pathogenic intronic SNV in BPs and PPT	(Shibata *et al.*, 2016)
MaxEntScan	Maximum entropy distribution	Other	Predicts 5' and 3' ss	(Yeo and Burge, 2004)
MMSplice	Modular NN, linear and logistic regression	ML	Predicts effects of variants on splicing	(Cheng *et al.*, 2019)
NetGene2	NN	ML	Predicts 5' and 3' ss	(Hebsgaard *et al.*, 1996)
NNSplice	NN	ML	Predicts 5' and 3' ss	(Reese, 1997)
RESCUE-ESE	Experimental and computational approach	-	Identification of SREs	(Fairbrother *et al.*, 2002)
SPIDEX	Bayesian modelling	ML	Predicts SNVs impact on splicing	(Xiong *et al.*, 2015)
SpliceAI	Deep residual neural network	DL	Predicts SNVs impact on splicing	(Jaganathan *et al.*, 2019)
SpliceAid2	Scanning against validated splicing sequences		Predicts SNVs impact on ESE, ESS, ISE, ISS	(Piva *et al.*, 2012)
Spliceator	CNN	DL	Predict 5' and 3'ss	(Scalzitti *et al.*, 2021)
SpliceFinder	CNN	DL	Classification of variants based on impact on splice sites	(Wang *et al.*, 2019)
SplicePort	FGA, and SVM	ML	Predict 5' and 3'ss	(Dogan *et al.*, 2007)
SpliceRover	CNN	DL	Prediction of 5' and 3' ss from primary sequence	(Zuallaert *et al.*, 2018)
SpliceSiteFinder-like	PWM	Other	Predict 5' and 3'ss	(Shapiro and Senapathy, 1987)
SVM-BP Finder	SVM		Prediction of BPs and PPT	(Corvelo *et al.*, 2010)

Table 1. List of bioinformatic tools available to predict core splicing signals. Abbreviations: BPs, branch site motifs; CNN, convolutional neural networks; DL, deep learning; FGA, feature generation algorithm; HMM, hidden Markov model; MDD, Maximal Dependence Decomposition;

ML, machine learning; MM, Markov model; NN, neural network; nt, nucleotides; PPT, Polypyrimidine tract; PWM, Position weight matrices; ss, splice site; SVM, Support vector machine. Adapted from (Riepe *et al.*, 2021).

The algorithms have been implemented over the years, and the available tools are based mainly on motif-based, machine learning, and deep learning algorithms. While machine learning approaches are needed that the features should be defined by the users before processing the models, novel deep learning (DL) algorithms have been implemented and involved the use of a convolutional neural network (CNN) and extract the features directly to the network (Eraslan *et al.*, 2019). In DL algorithms, the features used are defined directly from data during the training of the model, without the necessity to pre-specify these which differs from canonical machine learning. Therefore, deep learning algorithms allow the investigation of complex features in a huge region, such as the genomic region spanning a gene. In particular, for the identification of the splice sites, this is an advantage due to the number of genomic elements involved (see section 1.3.2.2) (LeCun, Y., Bengio, Y., Hinton, 2015).

In this study, SpliceAI, a deep learning-based tool widely used by clinical diagnostics laboratories was applied to assess genetic variants to lead splicing abnormalities. SpliceAI was especially created to investigate and characterise non-canonical splice variants that are not localized in the canonical AG or GT splice sites (Jaganathan *et al.*, 2019). The sequence variation outcomes were validated on pre-mRNA using the GENOCODE-annotated related to the pre-mRNA transcript as input (Harrow *et al.*, 2012; Jaganathan *et al.*, 2019). Each nucleotide is analysed in a pre-mRNA transcript to be a 5' and/or 3' splice sites. The algorithm examines a region of 50bp related to a presented sequence variation and then shows the predictive scores of the most substantial gain or loss of acceptor or donor potential splice sites as a consequence of the mutation, in contrast to other *in silico* tools that only investigate a small number of nucleotides window along exon-intron boundaries (Rowlands, Baralle and Ellingford, 2019).
SpliceAI was observed to have a sensitivity of 41% for identifying potential splice variants in introns >50 bp from exon using a 0.5 delta score (Δscore) threshold (Jaganathan *et al.*, 2019), whereas by decreasing the threshold to 0.2 (Riepe *et al.*, 2021) or 0.05 (Moles-Fernández *et al.*, 2021) the sensitivity values of ≥ 90% were achieved. This high sensitivity supports improvement for clinical evaluation of splice variants identified by WES and/or WGS by integrating the results into the diagnostic pipeline (Islam, 2021).
Overall, these bioinformatic tools aid in the identification of potential candidate variants, for which molecular transcript analysis can be performed in an efficient and pertinent number. However, they are not sufficient to be used alone for clinical purposes as they lack information

related to a real situation in cells, tissue, and organisms even when a number of tools are used in combination (Soukarieh *et al.*, 2016; Wai, Douglas and Baralle, 2019).

1.6.2 Minigene splicing system

The minigene splicing assay is one of the most common approaches to study RNA splice isoforms. This system permits the study and identification of the regulation of splicing of a particular exon, the splicing regulatory sequences localised in the intron and exons, and potential splicing factors which can influence it(Singh and Cooper, 2006). To interrogate the splicing pattern of a specific gene, the region of interest in the genomic DNA is amplified by PCR and cloned in an expression vector. The expression vectors include a ubiquitous transcriptional promoter and a gene segment for poly-A 3' end. A key point during the process is amplification of the region of interest along with one exon or more, and the flanking introns immediately upstream and downstream. This is important due to the potential *cis*-acting sequence located there, which can determine the inclusion or skipping of an exon. After cloning, the transfection of the constructs is performed in appropriate specific cell lines, where the spliceosome processes it. Finally, the mRNA splicing pattern is evaluated by RT-PCR where the primers are designed specifically to distinguish the processed transcripts derived from the minigene and the endogenous transcript.

The minigene system can also be deployed to test aberrant splicing of long exons. These long exons are difficult to analyse using RT-PCR due to the limited efficiency of reverse transcriptase. However, minigenes can accommodate long exons and be used for aberrant splicing analysis. The *BRCA1*exon 11 minigene (Raponi *et al.*, 2014)is an example of a long exon minigene and shows aberrant splicing pattern in various VUSs which was impossible to study with the RT-PCR method.

The potential application of minigene is not only limited to VUSs on the database, but can also be extended to VUSs which are yet to be discovered. For example, a saturation mutagenesis splicing assay (Ke *et al.*, 2018) can be applied to screen potential variants which affect the splicing of a specific gene of interest. These data are useful to verify newly discovered variants for their potential for aberrant splicing.

The minigene is not only a useful tool for "Precision diagnosis" but also for developing "Precision therapeutics". Splice switching oligos such as 2'MeO, 2'MOE and morpholinos (Havens and Hastings, 2016) and modified snRNAs (Dal Mas *et al.*, 2015)have been shown to correct aberrant splicing and can easily be tested in minigene assays. These examples show the usefulness of minigene in both diagnosis and therapeutics in the "Precision Medicine" era.

1.6.3 Direct approach for RNA analysis: RT-PCR and RNA sequencing

In clinical practice, blood and skin fibroblasts are the commonly available biospecimens which are frequently collected to investigate pre-mRNA maturation directly. These tissues express the majority of Online Mendelian Inheritance in Man (OMIM) genes in appropriate RNA quantities for performing functional studies (Aicher *et al.*, 2020; Wai *et al.*, 2020; Douglas and Baralle, 2021).

RT-PCR followed by Sanger sequencing or short read RNA-seq by massively parallel sequencing (MPS) are currently the two principal techniques applied for splicing analysis after the source of the RNA has been determined. Although RT-PCR analysis has a relatively short turnaround time, an RNA sample must be sourced from the patient. Therefore, accessibility to appropriate patient biopsy is key to performing RT-PCR using patient RNA. If the patient cannot be recruited to the study, or the patient is no longer alive, it will be impossible to source a patient biopsy. In cases where the patient is willing to be involved in the study, it still requires local, national, or international ethical committees' approval for handling of patient samples and data. These approval processes require time, which can hinder the efforts to diagnose patients with VUSs who may need urgent medical attention, for example, patients with CVDs. The minigene approach can overcome these issues as it is not necessary to wait for patient samples and/or lengthy approval processes.

RT-PCR has a lower detection limit than RNA-seq and multiple splice junctions can be detected in a single amplicon and thus identify the usage of different isoforms. While this approach has been used frequently for splicing analysis, it is known that biases can be introduced during the amplification (Maddirevula *et al.*, 2020). Therefore, laboratory expertise in splicing has a significant impact on diagnostic sensitivity because primer design and thermocycler conditions used may limit the detection of splicing abnormalities (Whiley *et al.*, 2014). Due to these limitations, RT-PCR can require other validation and cloning of amplicons to resolve all splicing events.

RNA-seq allows the investigation of multiple variants and allele-specific expressions due to the whole transcriptome analysis per sample. RNA-seq has been introduced in genetic diagnostic laboratories and has enhanced knowledge across diverse areas of human disorders, including diagnosis, prognosis and therapeutic strategies (Byron *et al.*, 2016).Even without a candidate genetic variant, implementation of RNA-seq has been demonstrated to increase the diagnostic rate by 8-36% in previously unsolved cases examined with WES. Therefore, RNA-seq does not require predetermined primers, in contrast to conventional techniques like quantitative reverse transcription polymerase chain reaction (RT-qPCR), which requires to design specific set of primers for each target gene. This is beneficial when studying transcriptomes for which the full set of targets is unknown (Gonorazky *et al.*, 2019; Murdock *et al.*, 2021).

RNA-seq starts with isolating RNA, converting RNA to complementary DNA (cDNA), library preparation, PCR amplification and finally sequencing with the NGS platform (Figure 7). Depending on the object of the study and budget availability, additional characteristics could be considered such as choice of biological sample, number of replicates, the type of RNA isolation protocol, strand-preserving libraries, single or paired-end reads, read length and sequencing depth (Kukurba and Montgomery, 2015; Conesa *et al.*, 2016). A critical point is the tissue selected for RNA isolation, ideally the tissue involved in the disease process. This is due to alternative splicing being tissue-specific, with splicing patterns being more similar between the same tissues of different species than between different tissues of the same organism (Lin *et al.*, 2014). Hence, the wrong selection of tissue type can result in missed and/or incorrect diagnoses due to tissue-specific expression of disease-causing isoform transcripts (Schoch *et al.*, 2020).

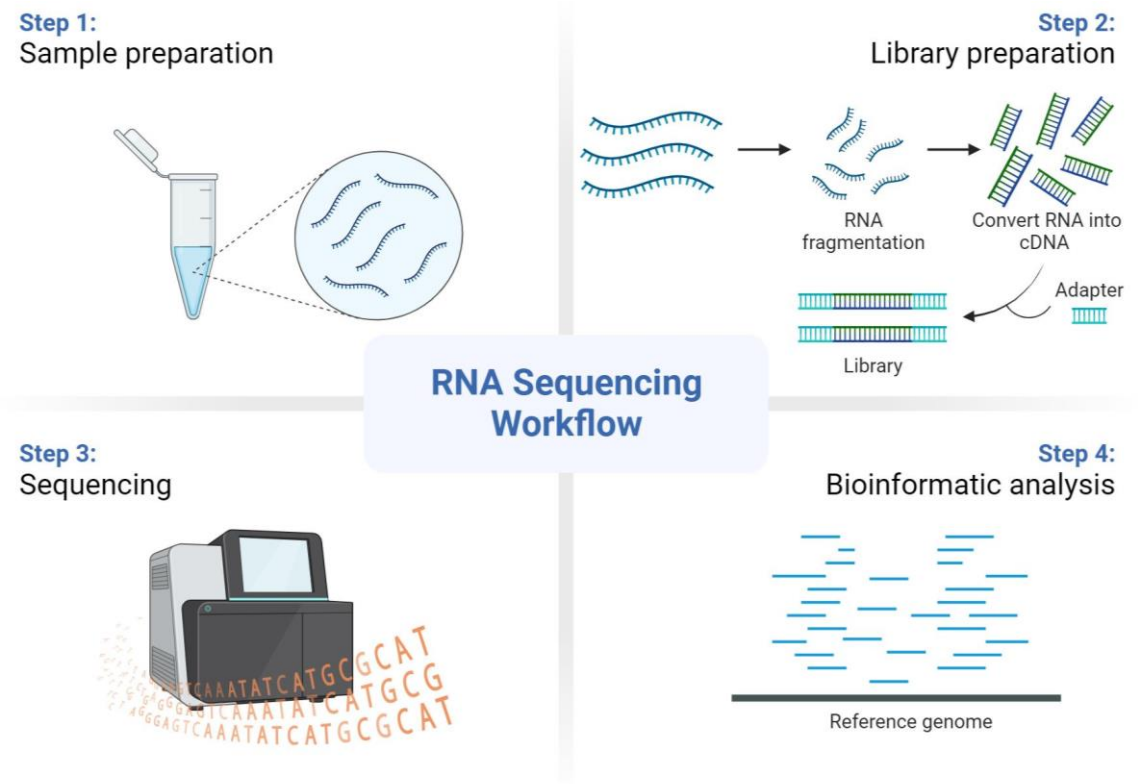


Figure 7.Schematic representation of RNA-seq workflow. The main steps include sample preparation, library preparation, sequencing, and bioinformatic analysis of the raw data. Drawing by BioRender.

A problem with RT-PCR and RNA-seq are genes which are not well expressed in tissues that can be easily sampled and analysed. In cases of genes with relatively high-level expression, the splicing pattern found in the source biopsy may not be the same as in affected organs or cells due to alternative splicing. Minigenes can solve this issue by using an appropriate cell line to detect aberrant splicing.

1.6.3.1 Computational tool for RNA-seq data analysis

Several computational tools and pipelines have been developed in order toanalyse RNA-seq data and identify aberrant splicing events and expression outliers in the absence of candidate variants. The biggest challenge of using RNA-seq in splicing analysis is that there is no single unified gold standard guideline for splicing read count to validate as “aberrant” (Conesa *et al.*, 2016; Xu *et al.*, 2016). Different RNA-seq workflows are developed based on the biological sample and the analyses required. For example, in RNA-seq of degraded RNA from formalin-fixed paraffin-embedded tissues, poly(A) enrichment of mRNA should not be performed due to the high quantities of degraded RNA which causes a strong 3’ bias in transcript coverage (Stark, Grzelak and Hadfield, 2019). In the case of RNA-seq from whole blood, depletion of human beta-globin transcripts is recommended in order to increase the read counts for non-globin genes (Harrington *et al.*, 2020).

RNA-seq data analysis starts with an accurate analysis of raw RNA-seq data. This first step is critical since downstream analysis depends on this because it is followed by alignment of sequencing reads against a genome or transcriptome. Different factors are considered when the aligner tool is chosen such as accuracy, detection of (annotated and novel) splice junctions, sensitivity, computational resources needed and speed. After selecting an aligner, each subsequent analysis should be performed in the same manner to ensure similar results (Au and Farkas, 2019). A common RNA-seq aligner is STAR (Dobin *et al.*, 2013), which has a higher rate of multi-mapped reads and a higher number of reliable recognised spliced junctions compared with other aligner tools (Dharshini, Taguchi and Gromiha, 2020). After the alignment step, gene expression or isoform quantification analysis is performed by counting the reads number. This is achieved by using the alignment file and a file containing the features of interest, followed by counting reads which align to each feature of interest. The counting reads step should also consider, other factors which include single- or paired-end reads, sequencing library preparation with or without a stranded protocol, how the features of interest are defined, and if the reads should be counted being part of multiple features (Au and Farkas, 2019). HTseq is a common counting tool which is considered one of the best tools for quantification (Anders, Pyl and Huber, 2015), while Salmon is one of the best for isoform quantification (Patro *et al.*, 2017). Several tools exist for identifying mis-splicing events, one of these is rMATS (Shen *et al.*, 2014). This tool uses an exon-based approach which detects alternative splicing events by comparing gene exon-inclusion levels defined with junction reads between samples, the focus on exon inclusion and exclusion allows greater accuracy in identifying individual alternative splicing events compared to whole isoform analysis (Conesa *et al.*, 2016). Cummings et al. (Cummings *et al.*, 2017) developed another tool to perform alternative splicing analysis where novel splice junctions can be discovered and normalised against overlapping annotated splice junctions.

Despite the existence of numerous computational methods for RNA-seq analysis, standard procedures for the alignment of sequencing reads to the reference genome, filtering and normalisation, and the sequencing depth necessary for splicing analysis have not yet been established. In addition, the reads generated by RNA-seq do not cover multiple splice junctions which leads to limited information regarding isoform and low confidence in mapping reads that span exon-exon junctions (split reads) to the reference genome. This may result in alignment errors or even filtering crucial sequencing reads important for diagnosis (Conesa *et al.*, 2016; Marco-Puche *et al.*, 2019).

To overcome these limitations third generation long read sequencing approaches have been developed by Oxford Nanopore and Pacific Biosciences (Pollard *et al.*, 2018; Wai, Douglas and Baralle, 2019). This new RNA-seq methodology can be used for sequencing RT-PCR amplicons or whole transcriptomes because it generates average read lengths up to 10 kb for complete transcript isoform. However, the introduction of long-read sequencing in routine practice for splicing analysis remains restricted due to high cost, sequencing error rate and no completed analysis pipelines (Mantere, Kersten and Hoischen, 2019).

To include RNA analysis in variant classification guidelines, standard technical platforms and best practices are required due to the complexity of splicing analysis.

1.7 Cardiovascular disorders phenotype and genetic background

Cardiomyopathies or cardiovascular disorders (CVDs) are the general terms used to indicate disorders affecting the heart and blood vessels. These negatively impact the normal function of the cardiovascular system and the heart loses its ability to pump blood to the body. CVDs represent one of the major causes of death worldwide, accounting for approximately 17.9 million deaths per year, as reported World Health Organization. The conditions affecting the cardiac muscle represent complex, multifactor, and polygenic conditions resulting from an interaction between the genetic background and environmental factors. Genetic cardiac disorders typically follow a pattern of autosomal dominant inheritance and are classified into four main groups; Hypertrophic cardiomyopathy (HCM), Dilated cardiomyopathy (DCM), Arrhythmogenic right ventricular cardiomyopathy (ARVC), and Restrictive cardiomyopathy (RCM) (Figure 8) (Richardson *et al.*, 1996).

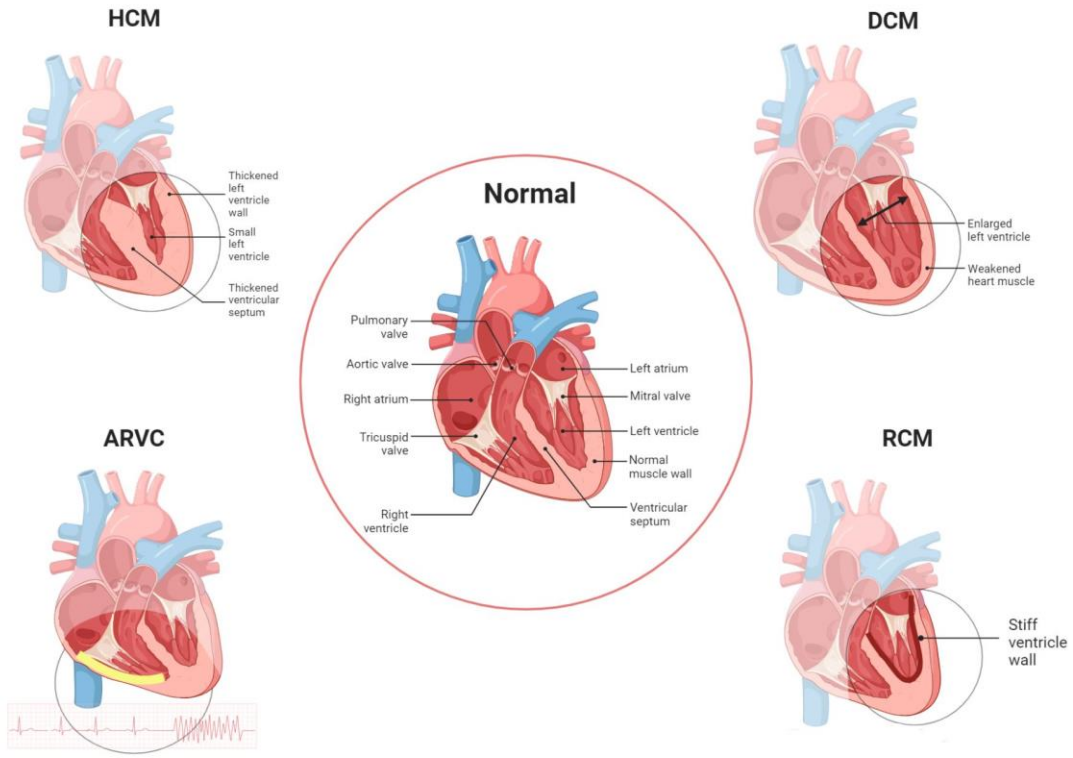


Figure 8.Schematic representation of cardiovascular disorders. The normal heart is represented in the middle with around the several abnormalities that the heart faces during the disease. Hypertrophic cardiomyopathy (HCM); Dilated cardiomyopathy (DCM), Arrhythmogenic right ventricular cardiomyopathy (ARVC), and Restrictive cardiomyopathy (RCM). Drawn by BioRender.

HCM and DCM result in an increase in cardiac mass with distinct ventricular remodelling patterns. HCM causes hypertrophy of the left ventricular wall (IVSd ≥ 1.5 cm) with a decrease of the ventricular chamber volume. In clinics is observed that systolic function is increased or at least preserved in affected patients, whereas diastolic function is decreased due to hypertrophy, interstitial fibrosis, and myocyte disarray. Diastolic dysfunction is responsible for heart failure and premature sudden cardiac death of patients affected by HCM.HCM is inherited in an autosomal dominant manner and shows also reduced penetrance and variable expressivity (Cirino and Ho, 1993). In contrast, DCM is characterised by a significant enlargement of the left ventricular chamber which causes systolic dysfunction. Patients with DCM need a heart transplant otherwise a sudden cardiac death is caused. Therefore, the DCM phenotype overlaps with the symptoms of other types of cardiac disorders, and this makes diagnosis more difficult (Pugh *et al.*, 2014). Variants in genes which cause DCM are typically inherited in a Mendelian autosomal dominant pattern, but a small number of cases with autosomal recessive, X-linked, or mitochondrial DNA inherited have been observed (Hershberger, Morales and Siegfried, 2010; Man *et al.*, 2013). DCM has clinical heterogeneity

with reduced penetrance (which means that not everyone with a genetic diagnosis will develop during their lifetime the clinical symptoms of the disease), and variable expression of symptoms across patients (which means that all patients do not present with the same severity of symptoms) (Katsuragi and Ikeda, 2018). Arrhythmogenic right ventricular cardiomyopathy (ARVC) is characterised by the loss of muscle cells, ventricular wall thinning, and fatty cells replacing muscle cells. These changes of cardiac muscle are usually clinically identified in the right ventricle, however, it is sometimes possible that the left ventricle is involved. Patients present with symptoms such as heart failure, abnormal heart rhythm, and sudden cardiac death. ARVC is typically inherited in autosomal dominant manner, but there are cases of autosomal recessive inheritance. This disorder has reduced penetrance and variable expressivity (Protonotarios and Elliott, 2019). Finally, restrictive diastolic dysfunction, such as restrictive filling and reduced diastolic volume of either or both ventriculars, characterises restrictive cardiomyopathy (RCM) with normal systolic function and wall thickness which is preserved.

Finally, there are disorders that affect the cardiovascular system, channelopathies or inherited arrhythmias, in which patients show a normal heart structure, but have an abnormal heart rhythm due to ion instability. Channelopathies are caused by genetic variants in genes which encode for ion channels. The function of ion channel in the cell is to allow the ion passage (e.g. Na^+, K^+, CL^-) into the cell membrane for heartbeat regulation. There are subtypes of channelopathies that include long QT syndrome (LQTS), Brugada syndrome (BrS), catecholaminergic polymorphic ventricular tachycardia (CPVT), and short WT syndrome (SQTS) (Campuzano *et al.*, 2015).

1.8 Cardiovascular disease resulting from variants of unknown significance

The NGS approaches mentioned above are being used in the field of genetic diagnosis of inherited cardiac disorders, where standard care procedures continue to be updated (Mogensen *et al.*, 2015). Currently, the Gene Cardiac Panel and NGS-based methods have identified more than 400 potentially causative mutations in 60 genes that are correlated with cardiomyopathies (Watanabe, Kimura and Kuroyanagi, 2018). Cardiac disorders are common in diseases caused by splicing abnormalities, and VUSs have been identified in key genes for normal heart function resulting, in dysfunction of heart muscle (Guo *et al.*, 2012). In primary arrhythmic diseases, the main genes affected by genetic mutations occur in cardiac ion channels (such as *KCNQ1*, *KCNH2*, *SCN5A*, *KCNJ2*), and intracellular Ca^{2+}-regulating gene (e.g. *RYR2*, *CASQ2* genes). In contrast inherited cardiomyopathies are linked with genes encoded for

sarcomeric and nuclear proteins (*TTN2-2*, *TNNI3*, *TTN*, *MYH7, FLNC, NFAT, MYBPC3*), as well as Ca^{2+} homeostasis regulators and desmosomal proteins (i.e. *PKP2*, *DSG2*) (Tse *et al.*, 2013; Hinson *et al.*, 2015; Protonotarios and Elliott, 2019).

To date, a small number of variants located in splice sites (Table 3) have been confirmed to directly cause human heart disease.

GENE	MUTATION(s) in	VARIANT EFFECT
Dilated cardiomyopathy		
TTN	17 different variants located in 5' and 3' ss	Skipping of exons leading to truncated TTN
DSP	Disruption of 5'ss exon1	Unknown
LMNA/c	Disruption of ss in exon 2	Skipping of exon 2
Hypertrophic cardiomyopathy		
TTN	1 variant in 5'ss	Skipping of exons leading to truncated TTN
TNNT2	Disruption of 5' ss and activation of cryptic ss in exon 15	Skipping of exon 15 and 13bp inserted from intron 15
MYBPC	Disruption of 3' ss and activation of cryptic 3'ss	Out-of-frame with PTC
Arrhythmogenic right ventricular cardiomyopathy		
DSG2	Disruption of 5' ss exon 4	Skipping exon 4
PKP2	Disruption of 5' ss and activation of cryptic 5'ss in exon 5 Disruption of 5' ss exon 7 and 11 Generation of 3' ss exon 13	Exon 5: out of-frame Skipping exon 7 and 11 Exon 13: out of-frame
JUP	Disruption of 3' ss and activation of cryptic ss exon 4	15bp deleted of exon 4
Long-QT syndrome		
KCNH2	Disruption of 5' ss exon 7 Disruption of BPS exon 9 Disruption of 5' ss exon 10	Retention intron 7 or skipping exon 7 Retention intron 9 Retention intron 10
KCNQ1	Disruption of 5' ss exon 7	Skipping exon 7
Brugada Syndrome		
SCN5A	Insertion of TGGG in intron 27 and formation of a cryptic ss	96bp deleted in exon 27
Congenital heart disease		
GATA4	Splice junction site exon/intron 1	*In silico* prediction showed altered binding affinity for RBPs SRSF6 and Myf1
NR2F2	Disruption of 5'ss exon3	*In silico* prediction showed skipping exon 3

Table 2. Splice site mutation reported in literature to cause aberrant splicing in human heart disorders. Abbreviation: bp, base pair; BPS, branch point site; PTC, premature termination codon; ss, splice site. Adapted from (Van Den Hoogenhof, Pinto and Creemers, 2016).

The thesis focuses on the analysis of significant cardiac genes harbouring genetic variants that may cause mis-splicing events in cardiac tissue causing heart failure.

1.8.1 Dystrophin (*DMD*)

DMD is located on chromosome X spanning over 3000kb and is transcribed into a 14kb mRNA (Koenig, Monaco and Kunkel, 1988; Nishio *et al.*, 1994). The gene includes 79 exons which encode for Dystrophin protein, a member of the spectrin superfamily. Dystrophin is characterised by an actin-binding domain and multiple spectrin repeats in N-terminal. In skeletal muscle, it forms a component of the dystrophin-glycoprotein complex (DGC), which bridges the inner cytoskeleton and the extracellular matrix. The *DMD* gene undergoes alternative splicing where three full-length dystrophin isoforms are produced and differ in their first exon. Each of the isoforms is controlled by a tissue-specific promoter in skeletal muscle and cardiomyocytes (Dp427m), brain (Dp427b), and Purkinje cerebellar neurons (Dp427p) (Muntoni, Torelli and Ferlini, 2003). Different mutations, such as deletions, duplications, point mutations and splicing mutations (principally identified at splice site consensus sequences), are seen and are related to Duchenne muscular dystrophy (DMD) cardiomyopathy (DCM) (Kamdar and Garry, 2016).

1.8.2 Desmoplakin (*DSP*)

The *DSP* gene encodes for desmoplakin protein which is one of the components of the desmosome structure, specialised for intercellular adhesion. The desmosomes are expressed in cells which undergo mechanical stress, in particular, the desmoplakin is a significant expression in the cardiac tissue. The protein presents three main domains, characteristics of the members of plakin superfamily, which consist of a globular N-terminal plakin domain, a central alpha-helix rod domain, and a C-terminal tail domain. Three different isoforms are produced by alternative splicing of the *DSP* gene: a long isoform (DSP-I), an intermediate isoform (DSP-Ia), and a short isoform (DSP-II) (Figure 9).

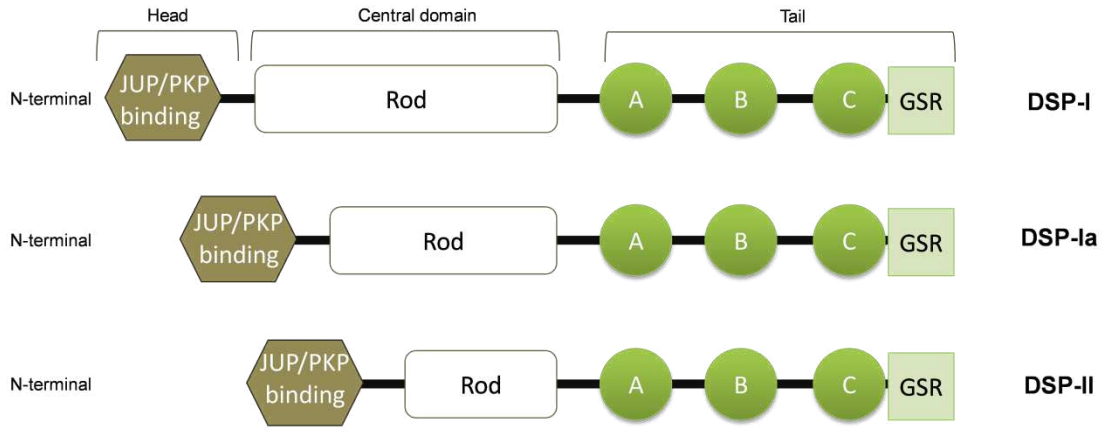


Figure 9. Schematic representation of the structure of DSP. The DSP protein includes three functional domains: JUP/PKP binding domain, the central rod domain (Rod), and the C-tail

which is composed of three PRDs (A-c) and a GSR domain that regulates DSP binding to intermediate filaments (IFs). The Rod domain differs in length in all the three isoforms (DSP-I, DSP-Ia, and DSP-II).

In terms of the N and C terminal domains, both isoforms are identical, except DSP-Ia which has only half of the rod domain while DSP-II lacks almost two-thirds of the rod domain(Yuan *et al.*, 2021). The most common cardiac isoform is DSP-I, which can also be detected in the skin. Although DSP-II was previously limited to the skin, but DSP-II transcripts have been identified in heart structure, such as the left atrium and ventricle, interventricular septum, left auricle, and apex, but significantly lower expression level than DSP-I. While DSP-Ia is expressed at low levels in epidermal keratinocytes and cardiac tissue, but it is the only isoform present in the aorta. DSP-Ia is the only isoform expressed in the aorta (Uzumcu *et al.*, 2006; Cabral *et al.*, 2010). Therefore, DSP regulates adipogenic and fibrogenic genes transcription, and controls ion channels and gap junctions to maintain an optimal electrical conductivity. Normal heart development and the preservation of its structural functions are regulated by DSP. According to studies, genetic variants in *DSP* gene have been linked to a number of hereditary cardiomyopathies, including dilated cardiomyopathy (DCM), arrhythmia cardiomyopathy, and left ventricular noncompaction.

1.8.3 Filamic C (*FLNC*)

FLNC encodes a structural protein, filamin C, expressed in striated muscles and is an actin-crosslinking protein in the sarcomere. In cardiomyocytes, the protein is located in the Z-disc and intercalated discs (Van der Flier and Sonnenberg, 2001; Estigoy *et al.*, 2009). It is one of 47 genes known to cause DCM (Tayal, Prasad and Cook, 2017), and in heart failure is shown to result in a differentially spliced transcript. *FLNC* contains a cassette exon encoding a flexible hinge region where in normal tissue, the cassette is largely spliced out (-H1 transcript), whilst in heart disease the inclusion of this cassette exon increases, with the inclusion ratio of the cassette being correlated to the state of disease. As a consequence, the function and structure of Z disks are altered in the cardiomyocyte (Kong *et al.*, 2010b). Splicing variants identified by whole exome sequencing (WES) in DCM patients have been observed. However, the pathogenic mechanisms that these variants have on splicing require further investigation, and consequently, there is a reduced expression level of cardiac FLNC protein (Begay *et al.*, 2016).

1.8.4 Lamin (*LMNA*)

LMNA maps to chromosome 1q21.1-21.2 and includes 12 exons spanning approximately 25kb. Nuclear lamins are encoded by alternative splicing events which produce the two main isoforms

lamin A and lamin C. Mutations in *LMNA* are common in familial cardiomyopathy (CMP) with around 6-8% of the cases representing idiopathic DCM (Hershberger and Siegfried, 2011), while the number of patients presenting with DCM and conduction defects is 33% and increasing (McNally and Mestroni, 2017).

1.8.5 Myosin binding protein C3 (*MYBPC3*)

Cardiac Myosin binding protein C (MyBP-C) is encoded by *MYBPC3* gene located on chromosome 11q11.2. Three isoforms have been identified in the adult human muscle: the cardiac isoform (cMyBP-C); the fast skeletal isoform encoded by the MYBPC2 gene which maps on chromosome 19q33.3, and the slow skeletal isoform produced by the MYBPC1 gene on chromosome 12q23.3(Carrier *et al.*, 1997). The cardiac isoform, cMyBP-C, has structural differences from the slow and fast skeletal isoforms. The distinct region includes 28 amino-acid insertions inside the C5 domain, several phosphorylation sites in the MyBP-C motif (M motif) between the immunoglobulin-like domain C1 and C2, and an additional immunoglobulin-like domain (C0 domain) at the N- terminal. The MyBP-C protein is oriented anti-parallel in the A-band of the cardiac sarcomere where titin, myosin and actin are binding in order to preserve the sarcomeric structure(Helms *et al.*, 2020). Pathogenic mutations in *MYBPC3* gene have been reported to be the major cause of HCM cases. The majority of *MYBPC3* mutations result in truncated protein products, while the remaining mutations include insertion/deletion, frameshift, or missense mutations of a single amino acid (Tudurachi *et al.*, 2023).

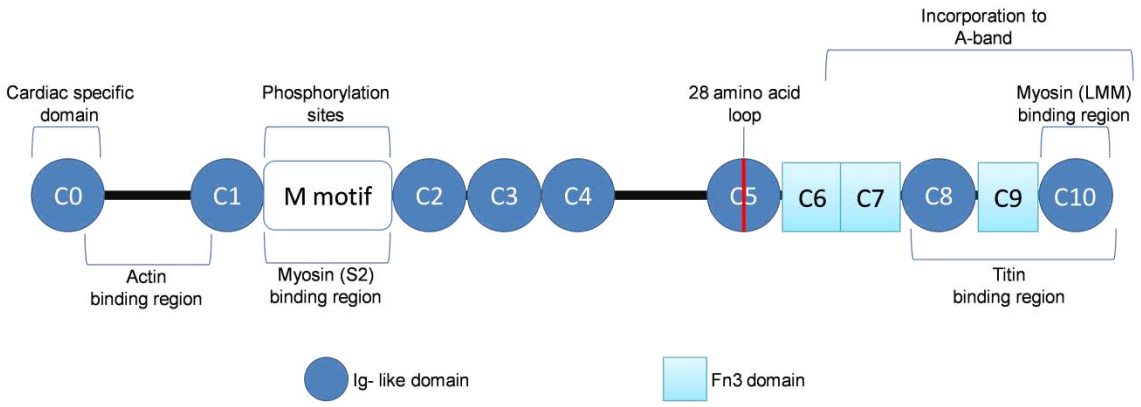


Figure 10. Schematic structure of cMyBP-C protein. cMyBP-C includes eight immunoglobulin (Ig)-like domains (C0, C1, C2, C3, C4, C5, C8, C10) and three Fibronectin type III (Fn3) domains (C6, C7, C9), with binding myosin and titin. The N-terminal C0 is specific for the cardiac isoform, as is M motif between C1 and C2 domain, which includes several phosphorylation sites.

1.8.6 Nexillin F-actin binding protein (*NEXN*)

The Nexillin F-actin binding protein is encoded by *NEXN* gene, located on chromosome 1 p31.1. It is a cardiac Z-disc protein which plays an important role in maintaining and stabilising the Z line within the sarcomere of cardiac tissue. The predominant mutations reported in the literature are missense and frameshift mutations, which are implicated in dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) (Wang *et al.*, 2010; Walsh *et al.*, 2017).

1.8.7 Polycystin-1 (*PKD1*)

PKD1 gene is located on chromosome 16p13.3 which encodes the protein polycystin-1 (PC-1). The encoded protein presents three functional domains: a large N-terminal extracellular stretch, several transmembrane domains, and a cytoplasmatic C-tail. PC-1, a membrane protein, regulates intracellular calcium homeostasis and calcium-permeable cation channels. Therefore, it is involved in cell-cell interactions and may influence G-protein-coupled signalling pathways(Su *et al.*, 2018). Pathogenic variants in *PKD1* or the associated gene *PKD2* are primarily recognised for their role in autosomal dominant polycystic kidney disease (ADPKD), studies have revealed its association with cardiovascular complications, The common complications associated with ADPKD include left ventricular hypertrophy, arrhythmias, and defects in cardiac valve (Spinelli, Giugliano and Esposito, 2021).

1.8.8 Ryanodine receptors 2 (*RYR2*)

RYR2 encodes a protein of the ryanodine receptors family (RyRs), ryanodine receptor 2. This protein plays a role in informing the channel that transports positively charged calcium atoms (calcium ions) in cells (Fill and Copello, 2002; Meissner, 2004). Interestingly, this gene shows one of the most complex isoforms in human genome, comprising > 100 exons, but only 13 splice variants have been detected in *RYR2* isoforms which causes alteration in Ca^{2+} release channels (George, Chang and Lai, 2005; Kimura *et al.*, 2005). However, their role in cardiac context is still unclear. *RYR2* mutations are associated with cardiac disorders with a strong impact.

1.8.9 Transmembrane protein 43 (*TMEM43*)

TMEM43 encodes transmembrane protein 43 (also named Luma) of 45kDa and is involved in maintaining nuclear envelope structure by organizing protein complexes (Bengtsson and Otto, 2008). *TMEM43* sequence variants are the main cause of Arrhythmogenic right ventricular cardiomyopathy (ARVC) (Baskin *et al.*, 2013).

1.8.10 Titin (*TTN*)

The majority of cardiovascular disease variants affecting splicing have been identified in the sarcomeric gene, *TTN*, where alternative splicing events are associated with heart failure (Kong *et al.*, 2010a). *TTN* encodes for the largest known protein in humans, titin, with 363 exons encoding 38,138 amino acids (Bang *et al.*, 2001). Titin is the third most abundant protein in sarcomeres (the basic unit of myofibrils in striated muscle) and is subjected to alternative splicing to generate different isoforms. The main isoforms produced are N2A, N2B, and N2BA which predominately differ in the presence of a specific domain in the transcript. The two isoforms expressed in adult cardiomyocytes are N2B and N2BA (Zhu, Chen and Guo, 2017). The difference is based on their length and stiffness, where N2BA is longer due to an insertion, inclusion of N2A domain and results in a larger molecular weight and less rigid protein compared to N2B (Freiburg *et al.*, 2000). Interestingly, during heart development, the expression ratio of isoforms is different and is reflected in the size of the sarcomere I-band region of titin and its passive tension (Neagoe *et al.*, 2003). Of the known human genetic variants that cause familial DCM, *TTN* mutations are the most common, truncating mutations accounting for 20-25%, of which 31% are splice site variants (Herman *et al.*, 2012).

1.8.11 Vinculin (*VCL*)

VCL maps to the long arm of chromosome 10 (10q22.1q23) and includes 22 exons encoding a cytoskeletal protein, vinculin, containing 1,066 aminoacids. Functional studies conducted on *VCL* indicate alternative splicing of exon 19, which produces a cardiac/smooth muscle-specific isoform, named metavinculin. Metavinculin is identical to vinculin except for an additional 68 amino acids in the C-terminal domain (Belkin *et al.*, 1988; KOTELIANSKY *et al.*, 1992). In cardiac muscle, both isoforms co-localise in costameres and intercalated discs (Chen, Choudhury and Craig, 2006; Thompson *et al.*, 2017). Genetic defects in *VCL* influence the cellular status of myocytes leading to the development of heart disease. Furthermore, several studies have observed a decrease in protein levels in intercalary discs of cardiomyocytes and not in the Z discs (Vasile, Edwards, *et al.*, 2006; Vasile, Ommen, *et al.*, 2006).

1.9 Potential therapeutic approaches for alternative splicing

Identification of VUS impacting splicing in relation to CVDs is important for several reasons. In a diagnostic setting, if the mutation is attributed as pathogenic and results in alternative splicing profiles, the resulting isoforms can be used as diagnostic biomarkers within the family. This is of particular importance when identifying family members at risk and allows for appropriate management and therapy to be administered(Le *et al.*, 2015). For example, a study conducted

on the heart of cardiomyopathy patients showed a high level of myomesin isoforms compared to normal heart tissue. As a consequence, there is activation of the foetal isoforms and morphology adaptation in the dilated heart. Therefore, the identification of myomesin isoforms at the early phases of the disease and their correlation with ventricular function can be used to implement the characterisation of DCM human patients. This could lead to prompt action in diagnosis in order to monitor pathophysiological mechanisms and prevent the progression of disease to irreversible ventricular remodelling and heart failure (Makarenko *et al.*, 2004; Schoenauer *et al.*, 2011).

With regards to RNA mis-splicing events and their association with disease, identification of a VUS as a splicing mutation may allow therapeutic approaches designed to target disease-causing RNAs, such as antisense oligonucleotides (ASOs), spliceosome-mediated RNA *trans*-splicing (SMaRT), bifunctional oligonucleotides, modified snRNAs, and small compounds (Figure 9).

Other therapeutic avenues, such as gene therapy to reverse the effects of RNA misprocessing have also been developed because of recent advancements in novel technologies. Zinc finger (ZF) proteins (ZFPs), transcription activator-like nucleases (TALENs), and CRISPR/Cas9 systems are examples of such strategies (Fernández, Josa and Montoliu, 2017; Knott and Doudna, 2018).

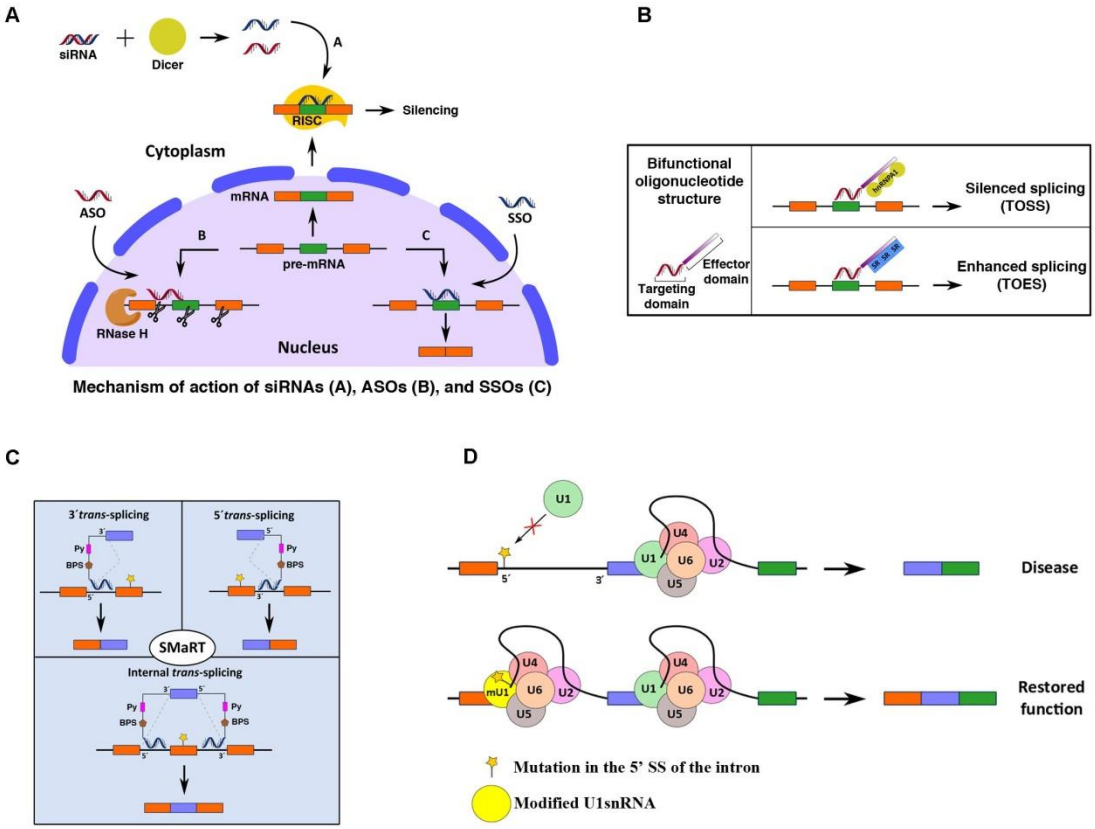


Figure 11.Representation of the strategies proposed for targeting mis-splicing events. (**1**) siRNA-, ASO-, and SSO- based strategies (a, b, and c in the figure respectively) for pre-mRNA

editing. (**2**) Bifunctional oligonucleotides include one small sequence which binds the pre-mRNA target sequence (targeting domain) and effector domain that recruits specific regulatory factors to influence splicing (TOSS and TOES). (**3**) Three different SMaRT approaches are illustrated: 5' *trans*-splicing, 3' *trans*-splicing, and IER, that target the 5' and 3' ends, or internal portion of a mutated target pre-mRNA, respectively. (**4**) Modified snRNA usage to correct splicing issues is represented schematically. In the cartoon, a modified U1 snRNA is represented to correct a mutation located in 5' ss. Reproduced from (Suñé-Pou *et al.*, 2020).

1.9.1 Antisense oligonucleotides (ASOs)

With respect to therapy, a promising approach is the use of Antisense Oligonucleotides (AONs or ASOs) to redirect splicing (Pros *et al.*, 2009; Bergsma *et al.*, 2018). This approach consists of a short synthetic single-stranded DNA or RNA molecule that is complementary to a specific pre-mRNA fragment. Their mechanism of action in modulating splicing is occluding the binding sites of specific proteins involved in splicing regulation (Siva, Covello and Denti, 2014).

ASOs can be designed to target the splice site blocking its recognition and redirecting splicing to an adjacent site. Alternatively, antisense oligonucleotides could influence the binding of *trans*-factor splicing at their target by covering the *cis*-elements (enhancers or silent elements), therefore splicing will be enhanced or blocked. A further mechanism is to base-pair across the base of regulatory structures, such as stem-loops, in order to stabilise/destabilise the RNA structure (Figure 12) (Havens, Duelli and Hastings, 2013; Suñé-Pou *et al.*, 2017).

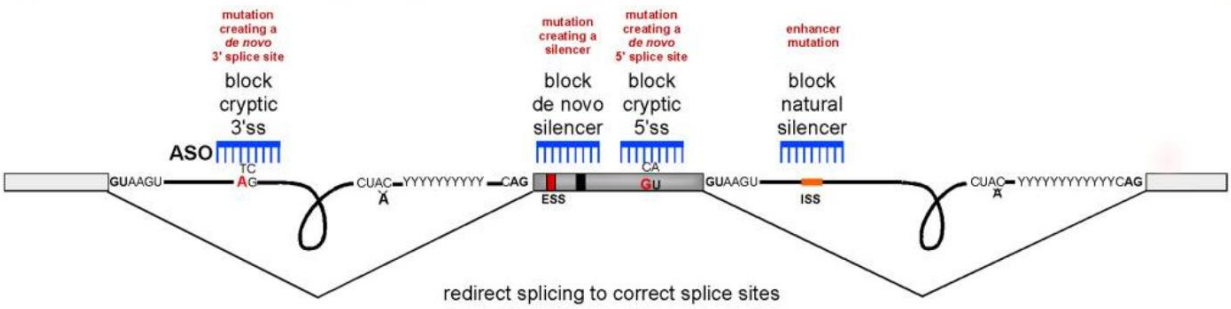


Figure 12.ASO-based correction of mutation-induced aberrant splicing. In blue is represented ASOs which can target different positions on pre-mRNA, while red indicates the correction from ASOs to rescue splicing and induce exon inclusion. Adapted from (Havens, Duelli and Hastings, 2013).

The sequence specificity of ASOs allows precise targeting of endogenous RNA and reliable recognition of different RNA isoforms and can inhibit only the mutated gene alleles, beneficial for treatment of genetic variants causing autosomal dominant disease. The specificity is

important to avoid non-specific side effects (Fiset and Gounni, 2001). Therefore, they are relatively non-invasive, as they do not directly alter the genome.

However, early use of antisense therapy had limitations, such as toxicity, delivery, and uptake of the drug. Chemical modifications have partly diminished these issues (Fiset and Gounni, 2001). The first ASO generation included changes to the phosphate backbones and/or sugar components of the oligonucleotides, such as the use of phosphorothioate deoxynucleotides (PS-ODN), where the oxygen in the sugar-phosphate backbone is replaced with a sulphur atom (Eckstein, 2014). Therefore, the molecule shows chirality, but maintains the overall charge and RNAseH should be activated for mRNA degradation. However, the ASOs' structure showed toxicity due to the binding of proteins on the cell surface or in the serum. As a consequence, the complement cascade is activated leading to hematologic changes such as reduced heart rate, blood pressure and cardiac dysfunction. In addition, they could not cross the lipid bilayer due to their polyanionic nature which leads to interaction with many cell types and not only the cell target when in circulation (Kuramoto *et al.*, 1992; Fiset and Gounni, 2001). Additional modifications in ASOs structure included the use of locked nucleic acid chemistry for bridging of the sugar furanose ring, alterations at 2' positions of the ribose sugar ring (2'-O-methylation [2'-MOE]) (second-generation ASOs), with increased target affinity and decrease of some PS-ODN side effects. Finally, third generation ASOs have been developed with the addition of phosphorodiamidate morpholino oligomers (PMO) or peptide nucleic acids (PNA) which avoid nuclease recognition. However, oligonucleotide toxicity remains a limitation (Crooke, 2017).

This approach was used in the case of Spinal muscular atrophy (SMA), a disorder characterised by an aberrant splicing of exon 7 in survival motor neuron 1 (SMN1) and survival motor neuron 2 (SMN2), caused by a C to T substitution in position +6 of the exon (Ottesen *et al.*, 2016). The only approved drug for SMA, Spinraza™/Nusinersen, consists of an antisense oligonucleotide which corrects SMN2 exon 7 splicing by blocking the intronic splicing silencer N1 (ISS-N1) located immediately downstream of exon 7. ZolgenSMA, is a s9-AAV which carries the SMN complementary DNA to SMA type 1. Another example can be found in the treatment of Duchenne muscular dystrophy resulting from frameshift mutations in the *DMD* gene. Exon 51 skipping is a frequent outcome in DMD patients. As a consequence, a treatment focused on restoring the reading frame of the protein due to the deletion of exon 51; two AON drugs (drisapersen and eteplirsen) have been developed and used in clinical trials with promising results (Veltrop and Aartsma-Rus, 2014).

1.9.2 Bifunctional oligonucleotides

Ectopic modulators of alternative splicing, such as bifunctional oligonucleotides, are used to regulate splicing patterns of specific genes. These oligonucleotides contain an antisense part that targets a unique sequence and a non-hybridizing tail or effector domain that recruits acting factors (targeted oligonucleotide enhancer of splicing [TOES] or target oligonucleotide silencer of splicing [TOSS]). Oligonucleotides that include binding site sequences for the splicing repressor hnRNP A17A2 have been used to reprogram AS via TOSS. The oligonucleotides are positioned upstream of a 5' ss to disrupt the binding site of the U1 snRNP binding and suppress the use of the splice site. An example was reported by Dikson et al. to restore *SMN2* splicing. The experiment consisted of a TOSS with hnRNPA1 tails to inhibit exon 8 inclusion in *SMN2*, promoting the inclusion of exon 7 and regaining the protein function. Analog to this, the bifunctional TOES, whose tail of enhancer sequences recruits activating proteins such as SR proteins, has been used to increase splicing of refractory exon 7 in *SMN2* in fibroblasts from patients with SMA(Dickson, Osman and Lorson, 2008). This approach is mechanistically distinct from ASO tools, although both can be used to promote exon 7 inclusion in SMA. While ASOs such as Spinraza block the binding of splicing factors to intron 7, resulting in exon 7 inclusion, the bifunctional TOSS/TOES is thought to drive exon 7 inclusion and restore protein expression (Suñé-Pou *et al.*, 2020).

1.9.3 Small interfering RNA (siRNAs)

Another approach for addressing aberrant splicing has been developed with the discovery of the siRNA pathway, which silences the expression of genes. RNA interference (RNAi) consistsof silencing or degradation of a specific mRNA by a double-stranded RNA (dsRNA). The approach is based on synthetic 21- to 25- nucleotide duplexes with overhanging 3' ends (siRNAs) whichnegatively regulate gene expression. Three main categories are included as small RNAs: siRNAs, microRNAs, or miRNAs, and piwi-interacting RNAs, or piRNAs. The structure, biological functions, associated effector proteins and origins of these small RNA species differ (Dana *et al.*, 2017).

To maintain genomic integrity, in a physiological environment, siRNAs block the action of foreign nucleic acids, including those of viruses, retrotransposons, transposons and transgenes, whereas miRNAs regulate the gene expression at post-transcriptional level as endogenous gene regulators. On the other hand, piRNAs have been associated with the silencing of retrotransposons at post-transcriptional and epigenetic levels as well as other genetic components in the germline that are activated during spermatogenesis(Meister and Tuschl, 2004; Siomi *et al.*, 2011).

Synthetic siRNAs, mimic endogenous Dicer products, are delivered into cells as a multiprotein complex known as the RNA-induced silencing complex (RISC) which binds them in the cytoplasm. The two strands of the siRNAs are separated and linked with the RISC complex hybridizes with the target mRNA.Argonaute-2 (Ago2), a RISC catalytic component, subsequently induces mRNA cleavage of the other strand (Figure 13) (Dana *et al.*, 2017).

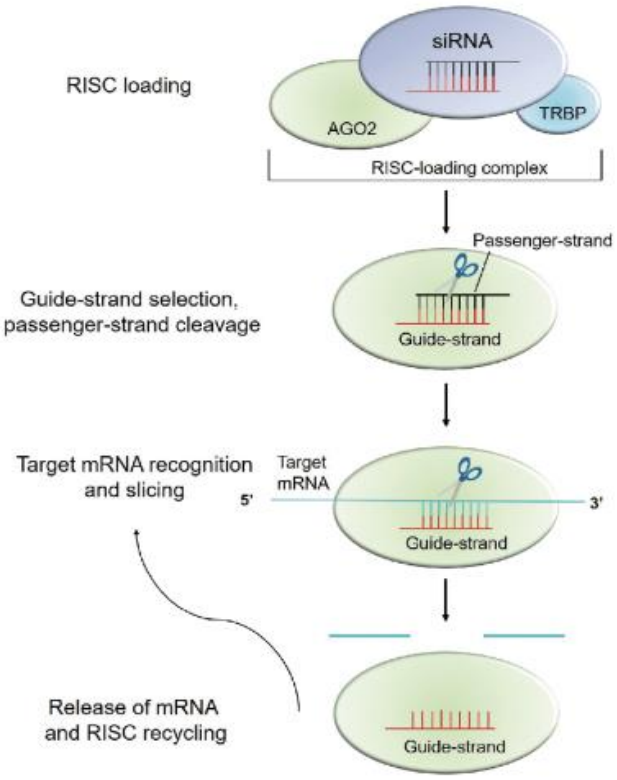


Figure 13.Schematic representation of siRNA silencing mechanism. Synthetic dsRNA introduced into the cytoplasm is processed by the RISC complex into 22-nt pieces with 2-nt single-stranded overhangs on the 3' ends. The structure of synthetic siRNA mimics that of Dicer products. The siRNA guide strand is loaded into the RNA induced silencing complex (RISC), and the passenger strand is cleaved by Argonaute-2 (Ago2). The activated RISC–guide-strand complex identifies and cleaves mRNA that is complementary to the guide strand, preventing translation and thereby silencing gene expression. Reproduced from (Saw and Song, 2020)

Targeting aberrant splicing isoforms for therapeutic purposes involves a strategy in which siRNAs are used to target exonic/intronic sequences near alternative exons or exonic/intronic junction sequences. These should cause the destruction of alternatively spliced transcripts and aberrant mRNAs without compromising the transcription of canonical mRNAs (Suñé-Pou *et al.*, 2017).

The necessity for safe and efficient delivery mechanisms is currently a major obstacle in achieving the full promise of siRNA-based therapies and their introduction into clinical treatment. Unmodified siRNAs cannot cross cell membranes to enter cells, they are unstable in

circulation and may be immunogenic (Whitehead, Langer and Anderson, 2009). Therefore, chemical changes and/or delivery materials are essential to safely transfer siRNA to the site of action. To overcome these difficulties of *in vivo* distribution, a wide range of strategies are being investigated including polymers, lipids, peptides, antibodies, aptamers, and small molecules (Dong, Siegwart and Anderson, 2019).

1.9.4 Trans-splicing: SMaRT

A method based on trans-splicing in which the complete coding sequence upstream or downstream of a target 5' or 3' splice site is replaced, is known as spliceosome-mediated RNA trans-splicing (SMaRT).The target mRNA, the spliceosome, and the pre-trans-splicing molecule (PTM), also referred to as the RNA trans-splicing molecule (RTM), are the three elements that compose this strategy.

The PTM includes an ASO that targets the endogenous intron of the mutated splice site, a synthetic splice site that directs splicing to bridge from the endogenous RNA to the PTM, and a copy RNA sequence that will be spliced to the endogenous RNA rather than to the mutated, inactive endogenous splice site(Wally, Murauer and Bauer, 2012).

SMaRT groups can be categorised according to the pre-mRNA region that is targeted: internal exon replacement (IER), which targets the internal component of the pre-mRNA, 5'-trans-splicing, which targets the 5' region, and 3’ *trans*-splicing, that targets the 3’ region.

This approach is an effective way to correct expression genes with genetic variations located in canonical splice sites, especially when the splice site mutation is in the first or last nucleotide of the intron. It is challenging to restore splice site activity because the base specificity at these positions is crucial for the catalytic phases of the splicing event. It has been proven that SMaRT is a useful method for correcting gene expression for these types of mutations (Suñé-Pou *et al.*, 2017)

One benefit of gene replacement is that the expression of the pre-mRNA of the target gene continues to be under endogenous control. Therefore, the transcription of the pre-mRNA is controlled by the endogenous promoter since just a small piece of the gene is replaced by the PTM. The tissue, temporal, and expression levels of the gene are unaffected because the *trans*-splicing molecule can only bind with an existing pre-mRNA. These features are promising when therapeutic RNA levels must be slightly increased, and some read-through of the mutant gene is acceptable. This is because 100% efficacy of the *trans*-splicing approach may not be achieved, due to limits on viral expression and delivery into the cells. So, a crucial step for the success of the treatment in clinics is the criteria used to select the appropriate delivery vector.

Different disorders, in cellular and mouse models including cystic fibrosis (CF), haemophilia A, SMA, retinitis pigmentosa (RP), frontotemporal dementia with parkinsonism-17 (FTDP-17), and tauopathies, have been exploited via SMaRT to correct defects in the splicing process. It has been reported that this method can successfully reprogram gene expression, giving prospects for future gene therapy applications but in order to maximise overall performance, optimization and a deeper understanding of the mechanism are needed (Havens, Duelli and Hastings, 2013).

1.9.5 Modified snRNAs

Another form of potential therapy addressing aberrant exon skipping is achieved by disrupting the exon definition sequences and their regulatory elements. The U1 snRNP plays a major role in splicing definition, during the last decade, research has focused on the refinement of the U1 snRNA particle to generate a possible splicing correction tool. Modified spliceosomal snRNAs have been generated with sequence modifications that bind by base-pairing to the mutant 5' ss in order to re-establish splicing. The complementary mutation allows the snRNA to effectively bind the mutant binding site of the pre-mRNA and restore normal splicing. The requirement to incorporate the snRNAs into an expression vector and deliver them to cells is a disadvantage of this strategy. This approach has similar limitations to trans-splicing approaches.

Many experimental examples have shown that increasing the complementarity of the U1 snRNA 5’ tail over the mutated 5’ ss of a specific exon can correct the aberrant splicing of coagulation factor VII gene exon 8 (Pinotti *et al.*, 2008). In 2009, Roca and Krainer (Roca and Krainer, 2009) reported that U1 snRNA splicing rescue is mediated by the recognition of both canonical and atypical splice sites with a high probability of producing deleterious off-target effects. For this reason, to reduce non-canonical interactions, a second generation of engineered U1s, called Exon-specific U1 snRNAs (ExSpeU1s), was developed and successfully tested showing the rescue of exon skipping caused by several types of splicing mutations. ExSpeU1s are modified U1 snRNAs that carry an engineered 5’ tail which interacts with intronic sequences downstream of the 5’ss of a specific exon by complementarity. The ExSpeU1 strategy has two important advantages: they do not interact directly with normal donor sites and can correct different types of splicing defects associated with exon skipping. ExSpeU1s have been demonstrated to have efficacy as splicing rescuers in numerous disease contexts including; Haemophilia B (Alanis *et al.*, 2012), Cystic Fibrosis, Familial Dysautonomia (Donadon *et al.*, 2018), Fanconi Anemia (Mattioli *et al.*, 2014), and SMA (Dal Mas *et al.*, 2015; Donadon *et al.*, 2019).

Modified versions of spliceosomal snRNAs have also been tested for their usefulness in restoration of base pairing to the mutated splice site. For instance, combination therapy with U1 and U6 snRNAs that have been mutation-adapted has been used to correct mis-splicing events

caused by genetic mutations in exon 5 of the *BBS1* gene, which causes developmental delay and retinal degeneration. Another example of a modified oligonucleotide is U7 snRNA, which participates in histone pre-mRNA maturation by identifying the sequence of the histone 3' untranslated region. Changes in the target sequence can be introduced to convert this snRNA into an antisense tool capable of blocking splicing signals and inducing exon skipping or inclusion. Using this method, synthetic U7 snRNAs have been created to improve the exon 23 skipping of mutant dystrophin in *DMD*. The modified snRNA strategy due to its molecular mechanisms, does not affect the physiological expression of the target gene, which is one of its key benefits (Sué-Pou et al., 2020).

1.9.6 Small-molecule compounds

This strategy applied to target disease-causing RNAs, uses small molecules which are often identified using high-throughput screening assay using cells that report alterations in a particular splicing event. These molecules can act by targeting splicing factors to directly modulate their function or indirectly by targeting RNA sequences and/or structures (such as hairpins or G-quadruplexes) in order to inhibit the recruitment of splicing factors to mutant sequences (Scotti and Swanson, 2016).

One drawback of small molecule therapeutics is the lack of specificity and information regarding the exact mechanism of action, which can lead to undesirable off-target effects. Despite this shortcoming, a major benefit of small molecules is that many are already approved and in use in clinical practice to treat diseases and conditions other than splicing defects. Therefore, these molecules have been deemed safe for use in humans, which vastly accelerates their development as a treatment for a disease. There are numerous classes of molecules currently being studied for their ability to therapeutically alter splicing (Havens, Duelli and Hastings, 2013).Compared with oligonucleotide-based therapeutics, these compounds are easier to deliver to target sites and normally have lower toxicity profiles (Suñé-Pou *et al.*, 2020).

A number of molecules have been discovered to impact numerous splicing phases, as shown in Figure 14.

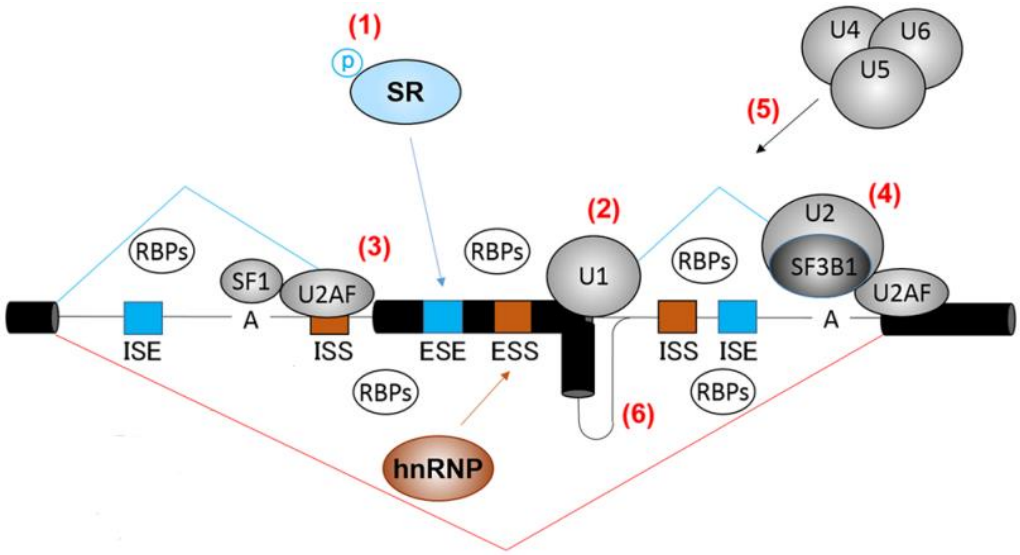
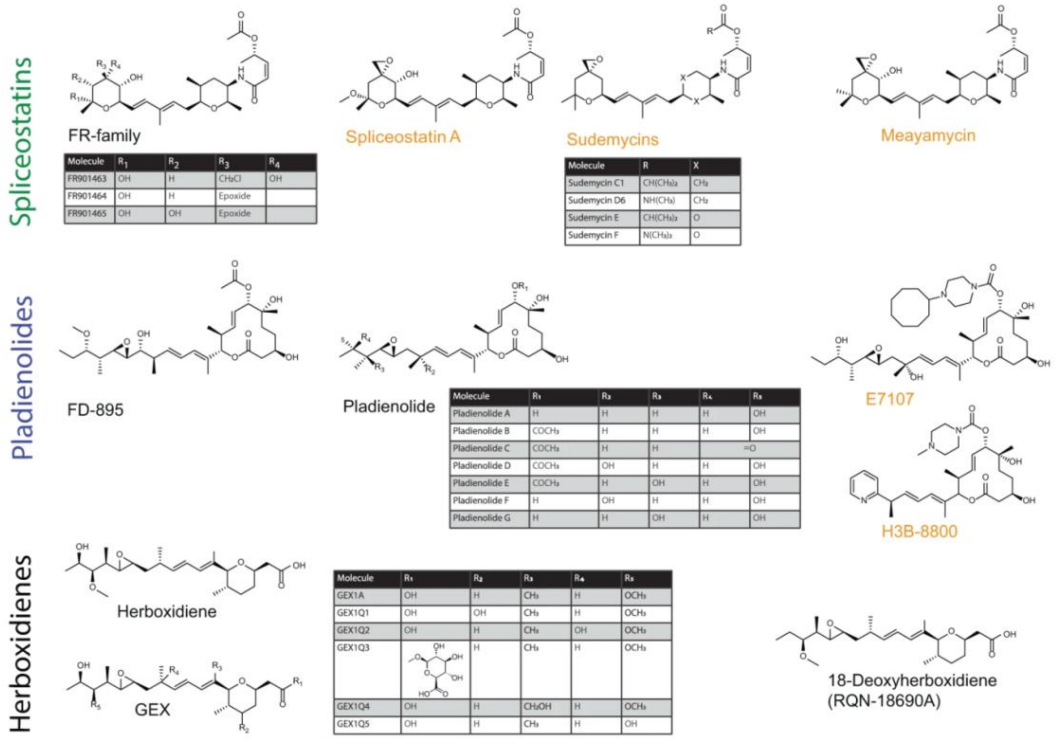


Figure 14.Schematic representation of splicing where chemical compounds are known to target (red numbers). (**1**) Phosphorylation of SR proteins. (**2**) 5' splice site recognition by U1 snRNP. (**3**) Polypyrimidine tract recognition by U2AF. (**4**) Branch point recognition by SF3B1. (**5**) Recruitment of U5, U4/U6 snRNP splicing complex to the spliceosome. (**6**) RNA structure. Adapted from (Ohe and Hagiwara, 2015)

SR proteins interact with exon splicing enhancers (ESE) and their activity is crucial for splicing regulation. Their phosphorylation is regulated by several kinases which could be influenced by RNA-binding protein kinase inhibitors. In general, blocking these kinases would prevent SR proteins from being phosphorylated and would result in exon-skipping. Inhibitors of RNA-binding protein kinases, as a therapeutic tool, can regulate the dysfunction of splicing events in inherited disorders (Rossi *et al.*, 1996; Fu and Ares, 2014).

Splicing modulators have been identified to interfere with spliceosome components and thus influence spliceosome assembly. For example, plant cytokinins have been found to modulate the binding of U1 small nuclear ribonucleoprotein (U1 snRNP) to the 5' splice site, an important step of pre-spliceosome formation in exon definition.

Another target is the splicing factor 3b, subunit 1 (SF3B1), which is a component of U2 small nuclear ribonucleoprotein (U2 snRNP) that base-pairing to the branch-point sequence. Mutations in SF3B1 have led to discovery of splicing inhibitors (SF3B1 inhibitors), which include splicing modulators from three families: spliceostatins, pladienolides and herboxidienes (Figure 15) (Kaida *et al.*, 2007; Kotake *et al.*, 2007; Hasegawa *et al.*, 2011). Compounds that target SF3B1 can be demonstrated to arrest the spliceosome complex in its open conformation, preventing the complete formation of the A complex (Cretu *et al.*, 2018).



Molecule	R_1	R_2	R_3	R_4
FR901463	OH	H	CH_2Cl	OH
FR901464	OH	H	Epoxide	
FR901465	OH	OH	Epoxide	

Molecule	R	X
Sudemycin C1	$CH(CH_3)_2$	CH_2
Sudemycin D6	$NH(CH_3)$	CH_2
Sudemycin E	$CH(CH_3)_2$	O
Sudemycin F	$N(CH_3)_2$	O

Molecule	R_1	R_2	R_3	R_4	R_5
Pladienolide A	H	H	H	H	OH
Pladienolide B	$COCH_3$	H	H	H	OH
Pladienolide C	$COCH_3$	H	H	=O	
Pladienolide D	$COCH_3$	OH	H	H	OH
Pladienolide E	$COCH_3$	H	OH	H	OH
Pladienolide F	H	OH	H	H	OH
Pladienolide G	H	H	OH	H	OH

Molecule	R_1	R_2	R_3	R_4	R_5
GEX1A	OH	H	CH_3	H	OCH_3
GEX1Q1	OH	OH	CH_3	H	OCH_3
GEX1Q2	OH	H	CH_3	OH	OCH_3
GEX1Q3		H	CH_3	H	OCH_3
GEX1Q4	OH	H	CH_2OH	H	OCH_3
GEX1Q5	OH	H	CH_3	H	OH

Figure 15. Spliceostatins, Pladienolides, and Herboxidienes families which inhibit SF3B1 are represented. The figure shows some reported synthetic and semi-synthetic derivates of natural products identified to target the SF3B1. Adapted from (Schneider-Poetsch, Chhipi-Shrestha and Yoshida, 2021)

Additionally, a biflavonoid isolated from the leaf of *Ginko Biloba*, named isoginkgetin, has been identified to inhibit splicing by preventing recruitment of U4/U5/U6 tri-snRNP association to the spliceosome, as a result, there is an accumulation of the pre-spliceosomal A complex (O'Brien *et al.*, 2008). These observations strengthen the idea that not only the mechanism of action but also the relative potency of the binding molecule determines the efficiency of a particular modulator.

Splicing modulators are excellent examples of the potential and the difficulties of chemical biology. They reveal the potency of natural products and also the need for medicinal chemistry to generate adequately stable and potent molecular effectors in sufficient amounts. In addition, they illustrate how molecules with different structures can share a binding site and mechanism of action, but that even minor changes in binding can lead to significantly different biological effects. Therefore, it is crucial to comprehend the physiological consequences beyond the interaction between the drug and target molecule because splicing modulators disturb a critical step in gene expression control, and even that may not be enough.

AIM

The field of RNA biology, especially RNA splicing, has seen significant progress over the past two decades. Knowledge of splicing regulation has greatly expanded, leading to the discovery of the essential splice factors and their function, unveiling the splicing code. In the heart, specific splicing factors have been identified and this has led to new insights into the mechanism of cardiac disorders. However, prior research indicates that splicing disruption may be more common in congenital heart disorders than previously assumed, given the lack of understanding regarding the number of splicing modulators, mutations, and splice isoforms implicated.

This thesis aims to analyse genetic variants, in particular VUSs, identified during diagnostic gene testing for inherited cardiac conditions and evaluate their impact on mRNA splicing. In addition, this project aims to investigate the splicing mechanism involved and explore potential therapeutic approaches in order to provide evidence to support their potential efficacy in treating splicing disorders.

The research objectives are:

1. To screen genetic variants identified in patients affected with CVDs that may influence mRNA processing by minigene assay to determine their effect on pre-mRNA splicing.

2. To determine the utility of whole blood as a clinically accessible specimen for RNA analysis of cardiac conditions in clinical practice via:

    - Determining the utility of RT-qPCR to identify nonsense-mediated decay.

    - Comparison of data from RT-PCR and short read RNA-seq for identifying splicing abnormalities and their application in diagnostic practice.

3. Evaluate the efficacy of small molecule splicing modulators to rescue mis-splicing events.

Chapter 2 Material and Methods

2.1 Chemical Reagents

General chemical reagents were purchased from ThermoFisher, Sigma-Aldirch, Bio-Rad, Roche, New England Biolabs, Promega, Gibco BRL (USA), and Invitrogen.

2.2 Common solutions and buffers

- PBS (Phosphate Buffered Saline): 137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, pH 7.4.
- 10X TBE: 108g/L TRIS, 55g/L Boric Acid, 40mL EDTA 0.5 M, pH 8.
- 5X DNA loading buffer: 100g sucrose, 40g UREA, 100µl TBE 5X, 1 % bromophenol blue bring at final volume 200mL with ddH_2O.
- Luria – Bertani medium (LB): 1% w/v DifcoBactotryptone, 0.5% w/v Oxoid yeast extract, 1% w/v NaCl in dH_2O, pH 7.5.

2.3 Synthetic oligonucleotides

Synthetic DNA oligonucleotides were purchased from Sigma-Aldrich and Integrated DNA Technologies (IDT). The DNA pellet was resuspended in ddH_2O at 1µg/µl final concentration. Primer was diluted 1:10 with ddH_20 (100ng/µl final concentration).

2.4 Oligonucleotides

Deoxyribonucleoside Triphosphate (dATP, dCTP, dGTP, dTTP, 100nM concentration) nucleotides were obtained from Promega (USA).

2.5 Bacterial culture

2.5.1 Bacterial culture

E. coli K12 strain DH5α competent cells were used to perform transformations with the plasmid of interest. Bacterial colonies were maintained in the short term as single colonies on agar plates containing the appropriate antibiotic at 4°C. When necessary, bacteria were amplified by an overnight incubation in Luria-Bertani medium at 37°C with shaking. When appropriate, the antibiotic Ampicillin (Sigma-Aldrich) was added to a final concentration of 100µg/mL to allow

selected bacterial growth. The following day, cultures of bacteria were centrifuged at 4000 rpm and the pellet was used as input for plasmid purification kits.

2.5.2 Preparation of bacterial competent cells

Transformation & Storage Solution (TSS) (Chung, Niemela and Miller, 1989) and the laboratory standard protocol were used to prepare bacterial competent cells. First, *E. coli* strains were spread on LB agar plate without antibiotic and incubated overnight at 37°C. The day after, a single colony was grown in 10mL of LB medium in 50mL falcon tube in a 37°C shaker overnight. After 24h, the pre-inoculum was transferred to 200mL of fresh LB medium (kept at room temperature) using aseptic technique and cells were grown at 37°C under agitation until OD_{600} was 0.3-0.4. The optical density was determined using a spectrophotometer. Cell growth was arrested by tube transferring the tube to ice and then pelleted by centrifugation at 3000 rpm for 10min at 4°C. The pellet was re-suspended at one-tenth of the initial volume in ice-cold transformation and storage solution (TSS, which consists of 10% w/v PEG, 5% v/v DMSO, 35mM $MgCl_2$, pH 6.5 in LB medium). Cells were aliquoted in 1.5mL tube and rapidly frozen in dry ice and stored at -80°C. The efficiency of competition was determined by transformation with 10pg of plasmid vector and considered satisfactory if resulted in more than 100 colonies.

2.5.3 Bacterial transformation

Transformation allows bacterial cells to take up exogenous genetic materials from the surroundings though the cell membrane. For re-transformation, 60µl of freshly thawed *E. Coli* DH5α competent cells were incubated on ice for 30min with 20ng of plasmid DNA. The cells-DNA mix was heat shocked in a water bath at 42°C for 90s and chilled on ice for 10 additional minutes. 60µl of preheated LB medium was added to bacteria and incubated for recovery at 37°C for 1h with gentle shaking. Cells were then spread onto agar plates containing the appropriate antibiotic concentration (100µg/mL Ampicillin) as a selection marker and the plates were incubated overnight at 37°C.For cloning purposes, ½ of the inactivated ligation reaction volume was transformed using the same heat shock and plating protocols described above.

2.5.4 Small scale preparation of plasmid DNA from bacterial cultures

Starting with an overnight culture of *E. coli* cells containing the plasmid of interest, the cells were pelleted and used for DNA extraction using the Wizard plus SV miniprep DNA purification kit (Promega) according to the manufacture's guidelines. The method is based on the alkaline lysis which separates chromosomic DNA, cellular proteins, and other contaminants from the DNA plasmid. The supernatant containing plasmid DNA was loaded onto a column, washed

several times, and eluted to achieve purified plasmid DNA. After purification, plasmid DNA was stored at -20°C for the subsequent reactions.

2.6 Quantification of nucleic acid concentration

The purity and concentration of nucleic acids were determined by measuring the absorbance ratios (A_{260}:A_{280}, and A_{260}:A_{230}) using NanoDrop (ND-1000 Spectrophotomer, ThermoFisher). An optical density of 1.0 at 260 nm is considered equivalent to a concentration of 50µg/mL for double stranded DNA (dsDNA) and 40µg/mL for single stranded DNA (ssDNA) and RNA. The ratio of values measured at 260 nm and 280 nm indicates the purity of the sample. The acceptable values for purity are 1.8 for DNA and 2 for RNA, variance indicates sample contamination.

2.7 Molecular cloning

2.7.1 Restriction enzyme digestion

Restriction enzymes were purchased from New England Biolabs Inc. (NEB) with the appropriate buffer recommended by the supplier. For analytical purposes, 500ng of DNA was digested in a volume of 20µl containing the appropriate units (U/µl) of the restriction enzyme per µg of DNA. The digested sample was incubated for 3h at the optimal temperature required by the enzyme used.

To prepare the vector and inserts for cloning, 5µg DNA was digested in 50µl reaction volume using the appropriate conditions needed by the restriction enzyme. The enzymatic activity was halted by heat inactivation. The main enzyme used in this study was *NdeI*, a Type II specific endonuclease isolated from *Neisseria denitrificans* that recognizes the 5'-CA^TATG-3' palindrome generating 2-base overhangs with low T_m.

2.7.2 Alkaline Calf Intestinal Phosphatase (CIP) reaction

Calf-Intestinal alkaline phosphatase (CIP) is an enzyme that non-specifically catalyses the dephosphorylation of 5´ and 3´ ends of DNA and RNA phosphomonoesters. In cloning, dephosphorylation prevents the self-ligation of linearised plasmid DNA. The enzyme acts on 5´ protruding, 5´ recessed and blunt ends. The vector dephosphorylation reaction is shown below (Table 3). The reaction was incubated for 1h at 37°C, and CIP heat inactivation was performed at 80°C for 15min.

Reagents	Volume/µl
Plasmid DNA	1µg
CIP Buffer 10x	5µl
CIP (1:100)	4µl
ddH_20 to a final volume of	50µl

Table 3 CIP reaction mix

2.7.3 T4 Ligase

T4 DNA Ligase, purchased from Roche, was used to join double stranded DNA. This enzyme catalyses the ATP-dependent formation of phosphodiester bonds between adjacent 3'hydroxyl and 5'phosphoryl termini of DNA. The reaction was performed with 20ng of digested vector and ligated with 5-10-fold molar excess of digested insert in a total volume of 30µl containing 1X ligase buffer and 1U of T4 DNA Ligase. The reaction mix was incubated overnight at 4°C and transformed into *E. Coli*DH5α competent cells after 8-16h.

2.8 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to size fractionated DNA samples and PCR products. This technique allows the separation of DNA molecules based on their molecular weight within a gel matrix. The % of agarose was established in relation to the length of the fragment, an agarose concentration of 1% (large and medium size fragments) or 2.5-3.5% (small fragments). The agarose powder was dissolved in 1X TBE Buffer which is also used for the run. To visualize the bands, Ethidium Bromide (EtBr) or Nancy-520 (Sigma-Aldrich) was added to the gel which absorbs UV light at 300-360nm and emits at 590nm or 520nm and emits at 560nm, respectively. The samples were resuspended with 1X DNA loading buffer before loading. Each electrophoresis run was performed in the presence of a DNA 1kb Plus Ladder marker (Invitrogen) to estimate the length of the DNA size fragment. Gels were electrophoresed at a constant voltage of 100V in 1X TBE running buffer.

2.9 Elution and purification of DNA fragments from agarose gel

This protocol was performed to purify PCR products or DNA fragments in preparation for sequencing. Following the electrophoresis run, DNA bands were visualized under UV light, and the required fragment was excised from the gel with a scalpel blade and transferred into a 1.5mL tube. DNA extraction was processed using QIAquick® Gel Extraction Kit (QIAGEN, Venlo, Netherlands) according to manufacturer's directions. Finally, samples were eluted in 20-50µL of

ddH_2O, measured by NanoDrop 1000 (Thermo Scientific) spectrophotometer to determine the concentration and a small portion was re-loaded ongel to check the quality of extraction.

2.10 DNA sequencing

Sequence analysis was performed through automatic sequencing service by Eurofins Genome.

2.11 Minigene construction

To determine the functional analysis of splicing pattern of the different cardiac genes, a minigene assay was used. Minigenes is an expression plasmid containing a global promoter and poly-A site in which an exogenous genomic sequence is inserted by multiple cloning site (MCS). A schematic representation of the minigene construction and its usage is reported in figure 16.

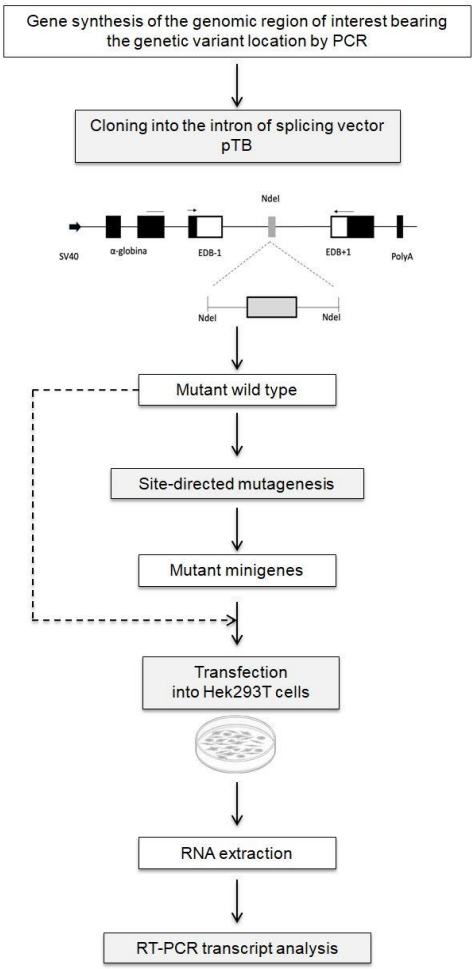


Figure 16. Workflow of the minigene protocol. The minigene assay includes (**1**) gene synthesis by PCR of the reference gene of interest; (**2**) cloning of the amplified fragment into the pTB vector or pcDNA3 vector and construction of the wild-type minigene; (**3**) site-directed mutagenesis to introduce the genetic variant of interest and create the minigene carrying the genetic variant; (**4**)

transfection in HEK293T cell line of the wild type and mutant minigenes, (**5**) RNA isolation; (**6**) splicing analysis by RT-PCR.

2.11.1 Preparation of insert DNA

To generate the reference (or wild-type - wt) minigene constructs, the relevant region from the gene of interest was generated by standard PCR amplification using genomic DNA as template. The oligonucleotides (Table 4) were designed with the specific restriction sites at their 5' ends that were used to clone the amplified fragment into the pTB and/or pcDNA3 expression plasmid.

Gene	Primer name	SEQUENCE 5'-3'
DMD	FW-int42 NdeI	ccggatatc**CATATG**gcgatccactctctcaggatgag
	RV-int43 NdeI	ccggatatc**CATATG**ttgtctgaattgttcatcattatttg
DSP	FW-int1 NdeI	ccg**CATATG**gaatgatatgctaagat
	RV-int2 NdeI	ccg**CATATG**gatggtgcacatgaagat
FLNC	FW-int21 NdeI	ccg**CATATG**ctaggcatgggccacggcaga
	RV-int24 NdeI	ccg**CATATG**agacgatgcgtgcaagaagg
	FW-int41 NdeI	ccg**CATATG**cgatgatgctgaagt
	RV-int43 NdeI	ccg**CATATG**ctctgcagcccataggcat
LMNA	FW-int9 NdeI	ccg**CATATG**gatgcttcctcaacagc
	RV-int10 NdeI	ccg**CATATG**ccagaagacaaatccag
MYBPC3	FW-ex7 BamHI	cgc**GGATCC**AGGCCATGGGCACCGGAGAC
	RV-ex10 XhoI	ccg**CTCGAG**CTCGGGGTCCGGAAACTGCGT
MYLK	FW-int20 NdeI	ccg**CATATG**gattctcctgcctcagcttc
	RV-int21 NdeI	ccg**CATATG**gagggtcatgtacacaag
NEXN	FW-int1 NdeI	ccc**CATATG**catgatcaactaccttcagct
	RV-int2 NdeI	ccc**CATATG**gagtataggtcaaactttgt
	FW-int8 NdeI	ccc**CATATG**cagcttctcgagtagctgggac
	RV-int9 NdeI	ccc**CATATG**gtgatgagtgagcattttgaag
PKD1	FW-ex31 BamHI	cgc**GGATCC**GTGGCTGGGAGCCCGAGC
	RV-ex34 EcoRI	ccg**GAATTC**AGGCTGCTGAGCAGGTCCGT
RYR2	FW-int18 NdeI	ccc**CATATG**gcatgcacctacagaggctgag
	RV-int19 NdeI	ccc**CATATG**tacaaccatatacctatac
SCN5A	FW-int20 EcoRV	ccg**GATATC**gtgttgtcagaaacg
	RV-int21 EcoRV	ccg**GATATC**cagtctgctgtggtgtgtg
TMEM43	FW-int9 NdeI	ccg**CATATG**ttgcgttcctggcctatgtg
	RV-int10 NdeI	ccg**CATATG**ttggttctcatccccaag
TTN c.669	FW-int4 NdeI	ccc**CATATG**catcgtagcaaccactggaattc
	RV-int6 NdeI	ccc**CATATG**gttgcatttcccaagatcac
TTN c.13282	FW-int45 EcoRV	ccg**GATATC**cacttcatcactttatttag
	RV-int46 EcoRV	ccg**GATATC**gacttagacactgac
TTN c.19501	FW-int76 NdeI	ccg**CATATG**gacaatttagtggccat
	RV-int78 NdeI	ccg**CATATG**gaatgagaggtacactaatg
TTN c.25808	FW-int103 NdeI	ccc**CATATG**gtagaacaagctattgtac
	RV-int106 NdeI	ccc**CATATG**cagtacaagaaaccatgcatg
TTN c.32875	FW-int140 PmeI	ccc**GTTTAAAC**gcttatgctcaatggttggctc
	RV-int144 PmeI	ccc**GTTTAAAC**gcacttcaggcacttaaaaac
TTN c.39574	FW-int143 NdeI	ccc**CATATG**gttccatatcctgtgttaaaagcag
	RV-int145 NdeI	ccc**CATATG**cataagctattttaaaatat
TTN c.41609	FW-int226 NdeI	ccc**CATATG**gtaacaaaccttagtagcaac
	RV-int228 NdeI	ccc**CATATG**cttcttggagactttagaag

TTN c.50346	FW-int233 NdeI RV-int237 NdeI	ccc**CATATG**gataggaagtagtctttctattc ccc**CATATG**cagtctggattgactttggtc
TTN c.57847	NdeI	Synthetised by GenScript
VCL	FW-int4 NdeI RV-int6 NdeI	ccc**CATATG**tatttatatcaaaatatgaaaatgt ccc**CATATG**acaaaaattagctgggcatggcagtg

Table 4. Sequence of primers used for the amplification of the reference sequence for the gene of interest. All primer sequences highlighted with bold letters indicate restriction sites for the primer specific enzymes.

PCRs were performed using a high-fidelity polymerase, TaKaRaTaq DNA Polymerase and Premix (TaKaRa) according to the manufacture's protocol, to avoid the introduction of nucleotide substitutions. The individual reaction used is detailed in table 5.

Components	Volume/µl
gDNA (100ng/µl)	xµl
10X PCR premix Taq DNA polymerase	12.5µl
FW primer (100ng/µl)	1µl
RV primer (100ng/µl)	1µl
ddH_20 to a final volume of	25µl

Table 5. Takara Taq DNA Polymerase and Premix PCR reaction for single sample.

Thermal cycling was conducted at 98°C for 30s as initial denaturation step, 35 cycles of the following were completed: 98°C for 10s, 55°C for 5s, and 72°C for 10s. A final elongation step at 72°C for 15s completed the amplification step.

The resulting DNA inserts were verified through separation by agarose gel electrophoresis and purified from agarose gel (see sections 2.8 and 2.9). The purified insert was digested with *NdeI* enzyme (New England Biolabs) (section 2.7.1) and ligated with T4 DNA Ligase (section 2.7.3) into the pTB or pcDNA3 vector.

2.11.2 Expression vector and features

To create the expression constructs for analysing splicing pattern*(in vitro* analysis), the digested and purified fragment was ligated into a digested and dephosphorylated expression vector (in this study pTB and pcDNA3) using the protocol described in section 2.7.1 and section 2.7.2. The vector pTB (Figure 17) is a 5998bp long plasmid that contains a portion of the alpha-globin gene (α-globin), consisting of exons 1, 2, 3 and related introns; the intron ('int fibronectin') in between exons 2 and 3 carries the unique *NdeI* site used in this study for subcloning purposes.

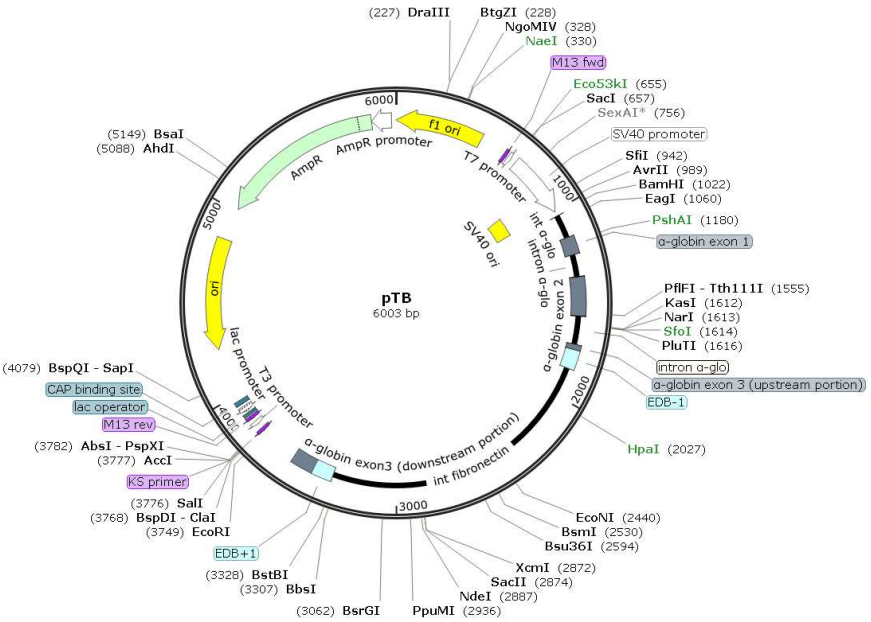


Figure 17. Schematic representation ofpTB vector.

The pcDNA3 vector was already present in Baralle's lab. This vector is commonly used in molecular biology, and it is designed for the expression of cloned genes in mammalian cells. The pcDNA3 vector (Figure 18) 5446bp long contains a strong eukaryotic promoter, typically the cytomegalovirus (CMV) which allows the expression of the inserted gene; the multiple cloning sites (MCS) where the gene of interest is inserted; the polyadenylation signal (poly-A) which ensures proper processing and stability of the mRNA produced from the inserted gene; a selectable marker that selects the cells that have successfully transfected; ampicillin-resistance cassette for the maintenance and selection of the plasmid in bacterial hosts.

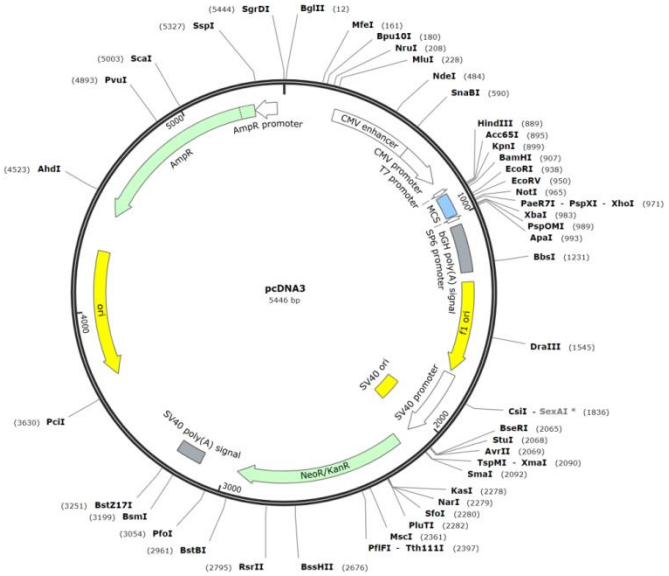


Figure 18. Schematic representation of pcDNA3 vector. The main components of the vector are reported as well as all restriction sites.

2.11.3 Colony PCR

After ligation (2.7.3), the vector-inserts were transformed into *E.coli*DH5α. Then, insert orientation was determined through colony PCR using specific primers, a pTB NdeI -100 forward (5'-GCTGATACTTACTTCAGATAT-3') that recognises the pTB vector or T7 primer specific for the pcDNA3 vector and the reverse primer used for the amplification of the insert. If the insert is in the correct orientation, an amplicon of a specific size will be amplified. Colony PCR uses as a template single and well-defined colonies which were individually picked and added directly to the PCR reaction. The PCR reaction was set up as shown in table 6.

Components	Volume/µl
10X reaction buffer	1.2µl
dNTPs 5mM	0.75µl
FW primer (100ng/µl)	0.36µl
RV primer (100ng/µl)	0.36µl
Taq DNA polymerase (Biolabs)	0.048µl
ddH_20 to a final volume of	12µl

Table 6. Colony PCR reaction

The thermocycling conditions for PCR screening were initial denaturation at 94°C for 2min, followed 35 cycles at 94°C for 30s, 55°C for 30s, 68°C 1min/Kb, and a final extension at 68°C 7min. PCR products were resolved on 1% agarose gels, run in 1X TBE buffer at 80-90V until the best resolution was achieved. A large yield of homogenous plasmid construct from positive colonies is produced following transformation, and final confirmation of the minigene was achieved by enzymatic restriction and sequencing.

2.11.4 Quick Change mutagenesis PCR method

DNA variants were introduced into the wt minigene by site-directed mutagenesis using Quick Change II Site-Directed mutagenesis kit (Agilent Technologies) in which the synthetic oligonucleotides used for the PCR reaction have the target modification within the sequence. A schematic representation of mutagenesis is represented in figure 19.

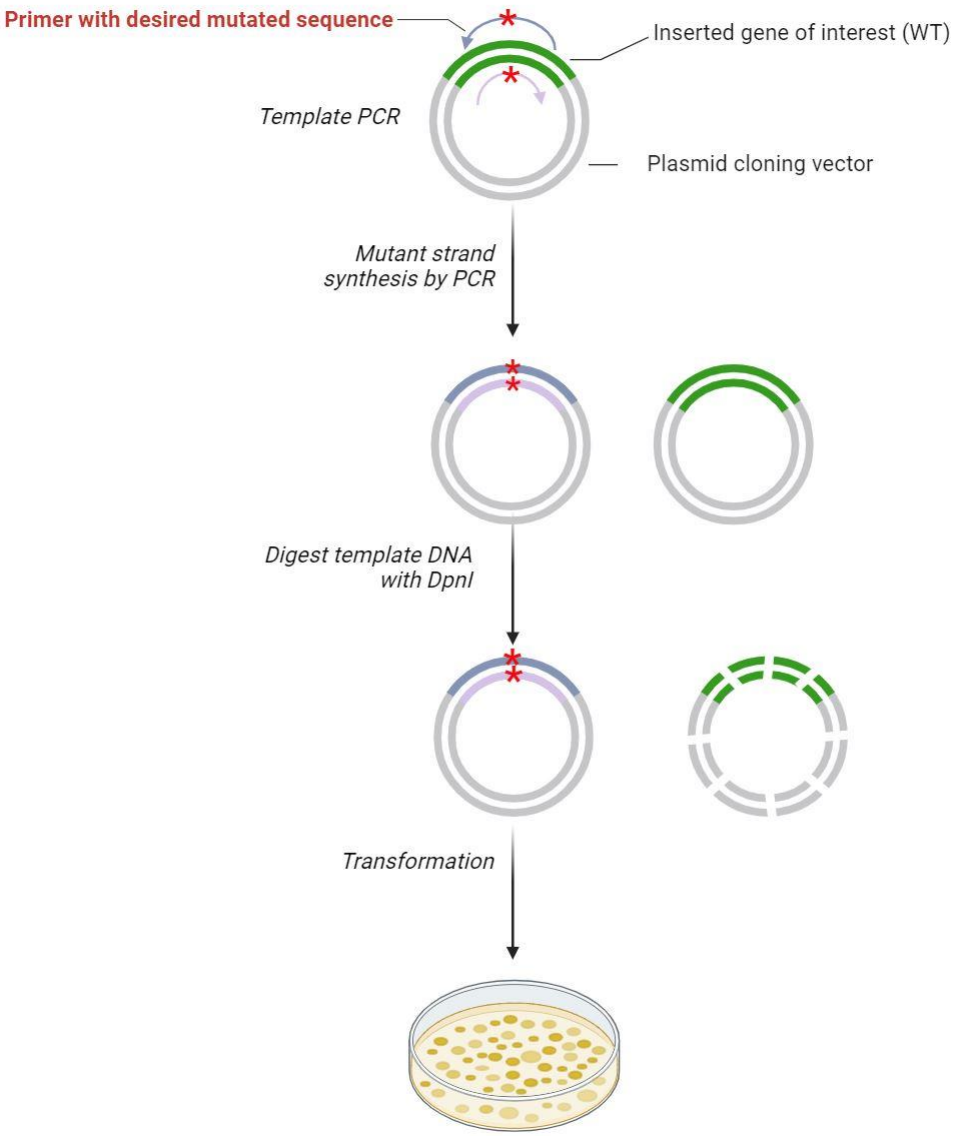


Figure 19. Workflow of Quick-change site directed mutagenesis. (**1**) Using the primers (blue and violet narrow) designed as recommended in the QuickChange™ protocol to insert the genetic variants of interest by PCR. (**2**) After PCR the parental DNA is digested with DpnI enzyme. (**3**) The newly synthesised DNA is transformed in bacterial culture. The green cycles represent the parental plasmid DNA, while violet and blue cycles represent the newly synthesised DNA. The cycles of violet and blue lines represent the DNA amplified using the parental DNA as templates while the cycles of gapped dash line are the DNA amplified using the newly synthesised DNA as templates. Arrows indicate the primers; where * indicates the location of the mutations.

The PCR reaction was performed in a final volume of 50µl by combing the following reagents listed in table 7. The oligonucleotides were designed to specifically insert the human genetic variants which are depictive in table 8.

Components	Volume/µl
Plasmid DNA 50ng	x µl
PfU Turbo buffer 10X	5µl
dNTPs 5mM	1µl
FW primer (100ng/µl)	1.25µl
RV primer (100ng/µl)	1.25µl
PfU Ultra HF DNA polymerase	0.25µl
ddH_20 to a final volume of	50µl

Table 7. Quick change site directed mutagenesis reaction.

Amplification was performed under the following conditions: initial denaturation 95°C for 30s, followed by 18 cycles 95°C for 30s, 55°C for 1min, 68°C for 8min, final extension for 68°C for 7min.

After PCR, 5µl of the reaction was loaded on 1% agarose gel to check the amplification and the PCR product was treated with 1µl of the DpnI enzyme for 3h at 37°C. This endonuclease recognises the specific 5'-Gm6ATC-3' sequence of methylated and hemi-methylated DNA used to digest the parental DNA template and to select for mutation-containing synthesised DNA. After digestion, 10µl of the nicked vector DNA containing the desired modifications was transformed into the *E. coli* DH5α strain. Plasmid DNA was extracted, and successful insertion of the specific variant was confirmed by sequencing.

Gene	Primer Name	SEQUENCE 5'-3'
DMD	FW-3C>A	CTGTTTTAAAATTTTTATATTA**A**AGAATATAAAAGATAGTCTAC
	RV-3C>A	GTAGACTATCTTTTATATTCT**T**TAATATAAAAATTTTAAAACAG
DSP	FW+3A>T	CATCGTGCAGCCTCGT**T**AGCTTTCCCTG
	RV+3A>T	CAGGGAAAGCT**A**ACAGGCTGCACGATG
	FW+5G>A	GTGCAGCCTGTAA**A**CTTTCCCTGTTC
	RV+5G>A	GAACAGGGAAAG**T**TTACAGGCTGCAC
FLNC	FW-1G>A	CTGGGCTCCA**A**GTGTCCTGCGGGAG
	RV-1G>A	CTCCCGCAGGACAC**T**TGGAGCCCAG
	FW+1G>A	GTCACCAGCCTCCAG**A**TTTGTGCCCAG
	RV+1G>A	CTGGGCACAAA**T**CTGGAGGCTGGTGAC
LMNA	FW-1G>A	CATGTCCCCACCA**A**GAAGTGGCCATGCGCA
	RV-1G>A	TGCGCATGGCCACTTC**T**TGGTGGGGACATG
MYBPC3	FW-1G>A	GGTCCCTCTCCATA**A**TGATAGCCATGAGGACAC
	RV-1G>A	GTGTCCTCATGGCTATCA**T**TATGGAGAGGGACC
MYLK	FW+22T>C	ATTGGCCACGGCACTATC**C**GGGGTCAAAGAAGTCATC
	RV+22T>C	GATGACTTCTTTGACCCC**G**GATAGTGCCGTGGCCAAT
NEXN	FW+9C>T	GCTGAGGTAAGTCT**T**AAAAGTAAAAAT
	RV+9C>T	ATTTTTACTTTT**A**AGACTTACCTCAGC
	FW+1G>A	GCAAGGAGAAATATG**A**TAAGACAGAAGCTAAC
	RV+1G>A	GTTAGCTTCTGTCTTA**T**CATATTTCTCCTTGC
PKD1	FW+5G>T	CTTCTCAGCATCAGGTGA**T**CTGGGGTGAGAG
	RV+5G>T	CTCTCACCCCAG**A**TCACCTGATGCTGAGAAG
RYR2	FW+3A>T	GAACCATGTCAGCAGGT**T**AATTCAGACAGAC
	RV+3A>T	GTCTGTCTGAATT**A**ACCTGCTGACATGGTTC

SCN5A	FW+5G>C	CATCGTAGACGTGA**C**TGTGGGCACCCGA
	RV+5G>C	TCGGGTGCCCACA**G**TCACGTCTACGATG
TMEM43	FW+8C>T	GGACTTCTCAGCAGAGGTGAGTG**T**TGTGCCCTACTCGTACGGTG
	RV+8C>T	CACCGTACGAGTAGGGCACA**A**CACTCACCTCTGCTGAGAAGTCC
TTN c.669	FW+4T>C	GACAAACCCGAATTGAAAAGGTA**C**TTTTCTATACTGTTGGTTTTC
	RV+4T>C	GAAAACCAACAGTATAGAAAA**G**TACCTTTTCAATTCGGGTTTGTC
TTN c.13282	FW+1G>A	CACTAACTGTGACAG**A**TAAATGCCATACATTG
	RV+1G>A	CAATGTATGGCATTTA**T**CTGTCACAGTTAGTG
TTN c.19501	FW+2T>C	GCTTGAACTGGCAGG**C**CAGTCCCTTTAT
	RV+2T>C	ATAAAGGGACTG**G**CCTGCCAGTTCAAGC
TTN c.25808	FW-1G>A	GACTGTTCTCTTTCAACA**A**AGCCTCCAGCTGTGGAAC
	RV-1G>A	GTTCCACAGCTGGAGGCT**T**TGTTGAAAGAGAACAGTC
TTN c.32875	FW+2T>C	CACCTGCCAAAGG**C**ACATGAACTTGCAATG
	RV+2T>C	CATTGCAAGTTCATGT**G**CCTTTGGCAGGTG
TTN c.39574	FW+7A>G	GACATCTTAAGTAAGT**G**TTACCAGGGGGAGTAC
	RV+7A>G	GTACTCCCCCTGGTAA**C**ACTTACTTAAGATGTC
TTN c.41609	FW-2A>C	GATCTTATTTTCATAT**C**GAAGTCATTAGAGATTGGCTG
	RV-2A>C	CAGCCAATCTCTAATGACTTC**G**ATATGAAAATAAGATC
TTN c.50346	FW+3A>G	GGCAAAGAAAGCAATGAAGGT**G**GGAAAATAACCATTCTC
	RV+3A>G	GAGAATGGTTATTTTCC**C**ACCTTCATTGCTTTCTTTGCC
TTN c.57847	+4delGTAA	Synthetised by GenScript
VCL	FW+4C>G	TGCCAGTTCTCATTTCAGGTA**G**TTCCTGCCTGTACTTTAT
	RV+4C>G	ATAAAGTACAGGCAGGAACTACCTGAAATGAGAACTGGCA

Table 8. Primer sequences used to perform Quick-change site directed mutagenesis. The mutated positions are underlined in red. The primer name and its 5'-3' sequence are indicated.

2.12 Cell culture

HEK293T cell line

The experiments were conducted using the HEK293T cell line (ATCC), transformed human embryonic kidney cells, which express a mutant version of the SV40 large T antigen. Initially, a vial (1mL) of HEK293T cells preserved in liquid nitrogen would be rapidly thawed. These were removed after gentle centrifugation and the cells seeded in 10mL of Dulbecco's modified Eagle's medium (DMEM: preheated glutamine, sodium pyruvate, pyridoxine and 4.5g/L glucose (Gibco®, Life Technologies) supplemented with 10% (v/v) Fetal Bovine Serum (FBS) (Euroclone) and a 1X mixture of antibiotic/antifungal (Sigma). Cells were kept in a humidified incubator at 37°C and 5% of CO_2.

For cell seeding, dishes containing a confluent monolayer of cells were washed with 1X PBS solution, treated with 2mL of 1X PBS/EDTA/trypsin solution (PBS containing 0.04% (w/v) EDTA and 0.1% (w/v) trypsin); at 37°C for 5min to detach the cells from the dish. Cells were collected and trypsin was inhibited by adding DMEM in corresponding trypsin volume used. Cells were pelleted by centrifugation for 5min at 1000rpm at room temperature and the cell pellet was resuspended in 10mL volume of DMEM seeding 1:10 in new P100 dishes with complete DMEM culture medium.

AC16 cell line

AC16 Human Cardiomyocyte cell line was purchased from Merck company. This cell line was generated from the fusion of adult human ventricular heart tissue with SV40 transformed, uridine auxotroph human fibroblasts, devoid of mitochondrial DNA (Litzkas, Jha and Ozer, 1984; Davidson *et al.*, 2005). Cells were thawed and maintained with DMEM/F12 (Sigma) supplemented with 2 mM L-Glutamine (Sigma), 12.5% FBS (Sigma) and 1X Penicillin-Streptomycin Solution (Sigma). For AC16 expansion and storage the manufacturer's instructions were followed. AC16 cells were transfected with 13 minigene constructs in order to compare the data obtained from HEK293T cells. Cell culture experiments were performed under sterile conditions and cell morphology was always checked before each experiment and the number of cell passages after thawing was limited to 10.

2.12.1 Transient transfection protocol

Transfection was performed for the expression of different minigenes to evaluate splicing patterns using Lipofectamine 2000 (Invitrogen). This reagent in solution forms small positively charged unilamellar liposomes (100-400nm) capable of forming complexes with the negatively charged exogenous DNA, thus favouring the endocytosis of the plasmid of interest inside the cell. The conditions under which the transfection was performed, and the amount of plasmid DNA used, are described briefly below. In addition, each minigene was transiently transfected in at least three independent transfections. One day before the transfection, HEK293Tor AC16 cells were sub-cultured at a seeding density of2.5 x 10^5cell/mL in 6 multiwell plate with the appropriate medium for each cell line and maintained at 37°C and 5% CO_2. 24h after plating, transient transfection was conducted with 0.7µg of construct DNA diluted with GibcoTM Opti-MEM (Concentration 1X, pH 7.0-7.4, with L-Glutamine, Phenol Red, Sodium Pyruvate) to a final volume of 125µl. In the meantime, a second mix including 117.5µl Opti-MEM and 7.5µl of Lipofectamine 2000 was combined with the DNA/Opti-MEM mix. The Lipofectamine/DNA mixture was incubated for 5min at room temperature to facilitate the formation of transfection micelles. Before adding the mixture, the media was replaced with antibiotic-free medium containing 10% FBS. RNA was harvested 24h after minigene transfection.

2.12.2 Total RNA extraction

To perform total cellular RNA extraction, transfected cells were rinsed from supernatant growth medium, resuspended, and lysed with 700µl of Trifast (Euroclone). Resuspended cells were transferred into 1.5mL microcentrifuge tube and 150µL chloroform was added. Samples were mixed, incubated for 10min at RT, and centrifuged at 13500rpm at 4°C for 10min for phase separation. The upper aqueous phase, containing the isolate RNA was transferred to a new

1.5mL centrifuge tube and 1:2 of cold isopropyl alcohol [100%] was added. RNA precipitation occurs during a 30min incubation at -20°C, samples were then centrifuged at 13500rpm at 4°C for 30min. Following this, the RNA pellet was washed with 200µL 70% ethanol and centrifuged for 10min, the solution was discarded, and the pellet was left-dried for 5minand resuspended in 30-50µL of autoclaved ddH_2O. The quality of RNA was checked by electrophoresis on a 0.8% agarose gel and quantified using NanoDrop 1000 (Thermo Fisher Scientific) spectrophotometer. Following extraction, RNA was stored at -80°C and thawed on ice prior to further application.

2.12.3 mRNA analysis by RT-PCR

2.12.3.1 Reverse transcriptase of RNA

Alternative splicing patterns can be analysed using RT-PCR analysis, in which cDNA synthesis is performed separately from PCR amplification. First-strand cDNA synthesis for quantitative PCR was transcribed using the Moloney murine leukemia (M-MLV) Reverse Transcriptase Kit (Invitrogen). The following components were mixed: 1µg of RNA, 200ng of random primer (Invitrogen) and sterile water to reach a final volume of 15µL. The mixture was then denatured at 75°C for 5min to eliminate any secondary template structures and cooled down on ice to prevent reforming of these structures. This was then mixed with a transcription mixture composed of 1x First Strand Buffer (Invitrogen), 10mM DTT, 5mM dNTPs each, and 100U of M-MLV reverse transcriptase enzyme called. The reaction mix was incubated at 37°C for 1hand PCR was performed.

2.12.3.2 cDNA analysis

PCR amplification was carried out using 3µl cDNA synthesis reaction as template, added directly to the PCR mixture following the Promega GoTaqG2 DNA polymerase basic protocol. The PCR reagents are detailed in table 9. The pair of primers used to analyse the splicing pattern in the minigene context is shown in table 10.

Components	Volume/µl
cDNA	3µl
Buffer 5X	5µl
$MgCl_2$ 25mM	3.75µl
dNTPs 5mM	1µl
FW primer (100ng/µl)	1µl
RV primer (100ng/µl)	1µl
GoTaq G2 DNA polymerase	0.12µl
ddH_20 to a final volume of	25µl

Table 9. GoTaq G2 Flexi reaction by Promega

Minigene	Primer name	SEQUENCE 5'-3'
pTB vector	α 2,3 FW	CAACTTCAAGCTCCTAAGCCACTGC
	BRA RV	GGTCACCAGGAAGTTGGTTAAATCA
pcDNA3 vector	T7 FW	TAATACGACTCACTATAGGG
	SP6 RV	ATTTAGGTGACACTATAGAATA

Table 10. Sequence of the primers used for PCR analysis of splicing.

The thermal cycler conditions were set as follows: 94°C for 2min followed by 35 cycles of 94°C for 30s, 58°C for 30s, 68°C for 1min/Kb and a final extension of 5min. PCR products were separated on a specific percentage of agarose gel, and the bands were excised and sequenced.

2.13 Splicing modulator treatment

2.13.1 Compounds

Kinetin solution was obtained from Sigma as well as 6-Benzyladenine, and Zeatin which were dissolved in DMSO at 20mM and 10mM, respectively. Branaplam was purchased from MedChemExpress and dissolved in DMSO at 20mM. Valproic acid (VPA), and Sodium butyrate (NaBu) were obtained from Selleckchem. VPA was dissolved in DMSO at 200mM, whereas NaBu was dissolved in ddH_2O at 200mM.

2.13.2 Cell culture and treatment

The HEK293T cell line was cultured in 6-well format 1.5 x 10^5 cell/mL in the presence of complete DMEM and transfected with the minigene construct after 24h. Compounds were added to the cell culture media 4h later. 6-Benzyladenine, Zeatin were tested at concentrations of 50μM, 100μM and 200μM. Kinetin solution was obtained from Sigma and cells were treated with 200μM and 300μM. Then, 500μM VPA and NaBu were added to the cultures. Branaplam was tested at 5μM. Drug-treated cells were incubated for 24h and 48h whereas 72h was tested only for kinetin with a drug renewal every 24h. After the incubation time was finished, cells were collected and pellet at 1000rpm for 5min at room temperature. Finally, RNA was extracted and processed for RT-PCR.

2.13.3 Cell viability assay

To perform cell viability, a resazurin assay was performed. This assay measures the reducing capacity of mitochondria in cells. HEK293T cell lines were seeded in 96-well plates at 1 x

10^5cells/mL in 100µL and incubated at 37°C, 5% CO_2 for 24h until they reached confluence. Cells were then treated with the compounds at different concentrations and incubated for 24h. Each condition was evaluated in three biological replicates. Then, the media was replaced with 400µL/well of 10% Alamar blue dye (Invitrogen), and cells incubated with dye at 37 °C, 5% CO_2 for 30min. The 96-well plate was exposed to an excitation wavelength of 560 nm and a tracking emission wavelength of 590nm, via multi-mode plate reader (Biotek). Therefore, the metabolic activity was measured at different incubation times to generate an adequate fluorescent signal above the background.

2.14 Biological material

2.14.1 Ethics and participant consent

Patients with genetic variants were identified through routine diagnostic genetic testing performed by Cattinara Hospital (Trieste, Italy), the Wessex Regional Genetics Laboratory, Salisbury, or through the Splicing and Disease research study at the University of Southampton. Ethical approval for this study was provided by the Health Research Authority (IRAS Project ID 49685, REC 11/SC/0269) and by the University of Southampton (ERGO ID 23056). Each patient from who samples were obtained submitted written informed consent to take part in splicing studies. All work was carried out in accordance with the Declaration of Helsinki.

2.14.2 Sample collection

Peripheral whole-blood samples (4-6ml) were collected in EDTA collection tubes to perform the RNA extraction on the day of blood collection. For longer term storage, blood samples (2.5ml) were collected in PAXgene blood tubes (QIAGEN), incubated at room temperature for 4h for RNA stabilization and then stored at -80°C.

Total RNA was isolated from EDTA collection tubes combining the RBC Lysis Buffer and RNeasy Mini Kit (QIAGEN) methods. The human blood (roughly 6mL) was transferred into a 50mL canonical RNase free tube and the volume was brought to 45mL with RBC Lysis Buffer 1X (Cold Spring Harb Protoc, 2016). The sample was then incubated at room temperature for 30min and spun down, and the supernatant was discarded. The next step was repeated with 1mL RBC Lysis buffer 1X until the resulting pellet was cleared. RNA was extracted using RNeasy Mini Kit (Qiagen), following the manufacturer's protocol. RNA quality control was performed on agarose gel electrophoresis and with NanoDrop (Thermo Fisher).

2.14.3 RNA purification from blood

RNA samples extracted from total blood were treated with TurboDNaseI free kit (Invitrogen) to remove DNA contaminants. This enzyme selectively digests DNA, leaving the RNA molecules intact. This is an important step to ensure future analyses of reverse transcribed transcripts are not compromised by amplification of genomic DNA. Each 20µl RNA sample was mixed with 1µl of DNaseI enzyme, 2µl of DNase buffer (10x), and incubated at 37°C for 30min. Then, the TurboDnase activity was inhibited by adding 2µl of DNase Inactivation reagent. The samples were left for 5min at room temperature, mixing occasionally. The RNA was collected into a fresh 1.5mL tube, after centrifugation for 2min at 13000rpm. Avoid introducing residual of the DNase Inactivation Reagent into solutions that may be comprised downstream reactions, the sample was then purified using the RNA clean & concentrator-5 kit (Zymo Fisher) according to the manufacturer's instructions.

2.14.4 RT-PCR for identifying splicing aberrations

RNA extracted from blood samples was converted to cDNA using MLV-RT enzyme (section 2.12.3.1) and random hexamers. Following cDNA synthesis, standard PCR was used to evaluate the alternative splicing of cardiac genes. Primer pairs (table 11) were designed manually to amplify the region surrounding each variant. For each variant, whenever possible, primer sequences were positioned at least two exons up- and downstream of the target variant. PCR experiments were conducted using GoTaq G2 Polymerase PCR system (Promega, UK) according to the manufacturer's instructions, and cycling parameters were determined for each gene. RT-PCR products were evaluated by agarose gel electrophoresis against the control sample and amplicons were gel-purified by QIAquick Gel extraction kit (QIAGEN, section 2.9) and sent for Sanger sequencing.

Gene	Primer name	SEQUENCE 5'-3'
DMD	FW-ex41	CTGAGGGCTTGTCTGAGGATG
	RV-ex45	TCTGACAACAGTTTGCCGCT
DSP	FW-ex1	GGCACCAGCAGGATGTACTA
	RV-ex4	GCTCGCATTTGCTCTTGGAG
FLNC c.3791	FW-ex21	GCCATTCTCCATCGACACCA
	RV-ex24	GAAGGGGATGTACTCCACGGT
FLNC c.7251	FW-ex41	CCCCTTTCAGTTCACTGTGG
	RV-ex45	GACCGTAGGCTGACACCAAG
LMNA	FW-ex8	GCAGATCAAGCGCCAGAATG
	RV-ex11	GACACTGGAGGCAGAAGAGC
NEXN	FW-ex7	CGACCATCTCTCAAGGAAGCA
	RV-ex11	GCTCTCCTTCTTGCTCGCTC
TTN c.669	FW-ex4	CTGAGACAGCACCACCCAACTTC
	RV-ex6	CTGACCGGAGATGGGGTCGGTGC

TTN c.19501	FW-ex76	CTGGTGCAAGGTGAATCGTC
	RV-ex80	CTCACAACTCTCTGCACGGT
TTN c.32875	FW-ex138	GGTGCCAGTGATACCAGTCAA
	RV-ex144	CCTGAGGTAGAGCTACAGGA
TTN c.41609	FW-ex226	CAACTTCCTGTGCGTTTTG
	RV-ex228	GGTGAGGGCCTTAGAAGTTC
TTN c.57847	FW-ex294	CTGGAAAGCCTCCTCCAACC
	RV-ex296	CAGCACTGACACGGTACTCA
VCL	FW-ex3	CTCAGTGCCTGCTCGAGATT
	RV-ex7	GTCTATGGAGGCCAATGCTCT

Table 11. Sequences of the primers used to perform q-PCR. The endogenous target genes, the primer name and its 5'-3' sequence are indicated. Abbreviations: ex, exon; FW, forward; RV, reverse.

2.14.5 Real time PCR for gene expression level

cDNA was used as template for Real time PCR (qPCR) to assess the expression levels of transcripts of interest using housekeeping gene *GAPDH* and *HuPo* as normalise. All amplicons were done using the iQ SYBR Green Supermix on CFX 96-Real-time PCR detection system (BioRad). Primer sequences were selected for each target gene whenever possible were designed downstream of the identified variants. Primer's specificity for each gene was tested with standard basic local alignment search tool (BLASTs) covered on the National Center of Biotechnology information homepage (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Table 12 depicts the primer sequence used to perform qPCR. Primers were tested for their ability to produce a single PCR amplification product of correct size observed upon agarose gel analysis.

Gene	Primer name	SEQUENCE 5'-3'
DSP	FW-ex18	TGACAGACCGCTGGCAAAGGAT
	RV-ex19	GGCGTTTAGCATCATAGAGCCAC
LMNA	FW-ex4	ATGAGGACCAGGTGGAGCAGTA
	RV-ex5	ACCAGGTTGCTGTTCCTCTCAG
NEXN	FW-ex11	AAGAGCGAGCAAGAAGGAGAGC
	RV-ex12	GTGAGTAAATGGAGCCTCGCTTT

Table 12. Sequences of the primers used to perform qPCR. The endogenous target genes, the primer name and its 5'-3' sequence are indicated. Abbreviations: ex, exon; FW, forward; RV, reverse.

Thermal cycling conditions were 98°C for 30s, 95°C for 10s, followed by 39 cycles of 95°C for 30s, Ta (relevant annealing temperature; range 55-57°C) for 30s, and 72°C for 1min. A melting curve analysis was added after the final PCR cycle to evaluate the presence of non-specific PCR products and primer dimmers. The threshold cycle (Ct) represents the PCR cycle at which an increase in SYBR Green fluorescence above a base line signal can be first detected. For each

sample, the amount of mRNA was quantified relative to 500ng of total RNA applied by real-time PCR.

Relative gene expression was performed using the following equation for using multiple reference genes (Vandesompele *et al.*, 2002; Hellemans *et al.*, 2008) (https://toptipbio.com/qpcr-multiple-reference-genes/):

$$Relative\ gene\ expression = \frac{(\mathrm{E_{GOI}})^{\ \Delta \mathrm{Ct\ GOI}}}{\mathrm{Geomean}[(\mathrm{E_{Ref}})^{\ \Delta \mathrm{Ct\ Ref}}]}$$

2.15 Statistical analysis

2.15.1 Splicing analysis of minigene transcripts

RT-PCR was used to analyse the splicing pattern of the minigenes. Densitometry was used to calculate the ratio of WT/MU transcripts with the integrated density value (IDV) for each band quantified using ImageJ software. The level of exon inclusion was calculated as the relative density value of the WT band and expressed as a percentage, which was calculated as reported as follows:

$$\%\ exon\ inclusion = \frac{\text{band of interest (WT transcript)}}{\text{Total bands}}\ \text{x}\ 100$$

All data were analysed and plotted using GraphPad Prism Software (Motulsky, 2021). Statistical comparison was performed by unpaired Student t-test. All data are presented as mean ± standard deviation (SD). Star code for significance is reported as: p> 0.05 was scored as not significant. *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, and ****p ≤ 0.0001.

2.16 In silico splicing prediction: SpliceAI tool

The SpliceAI tool was used to predict possible mis-splicing events caused by genetic variants identified in CVD patients. The default rules were set up to interpret the SpliceAI Δ-score which output consists of acceptor loss, acceptor gain, donor loss, and donor gain; for each the Δscore ranging from 0 to 1 within a window around +/- 500 nucleotides of the interrogated variant (Jaganathan *et al.*, 2019).

2.17 Transcriptome sequencing and computational analysis

2.17.1 RNA sequencing

RNA isolated from 17 samples of whole blood (4 healthy controls and 13 CVD patients) was sent to Novogene (UK) for sequencing after confirming the isolated RNA was high-quality (RIN score >6) and with a concentration greater than 20ng/µl (sample requirement by Novogene (UK)). RNA-Seq was conducted according to the Novogene Sequencing protocol using the NEBNext Globin and rRNA Depletion Kit and NEB Next Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs, MA). For each sample was generated at least 70 million 150bp paired-end reads.

Each step was performed by Novogene from RNA sample to the final data including sample testing, library preparation, and sequencing. Quality control (QC) was carried out at each step of the workflow to ensure the accuracy of the data (Figure 20).

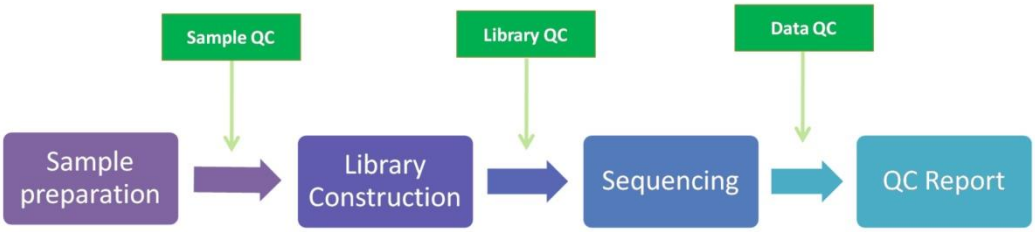


Figure 20. RNA-seq sample processed by Novogene. QC, quality control

In more detail, the sequencing workflow was as follows: quality control of the RNA sample (preliminary RNA quantity, RNA degradation and contamination, RNA integrity and final concentration), ribosomal RNA (rRNA) removal from total RNA using the ethanol precipitation method, mRNA fragmentation, first strand cDNA synthesis using random hexamer primers, second strand cDNA synthesis using dTTPs in the reaction buffer to replace dUTP, terminal end repair, A-tailing addition, sequencing adapter ligation, fragment size selection, second cDNA strand degradation followed by USER enzyme digestion, PCR amplification, purification, QC report (preliminary library concentration insert size, precise library concentration) on the first strand cDNA library and finally sequencing was performed.

2.17.2 Index genome and alignment of RNA-seqreads

The RNA-seq raw reads data analysis was performed using the University of Southampton's IRIDIS5 high-performance computing cluster and the pipeline was written in Bash containing several tools used to perform the analysis. First, RNA-seq read mapping was performed using STAR version 2.7.10a and all sequenced reads were aligned to the reference human genome.

The reference genome was generated with STAR's genome generating using GRCh38.primary_assembly.genome.fa and gencode.v30.annotation.gtf, which were downloaded from GENCODE (https://www.gencodegenes.org/human/) with –sjdbOverhang 149 and all other settings as default. RNA-seq reads were aligned in two pass Basic mode using the following criteria as a default: --out SAM mapqUNisque 60, --outFilterTypeBySJout, --out Reads Unmapped Fastx, --out SAM type BAM unsorted. Sam tools 1.9 version was used to sort the BAM file which were generated by STAR. The sorted BAM file contains the information related to the annotated duplicate reads.

2.17.3 Quality analysis of output files

Quality control (QC) was performed on aligned data using RSeQC (v2.6.4) (https://rseqc.sourceforge.net/): bam_stat.py, infer_experiment.py, geneBody_coverage.py, junctional_annotation.py and junction_saturation.py. Then, results were compiled and visualized with MultiQC version 1.14 (Ewels *et al.*, 2016) which was used to generate a quality control (QC) report on the output files of the FastQC, STAR, and HTSeq analysis of each individual sample.

2.17.4 Alternative splicing detection

The splicing effect of the interested genetic variants was visualized using the Integrative Genomics Viewer (Braos Institute, MA) and a Sashimi plot was generated (Robinson *et al.*, 2011, 2023).

Chapter 3 Results

3.1 Genetic variants identified in cardiac disorders

DNA sequence variants were identified on a cohort of patients affected with potential inherited cardiac disorders who underwent exome sequencing by the Cardiology Unit of Cattinara Hospital (Trieste, Italy) and the Wessex Regional Genetics Laboratory (Salisbury, UK) for which no known cause had yet been identified (Table 13). However, evaluating genetic variants associated with phenotypically concordant genes remains a challenge. This is due to the absence of supportive functional studies on messenger RNA (mRNA) or proteins that are not typically performed in clinical diagnostic laboratories. Following genetic assessment, several heterozygous variants associated with cardiac genes were observed in a total of 18 genes, including *DMD, DSP, FLNC, LMNA, MYBPC3, MYH6, MYH7, MYLK, NEBL, NEXN, OBSL1, PKD1, RYR2, SCN5A, TMEM43, TNNT2, TTN,* and *VCL*. The sequence variants were clinically classified as Pathogenic (n=5), likely pathogenic (n=6) or as variants of unknown significance (n=25) according to the ACMG/AMP guidelines described in section 1.2. The majority of genetic variants were present in *TTN* (25%), followed by *FLNC* (23%), and *LMNA*, *NEXN* (6.8% for each gene). *FLNC* c.3791-1G>A was the most recurrent splice-site variant in the cohort patients, detected in 6 different patients. In addition, the c.5842+2 T>A variant in *FLNC,* the c.1053+1 G>A variant in *NEXN*, and c.32875+2 T>C in *TTN*, were each identified in 2 unrelated patients.

Gene	Method	Disease	Identified Variant	Status	ACMG/AMP
DMD	ILLUMINA	CMPD	c.6118-3C>A	Hemizygous	VUS
DMD	ILLUMINA	ARVD	c.10223+3G>A	Hemizygous	VUS
DSP	ILLUMINA	CMPD	c.273+3A>T	Heterozygous	VUS
DSP	ILLUMINA	CMPD	c.273+5G>A	Heterozygous	VUS
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP
FLNC	ILLUMINA	CMPD	c.3791-1G>A	Heterozygous	LP

FLNC	ION Torrent	CMPD	c.5842+2T>A	Heterozygous	LP
FLNC	Sanger sequencing	CMPD	c.5842+2T>A	Heterozygous	LP
FLNC	ION Torrent	CMPD	c.7251+1G>A	Heterozygous	P
FLNC	ION Torrent	CMPD	c.7252-2A>G	Heterozygous	P
LMNA	ILLUMINA	ARVD	c.-1C>A	Heterozygous	VUS
LMNA	ILLUMINA	CMPD	c.1608+1G>T	Heterozygous	P
LMNA	ILLUMINA	CMPD	c.1609-1G>A	Heterozygous	LP
MYBPC3	ION Torrent	CMPI	c.852-1G>A	Heterozygous	LP
MYH6	ION Torrent	CMPD	c.5163+1G>A	Heterozygous	VUS
MYH7	ION Torrent	CMPI	c.530+1G>A e	Heterozygous	P
MYLK	ION Torrent	N.A.	c.3703+22T>C	Heterozygous	VUS
NEBL	ION Torrent	CMPI	c.2519-2A>-	Heterozygous	VUS
NEXN	ION Torrent	CMPD	c.27+9 C>T	Heterozygous	VUS
NEXN	ILLUMINA	CMPD	c.1053+1G>A	Heterozygous	LP
NEXN	ILLUMINA	CMPD	c.1053+1G>A	Heterozygous	LP
OBSL1	ION Torrent	CMPD	c.4609+1G>T	Heterozygous	VUS
PKD1	N.A.	N.A.	c.10405+5G>T	Heterozygous	VUS
RYR2	ILLUMINA	CMPI	c.1961+3A>T	Heterozygous	VUS
SCN5A	ILLUMINA	CMPD	c.393-5C>A	Heterozygous	LP
SCN5A	N.A.	N.A.	c.3840+5G>C	Heterozygous	VUS
TMEM43	ILLUMINA	CMPD	c.882+8C>T	Heterozygous	VUS
TNNT2	ION Torrent	CMPI	c.821+1G>A	Heterozygous	P
TTN	ILLUMINA	CMPD	c.669+4T>C	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.13282+1G>A	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.19501+2T>C	Heterozygous	VUS

TTN	ION Torrent	CMPD	c.25808-1G>A	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.32875+2T>C	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.32875+2T>C	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.39574+7A>G	Heterozygous	VUS
TTN	ION Torrent	CMPD	c.40153+1G>A	Heterozygous	VUS
TTN	ILLUMINA	CMPD	c.41609-2A>C	Heterozygous	VUS
TTN	ILLUMINA	CMPD	c.50346+3A>G	Heterozygous	VUS
TTN	ILLUMINA	CMPD	c.57847+4delGTAA	Heterozygous	VUS
VCL	ILLUMINA	CMPD	c.622+4C>G	Heterozygous	VUS

Table 13. Genomic variants identified in patients. Abbreviations: ARVD, *Arrhythmogenic Right Ventricular Cardiomyopathy*; CMPD*, Dilated Cardiomyopathy*; CMPI, *Hypertrophic Cardiomyopathy*; N.A., not available; LP, Likely Pathogenic; P, Pathogenic; VUS, Variant of Unknown Significance.

The variants reported in Table 13 are predominantly located 1-10bp upstream of splice acceptor sites or 1-5bp downstream of splice donor sites (Figure 21) and consequently may influence mRNA processing.

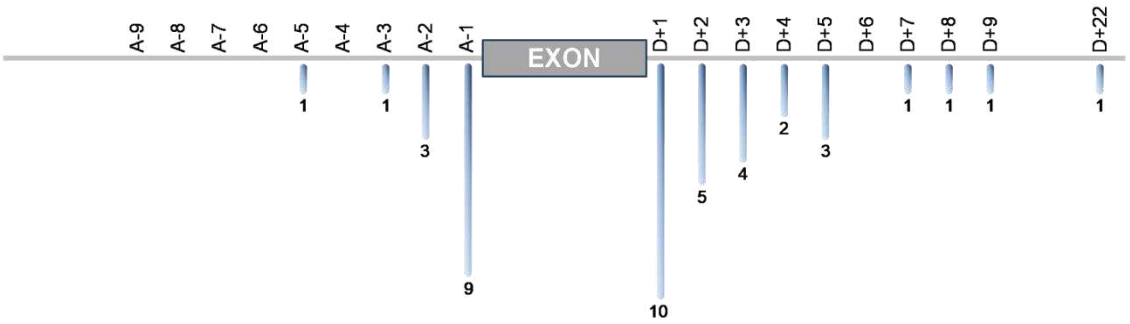


Figure 21.Schematic representation of site variants identified in cardiac disorders. The number indicates the number of variants located in the consensus splice site sequences.

3.2 *In silico* analysis and splicing prediction

Computational predictors have been developed to provide valuable insights about variant implication, even in the absence of experimental tests of variant function. In this study, SpliceAI was used to investigate the role of the variants listed in Table 13 on RNA splicing (Table 14).

GENE	Variants	ΔPosition/Δ score			
		Acceptor loss	Acceptor gain	Donor loss	Donor gain
DMD	NM_004006.3:c.6118-3C>A	0.42 (-3bp)	0.01 (+42bp)	0.00	0.00
DMD	NM_004006.3:c.10223+3G>A	0.01 (129bp)	0.01 (139bp)	0.11 (104bp)	0.04 (3bp)
DSP	NM_004415.4:c.273+3A>T	0.04	0.00	0.42 (-3bp)	0.81 (58bp)
DSP	NM_004415.4:c.273+5G>A	0.00	0.00	0.22 (-5bp)	0.00
FLNC	NM_001458.5: c.3791-1G>A	1.00 (1bp)	0.78 (2bp)	0.00	0.00
FLNC	NM_001458.5: c.5842+2T>A	0.00	0.00	1.00 (-2bp)	0.46 (9bp)
FLNC	NM_001458.5:c.7251+1G>A	0.00	0.00	0.99 (-1bp)	0.56 (12bp)
FLNC	NM_001458.5:c.7252-2A>G	1.00 (2bp)	0.67 (47bp)	0.00	0.00
LMNA	NM_170707.4:c.-1C>A	N.A.	N.A.	N.A.	N.A
LMNA	NM_170707.4:c.1608+1G>T	0.00	0.00	1.00 (-1bp)	0.66 (-30bp)
LMNA	NM_170707.4: c.1609-1G>A	1.00 (1bp)	0.99 (2bp)	0.00	0.00
MYBPC3	NM_000256.3:c.852-1G>A	0.81(-1bp)	0.55 (-7bp)	0.01 (-54bp)	0.01 (171bp)

MYH6	NM_002471.4:c.5163+1G>A	0.00	0.00	1.00 (+1bp)	0.51 (-7bp)
MYH7	NM_000257.4:c.530+1G>A	0.61 (+28bp)	0.01 (-32bp)	0.96 (+1bp)	0.00
MYLK	NM_053025.4:c.3703+22T>C	0.00	0.00	0.00	0.13
NEBL	NM_006393.3:c.2519-2A>-	0.05 (-1bp)	0.00	0.00	0.00
NEXN	NM_144573.4:c.27+9C>T	0.00	0.00	0.00	0.00
NEXN	NM_144573.4:c.1053+1G>A	0.00	0.00	0.98 (-1bp)	0.00
OBSL1	NM_015311.3:c.4609+1G>T	0.00	0.00	0.99 (1bp)	0.35 (8bp)
PKD1	NM_001009944.3:c.10405+5G>T	0.32 (114bp)	0.02 (10bp)	0.37 (5bp)	0.01 (-5bp)
RYR2	NM_001035.3:c.1961+3A>T	0.00	0.02 (-48bp)	0.96 (-3bp)	0.77 (28bp)
SCN5A	NM_001099404.2:c.393-5C>A	0.06 (-5bp)	0.25 (-2bp)	0.00	0.00
SCN5A	NM_001099404.2:c.3840+5G>C	0.12 (-175bp)	0.00	0.99 (5bp)	0.54 (-42bp)
TMEM43	NM_024334.3:c.882+8C>T	0.00	0.00	0.00	0.00
TNNT2	NM_001001430.2:c.821+1G>A	0.23 (41bp)	0.00	0.95 (+1bp)	0.35 (-12bp)
TTN	NM_001267550.2:c.669+4T>C (Meta)	0.00	0.00	0.00	0.00
TTN	NM_003319.4:c.13282+1G>A (N2B)	0.00	0.00	1.00 (1bp)	0.10 (42bp)

TTN	NM_003319.4:c.19501+2T>C (N2B)	0.00	0.00	0.48 (2bp)	0.16 (-18bp)
TTN	NM_003319.4:c.25808-1G>A (N2B)	1.00 (-1bp)	0.99 (-3bp)	0.00	0.00
TTN	NM_001256850.1:c.32875+2T>C (N2AB)	0.00	0.00	0.88 (2bp)	0.39 (-48bp)
TTN	NM_003319.4:c.39574+7 A>G (N2B)	0.00	0.00	0.00	0.00
TTN	NM_003319.4:c.40153+1G>A (N2B)	0.00	0.00	0.97 (+1bp)	0.02 (+31bp)
TTN	NM_001267550.2:c.41609-2A>C (Meta)	0.99 (-2bp)	0.81 (-5bp)	0.00	0.00
TTN	NM_001256850.1:c.50346+3A>G (N2BA)	0.00	0.01 (-45bp)	0.03 (+3bp)	0.00
TTN	NM_001267550.2:c.57847+4delGTAA	0.00	0.00	0.98 (+5bp)	0.06 (+2bp)
VCL	NM_014000.3:c.622+4C>G	0.00	0.00	0.00	0.01 (-bp)

Table 14. Splicing predictions of the splice site variants by SpliceAI. SpliceAI predicts both acceptor and donor gain and acceptor and donor loss. SpliceAI Δscore was evaluated as 0.2 (high recall), 0.5 (recommended), and 0.8 (high precision). The position representing the possible position of splice sites within the window (+/- 500bp by default) is also reported. Negative values correspond to the left of the variant in genomic coordination, regardless of the transcript orientation.

According to the ACMG/AMP classification guidelines (Table 13), and the analysis conducted with SpliceAI (Table 14), of the 6 likely pathogenic intronic variants, 5 variants were predicted to disrupt canonical splice sites. While for the 25 variants of unknown significance, 15 variants were predicted to induce aberrant splicing. On the other hand, all the variants (n=5) were

classified as pathogenic, the *in silico* prediction confirmed their potential role to disrupt mRNA processing.

3.3 Functional analysis of splice variants by minigene assay

To determine the pathogenicity of the variant, a total of 24 variants (6 variants located in the region of the acceptor site and 18 located at the donor site) were selected and a cell-based minigene assay strategy was used when blood samples were not available (Figure 22). In 13 cases, the effect of variants on RNA processing was analysed and compared in both peripheral blood and a minigene assay.

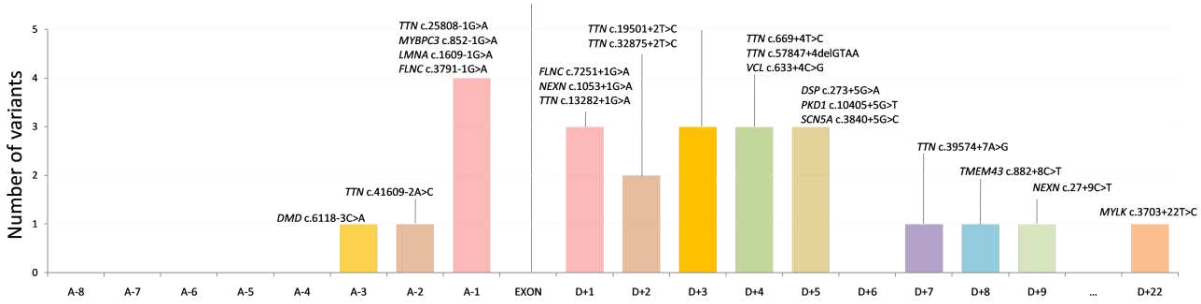


Figure 22. Schematic location of the analysed variants by minigene assay relative to the intron-exon boundary. The gene name and the variant are reported. A (acceptor splice site), D (donor splice site).

To generate the wt minigene, depending on the primer sets used (see section 2.11.1), the amplified genomic region of interest includes at least the nearest exon for each variant and approximately 300bp of the intronic regions flanking, which should be enough and contain the most important regulatory elements (Riedmayr *et al.*, 2018). In some cases, to preserve exon-intron boundaries sufficient for correct mRNA splicing, more exons were included. Then, the genomic fragment from the gene of interest was cloned into pTB or pcDNA3 vectors, and the correct orientation of the inserts was checked by colony PCR (see section 2.11.3) and Sanger sequencing. The mutagenesis of the wild-type minigene constructs was required to introduce the variants of interest. The set of primers used are listed in Table 8 and the insertion of the variants was confirmed by Sanger sequencing. To characterise the effect on the splicing of the variant under study, both wt and mutant minigene constructs were transiently transfected in HEK293T cells, which have a high level of transfection efficiency. The splicing profile is analysed by PCR of the retrotranscribed RNA; this method is direct and detects the possible alternative transcript isoforms. This method permits *in vitro* validation of the accuracy of the bioinformatic predictions and validates the possible mis-splicing events.

3.3.1 Acceptor splice site

***DMD*: c.6118-3 C>A**

The variant in *DMD,* c.6118-3 C>A, occurs in the region of the 3' splice site of exon 43, outside of the canonical conserved AG dinucleotide. This variant was classified as a VUS (Table 14), while SpliceAI analysis showed that the substitution potentially disrupted the acceptor splice site of intron 42 (Table 14).

To investigate the functional impact, wild-type and mutant minigene constructs were created by amplifying a region encompassing exon 43 together with 213bp of upstream intronic sequence and 171bp of downstream exonic sequence and cloning into the pTB minigene described in section 2.14.2 (Figure 23A). The variant c.6118-3 C>A was then inserted by quick change mutagenesis PCR as described in section 2.14.4. Upon transient transfection in HEK293T cells, the splicing pattern was analysed by extracting RNA and performing RT-PCR. Given the size of the insert and the relative position of primers over pTB exons, a correctly spliced transcript would have been 420bp long. As can be seen in figure 23B, both the wild-type and mutated minigenes showed one amplicon of the expected size for the correct mRNA transcript, which was confirmed by Sanger sequencing.

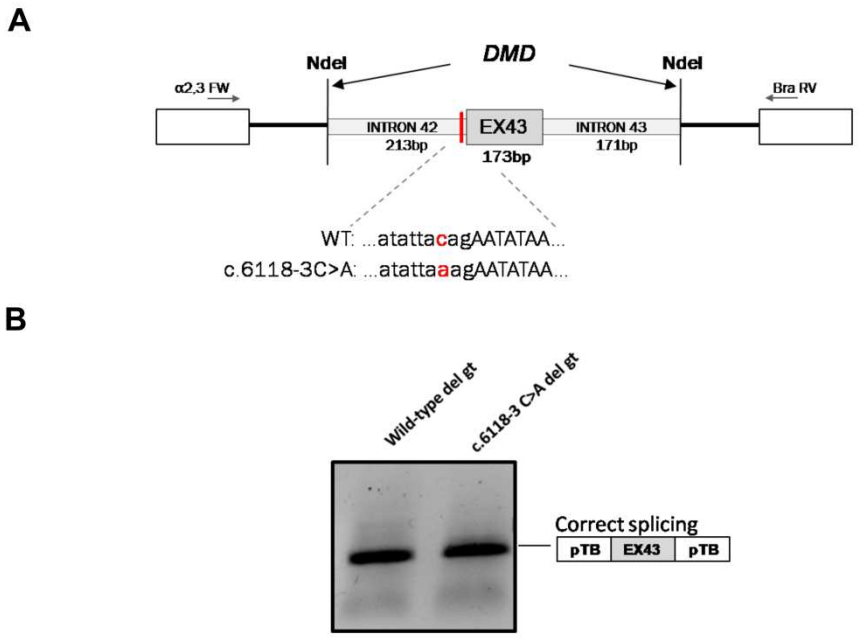


Figure 23. Minigene analysis of pTB *DMD* c.6118-3C>A in HEK293T cells. (**A**) The structure of the minigene harbouring*DMD* exon 43 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of variant studied in the assay is indicated in red. (**B**) RT-

PCR from HEK293T cells transfected with the minigene constructs. A schematic representation of transcripts is presented to the right of the gel image.

***FLNC*: c.3791-1 G>A**

The ACMG criteria characterised the *FLNC* c.3791-1 G>A variant as LP and predicted by SpliceAI to disrupt the canonical acceptor site (1.00 Δscore). The wt minigene was generated by amplifying the region containing exons from 22 and 24 and the respective introns, and subsequently, the variant of interest was introduced by mutagenesis (Figure 24A). HEK293T cells were transfected with both minigenes, and RT-PCR was used to assess the splicing profile. The expected band should be 745bp long which corresponds to the correct splicing of *FLNC* exons joined with α2-3 exon and fibronectin exon. As can be seen in the figure 24B, the wt and mutated minigene produced many transcripts between approximately 650bp, and 250bp in size. sequencing from the wild-type constructs revealed that the upper band (745bp) corresponded with the correctly spliced transcript, while exon 24 skipping was observed in the middle band (583bp), and the lower band corresponds with the empty vector (250bp). On the other hand, the upper band of the mutant constructs the sequencing confirmed the exon 22 skipping (174bp), while the 300bp size corresponds with exon 22 and exon 24 skipping, and the 250bp band is the empty vector, supporting the assignment of the variant as pathogenic.

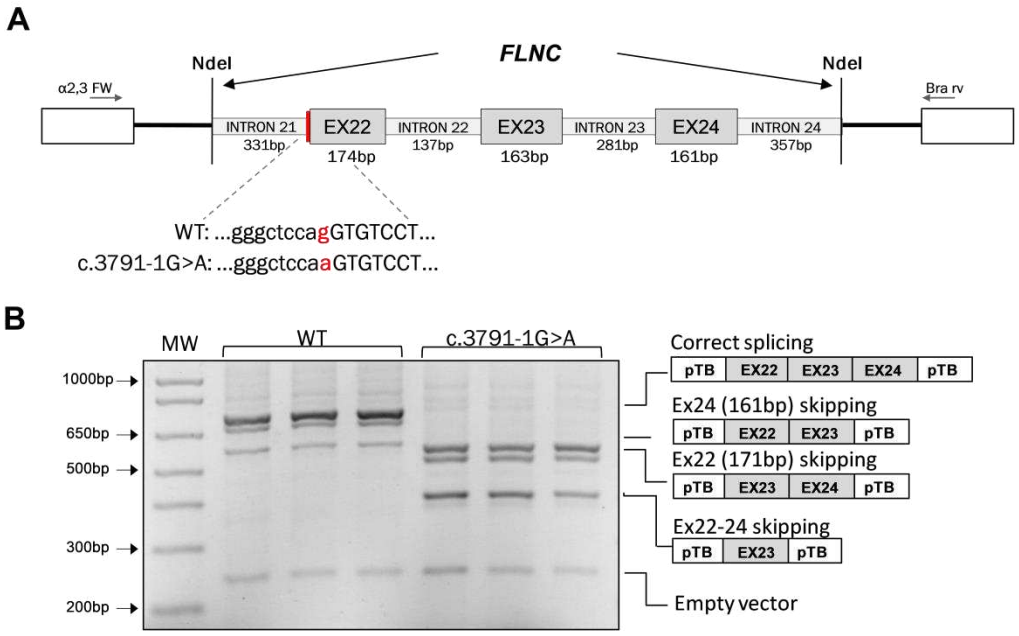


Figure 24. Minigene analysis of pTB *FLNC* c.3791-1G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *FLNC* exons (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after

transfection of HEK293T cells with wild-type and mutated constructs. A schematic representation of transcripts is presented to the right of the gel image.

***LMNA*: c.1609-1 G>A**

The *LMNA* c.1609-1G>A variant was classified according to ACMG guidelines as LP and predicted to disrupt the canonical acceptor site (1.00 SpliceAI Δscore). The mRNA splicing profile was analysed by generating wild-type and mutant minigenes (Figure 25A). The minigenes were transfected into HEK293T cells and the splicing outcome was analysed by RT-PCR. Figure 25B shows that the wild-type *LMNA* minigene resulted in two products: 337bp and 247bp, which correspond to correct splicing of exon 10 and its skipping, respectively. While the mutant *LMNA* c.1609-1G>A minigene results only in the product with exon 10 skipping. Sanger sequencing of the lower bands confirmed this hypothesis and revealed that there was skipping of *LMNA* exon 10 (90bp) in both the wild type and mutated minigene, with the 337bp correctly spliced transcript only observed in the wild type. Skipping of exon 10 is predicted to result in an in-frame deletion of 90 nucleotides due to the absence of exon 10.

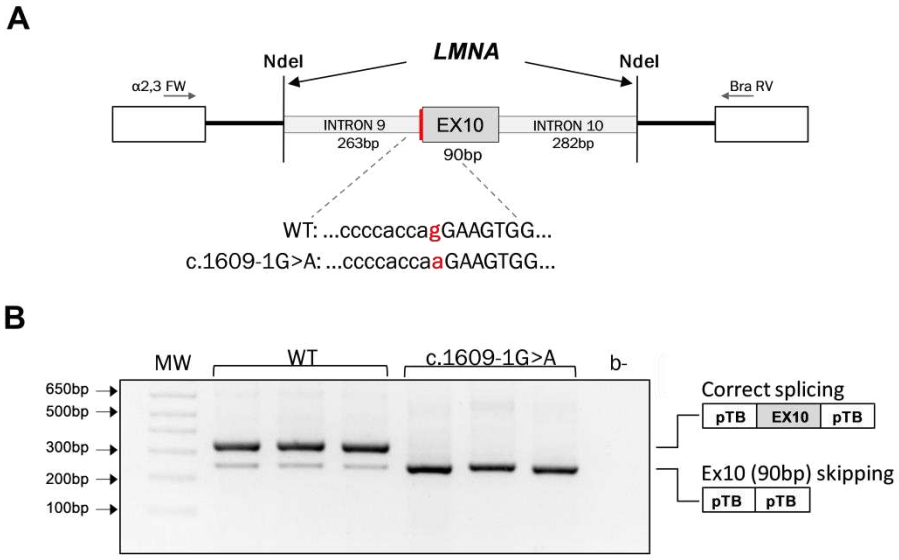


Figure 25. Minigene analysis of pTB *LMNA* c.1609-1G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *LMNA* exon 10 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wild-type and mutated constructs. A schematic representation of transcripts is presented to the right of the gel image.

***MYBPC3*: c.852-1 G>A**

To assess the influence on splicing of c.852-1 G>A variants in *MYBPC3* gene wild type and mutant minigene constructs were generated (Figure 26A). The variant was classified as VUS according to ACMG/AMP guidelines and predicted to induce mis-splicing by SpliceAI. The region of interest includes exon 7 to exon 10 which were amplified by PCR and subsequently cloned in the pcDNA3 vector. The constructs were transiently transfected in HEK293T cells and their splicing profile was analysed by total RNA extraction and RT-PCR. Given the size of the insert and the relative position of the primers T7 and SP6, a correctly spliced transcript would have been 250bp long. As can be seen in figure 26B, the wild-type construct generated two bands (one of them is really faint) around 250bp long which corresponds with the correct spliced transcript. While the construct carrying the variant gave two bands with different sizes than expected, with an approximate size of 270bp and 300bp. This suggests that a cryptic 3'ss is activated immediately upstream of the canonical acceptor site. Sanger sequencing should be performed to confirm this hypothesis.

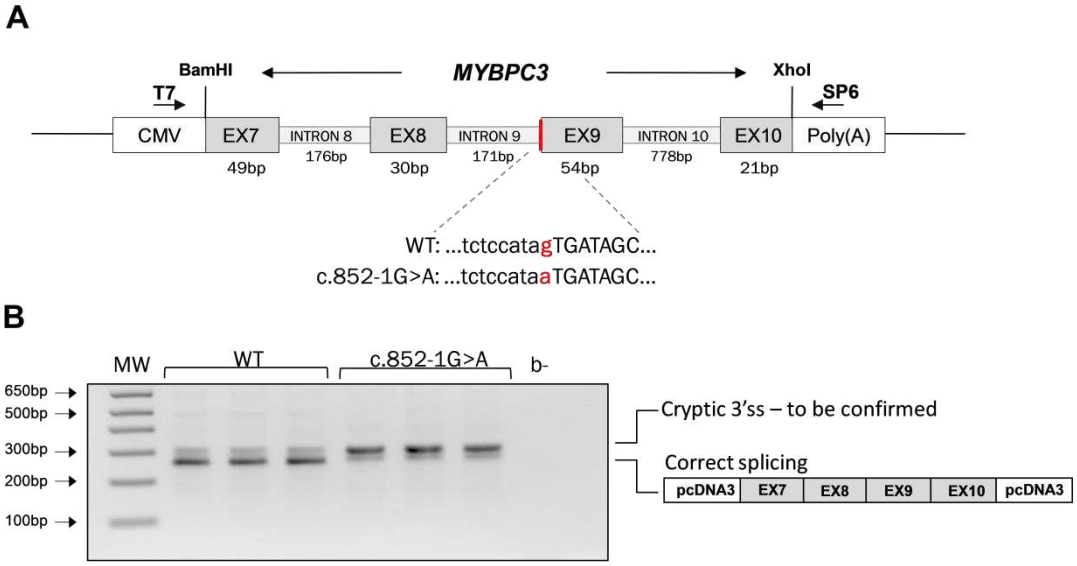


Figure 26. Minigene analysis of pcDNA3 *MYBPC3* c.852-1G>A in HEK293T cells. (**A**) The structure of the minigene harbouring*MYBPC3* gene exon 7 to exon 10 (grey box)inserted into the pcDNA3 vector. White boxes and the flanking lines indicate the CMV and Poly(A) of pcDNA3 vector. The T7 and SP6 primers for RT-PCR are represented as arrows. The location of variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wild-type and mutated constructs. A schematic representation of transcripts is presented to the right of the gel image.

***TTN*: c.25808-1 G>A and c.41609-2 A>C**

Two heterozygous variants, c.25808-1 G>A, and c.41609-2 A>C were identified through clinical genetic testing in *TTN*. These variants were clinically reported as VUS and predicted to result in the disruption of the acceptor sites according to SpliceAI scores (Table 14).

The variant c.25808-1 G>A is located at the end of intron 104 and causes the disruption of the natural AG dinucleotide. To analyse the functional impact on splicing, wt and mutant constructs containing exons 104 to 106 of *TTN* were created by amplification of these regions and subsequent cloning into the pTB as outlined in section 2.14 (Figure 27A). Mutagenesis of the wt minigene was performed to introduce the splice-site variant. The plasmids were transfected into HEK293T cells and their splicing pattern was assessed by extraction of the RNA and RT-PCR. As shown in figure 27B, a single amplicon was observed for the wt and mutated minigenes. However, sequencing of the PCR products revealed that whilst the transcript produced by the control minigene showed a normal splicing pattern, the minigene carrying c.25808-1 G>A used this as an alternative AG splice acceptor site in exon 105 at position c.25808. This produced a PTC at position 8625 that would in turn likely result in mRNA degradation and consequently reduce protein levels.

The same methodology was used to assess the impact of the *TTN* c.41609-2A>C variant. Wild type and mutant minigenes were generated including exons 226-228 of *TTN* with its flanking introns, cloned into the pTB plasmid using NdeI restriction enzyme sites (Figure 27C). Following transient transfection, total RNA was extracted, and the splicing profile was analysed through RT-PCR. A single transcript of the approximate size of 965bp was observed with both wt and mutant constructs (Figure 27D). Sanger sequencing of PCR products identified a usage of a cryptic 3' ss in the first nucleotides of exon 226 (c.41609) which removes 3bp of exon 226. This event is in-frame, and 1 amino acid is removed from the Titin protein (p.13870Glu).

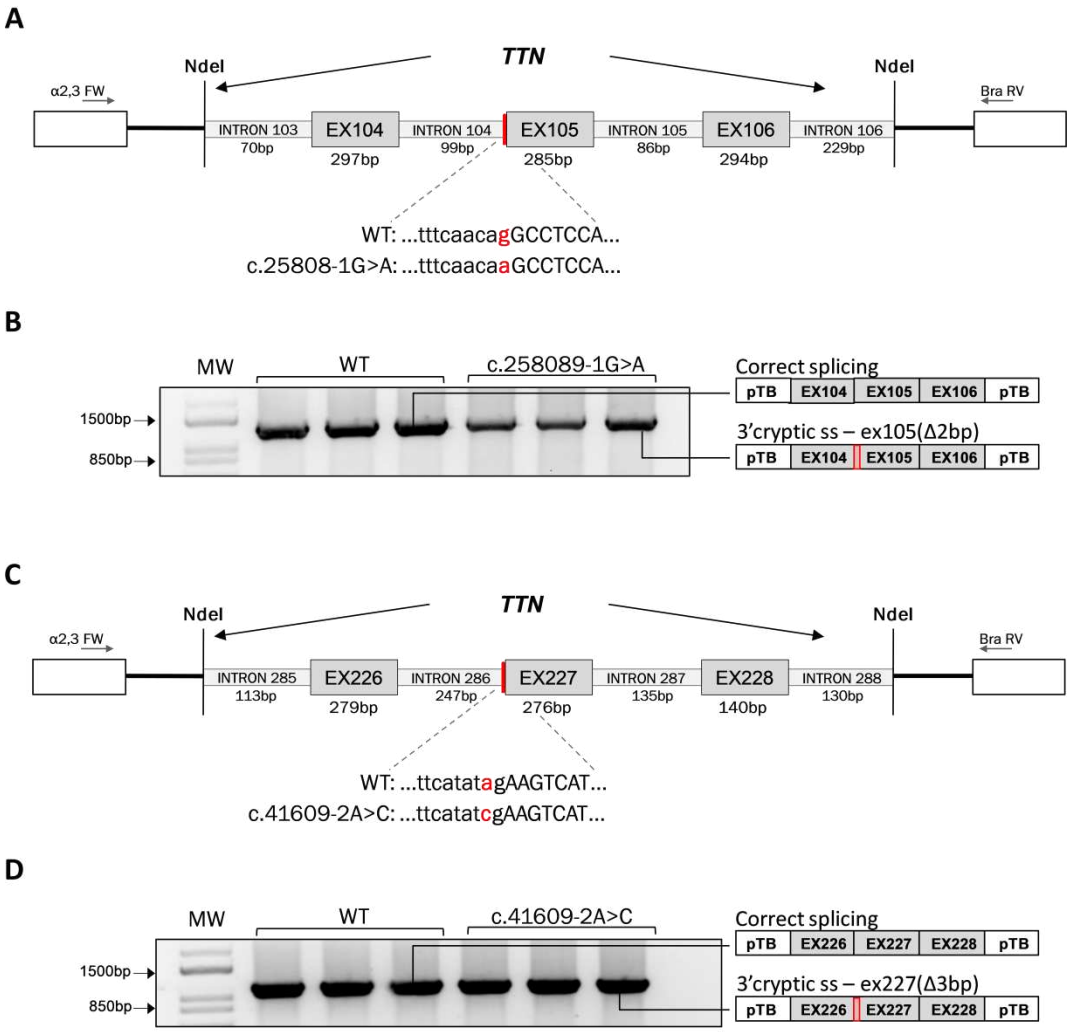


Figure 27. Minigene analysis of pTB *TTN* c.25808-1 G>A and c.41609-2 A>C in HEK293T cells. (**A**) The structure of the *TTN* minigene harbouring *TTN* c.25808-1 G>A. Exons and flanking introns are indicated as grey and light grey boxes respectively. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutant constructs. (**C**) The structure of the *TTN* minigene harbouring c.41609-2 A>C variant. Exons and flanking introns are shown as grey and light grey boxes, respectively. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of variant is indicated in red. (**D**) RT-PCR from HEK293T cells transfected with both wt and mutant minigene constructs. A schematic representation of transcripts is reported to the right of the gel image.

3.3.2 Donor splice site variants

***DSP*: c.273+5G>A, and c.273+3 A>T**

To assess the pathogenicity of the putative causative c.273+5 G>A variant, a splicing assay was performed using minigene as previously described. A construct containing the wt sequence was compared to a construct including the variant, which was introduced by site-directed mutagenesis (Figure 28A). After independent transfection into HEK293T cells, *DSP* transcripts were amplified by RT-PCR to determine mis-splicing events. As it can be seen from figure 28B, the wt minigene resulted in one product, while the minigene carrying the variant resulted in two distinct products of 350bp and 411bp respectively. The 350bp product corresponded to correct splicing of exon 2, while the 411bp fragment was the result of using a cryptic 5' ss. The sequencing revealed a retention of 61 nucleotides that cover *DSP* intron 2, which encodes a PTC and may be targeted by NMD.

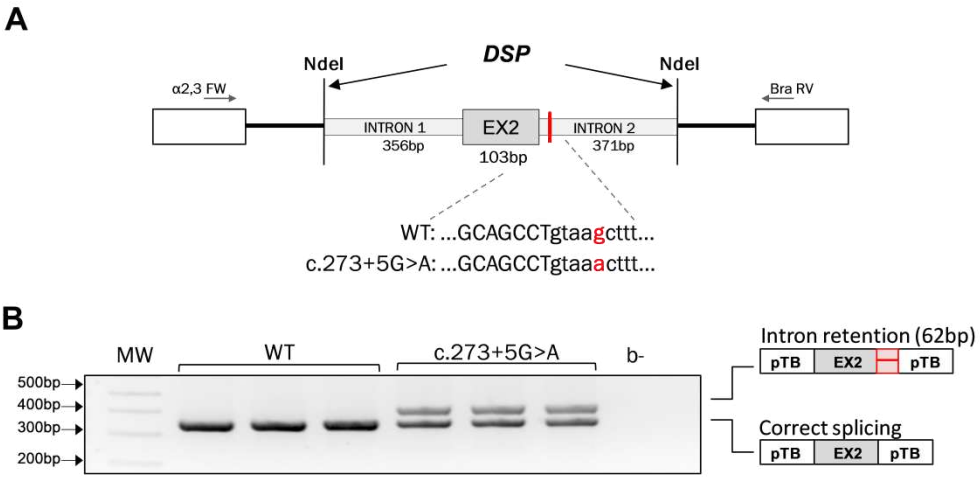


Figure 28. Minigene analysis of pTB *DSP* c.273+5 G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *DSP* exon 2 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of transcripts is represented to the right of the gel image.

Another genetic variant located in the same region of the previous, was identified in a patient who carried a nucleotide substitution in position +3 of intron 2 with an A became a T. The variant was classified as VUS and prediction confirmed that can disrupt the native 5' ss. The same methodology based on the minigene assay was performed to analyse the splicing pattern (Figure 29A). RT-PCR reported a splicing error with a similar splicing profile to the previous

variant which seems to activate a cryptic 5' ss. As it can be seen from figure 29B, the wt minigene resulted in one product, while the minigene carrying the variant resulted in three distinct products of 400bp, 350bp and 250bp respectively. The 350bp product corresponded to correct splicing of exon 2, the 400bp fragment was the result of using a cryptic 5' ss. Whereas the 250bp corresponds to the size of pTB exons as a result of exon 2 skipping. Sanger sequencing confirmed this hypothesis and the usage of a 5' cryptic ss at 61nt downstream of the canonical donor site.

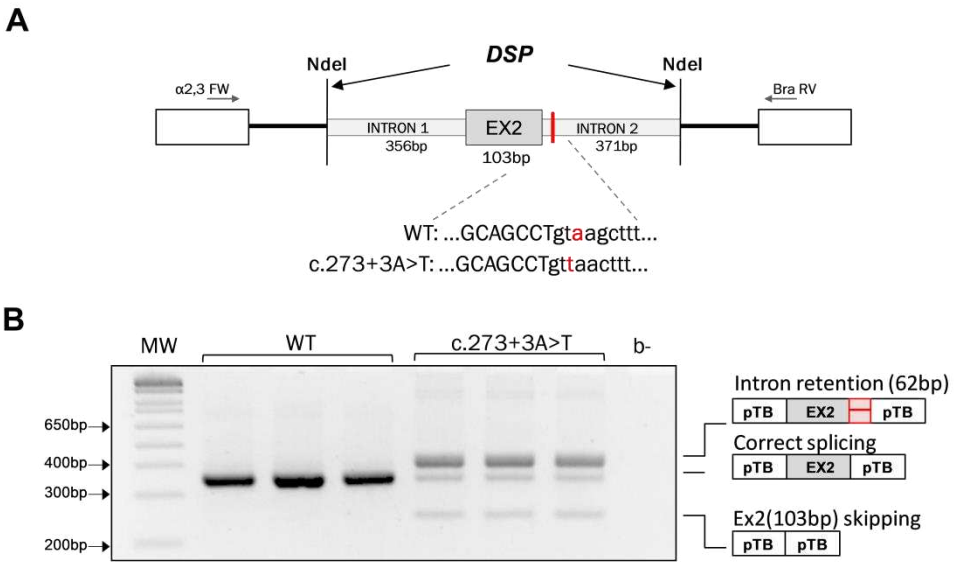


Figure 29. Minigene analysis of pTB *DSP* c.273+3 A>T in HEK293T cells. (**A**) The structure of the minigene harbouring *DSP* exon 2 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of variant studied in the assay is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of transcripts is represented to the right of the gel image.

FLNC: c.7251+1G>A

The *FLNC* c.7251+1 G>A variant is located at the first nucleotide of intron 43 and causes disruption of the GT dinucleotides which become AT. The variant was clinically classified as Pathogenic and predicted to disrupt the donor site by SpliceAI (1.00 Δscore). To determine the pathogenicity of the variant, wt and mutant *FLNC* minigenes were generated (Figure 30A). To evaluate the splicing outcome both minigenes were transiently transfected into HEK293T cells and RT-PCR was performed. Given the size of the insert and the relative position of primers along pTB vector, a correctly spliced transcript would have been 461bp long. As can be seen in figure 30B, two distinct bands were observed in the wild-type *FLNC* minigene. The upper band corresponds to the use of an alternative acceptor splice site in the intron, while the lower band

results in the correct splicing of exon 42 and exon 43 joining with the vector exons. In contrast, a different splicing profile was observed in the *FLNC* c.7251+1 G>A minigene, with three distinct products observed. The upper band of 472bp corresponds to exon 43 skipping and the usage of an alternative 3'ss in the fibronectin intron, the middle band of 385bp corresponds to a transcript lacking exon 43, and the lower band of 250bp corresponds to an uncertain transcript, possibly the pTB exons joined together. To clarify the nature of the fragments, DNA sequencing of the bands was performed. Sequencing confirmed the hypothesis.

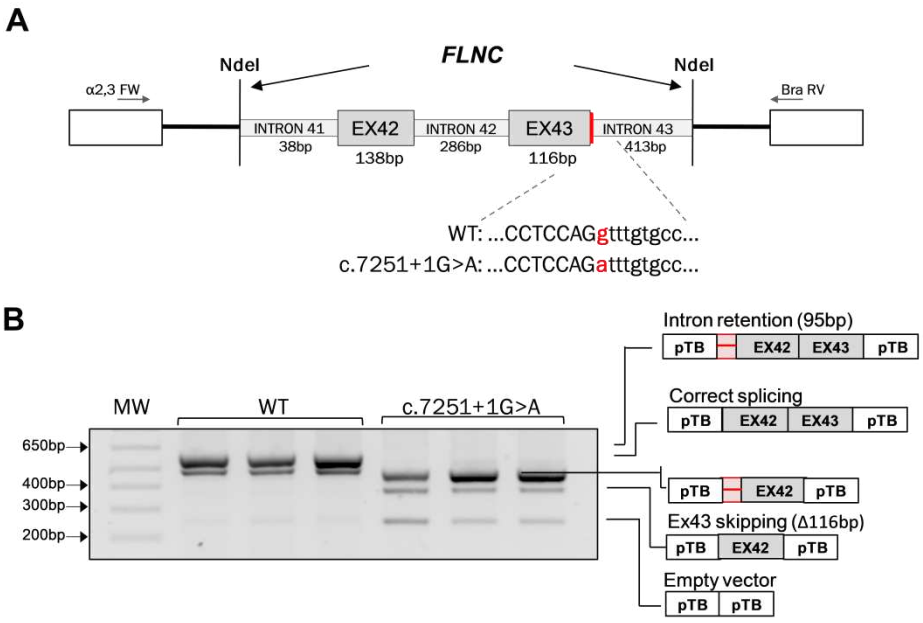


Figure 30. Minigene analysis of pTB *FLNC* c.7251+1 G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *FLNC* exon 42 and exon 43 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293Tcells with wt and mutant constructs. A schematic representation of transcripts is presented to the right of the gel image.

***MYLK:* c.3703+22 T>C**

The MYLK c.3703+22 T>C variant was predicted to not interfere with splicing by SpliceAI and classified as VUS based on ACMG/AMP criteria. Wild-type and mutant minigene constructs were generated (Figure 31A). Given the size of insert and the relative position of the primers over the α,2-3 fibronectin and β-globin exons, a correctly spliced transcript would have been 298bp long. As can be observed in figure 31B, the PCR products from the wt and the mutated minigene gave one single transcript of the expected size (298bp). Sanger sequencing revealed this corresponds to the correctly spliced transcript. Therefore, these data support the classification of non-pathogenic for this variant.

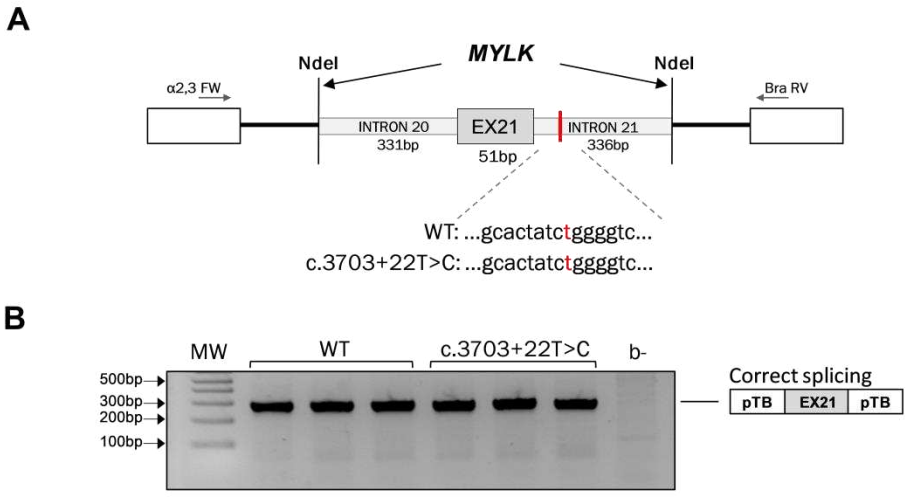


Figure 31. Minigene analysis of pTB *MYLK* c.3703+22 T>C in HEK293T cells. (**A**) The structure of the minigene harbouring *MYLK* exon 21 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR from HEK293T cells transfected with the minigene constructs. A schematic representation of the transcripts is reported to the right of the gel image.

NEXN: c.27+9C>T, and c.1053+1G>A

The variant *NEXN* c.27+9 C>T was reported as VUS by ACMG/AMP criteria and predicted to have no effect by SpliceAI. However, to confirm or refute the prediction the minigene assay was performed. Wild type and mutated *NEXN* minigenes covering exon 2 and part of the adjacent intron sequence were generated (Figure 32A). The minigenes were transfected into HEK293T cells and splicing profile was analysed by RT-PCR (Figure 32B). The wt minigene resulted in one distinct product of 470bp, which was also observed in mutant minigene. The correct splicing of *NEXN* exon 2 was confirmed by Sanger sequencing.

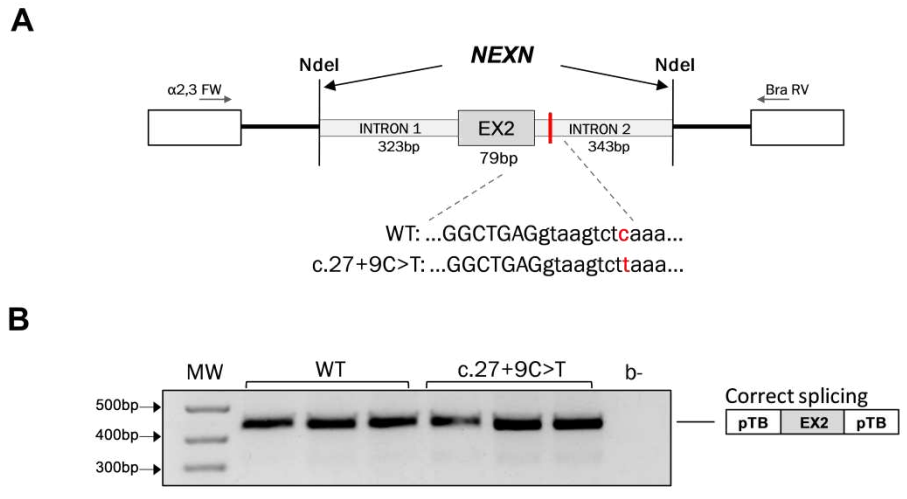


Figure 32. Minigene analysis of pTB *NEXN* c.27+9 C>T in HEK293T cells. (**A**) The structure of the minigene harbouring *NEXN* exon 2 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of transcripts is shown to the right of the gel image.

The variant in *NEXN,* c.1053+1 G>A, is located in the donor site of exon 9. The ACMG/AMP guidelines classified this variant as LP, while SpliceAI analysis showed that the substitution potentially disrupted the donor splice site of intron 9 (Table 14).To investigate the functional impact, wt and mutant minigene constructs were created by amplifying a region encompassing exon 9 together with 453bp of upstream intronic sequence and 497bp of downstream exonic sequence and cloning into the pTB minigene described in section 2.14.2 (Figure 33A). The variant c.1053+1 G>A was then inserted by quick change mutagenesis PCR as described in section 2.14.4. Upon transient transfection in HEK293T cells, the splicing pattern was analysed by extracting RNA and performing RT-PCR. Given the size of the insert and the relative position of primers over pTB exons, a correctly spliced transcript would have been 439bp long. As can be seen in figure 33B, the wt minigene showed one amplicon of the expected size for the correct mRNA transcript, which was confirmed by Sanger sequencing. In contrast, the minigene carrying the variant showed two distinct bands of unexpected size around 500bp and 250bp, respectively. Sanger sequencing revealed that the 500bp band corresponds with the usage of a cryptic 5' ss at 168nt downstream of the natural donor site which is disrupted by the nt substitution, and exon 9 skipping in the band with size of 250bp.

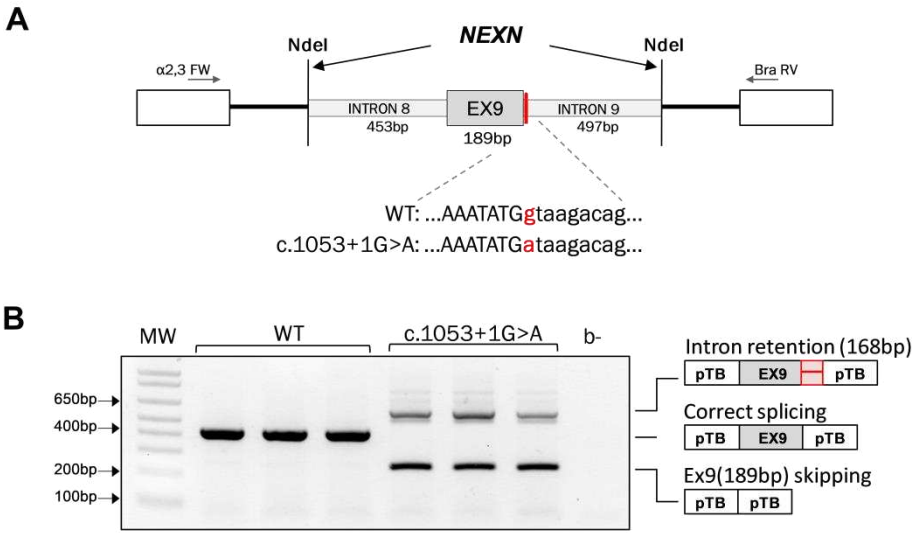


Figure 33. Minigene analysis of pTB *NEXN* c.1053+1 G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *NEXN* exon 9 (grey box) and the relative flanking introns (light grey

between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of transcripts is shown to the right of the gel image.

***PKD1*: c.10405+5 G>T**

The variant in *PKD1*, c.10405+5 G>T, occurs in the region of the 5' splice site of exon 33. This variant was classified as VUS according to the ACMG criteria, while the nt substitution was predicted to potentially influence the normal mRNA processing (0.32 Δscore). To clarify the role of the variant, a wt minigene construct was generated by amplifying the region encompassing from exon 31 to exon 34 and cloning in pcDNA3 vector. Then, the specific variant was introduced by quick change site directed mutagenesis (Figure 34A). Transient transfection was performed in HEK293T cells, and RT-PCR was used to assess the splicing pattern. As can be seen in Figure 34B, the splicing pattern from the wt and mutant minigenes clearly differ, thus suggesting a splicing aberration due to the analysed variant. The wt minigene produced two transcripts (one of them is really faint) of the approximate of 550bp and 400bp size, while the minigene carrying the variant gave rise to two transcripts of approximate size of 480bp and 400bp. Sanger sequencing revealed that the upper bands (550bp) correspond with the correctly spliced transcript. In contrast, the lower band in common in both wt and mutant minigenes revealed the exon 33 skipping, while the middle band (480bp) in the mutant minigene corresponds with the intron 31 retention (87bp) and exon 33 skipping. Therefore, these data support the pathogenicity role of the variant which causes mis-splicing events in *PKD1* gene.

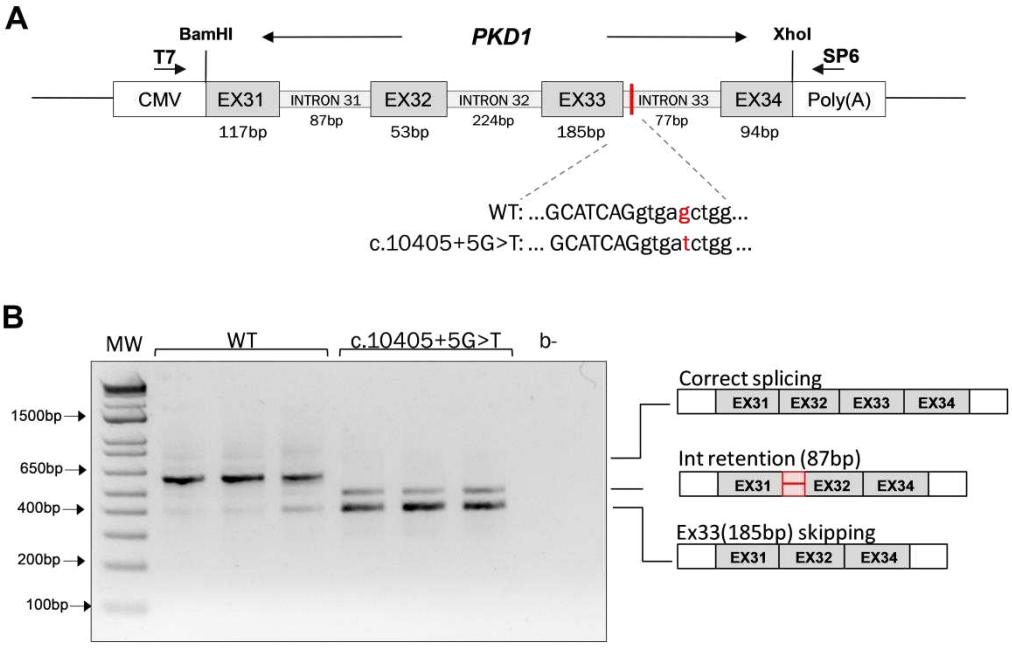


Figure 34. Minigene analysis of pcDNA3 *PKD1* c.10405+5 G>T in HEK293T cells. (**A**) The structure of the minigene harbouring *PKD1* exon 31 to exon 34 (grey boxes) and the relative introns (black line) inserted into pcDNA3 vector. White boxes and the flanking lines indicate the pcDNA3 vector. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of transcripts is shown to the right of the gel image.

***RYR2*: c.1961+3 A>T**

The *RYR2* c.1961+3 A>T variant is located at the beginning of intron 19. This heterozygous variant was clinically evaluated as VUS, but SpliceAI prediction suggested disruption of the donor site. In order to analyse the effect of this variant on mRNA splicing, minigenes were generated for the wt and mutated sequence (Figure 35A), by inserting a genomic region which included exon 19 (134bp) flanked by 248bp intronic sequence upstream and 302bp intronic sequence downstream into the pTB vector (see section 2.14). After transfection into HEK293T cells, RT-PCR analysis was performed. The size of a correctly spliced transcript should be 381bp due to the length of the exon 19 (134bp) and the relative position of primers over the pTB vector. The size of the PCR products obtained from the wt and mutated minigenes clearly differed, thus suggesting a splicing error induced by the variant (Figure 35B). Sequencing of the amplicons showed that a cryptic 5'ss was used 31bp downstream of the canonical GT, compared to the transcript originating from the wt minigene. To clarify the nature of the additional bands on the gel which were suspected to be heteroduplexes, DNA sequencing of the eluted bands was performed. The resulting electropherogram clearly shows a sequence profile with over-impressed double peaks, consistent with the hypothesis of heteroduplex formation. Abnormal splicing events include a retention of 31nt in the mature RNA transcript, which encodes a PTC immediately downstream of exon 19 and resulting mRNA may likely undergo NMD.

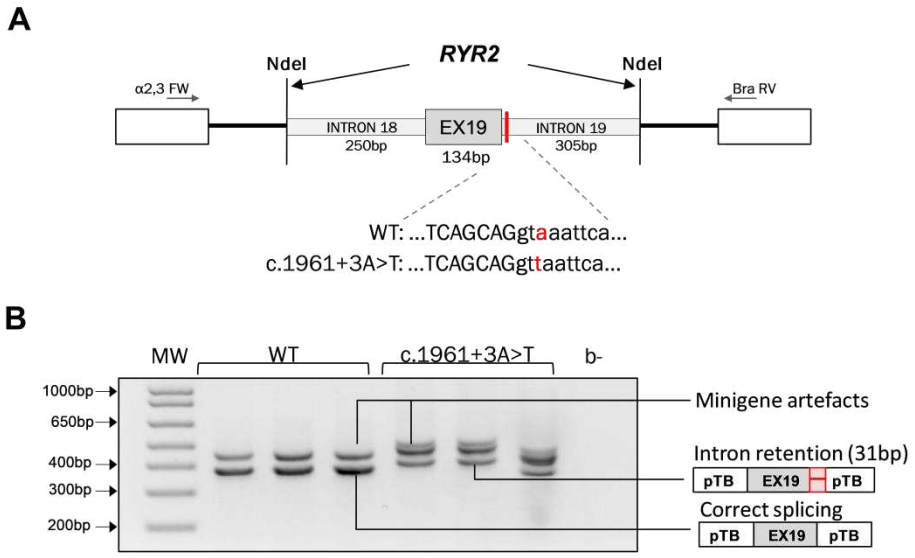


Figure 35. Minigene analysis of pTB *RYR2* c.1962+3 A>T in HEK293T cells. (**A**) The structure of the minigene harbouring*RYR2* exon 19 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of the transcripts is shown to the right of the gel image.

***SCN5A*: c.3840+5 G>C**

The *SCN5A* c.3840+5 G>C variant is located inside the consensus sequence of the donor splice site which can induce a mis-splicing event. The variant was classified as VUS following the ACGM/AMP criteria and predicted to disrupt the normal exon-exon joined by SpliceAI (Δscore 0.99). The minigene constructs were generated and include the region of interest exon 21, and its flanking introns (Figure 36A). The possible pathogenicity of the variants was analysed by RT-PCR after transient transfection of HEK293T cells. The expected size should be 421bp long, one single band was observed from the wild-type minigene, with a different pattern detected in the mutated construct (Figure 36B). The variant appears to induce the activation of cryptic 5' ss due to the size of the band, which is longer than the correctly spliced transcript. The hypothesis was confirmed after performing Sanger sequencing, which identified the activation of two different cryptic 5'ss, 51nt and 46nt downstream of the canonical donor site. As a result, the intron retention causes the introduction of a PTC into the transcript which can be degraded by NMD.

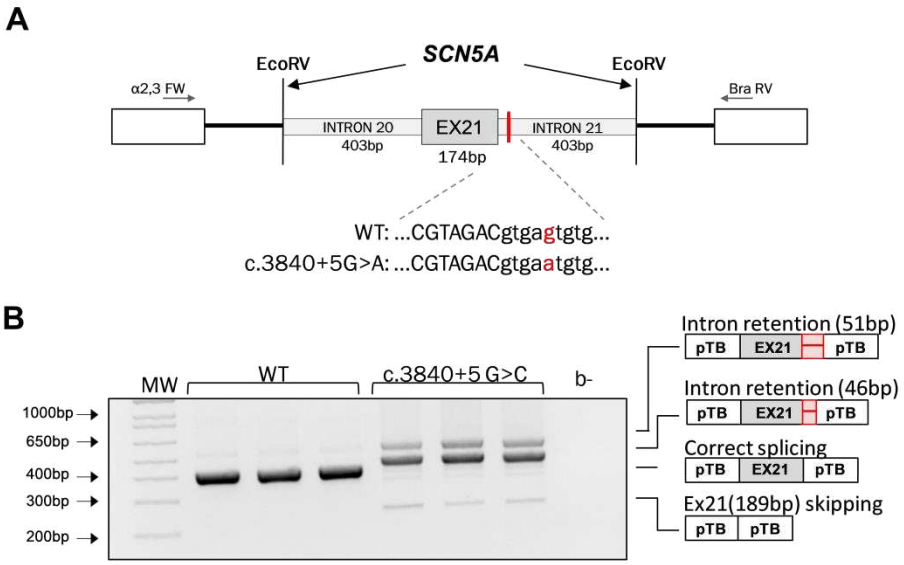


Figure 36. Minigene analysis of pTB *SCN5A* c.3840+5 G>C in HEK293T cells. (**A**) The structure of the minigene harbouring *SCN5A* exon 21 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The variant location is indicated by a red line. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of the transcripts is represented to the right of the gel image.

***TMEM43*: c.882+8 C>T**

The *TMEM43* c.882+8 C>T variant is located in the vicinity of the exon-intron junction (exon10-intron10) changing C to T 8 nucleotides downstream of the native 5' ss. This variant was labelled as a VUS by ACMG/AMP criteria and no predicted score was given by SpliceAI. To evaluate if this variant could be pathogenic, the minigene constructs for the wt and variant were created by cloning a region composed of exon 10 with 268bp of upstream intronic sequence and 243bp of downstream intronic sequence into the pTB vector as described in section 2.14 (Figure 37A). After transient transfection in HEK293T cells, mRNA was extracted, and RT-PCR was performed. The results showed an amplicon of similar size bands on gel electrophoresis in both wt and mutated minigenes, indicating that no abnormal splicing events had occurred (Figure 37B). Sanger sequencing of gel-purified bands confirmed the hypothesis.

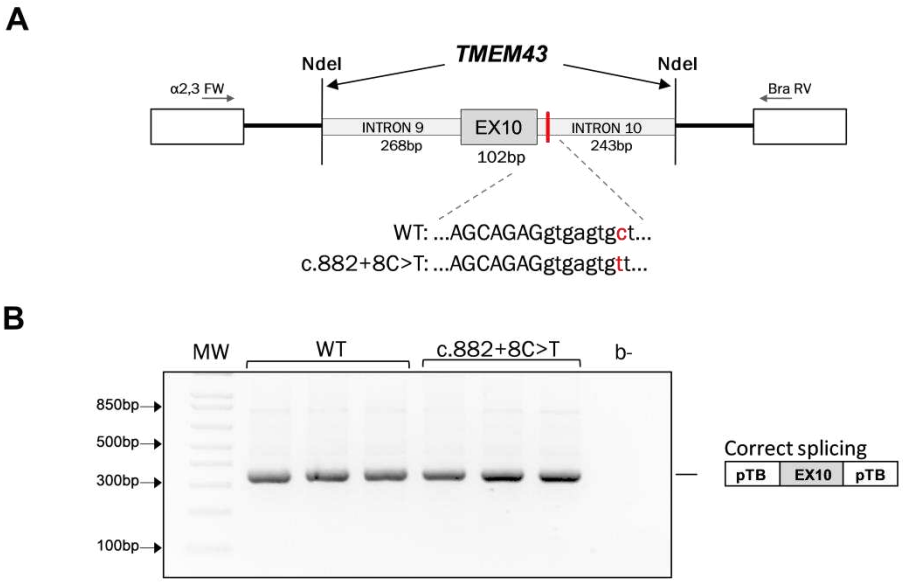


Figure 37. Minigene analysis of pTB *TMEM43* c.882+8 C>T in HEK293T cells. (**A**) The structure of the minigene harbouring *TMEM43* exon 10 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. The location of the variant is indicated in red. (**B**) RT-PCR from HEK293T cells transfected with the minigene constructs. A schematic representation of the transcript is reported to the right of the gel image.

TTN: c.669+4 T>C, c.13282+1G>A, c.19501+2 T>C, c.32875+2 T>C, c.39574+7 A>G, c.50346+3 A>G and c.57847+4delGTAA

For *TNN*, seven variants (c.50346+3 A>G, c.669+4 T>C, c.57847+4delGTAA, c.32875+2 T>C, c.39574+7 A>G, c.13282+1G>A, and c.19501+2 T>C) located in the vicinity of consensus donor splice sites were studied. The variant c.50346+3 A>G is located three nucleotides downstream of exon 234 and had a zero SpliceAI score which corresponds to not disruptive for splicing, and was classified as VUS based on ACMG/AMP criteria. A wt pTB construct was generated including the genomic region of interest spanning from exon 233 to exon 236 (with part of the upstream and downstream introns included) (Figure 38A). The variant c.50346+3 A>G was generated by quick change mutagenesis using the wt construct as template. For this experiment, minigene constructs were transiently transfected into HEK293T cells. After 24 hours total RNA was extracted, and the splicing pattern analysed by RT-PCR amplification. A single transcript was observed in both wt and mutated minigenes of the approximate size of 1000bp (Figure 38B). Sanger sequencing revealed that the bands were the correctly spliced transcript. Similarly, the c.669+4 T>C variant was also analysed. This variant was classified as a VUS and no effect on splicing was predicted by SpliceAI. The wt minigene was generated by

inserting a genomic region spanning from exon 4 to exon 6 flanked by 418bp intron 3 and 288bp intron 6 cloned into pTB vector using NdeI restriction sites (Figure 38C). Transfection of these constructs was performed in HEK293T cells, followed by RT-PCR. The analysis of PCR products on agarose gel showed that the variant did not cause aberrant splicing (Figure 38D). The results were confirmed by Sanger sequencing of eluted bands.

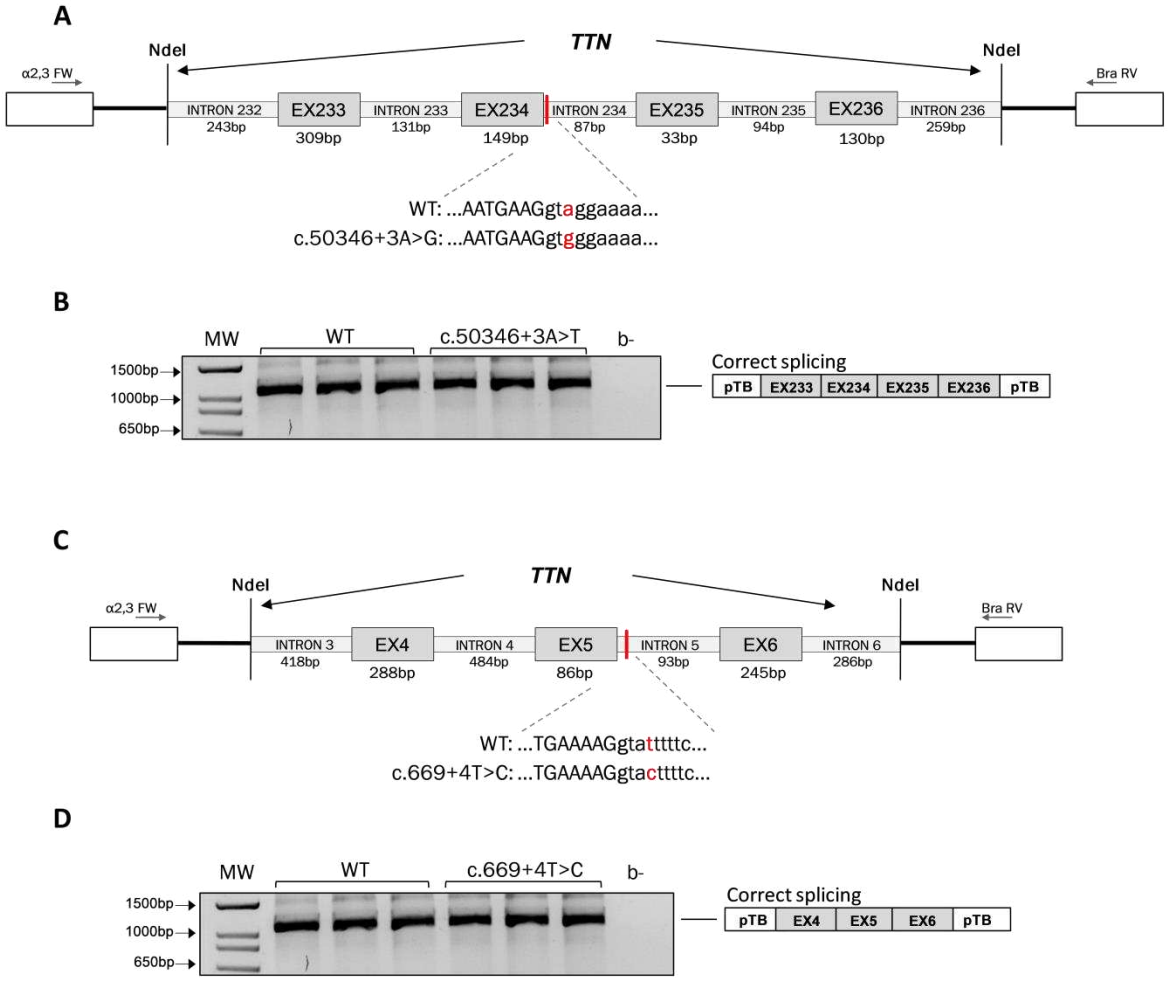


Figure 38. Minigene analysis of pTB *TTN* c.50346+3 A>G and c.669+4 T>C in HEK293T cells. (**A**) The structure of the *TTN* minigene harbouring exon 233-236 (grey boxes) and the relative flanking introns (light grey boxes between NdeI site) inserted into pTB vector. The position of the variant is indicated in red. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR from HEK293T cells transfected with both wt and mutant minigene constructs. (**C**) The structure of the *TTN* minigene harbouring exons 4 to 6 (grey boxes) and the relative flanking introns (light grey boxes between NdeI site) inserted into pTB vector. The variant position is indicated in red. White boxes and the flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**D**) RT-PCR from HEK293T cells transfected with both wt and mutant minigene constructs. A schematic representation of the transcript is reported to the right of the gel image.

The variant *TTN* c.13282+1 G>A was classified as VUS according to ACMG/AMP criteria. The effect of the splice-site variant was calculated by the SpliceAI algorithm which predicted disruption of the canonical donor site (Δscore 1.00). The splicing profile was evaluated by generating the wt and mutated minigenes which contained exon 46 and its flanking introns (Figure 39A). The resulting constructs were transiently transfected into HEK293T cells and *TTN* alternative splicing was evaluated by RT-PCR analysis. It was observed that the mutant minigene carrying the variant c.13282+1G>A induced aberrant splicing compared to the wt minigene, where the single fragment observed corresponds to the expected transcript including the *TTN* exon 46. In the mutated *TTN* c.13282+1 G>A minigene, three distinct unexpected fragments were observed (Figure 39B) and sequencing revealed that the upper band corresponded to a cryptic 5' splice site in intron 46 located 132bp downstream of the canonical donor site, while the middle fragment results from a deletion of exon 46 due to the activation of an alternative splice site at +197bp of exon 46. Finally, the lowest band results from exon 46 skipping.

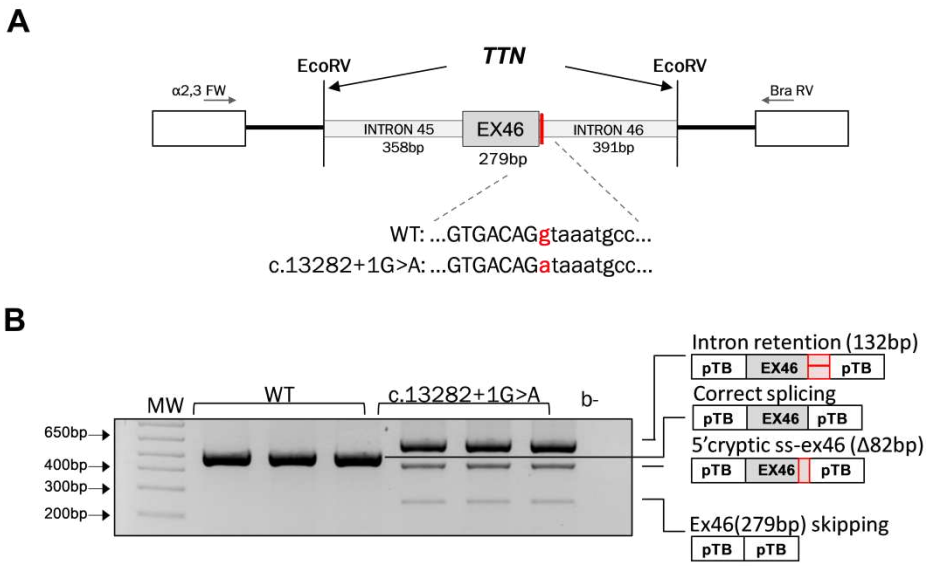


Figure 39. Minigene analysis pTB *TTN* c.13282+1 G>A in HEK293T cells. (**A**) The structure of the minigene harbouring *TTN* exon 46 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the c.13282+1 G>A variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of the transcripts is reported to the right of the gel image.

The *TTN* c.19501+2 T>C variant has been predicted to potentially disrupt the natural donor site by SpliceAI and has been classified as a VUS in accordance with the criteria outlined by the ACMG/AMP. In order to elucidate the pathogenicity associated with this variant, minigenes

representing both the wt and mutated sequences were generated, as illustrated in Figure 40A. Considering the insert size and the primer positions relative to α2-3 globin and fibronectin exons, a correctly spliced transcript should be 639bp in length. The splicing pattern analysed by RT-PCR, revealing the generation of multiple transcripts ranging from approximately 250bp to 650bp (Figure 40B). The observed differences in band sizes between wt and mutated minigenes strongly suggested a splicing alteration induced by the variant. Therefore, it was not expected to obtain all of these transcripts probably this is associated with the size of the cloned insert. The band of 650bp derived from the wt minigene seems to correspond with the expected transcripts, while the lower band resulted 120-100bp shorter. Sanger sequencing revealed that the upper band did not align with the expected transcripts due to the exon 77 skipping and retention of intronic regions from the pTB vector, while the lower band corresponded to the correctly spliced transcript with exon 78 joined with pTB exons but the skipping of exon 77. In contrast, Sanger sequencing of the bands derived from the mutant minigene revealed that the upper band corresponded to the exon 77 skipping and activation of a cryptic 5'ss within exon 78, resulting in a deletion of 175bp. The middle band potentially indicated minigene contamination, evidenced by the presence of double peaks in the electropherogram, while the lower band corresponds to the empty vector. Importantly, all transcripts derived from both wt and mutant minigenes showed exon 77 skipping, thus suggesting the lack of information upstream of the indicated exon. However, these findings support the pathogenicity of the variant.

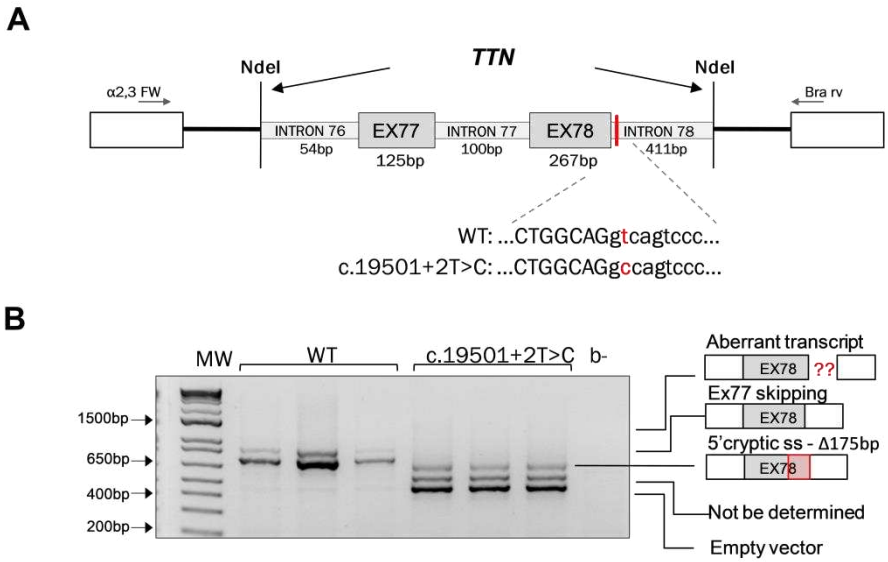


Figure 40. Minigene analysis of pTB *TTN* c.19501 T>C in HEK293T cells. (**A**) The structure of the minigene harbouring *TTN* from exon 77 to exon 78 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the c.19501+2 T>C variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR

from HEK293T cells transfected with the minigene constructs. A schematic representation of the transcript is presented to the right of the gel image.

The *TTN* c.32875+2 T>C variant was predicted to potentially disrupt the natural donor site by SpliceAI and classified as VUS by ACMG/AMP guidelines. To investigate the pathogenicity of this variant, wt and mutated minigenes were generated (Figure 41A). Given the size of insert and the relative position of primers over α2-3 globin and fibronectin exons, a correctly spliced transcript would have been 795bp long. The splicing profile analysed by RT-PCR showed that both wt and mutated minigenes gave rise to one transcript of the approximate size of 800bp and 700bp, respectively (Figure 41B). The size of the bands obtained by wt and mutated minigenes clearly differed, thus suggesting a splicing alteration caused by the variant. Sanger sequencing of the lower bands confirmed this hypothesis and revealed that there was a skipping of *TTN* exon 141 (84bp) in the mutated minigene compared to the 795bp correctly spliced transcript in the wild type minigene.

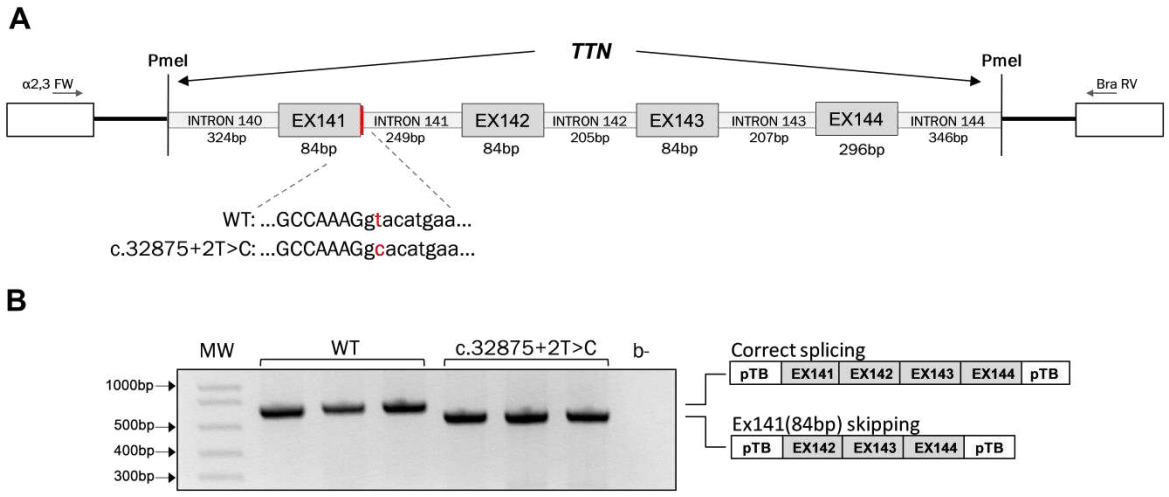


Figure 41. Minigene analysis of pTB *TTN* c.32875+2 T>C in HEK293T cells. (**A**) The structure of the minigene harbouring *TTN* from exon 141 to exon 144 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the c.32875+2 T>C variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR from HEK293T cells transfected with the minigene constructs. A schematic representation of the transcript is presented to the right of the gel image.

The *TTN* c.39574+7 A>G variant was classified as VUS by ACMG/AMP criteria and SpliceAI did not predict any effect on splicing (Table 13 and 14, respectively). Minigene experiments using the pTB vector were undertaken. These included the human *TTN* genomic region spanning from exon 144 to exon 145 with parts of upstream and downstream introns included (90bp and 92bp, respectively) (Figure 42A). HEK293T cells were transfected with these constructs. The splicing

pattern derived from the constructs was then examined by RT-PCR analysis, where a common band was observed in both wt and mutated constructs corresponding to the expected wild-type transcript. An additional band was observed only in the mutant construct which is an aberrant transcript in which exon 144 is skipped (ΔE144 – 306bp) (Figure 42B). These results were confirmed by Sanger sequencing. The exon skipping is an in-frame event, removing 102 amino acids (p.Tyr13090_Ile13192) and a substitution of 1 amino acid at position 13090 (p.13090Phe).

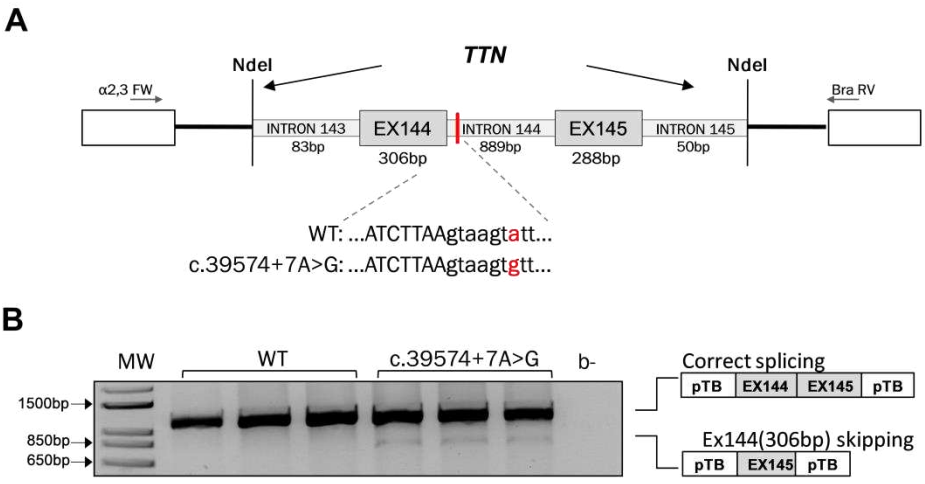


Figure 42. Minigene analysis pTB *TTN* c.39574+7 A>G in HEK293T. (**A**) The structure of the minigene harbouring *TTN* exon 144-145 (grey boxes) and its flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the c.39574+7 A>G variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutated constructs. A schematic representation of the transcripts is reported to the right of the gel image.

The variant c.57847+4delGTAA *TTN* deletes 4nt downstream of the native 5' ss. The deletion was classified as VUS by ACMG/AMP criteria and predicted to cause the loss of the canonical 5' ss by SpliceAI (Table 13 and 14, respectively). To investigate the effect on mRNA processing, a minigene analysis was performed using wt and variant constructs. The engineered wt construct had exon 295 flanked by 303bp intron 294 and 327bp intron 295 cloned in the pTB vector, as described in section 2.14 (Figure 43A). The wt and mutated minigenes were transiently transfected in HEK293T cells and the splicing profile was assessed by RT-PCR. As can be seen in figure 43B, the mutated minigene showed two transcripts of approximately 550bp and 1300bp, in contrast to a single transcript from the wt minigene. The size of the bands obtained by wt and mutated minigenes clearly suggest a splicing alteration due to the variant. Sanger sequencing of the upper band revealed intron retention of 779bp in the transcript compared to the 550bp of the lower band which was the correctly spliced transcript. Retention of intron 295

leads to the introduction of a PTC in the mRNA immediately after exon 295 which would likely result in its degradation via NMD.

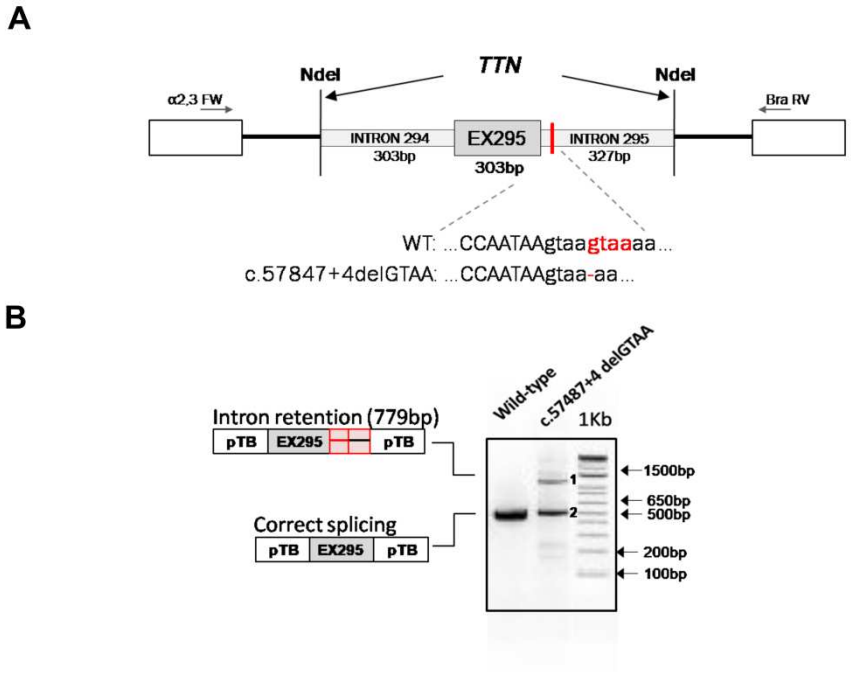


Figure 43. Minigene analysis pTB *TTN* c.57847+4delGTAA in HEK293T cells. (**A**) The structure of the minigene harbouring *TTN* exon 295 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR results after transfection of HEK293T cells with wt and mutant constructs. A schematic representation of the transcripts is reported to the right of the gel image.

***VCL*: c.622+4 C>G**

The variant c.622+4 C>G occurs in position +4 of exon 5 (+4 C>G) in *VCL*. It was previously classified as VUS and SpliceAI prediction reported it to not be splicing disruptive. The wt minigene was generated by amplifying a region that contained the exon 5 flanked by 245bp of intron 4 and 262bp of intron 5 and cloned into pTB vector using NdeI restriction site (see section 2.14) (Figure 44A). Quick change mutagenesis was performed to introduce the variant as described in section 2.14.4. Transient transfection experiments were performed in HEK293T cells and RNA was extracted 24h post transfection. RT-PCR analysis showed that the splicing pattern of both minigenes behaved in an identical manner (Figure 44B). The gel electrophoresis showed an upper band in both wt and mutant constructs, where cryptic splice sites, 5' and 3' respectively, were activated in the intron of the *VCL* gene, as a result there is an inclusion of a small portion of VCL intron 5. This event is an artefact of the minigene system. The lower band

corresponds to the canonical VCL mRNA. Sanger sequencing of the lower band confirmed canonical splicing of VCL exon 5.

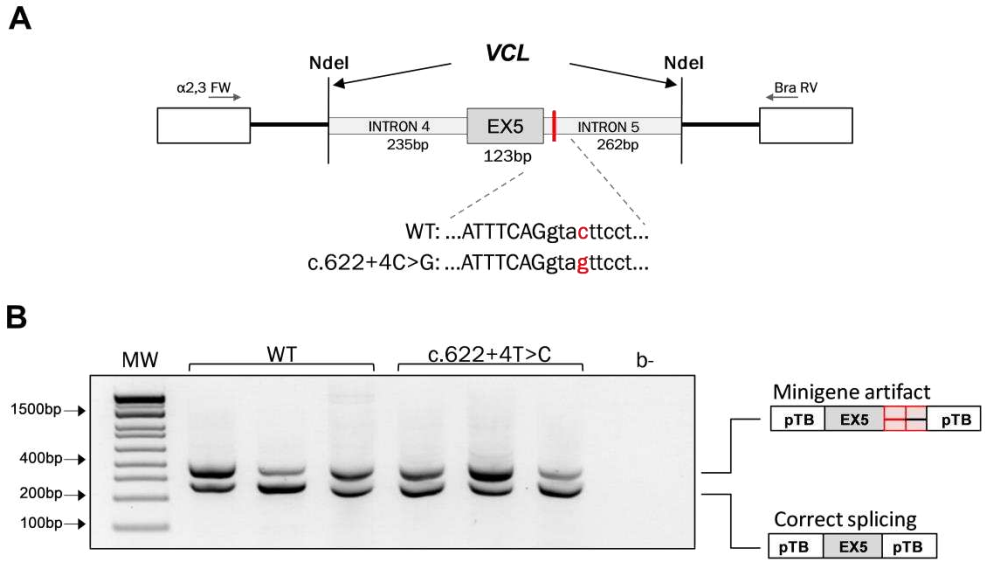


Figure 44. Minigene analysis pTB *VCL* c.622+4 C>G in HEK293T cells. (**A**) The structure of the minigene harbouring *VCL* exon 5 (grey box) and the relative flanking introns (light grey between NdeI site) inserted into pTB vector. The position of the c.622+4 C>G variant is indicated in red. White boxes and flanking lines indicate the pTB vector exons and introns, respectively. The forward and reverse primers for RT-PCR are represented as arrows. (**B**) RT-PCR from HEK293T cells transfected with both wt and mutant minigene constructs. A schematic representation is presented to the right of the gel image.

3.4 Assessment of AC16 splicing profile by minigene assay

In this study, to confirm results observed in the HEK293T cell line, the splicing of specific cardiac genes was examined in a physiologically relevant cell type. This makes it possible to visualise the splicing pattern in a context close to the heart organ. In human cardiomyocyte cell line (AC16), 13 minigene constructs containing both wild-type and mutated minigenes were transiently transfected in three independent experiments and RT-PCR was carried out. The analysed minigene had been previously used to study the variant effect in HEK293T cell line. The chosen minigene constructs correspond to the genetic variants also evaluated in blood samples. This comparative analysis provides significant information regarding the regulatory mechanism of splicing into cardiology physiology and how it may modulate splicing events in CVDs.

3.4.1 Acceptor splice site

***DMD:* c.6118-3 C>A**

To confirm the previous data that c.6118-3 C>A variant in *DMD* gene did not induce splicing aberrations, a minigene assay was carried out in AC16 cells. As can be seen in figure 45, both wt and mutant constructs produced several transcripts of approximately size from 500bp to 320bp. The longest band, in both wild-type and mutant constructs, corresponds to 80bp intron retention caused by an activation of a cryptic splice site inside the fibronectin intron. The size of the faint bands (approximately 450bp), in mutant minigene, can suggest the activation of a cryptic splice site with a partial nt retention in the transcript. The other band of 420bp long, in both wt and mutant minigene, corresponds with the correctly spliced transcript. Finally, the lowest band (300bp) in mutant construct due to the size can suggest the activation of a cryptic splice site inside the exon 43 of *DMD* gene. Sanger sequencing should be performed to confirm the hypothesis.

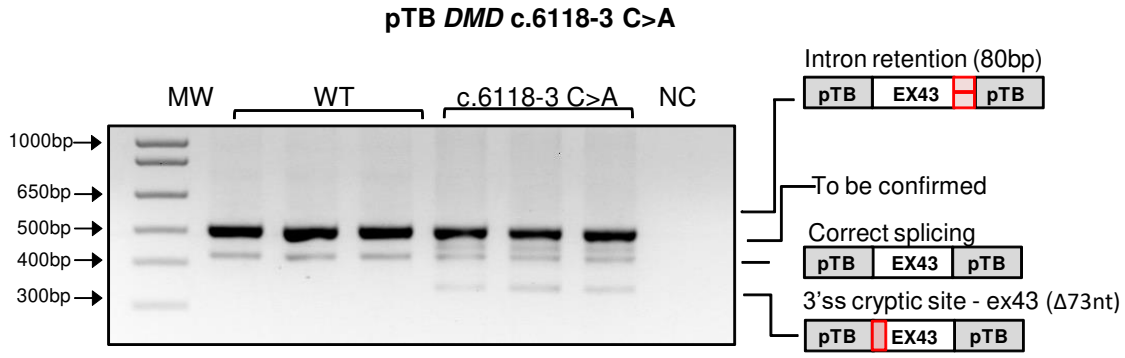


Figure 45. Minigene analysis of pTB *DMD* c.6118-3 C>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wild type (WT) and mutant minigenes. A schematic representation of the transcripts is reported to the right of the gel image. MW, molecular weight; NC, negative control.

***FLNC:* c.3791-1 G>A**

To validate previous findings regarding the impact of the c.3791-1G>A variant in *FLNC* gene on splicing, a minigene assay was conducted in AC16 cells. As can be seen in figure 46, a consistent pattern of bands comparable to that observed HEK293T cells was obtained. RT-PCR revealed the presence of many transcripts of approximately size from 750bp to 250bp. Therefore, the wt minigene construct produced the correctly spliced transcripts, whereas the minigene carrying the variant was found to disrupt the canonical 3' ss, resulting in the exclusion of exon 22 (174bp) from the transcript. Additionally, in AC16 cells, supplementary bands corresponding to construct artefacts were identified. Despite the apparent complexity of the

results, these data suggest the potential pathogenicity of the variant. Further specific investigations may be desirable, as the multiplicity of bands obtained may be attributed to the size of the insert cloned into the minigene construct.

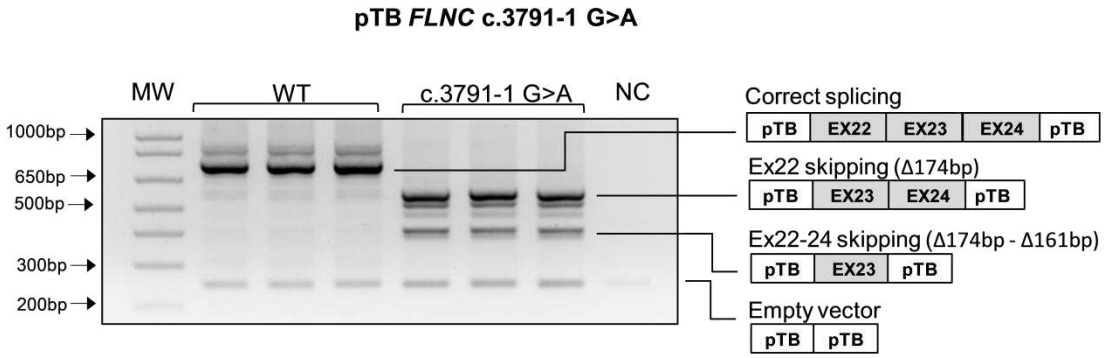


Figure 46. Minigene analysis of pTB *FLNC* c.3791-1 G>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is reported to the right of the gel image. MW, molecular weight; NC, negative control.

***LMNA:* c.1609-1 G>A**

The minigene pTB *LMNA* carrying the variant c.1609-1 G>A was previously observed to induce exon 10 skipping in the HEK293T cell line. Results after performing RT-PCR, indicated a similar aberrant splicing profile in the wt and mutant minigene constructs (Figure 47). This data confirmed the pathogenicity of c.1609-1 G>A variant.

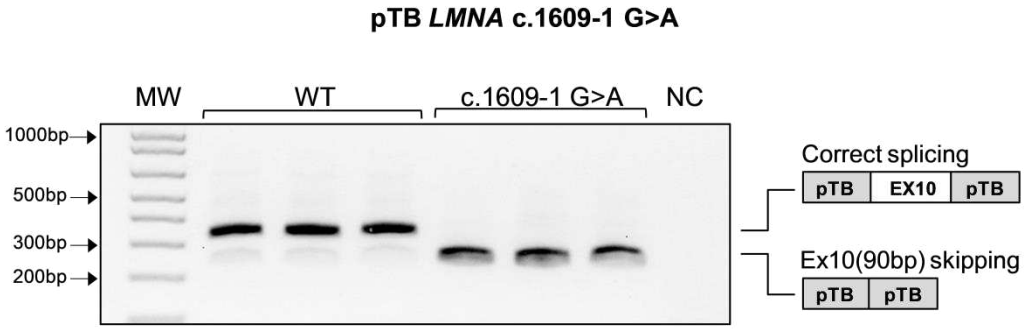


Figure 47. Minigene analysis of pTB *LMNA* c.1609-1 G>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is reported to the right of the gel image.

***TTN:* c.41609-2 A>C**

The minigene pTB *TTN* carrying the variant c.41609-1 A>C which disrupts the canonical AG acceptor splice site was included in the splicing profile analysis in AC16 cell line. In line with the outcome observed in HEK293T cells, in which the variant induces the disruption of the

canonical acceptor splice site and the activation of a cryptic 3'ss inside the first nt of exon 227, the same splicing outcome was observed after performing RT-PCR from RNA extracted from the AC16 cell line (Figure 48).

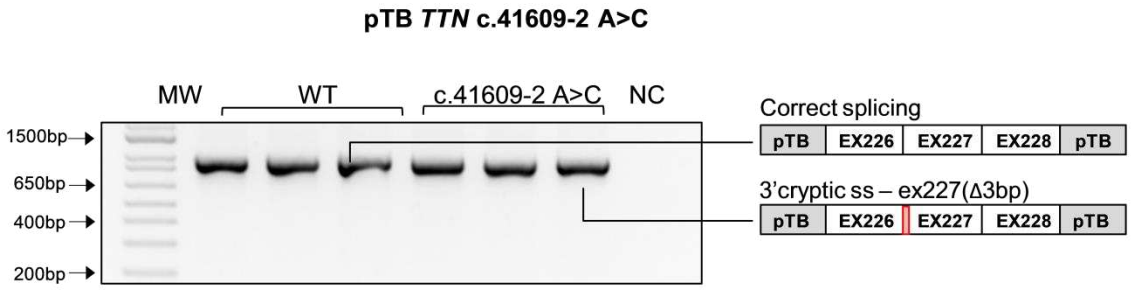


Figure 48. Minigene analysis of pTB *TTN* c.41609-2 A>C in AC16 cells. Agarose gel of RT-PCR from AC16 transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

3.4.2 Donor splice site

***DSP:* c.273+5 G>A**

The *DSP* c.273+5 G>A variant analysed using the minigene assay demonstrates that the variant induces splicing disruption by activating a cryptic 5'ss, 62nt downstream of the canonical GT site. As can be seen in figure 49, the mutant minigene results in two transcripts of approximate size 300bp and 330bp, with the300bp band comparable with the single transcript observed in the wt constructs. This splicing pattern was in accordance with the splicing outcomes observed in HEK293Tcells.

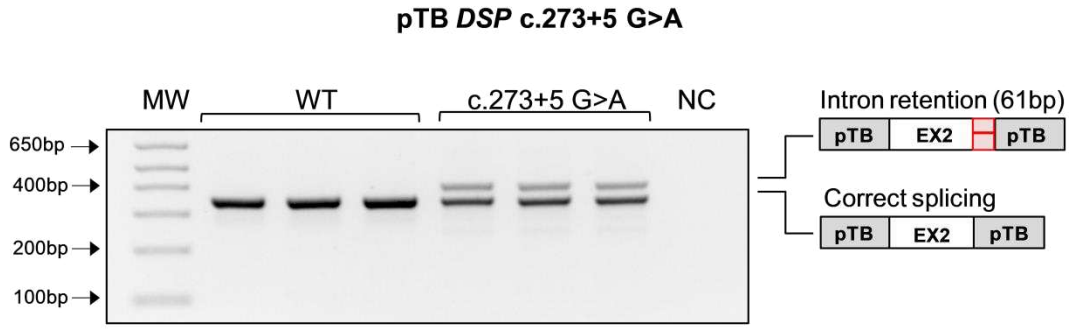


Figure 49. Minigene analysis of pTB *DSP* c.273+5 G>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

***FLNC:* c.7251+1 G>A**

A minigene assay was performed in AC16 cells to confirm earlier results about the effect of the c.7251+1 G>A variant in the *FLNC* gene on splicing. Figure 50 illustrates the consistent band pattern that was achieved, which is similar to the reported HEK293T cells. A large number of transcripts, ranging in size from 600bp to 250bp, were detected by RT-PCR. Consequently, it was confirmed that the minigene carrying the variation disrupted the canonical 5'ss, resulting in the exon 43 (116bp) skipping from the transcript, whereas the wt minigene construct produced the correctly spliced transcripts. Furthermore, additional bands that correspond to minigene artefacts were detected in AC16 cells. These data indicate the potential pathogenic role of the variant. Therefore, the size of the insert cloned into the minigene construct may have an impact on the number of bands obtained, so further focused analysis may be needed.

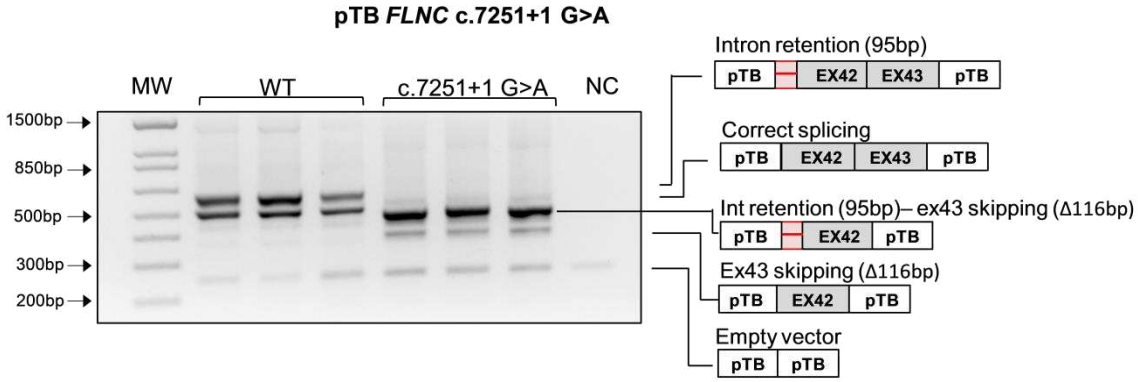


Figure 50. Minigene analysis of pTB *FLNC* c.7251+1 G>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

***NEXN:* c.1053+1 G>A**

The minigene pTB *NEXN* c.1053+1 G>A was shown to disrupt the canonical 5'ss and induce splicing disruption by exon 9 skipping and activation of a cryptic 5'ss in HEK293T cells. After transfection of both wt and mutant constructs in the AC16 cell line, the same aberrant splicing event previously identified was observed (Figure 51).

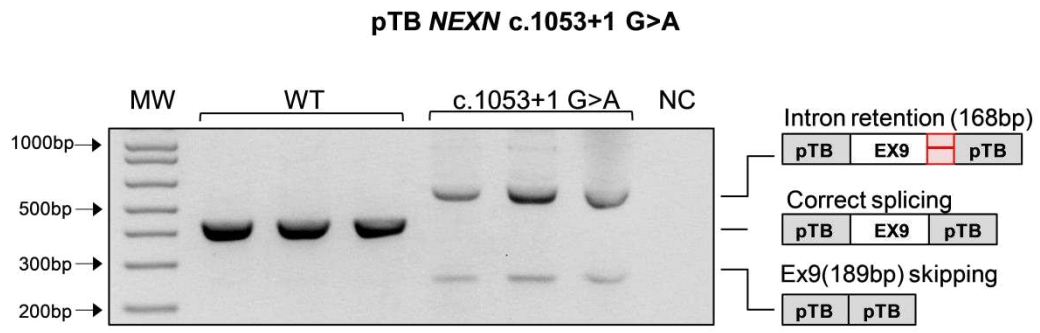


Figure 51. Minigene analysis of pTB *NEXN* c.1053+1 G>A in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

***TTN:* c.669+4 T>C, c.19501+2 T>C, and c.57847+4delGTAA**

The splicing pattern of the c.699+4 T>C variant was assessed in AC16 cells through the transfection of both wt and mutant minigene constructs. Previous analysis in HEK293T cells indicated an absence of mis-splicing events associated with this variant. Considering the size of the *TTN* exons and the primer's position, the correctly spliced transcript should be 886bp in size. However, the generation of multiple transcripts in both constructs was observed after RT-PCR (Figure 52). The upper bands correspond to the size of the expected transcripts, while a band approximately 20bp shorter was not previously observed. Notably, a band around 250bp corresponds to the empty vector, while the lowest band suggests primer duplex. These outcomes are not in accordance with the splicing events detected in the HEK293T cell line, where only one band representing the correctly spliced transcripts was observed. Further, analysis should be performed to clarify the nature of the detected transcripts in AC16 cells.

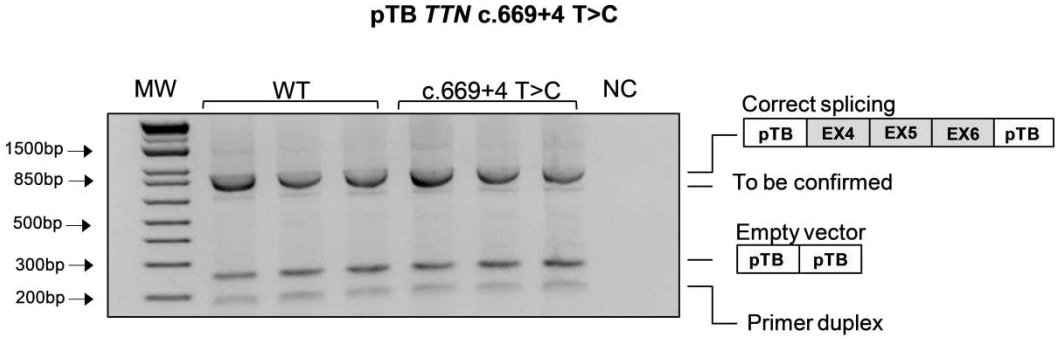


Figure 52. Minigene analysis of pTB *TTN* c.669+4 T>C in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

The c.19501+2 T>C variant was analysed by transfecting both wt and mutant minigene constructs into the AC16 cell line. In the wt minigene construct, two transcripts were observed, which are consistent with the results from the HEK293T cell. In particular, the upper band was related to minigene e artefacts, whereas the lower band (500bp) corresponded to the correctly spliced transcript. On the other hand, analysis of the splicing pattern from the mutant minigene detected an additional band of approximately size of 550bp, previously not identified in HEK293T cells. While the other three bands correspond to the activation of a cryptic 5'ss, a minigene artefact, and the empty vector, respectively. Sanger sequencing needs to be performed in order to define the nature of the additional observed band. Based on the data, to possibly eliminate artefact bands, it may be advantageous to modify the insert for cloning into the vector.

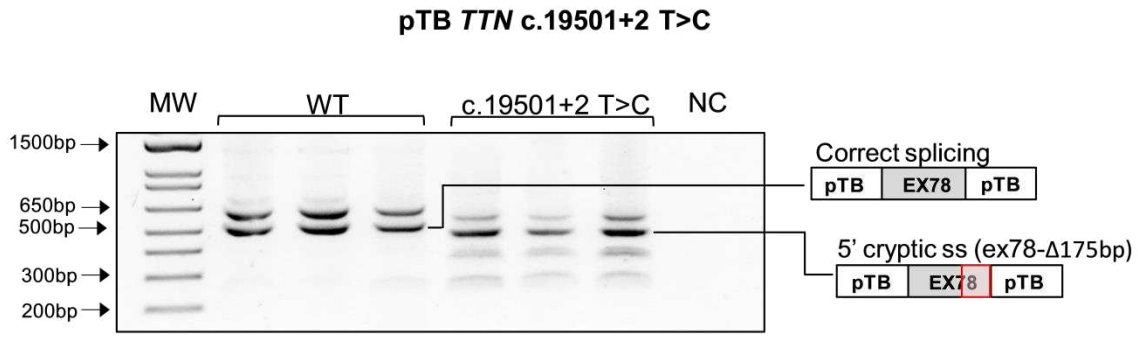


Figure 53. Minigene analysis of pTB *TTN* c.19501+2 T>C in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

The c.57847+4delGTAA variant was analysed by transfecting the wt and mutant minigene constructs into the AC16 cell line. As can be seen in Figure 54, the wt construct produces one single transcript of the expected size of 500bp, which corresponds with the canonical transcript. In contrast, the mutant minigene generated three different bands which demonstrates the splicing aberration caused by the presence of the variant. The splicing pattern is in accordance with the splicing events detected in the HEK293T cell line.

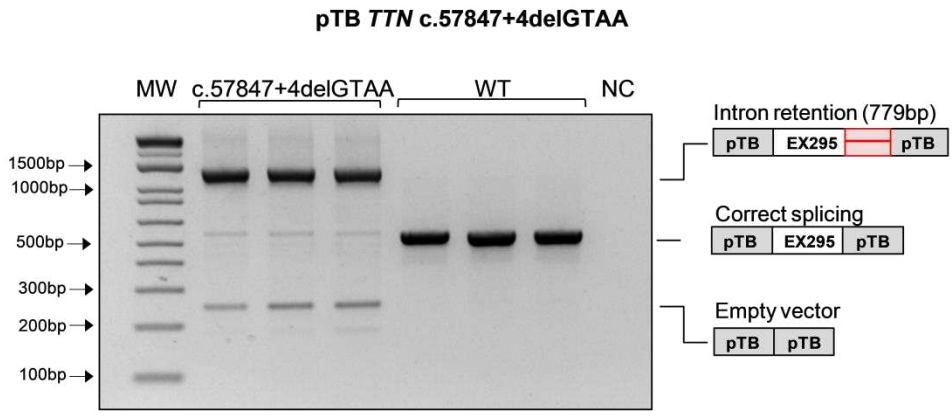


Figure 54. Minigene analysis of pTB *TTN* c.57847+4delGTAA in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

***VCL:* c.622+4 C>G**

Thec.622+4 T>C variant located in exon 5 of *VCL* showed no significant impact on mRNA processing in HEK293T cells, and a similar splicing pattern was observed after transfection of both wt and mutant minigene in AC16 cell line. As shown in Figure 55, two different transcripts were observed in both minigenes of approximate size 400bp and 500bp. The upper bands correspond with the activation of a cryptic donor splice site inside the fibronectin intron. In contrast, the lower band was the 370bp correctly spliced transcript.

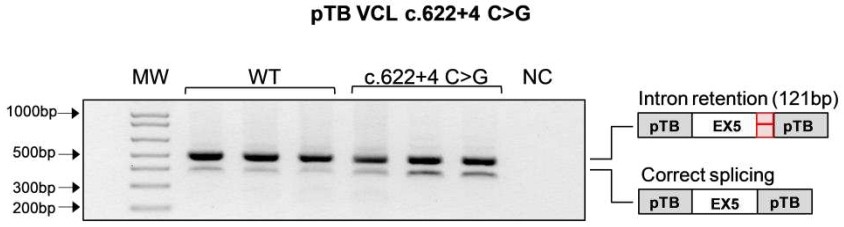


Figure 55. Minigene analysis of pTB *VCL* c.622+4 C>G in AC16 cells. Agarose gel of RT-PCR resulting from AC16 cells transfected with both wt and mutant minigenes. A schematic representation of the transcripts is presented to the right of the gel image.

3.5 RNA analysis of splicing variants from patients' blood

The patients with CVD diagnosis with identified genetic variants as reported in table 15, were invited to participate in the study. A total of 13 participants agreed to venous sampling to allow direct investigation of splicing profiles of the relevant gene in RNA extracted directly from blood. Of this,10 (77%) were diagnosed with an inherited dilated cardiomyopathy, and 3 (23%) with an inherited arrhythmia syndrome. Most of the identified variants had a classification as VUS (9,

69%), while three participants (23%) had variants with a classification of LP, and only one variant (8%) was classified as Pathogenic. Participants were primarily European ethnicity (13, 100%), of whom 6 were female (46%), and 7 male (54%) with 54% having a family history of cardiac condition or sudden cardiac death, and 46% with no family history.

Characteristics Participants	***n*** **(%)**
Sex	
Female	6 (46%)
Male	7 (54%)
Ethnicity	
European	13 (100%)
Age category, years	
14-25	5 (38%)
26-35	0
36-45	4 (31%)
46-55	4 (31%)
Diagnosed heart condition	
Hypertrophic cardiomyopathy (HCM)	0
Dilated cardiomyopathy (DCM)	10 (77%)
Left ventricular non-compaction (LVNC)	0
Arrhythmogenic left ventricular cardiomyopathy (ALVC)	3 (23%)
Long QT syndrome (LQTS)	0
Sudden cardiac arrest (SCA)	0
Family history of the condition or sudden cardiac death	
Yes	7 (54%)
No	6 (46%)
Type of classification	
Path	1 (8%)
LP	3 (23%)
VUS	9 (69%)
LB	0

Table 15. Cohort characteristics. Path, pathological; LP, Likely pathogenic; VUS, Variant with unknown significance; LB, likely benign.

From the 24 genetic variants analysed using minigene assay, a cohort of 13 individuals related to a diagnosis of HCM or DCM were recruited to collect total blood. Along with 7 healthy controls. Direct RNA analysis from blood offers the opportunity to directly characterise splicing events in affected individuals *in vivo*. Different RNA extraction techniques were analysed and compared prior to performing RNA analysis from patients' whole blood. First, blood samples were collected in EDTA tubes and RNA was preserved with or without RNA Later. The setup of the RNA extraction method was performed under different conditions: phenol/chloroform precipitation or using the RNAeasy mini kit (QIAGEN). The results showed that the highest RNA yield was obtained by a procedure that combined lysis in RBC buffer following RNA extraction kit

(QIAGEN), when RNA extraction was performed on the day of sample collection in EDTA tubes. In addition, to increasing the sensitivity of gene detection and avoiding haemoglobin RNA (hgbRNA) contamination from red blood cells, RNA was treated with DNase Turbo (Invitrogen) and subsequently purified.

Three techniques, RNA Later, EDTA tubes, and PAXgene tubes, were compared for long-term storage of whole blood samples. The most efficient technique for RNA preservation was determined to be to use PAXgene tubes before performing RNA extraction from frozen blood.

Initially, the expression of the reported genes in blood was checked using the GTEx Portal, the median level of gene expressed as Transcripts Per Milion (TPM) is reported in table 16. Then, RNA isolated from control and patient blood samples was subjected to RT-PCR, RT-qPCR, and RNA-Seq. The transcripts detected in blood were compared to human heart RNA, which was used as a control.

GENE	VARIANT	Median TPM
DMD	c.6118-3 C>A	0.03211
DSP	c.273+5 G>A	0.2894
FLNC	c.3791-1G>A c.7251+1 G>A	0.3038
LMNA	c.-1 C>A c.1609-1 G>A	6.029
NEXN	c.1053+1 G>A	0.5512
TTN	c.669+4 T>C c.19501+2 T>C c.32875+2 T>C c.41609-2A>C c.57847+4delGTAA	0.2851
VCL	c.622+4 C>G	18.58

Table 16. Lists of genes analysed directly from patients' blood samples. The variants of interest are reported as well as the median TPM value from GTEx Portal.

3.5.1 Splicing profile analysis from total blood by RT-PCR

***DMD:* c.6118-3 C>A**

The hemizygous *DMD* c.6118-3 C>A variant, was established to produce normal splicing after performing minigene assay. To confirm the outcome this variant was analysed by RT-PCR on RNA directly extracted from blood. The primers were designed to amplify the genomic region from exon 41 to exon 45. However, no amplification signals were obtained in blood samples. However, in the human heart RNA sample amplification was observed, with the band corresponding with the expected size (683bp) (Figure 56). Based on these data, the blood tissue cannot be used to provide information for variants located on *DMD* gene due to the low gene expression level (see Table 15).

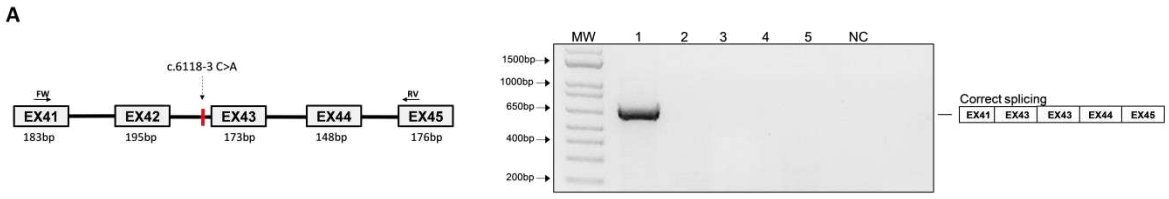


Figure 56. *DMD* c.6118-3 C>A variant splicing analysis from blood. (**A**) Schematic representation of *DMD* gene showing the region of the c.6118-3 C>A variant (red line). FW and RV indicate the location of the primers used for the amplification. On the right agarose gel electrophoresis of the resulting cDNA amplification. 1. Human heart RNA; 2. CRT1; 3. CRT2; 3. CRT3; 4. CRT4; 5. Patient; NC, negative control. A schematic representation of the resulting transcript is presented to the right of the gel image.

***FLNC:* c.3791-1 G>A**

The *FLNC* c.3791-1 G>A variant was previously observed by minigene assay to induce mis-splicing events with exon 22 skipping. The splicing pattern was investigated by performing RT-PCR directly from blood RNA. The primers were designed to amplify the genomic region from exon 21 to exon 23 (Figure 57A). However, no amplification signals were obtained in blood samples except from the human heart RNA sample, in which the band corresponds with the expected size (683bp) (Figure 57A). Based on these data, the blood tissue cannot be used to provide information for variants located in *FLNC* due to the low gene expression level in blood (see Table 15).

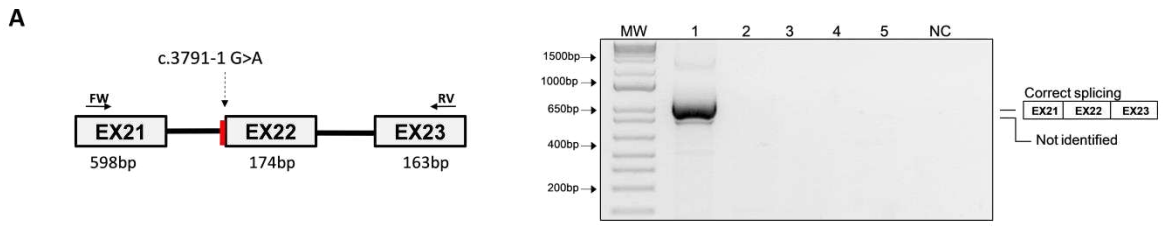


Figure 57. *FLNC* c.3791-1 G>A variant splicing analysis from blood. (**A**) Schematic representation of *FLNC* showing the region of the c.3791-1 G>A variant (red line). FW and RV indicate the location of the primers used for the amplification. On the right agarose gel electrophoresis of the resulting cDNA amplification. 1. Human heart RNA; 2. CRT1; 3. CRT2; 3. CRT3; 4. CRT4; 5. Patient; NC, negative control. A schematic representation of the resulting transcript is presented to the right of the gel image.

***LMNA:* c.1609-1 G>A**

Direct RNA analysis was also performed for the *LMNA* c.1609-1 G>A variant due to the availability of patient blood. RT-PCR was conducted on the day of sample collection and amplification of the region spanning from exon 8 to exon 11 was performed using the primers indicated (Figure 58A). Analysis of amplicons by agarose gel electrophoreses showed two transcripts in the patient sample compared to one in the controls (Figure 58A). DNA sequencing of the amplicons revealed wild-type splicing, as well as a smaller band corresponding to an amplicon in which exon 10 is skipped in the mutant allele (Figure 58B). The sequencing data confirmed the results observed by minigene assay.

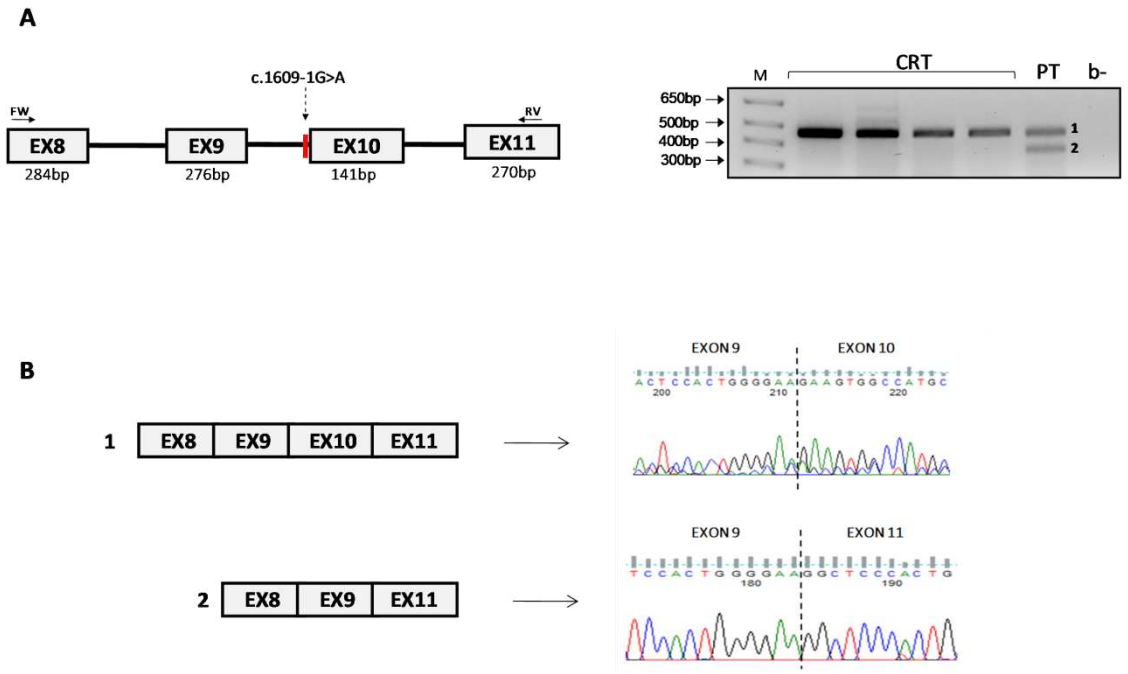


Figure 58. *LMNA* c.1609-1G>A variant splicing analysis from blood. (**A**) Diagram of *LMNA* showing the region of the c.1609-1 G>A variant. FW and RV indicate the location of the primers used in the amplification. On the right agarose gel electrophoresis of amplicons from control blood and from the patient *FLNC* c.1609-1 G>A variant. The upper band indicates the wild-type transcript (band1) and the lower band indicates exon 10 skipping (band 2). (**B**) Sequencing analysis of the PCR products.

***TTN*: c.41609-2 A>C**

To further confirm the functional impact of *TTN* c.41609-2 A>C, analysis was also performed using RNA extracted from a blood sample. Amplification of the region spanning exons 226-228 (Figure 59A) revealed the presence of one cDNA fragment in the patient carrying the substitution. DNA sequencing of the band showed a double peak in the region where the variant occurs indicating that this derives both from the mutant and wild-type transcript. To confirm the hypothesis, the PCR products were subsequently subcloned into pGEM-T easy vector and used to transform bacteria. 10 clones from each amplification were then sequenced using the universal oligo T7 primer. The result of sequencing of the patient sample showed the presence of the two cDNA fragments confirming previous results (Figure 59B).

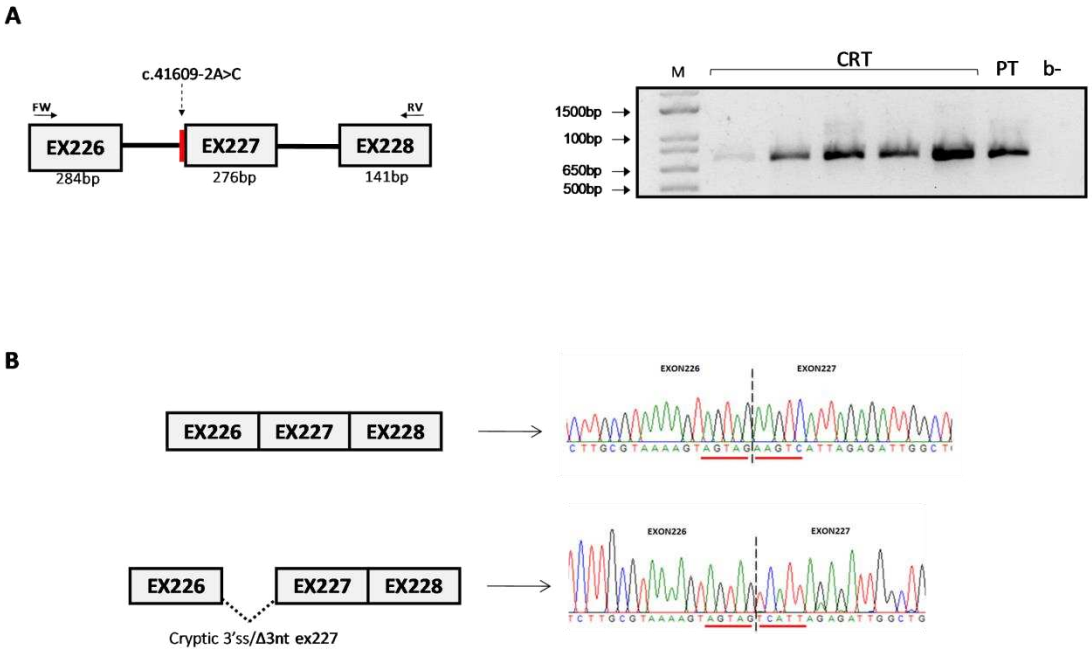


Figure 59. *TTN* c.41609-2 A>C variant splicing analysis from blood. (**A**) Diagram of *TTN* showing the region of the c.41609-2 A>C variant. FW and RV indicate the location of the primers used in the amplification. On the right agarose gel electrophoreses of amplicons from control blood (left panel) and from the blood of the patient carrying the *TTN* c.41609-2 A>C variant (right panel). (**B**) Sequence analysis of the cloned band in pGEM-T easy system from patient.

***DSP:* c.273+5 G>A**

The same methods were used to analyse the variant *DSP* c.273+5 G>A. This variant is located +5bp downstream of the canonical donor site of exon 2 and was clinically classified as a VUS (Table 14) with a SpliceAI score of 0.22 indicating loss of the natural canonical site (Table 10). After RNA extraction from blood, PCR amplification was performed with specific primers encompassing the genomic region from exon 1 to exon 4. As seen in Figure 60, the variants did

not lead to aberrant splicing. The outcome is in contrast with the splicing profile obtained from the minigene assay, where abnormal splicing was detected (see section 3.3.2).

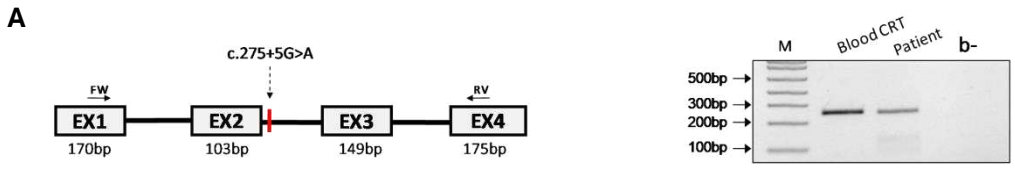


Figure 60. Molecular characterisation of the splicing effect of *DSP* c.273+5 G>A variant in blood. (**A**) Schematic representation of the *DSP* genomic region of interest amplified through PCR. The primers used are indicated as arrows. (**B**) RT-PCR from blood.

FLNC: c.7251+1 G>A

The variant *FLNC* c.7251+1 G>A is located at the donor site GT→AT and as this nucleotide is usually very highly conserved (see section 1.2.2.2), it is predicted to lead to aberrant splicing. Clinical classification was likely pathogenic, and the VUS had a 0.99 Δscore indicating disruption to the canonical splice donor site (Table 13 and 14, respectively). To test this hypothesis, RT-PCR was performed using specific primers which amplified the genomic region from exon 41 to exon 45 (Figure 61A). As shown in figure 61B, a variety of different sized bands were seen on electrophoresis agarose gel derived from the PCR products. DNA sequencing identified that the band with high intensity corresponds to the canonical transcript but further experiments to determine the results of the other bands will need to be undertaken.

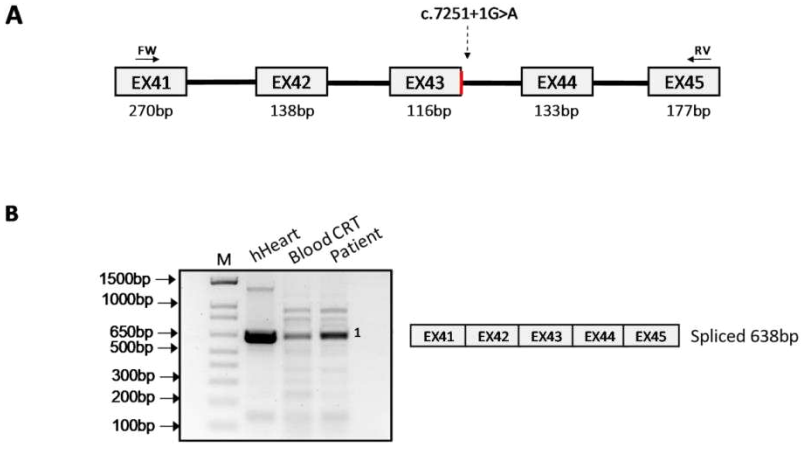


Figure 61. Molecular characterization of the splicing effect in blood of variant c.7251+1 G>A in *FLNC*. (**A**) A schematic representation of the *FLNC* genomic region of interest amplified through PCR. The primers used are indicated as arrows. (**B**) RT-PCR from blood.

***NEXN:* c.1053+1 G>A**

The heterozygous *NEXN* c.1053+1 G>A variant was classified as likely pathogenic clinically (Table 14) and is predicted to disrupt the natural donor site GT with a high SpliceAI score (0.98 Δscore). To evaluate how the variant impacts on splicing. Oligonucleotides were designed which amplified the human *NEXN* genomic region from exon 7 to exon 11 (Figure 62A). Amplified products were resolved by agarose gel electrophoresis. As shown in figure 47 A, no apparent difference in splicing between patient and control samples was seen. The bands were eluted from agarose gel and sequencing confirmed these results (Figure 62B).

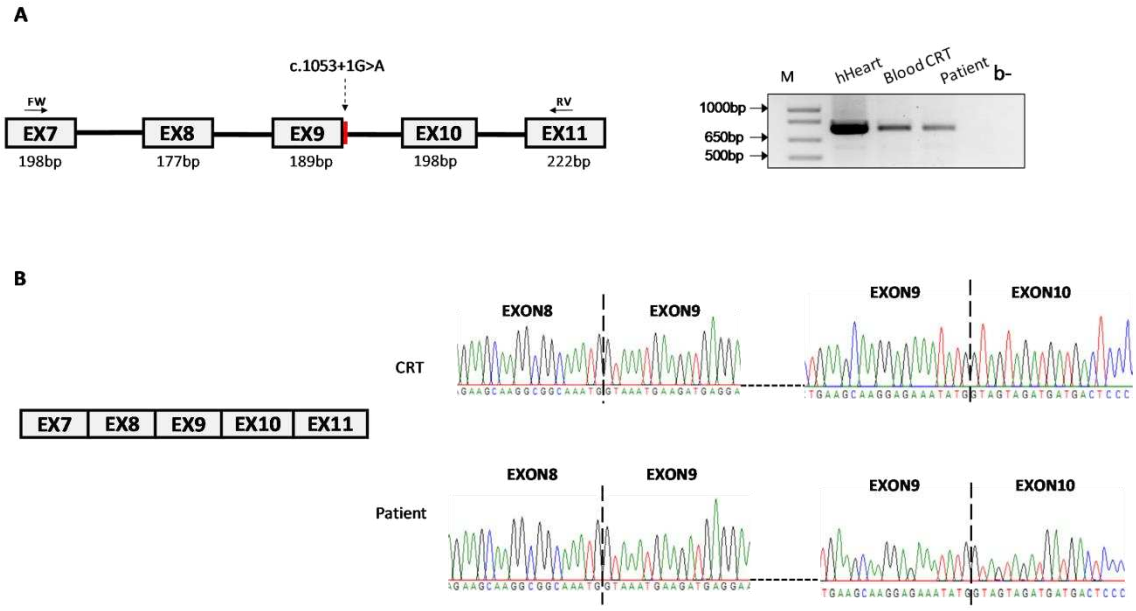


Figure 62. Molecular characterization of the splicing effect of *NEXN* c.1053+1 G>A variant in blood. (**A**) Schematic representation of the *NEXN* genomic region of interest amplified through PCR. The primers used are indicated as arrows. (**B**) Sequencing of PCR products from control and patient samples.

***TTN*: c.669+4 T>C, c.19501+2 T>C, c.32875 +2 T>C, and c.57847+4delGTAA**

The *TTN* variant c.669+4, previously analysed with a minigene assay and predicted to not affect splicing by SpliceAI (0.00 Δscore), was also tested by RT-PCR on patient whole blood. Total RNA was extracted on the day of collection and used for cDNA amplification PCR with oligonucleotides spanning the genomic region of interest from exon 4 to exon 6 (Figure 63A). PCR products were resolved on agarose gel. The expected band (size 645bp) corresponds to the upper band which is common to all samples. There is an aberrant lower band in the patient sample. PCR products were then sent for sequencing, confirming the upper band as the wild-type transcript, but the electropherogram for the lower band was of poor quality and difficult to

read. The results are interesting as they are different to those seen with the minigene assay where only one band was visible in both constructs.

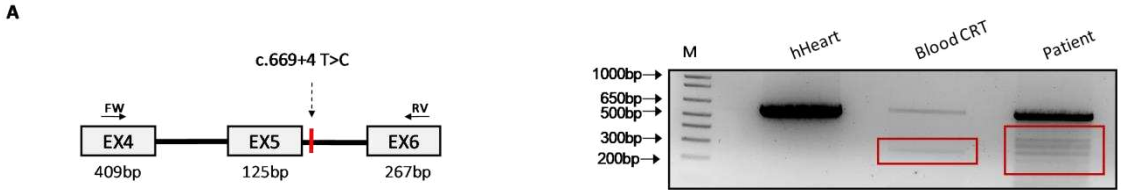


Figure 63. Molecular characterization of the splicing effect of *TTN* c.669+4 T>C from blood sample. (**A**) Schematic representation of the genomic region of interest amplified through PCR. The primers used are indicated as arrows. (**B**) RT-PCR from blood. One major band was identified in human heart control, common with the other two samples. In the patient sample additional bands are seen.

The variant of *TTN,* c.19501+2 T>C, was clinically classified as VUS and predicted to disrupt the canonical donor site (0.48 Δscore). RT-PCR analysis was performed from RNA isolated from whole blood. As shown in figure 64A, control samples generated a major transcript at 1374bp corresponding to the WT transcript. In contrast, the patient sample generated two products: the longer one corresponding to wild-type transcript and the shorter one originating from truncation of exon 78. DNA sequencing performed from DNA eluted from agarose gel identified that the 1183bp product corresponds to the canonical transcripts while the data from the lower product was not able to be read. Subsequently, the PCR products were cloned into pGEM T-easy vector and used to transform bacteria. 5 clones from each amplification were then sequenced using the universal oligo T7 primer (see section 2.10.3). The results of sequencing show the canonical transcript in blood control, while in patient clones 32 identical canonical transcripts were identified, with usage of a cryptic 5' splice site in exon 78 and with a deletion of 175nt, and another with usage of a cryptic donor site in exon 78 with a deletion of 57 nucleotides (Figure 64B).

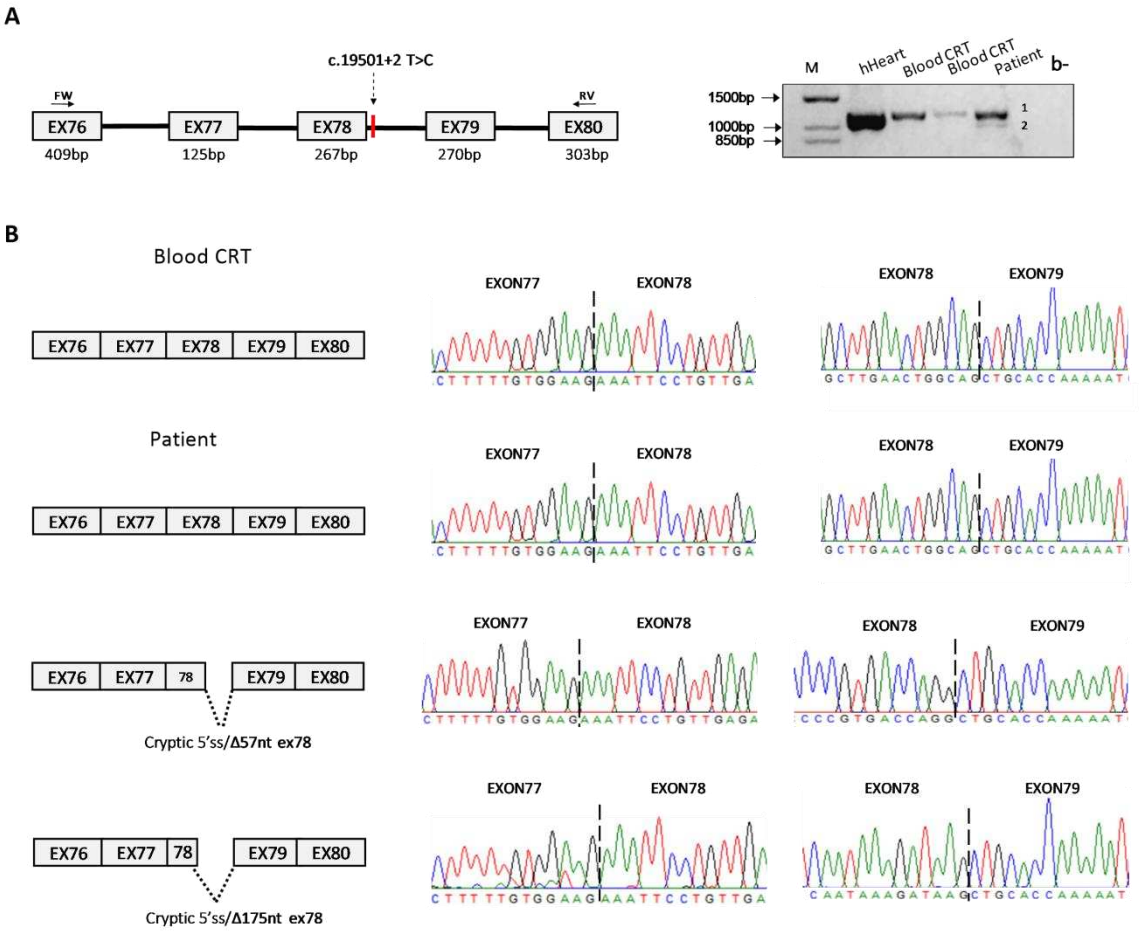


Figure 64. Molecular characterization of the splicing effect in blood of variant c.19501+2 T>C in *TTN* gene. (**A**) Schematic representation of *TTN* showing the region of the c.19501+2 T>C variant. FW and RV indicate primer location. Agarose gel is shown on the right. (**B**) Sequencing of PCR products from control and patient samples.

The variant of *TTN,* c.32875+2 T>C, was classified as VUS and predicted to disrupt the canonical donor site (0.88 Δscore). RNA analysis was performed from RNA isolated from whole blood. As shown in figure 65A, expected PCR amplicons of 430bp were obtained in control blood samples but not visualised in a patient sample. Therefore, a second band, common in all samples, of 400bp suggests a deletion of approximately 30nt. Subsequently, the PCR products were cloned into pGEM T-easy vector and used to transform bacteria. Sanger sequencing from bacterial clones should be performed to confirm the sequence of the transcripts, no conclusive data can be done.

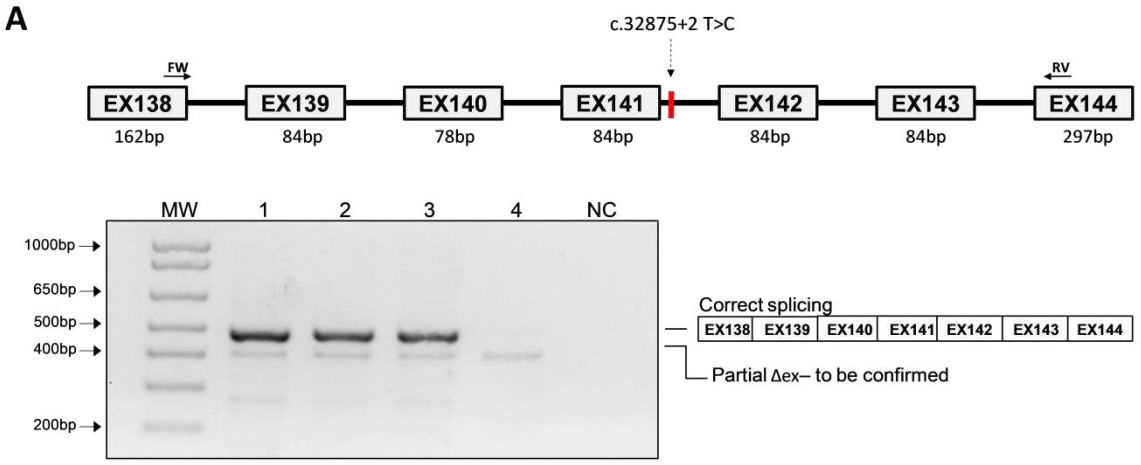


Figure 65. Molecular characterization of the splicing effect in blood of variant c.32875+2 T>C in *TTN* gene. (**A**) Schematic representation of *TTN* gene showing the region of the c.32875+2 T>C variant (red line). FW and RV indicate the location of the primers used for the amplification. On the bottom, agarose gel electrophoresis of the resulting cDNA amplification. 1. CRT1; 2. CRT2; 3. CRT3; 4. Patient; NC, negative control. A schematic representation of the transcript is presented to the right of the gel image.

The variant of *TTN,* c.57847+4delGTAA, was clinically classified as VUS and predicted to disrupt the canonical donor site (0.98 Δscore). Results from the minigene assay confirmed the pathogenicity of the variant which induced splicing abnormalities. Therefore, RT-PCR from RNA isolated from whole blood (controls and patient) showed a single transcript band of the expected amplicons of 720bp (Figure 66A). While in the human heart RNA sample, a lower band of approximately of 700bp was detected which may correspond with the other titin isoform expressed in the heart. Subsequently, the PCR products were cloned into pGEM-T easy vector and used to transform bacteria. The clones from each amplification were then sequenced using the universal oligo T7 primer (see section 2.10.3). The results of sequencing did not clarify the origin of the transcripts.

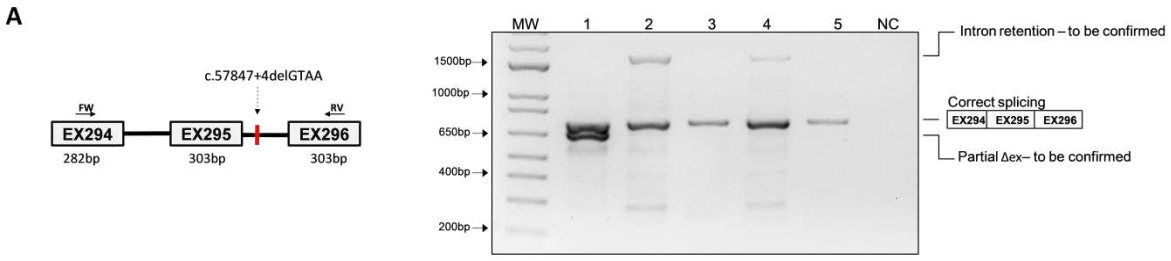


Figure 66. Molecular characterization of the splicing effect in blood of variant c.57847+4delGTAA in *TTN* gene. (**A**) Schematic representation of *TTN* gene showing the region of the c.57847+4delGTAA variant (red line). FW and RV indicate the location of the primers used

for the amplification. (**B**) Agarose gel of the resulting cDNA amplification. 1. Human heart RNA; 2. CRT1; 3. CRT2; 3. CRT3; 4. CRT4; 5. Patient; NC, negative control. It is indicated a schematic representation of the resulting transcript on the right of the gel.

***VCL:* c.622+4 C>G**

The heterozygous variant c.622+4 C>G identified in *VCL* was also tested in RNA isolated from whole blood of the patient followed by RT-PCR. RNA extraction was performed on the same day of blood collection according to the protocol reported in material and methods (section 2.15.2). The genomic region was amplified with specific primers (Figure 67A). PCR products separated on agarose gel showed one single transcript corresponding to the wild-type transcript in both control and patient samples (Figure 67A). The results identified in blood were in line with the findings in the minigene assay. The result was confirmed by DNA sequencing (Figure 67B).

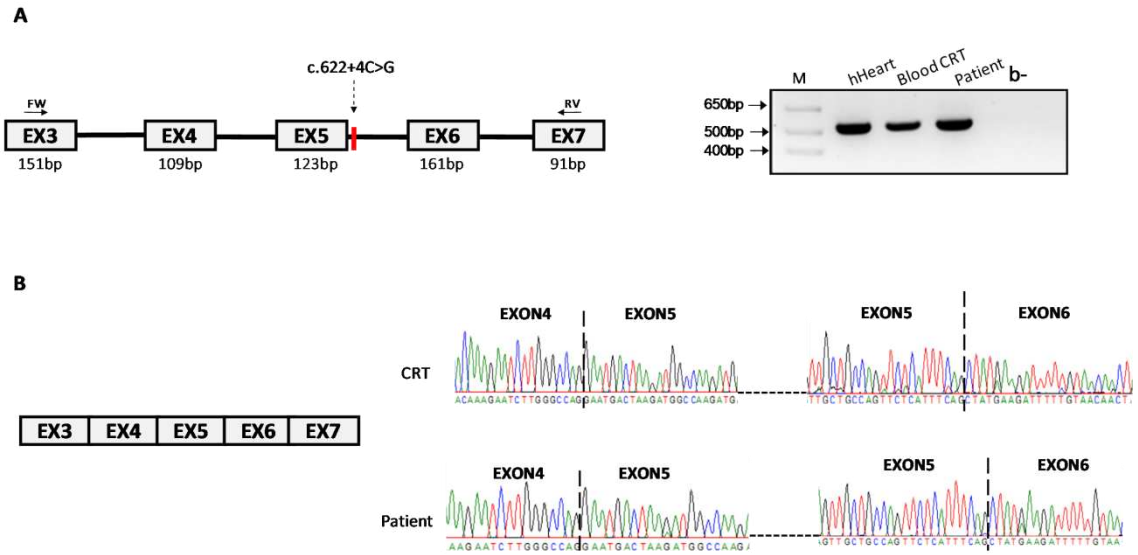


Figure 67. Molecular characterization of the splicing effect of *VCL* c.622+4 C>G from blood sample. (**A**) A schematic representation of the genomic region of interest amplified through PCR. The primers used are indicated as arrows. RT-PCR from blood is shown on the right. One major band was identified. M indicates the molecular weight marker. (**B**) Sequencing of PCR products.

3.5.2 Quantitative RT-PCR analysis of gene expression in blood

In several cases, the aberrant transcript isoforms resulting from the minigene analysis carrying the genetic variants were not detectable in the blood samples. This is likely because these isoforms result in NMD due to the insertion of a PTC as previously mentioned. To investigate this and quantify the levels of gene expression in samples that resulted in aberrant splicing, reverse transcriptase quantitative PCR (RT-qPCR) was performed in patient blood samples after RNA

isolation. The mRNA expression level of each gene was compared with 7 healthy individuals, which were used as a control group.

Before determining the gene expression in patient samples, a crucial step in RT-qPCR is the internal control gene choice. Three housekeeping genes (*GAPDH,HNRPT*, and *HuPo*) were selected after reviewed studies validating housekeeping genes in whole blood (Dheda *et al.*, 2004; Martínez-Sánchez *et al.*, 2019). The gene stability was evaluated using RefFinder, a comprehensive web-based tool that integrates the geNorm algorithm, which calculates the expression stability value (M) (Xie, Wang and Zhang, 2023). The M value is the average of the pairwise variation of a particular gene with all other candidate reference genes (Vandesompele *et al.*, 2002). The geNorm algorithm considered stable genes *GAPDH* and *HuPo*, which were selected to be used as reference genes (Figure 68). For gene expression of the target gene, the geometric mean of these two reference genes was used for normalization.

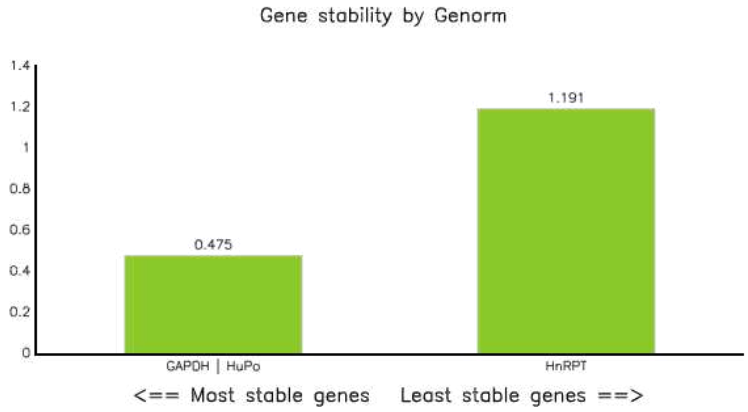


Figure 68. Stability of candidate housekeeping genes by geNorm. The stability value (M) is reported on y axis, the housekeeping genes were considered stable with less M value. The graph is reproduced from RefFinder (http://leonxie.esy.es/RefFinder/).

As reported in the analysis of *DSP* c.275+5 G>A, a mis-splicing event was observed in the minigene assay, which was not observed in the blood analysis. To confirm the hypothesis that NMD was occurring, the mRNA expression level of *DSP* gene was measured by RT-qPCR in patient blood compared to controls. As shown in Figure 69A, the *DSP* gene expression level displayed a decrease in expression level in the patient sample. *LMNA* gene expression levels in a patient carrying the mutation c.1609-1G>A were also quantified. In this case, the minigene carrying the genetic variant wasidentified in an isoform in which exon 10 was skipped in the transcript which was also observed in patient blood RT-PCR analysis. As depicted in figure 69B, *LMNA* mRNA, derived from the patient carrying the nucleotide substitution that disrupts the acceptor splice site (c.1609-1G>A) was greatly reduced. This result was unexpected, as exon 10 skipping results in a transcript that is in frame. This may indicate that the mRNA isoform without exon 10 could result in lower stability. Finally, the effect of c.1053+1G>A variant in the *NEXN*

gene was evaluated. RT-PCR of blood samples did not identify a mis-splicing event, and the minigene assay is currently ongoing. However, as the mutation disrupts the canonical donor splice site, it was assumed that the lack of aberrant splicing in the blood sample was likely due to NMD. Quantification of the mRNA levels confirmed that *NEXN* mRNA in the patient was reduced compared to the healthy group (Figure 69C).

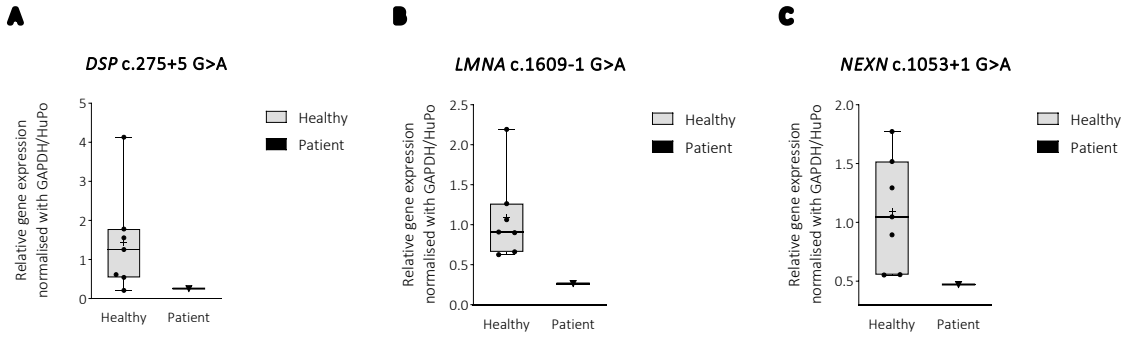


Figure 69. Box plot of genes validated by RT-qPCR. Grey bars represent healthy samples, back bars, and patient samples. The horizontal line in the middle of each box indicates the median, and the top and bottom border of the box mark the 75th and 25th percentiles, respectively. The whiskers above and below the box mark the interquartile range. Individual outliers are indicated with dots and triangles, healthy and patient, respectively. The bars beyond the whiskers are the outlier's maximum and minimum. *GAPDH* and *HuPo* were used as reference genes for normalisation. (**A**) RT-qPCR of *DSP* gene in control group and patient. (**B**) RT-qPCR of *LMNA* gene in the control group and patient. (**C**) RT-qPCR of *NEXN* gene in control group and patient.

3.6 RNA-seq analysis

3.6.1 Flowchart of RNA-seq data analysis pipeline

RNA sequencing (RNA-seq) is an emerging technology in clinical genetics for evaluating genetic variant effects and clarifying the effects related to specific nt substitutions through the entire genome. In this thesis, RNA-seq was evaluated for genetic variants predicted to impact pre-mRNA splicing. The pipeline designed for the experimental goal of the study is reported in figure 70, and consists of the following steps:

1. Quality control (QC) of the raw sequence reads by removing reads that are of poor quality or that do not map to the reference genome.

2. Reads alignment to the reference genome which determines the origin of the genes and/or transcripts.

3. Quantification of the aligned reads by counting how many reads mapped to that gene or transcripts.

4. Visualisation of spicing profile by Integrated Genome Viewer (IGV) which shows the read coverage and transcript isoform.

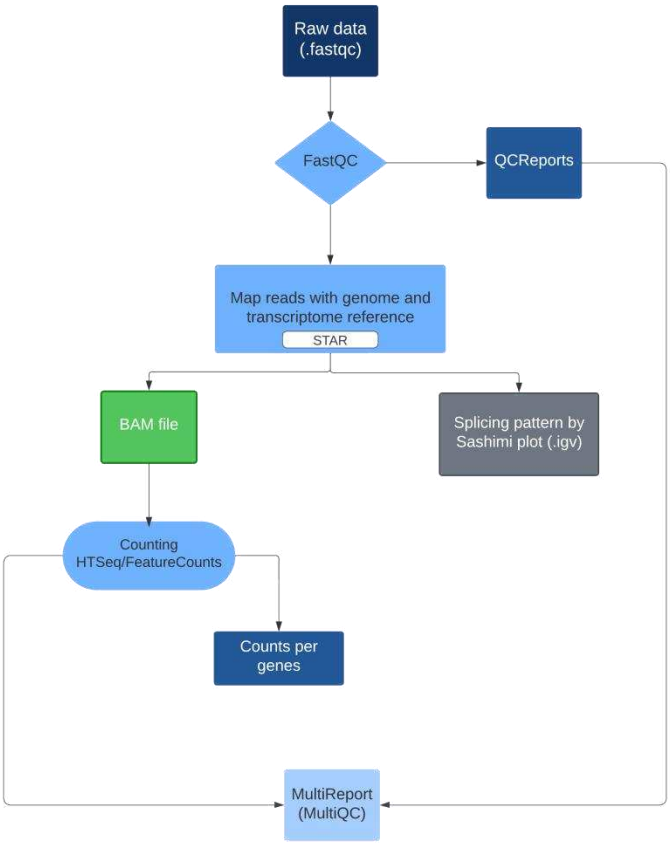


Figure 70. Flowchart of the RNA-seq analysis pipeline.

3.6.2 Quality and quantity of isolated RNA

Isolated RNA analysis of blood from healthy volunteers and patients was analysed by Novogene before bioinformatic analysis of the raw sequencing data.

3.6.3 MultiQC analysis on output files of computational tool

The average unique read percentage in the raw RNA-seq data was 44.9% (min value of 24.4% and max value of 56.9%) and the average duplicate read percentage was 55.1% (min value of 43.1% and max value of 74.6%). The mean base pair quality score in the sequence read as well as the mean sequence quality were above a phred score of 30, indicating high quality reads. Adapter content was below 5% and slowly increased at read position 100 (see Appendix A). The majority of reads were uniquely mapped after STAR alignment; the average percentage of uniquely mapped reads was 81.69% (min value of 57% and max value of 87%), with 12%

mapped to multiple loci (min value of 9.6% and max value of 22.5%); 0.135% were mapped to many loci (min value of 0.1% and max value of 0.4%), and 6.18% of the reads were unmapped (min value of 3.1% and max value of 22.6%) (see Appendix B). Average assigned read percentage to genomic features was 32,38% (min value of 14.6% and max value of 42.2%), 1.1% were ambiguously mapped (min value of 0.6% and max value of 1.3%), 35.95% are aligned but not unique (min value of 26.6% and max value of 685) and 30.52% are aligned but to no feature (min value of 16.8% and max value of 37.2%) (see Appendix C).

3.6.4 Visualisation of aberrant splicing profiles

Considering the ubiquitous expression of cardiac genes, short-read RNA sequencing (RNA-seq) was conducted on blood samples derived from affected individuals for detection of splicing impairment. The diagnostic utility of RNA-seq was evaluated for 13 cases and 4 healthy controls. For 3 of 13 cases, aberrant splicing was observed by blood RT-PCR. In contrast 9 of 13 cases identified by minigene assay to cause mis-splicing events. However, Sashimi plots were generated to visualise the region of the variants and splicing outcomes previously observed in blood.

For 4 cases (*DSP* c.273+5 G>C, *DMD* c.6118-3C>A, *FLNC* c.3791-1G>A, and *FLNC* c.7251+1G>A and), RNA-seq was not informative. The case of *DSP* gene could not be analysed using RNA-seq data because of the low read coverage in the first portion of the gene, as reported in figure 71A. While RNA-seq failed due to the absence of read coverage in the region of the gene of interest for *FLNC* and *DMD* gene (Figure 71B-D).

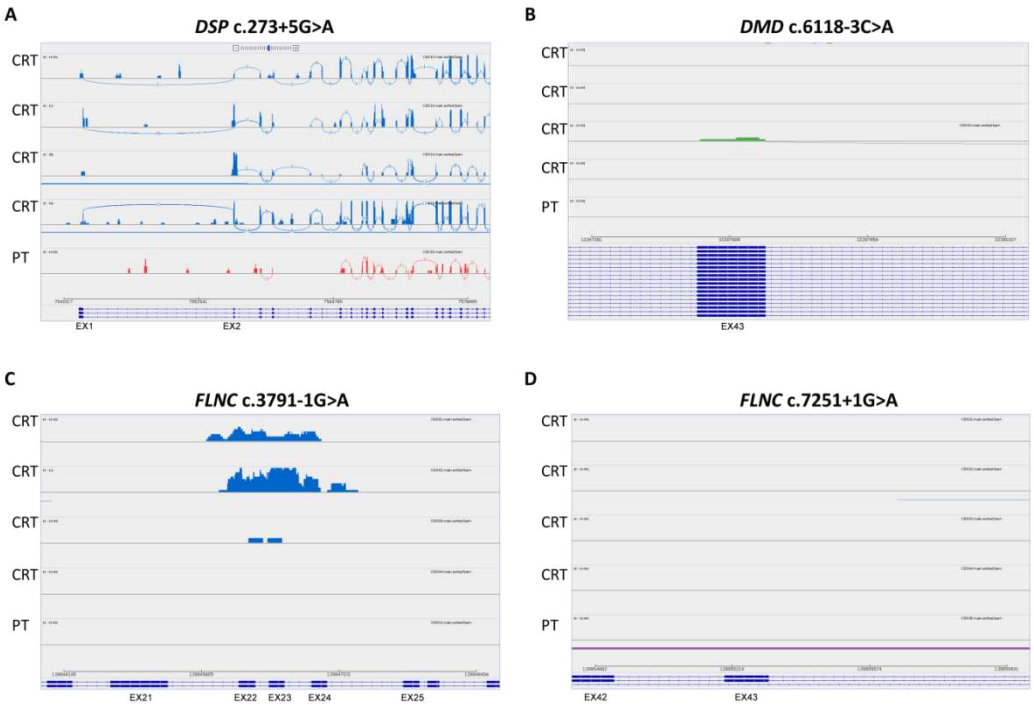


Figure 71. Visualisation of *DSP*, *DMD*, and *FLNC* genes by IGV. IGV-Sashimi plots of patient (PT) and controls (CRT) (n=4) are shown which represent no read coverage and the transcript isoforms of the respective genes. (**A**) Sashimi plot from *DSP* c.273+5G>C. (**B**) Sashimi plot from *DMD* c.6118-3 C>A. (**C**) Sashimi plot from *FLNC* c.3791-1 G>A. (**D**) Sashimi plot from *FLNC* c.7251+1 G>A.

The c.1609-1G>A variant in *LMNA* was previously known to induce splicing errors by RT-PCR from blood and minigene assay. The view of *LMNA* from the IGV showed clear evidence for mis-splicing, exon 10 skipping, in patient RNA-seq compared to the control samples (Figure 72A), as well as the corresponding variant was mapped to the 3' ss of exon 10 (Figure 72B).

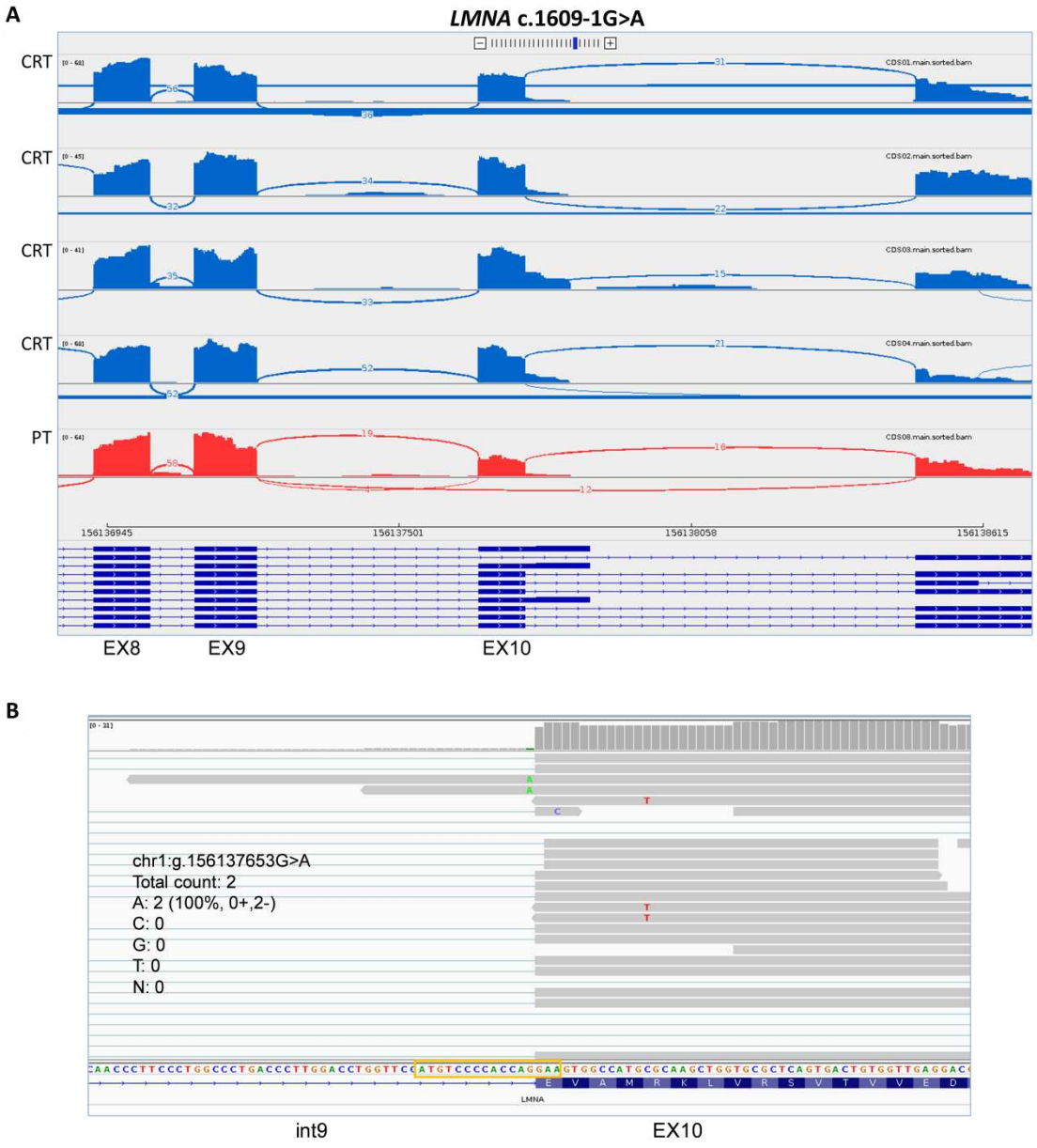


Figure 72. Visualisation of *LMNA* c.1609-1G>A by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by

the size of the sashimi plot curve. (**A**) Sashimi plot from *LMNA* c.1609-1 G>A. (**B**) The orange box highlights the intron-exon boundaries where the variant was mapped.

Another variant c.-1C>A in *LMNA* located in 5'UTR was investigated through RNA-seq analysis, where no different usage in exon-intron junction was observed in the region of the gene containing the variant, which was detected by reads mapping (Figure 73A-B).

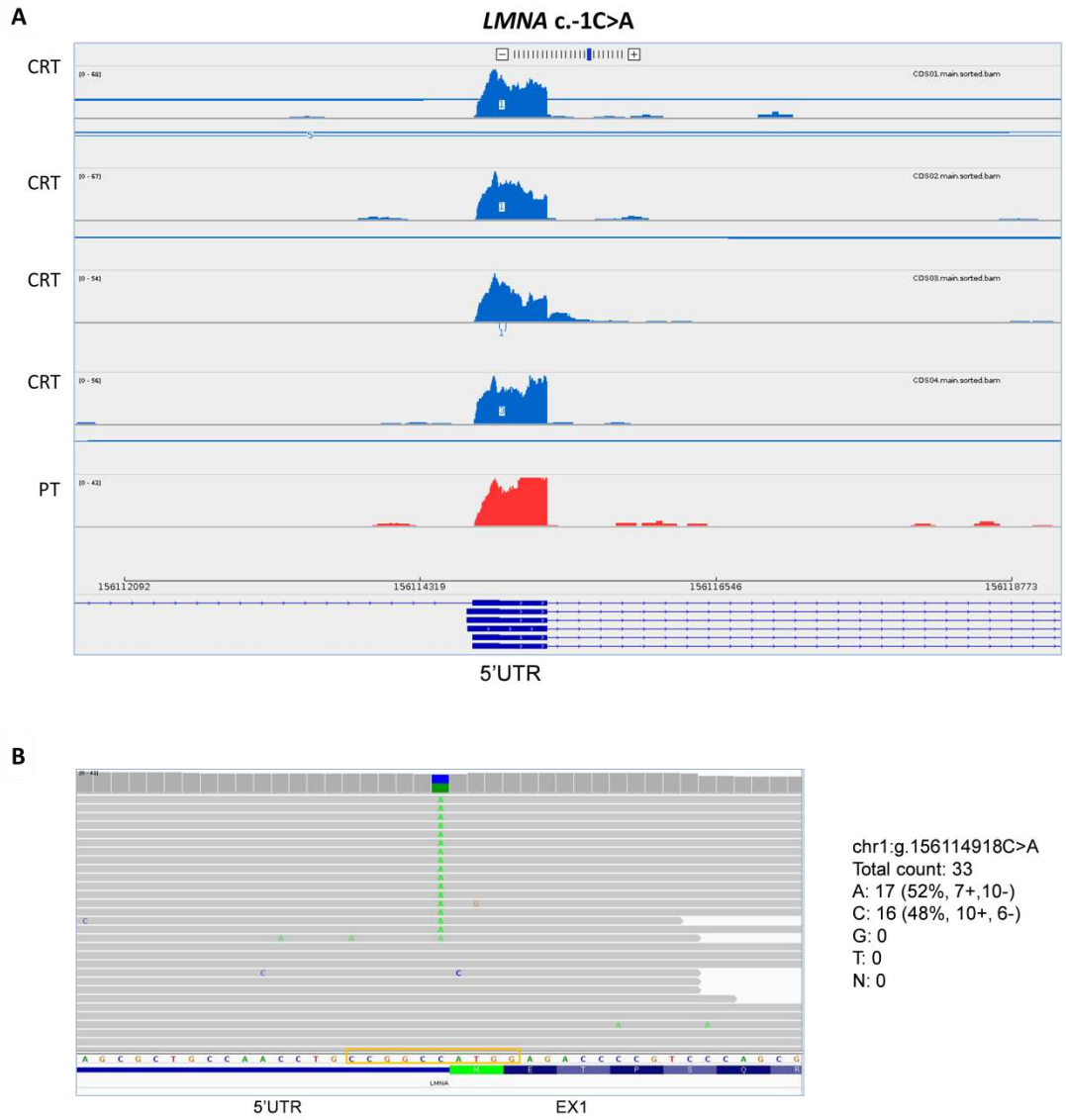


Figure 73. Visualisation of *LMNA* c.-1C>A by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *LMNA* c.-1C>A. (**B**) The orange box highlights the 5'UTR and exon 1 where the variant was mapped.

Interestingly, for *NEXN,* the mutation c.1053+1G>A disrupted the natural 5' ss and created a potential donor site inside the intron 9 which was reported to introduce a PTC. This event was also observed by analysing the gene region containing the variant from RNA-seq data (Figure

74A). RNA-seq identified the pathogenic variant located at the 5' ss in patient sample (Figure 74B).

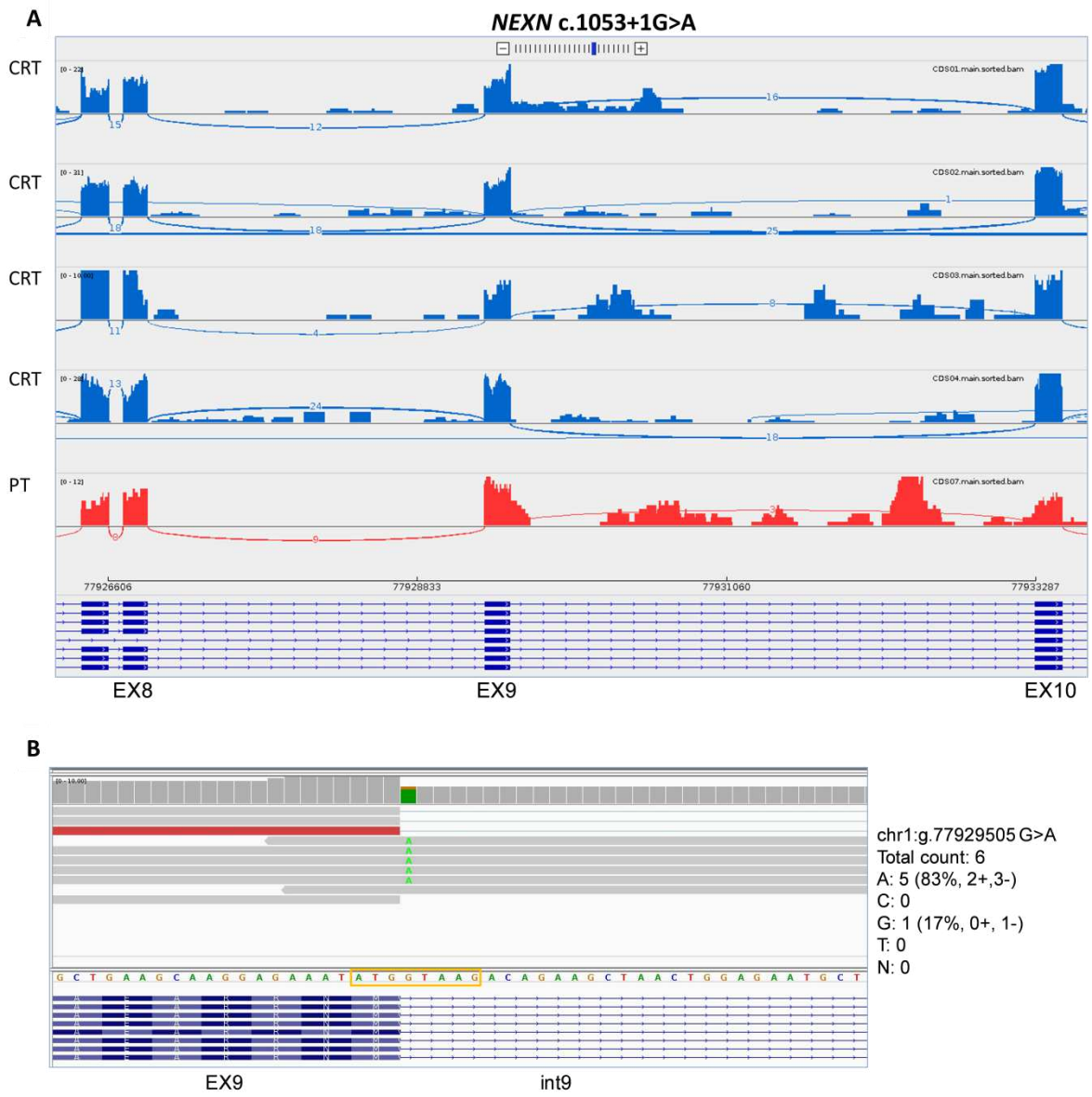


Figure 74. Visualisation of *NEXN* c.1053+1 G>A gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *NEXN* c.1053+1 G>A. (**B**) The orange box highlights the intron-exon boundaries where the variant was mapped.

Five variants in *TTN* gene (c.669+4 T>C, c.19501+2 T>C, c.32875+2 T>C, c.41609-2 A>C, and c.57847+4delGTAA), were analysed by RNA-seq. The c.669+4 T>C variant based on previous data was observed to not cause splicing abnormalities. RNA-seq sashimi plot of the interested region identified the correct usage of exon-exon splice junctions, as well as the interested variant was not detected in patient sample (Figure 75A-B).

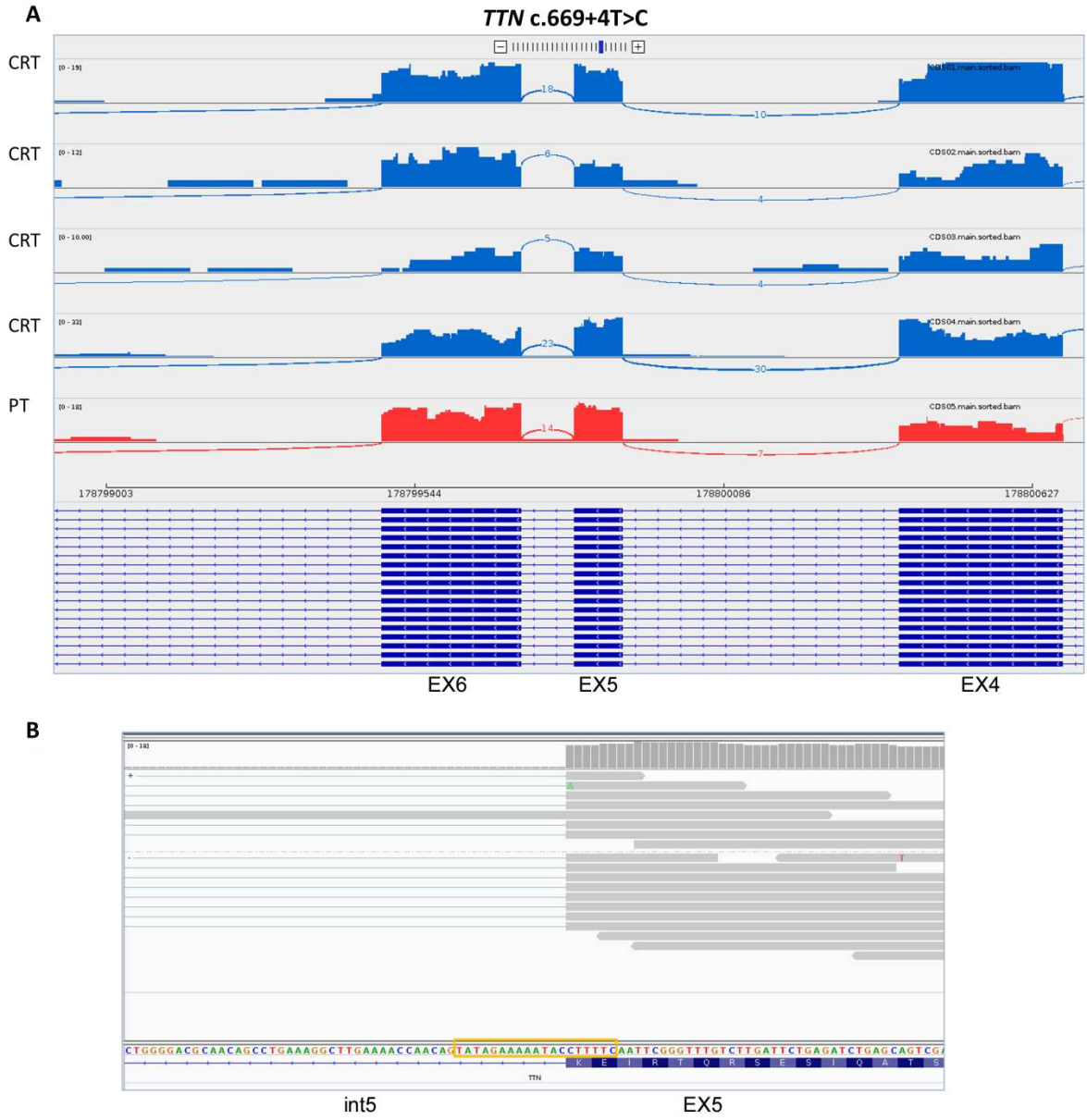


Figure 75. Visualisation of *TTN* c.669+4 T>C gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *TTN* c.669+4T>C. (**B**) The orange box highlights the intron-exon boundaries of interest.

The variant c.19501+2 T>C in *TTN,* was previously observed to activate two cryptic 5' donor sites inside exon 78 (57nt and 175nt upstream of the canonical GT), while minigene carrying the variant detected the activation of a cryptic 5' ss 175nt upstream the canonical 5'ss of exon 78. It is important to highlight that informative RNA-seq data was obtained for which the 5' cryptic site 57nt upstream the canonical GT is not related to the variant due to its presence in healthy control samples. Whereas in patient sample, RNA-seq confirmed the usage of a cryptic 5' ss 175nt upstream the natural donor site (Figure 76A). However, the variant of interest was

identified by reads mapping as can be seen in Figure 76B.

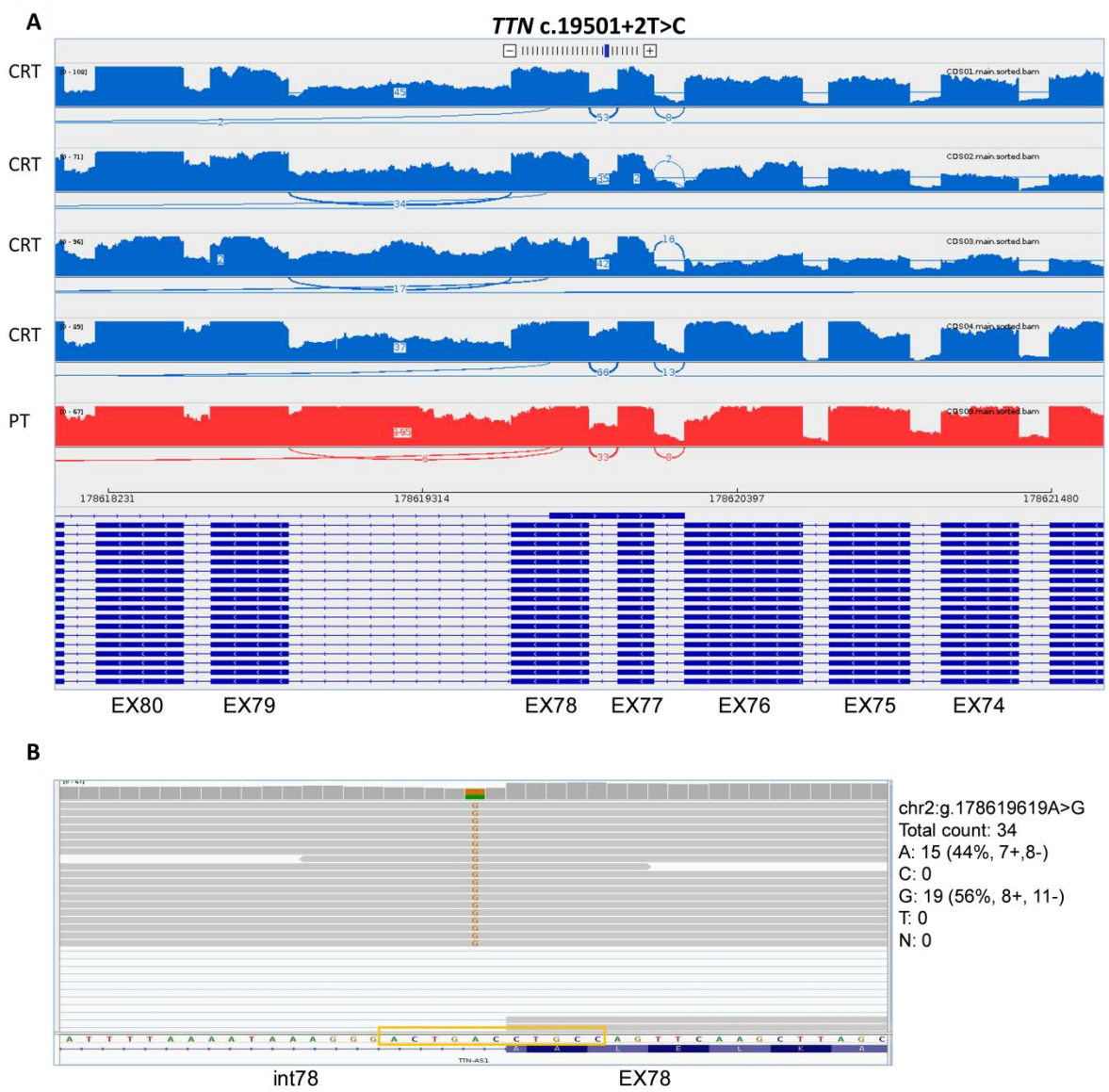


Figure 76. Visualisation of *TTN* c.19501+2 T>C gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *TTN* c.19501+2T>C. (**B**) The orange box highlights the intron-exon boundaries where the variant was mapped.

Regarding variant c.32875+2 T>C in *TTN*, minigene assay identified that the variant induces exon 141 skipping while data from RT-PCR was not conclusive, but a decreased mRNA expression was detected in patient's blood sample. Therefore, RNA-seq sashimi plot identified exon skipping in patient's sample compared to the healthy control (Figure 77A). The allele carrying the variant was not detected through the read coverage, and the splicing junction usage clearly identified the canonical mRNA splicing as well as exon 141 skipping (Figure 77B).

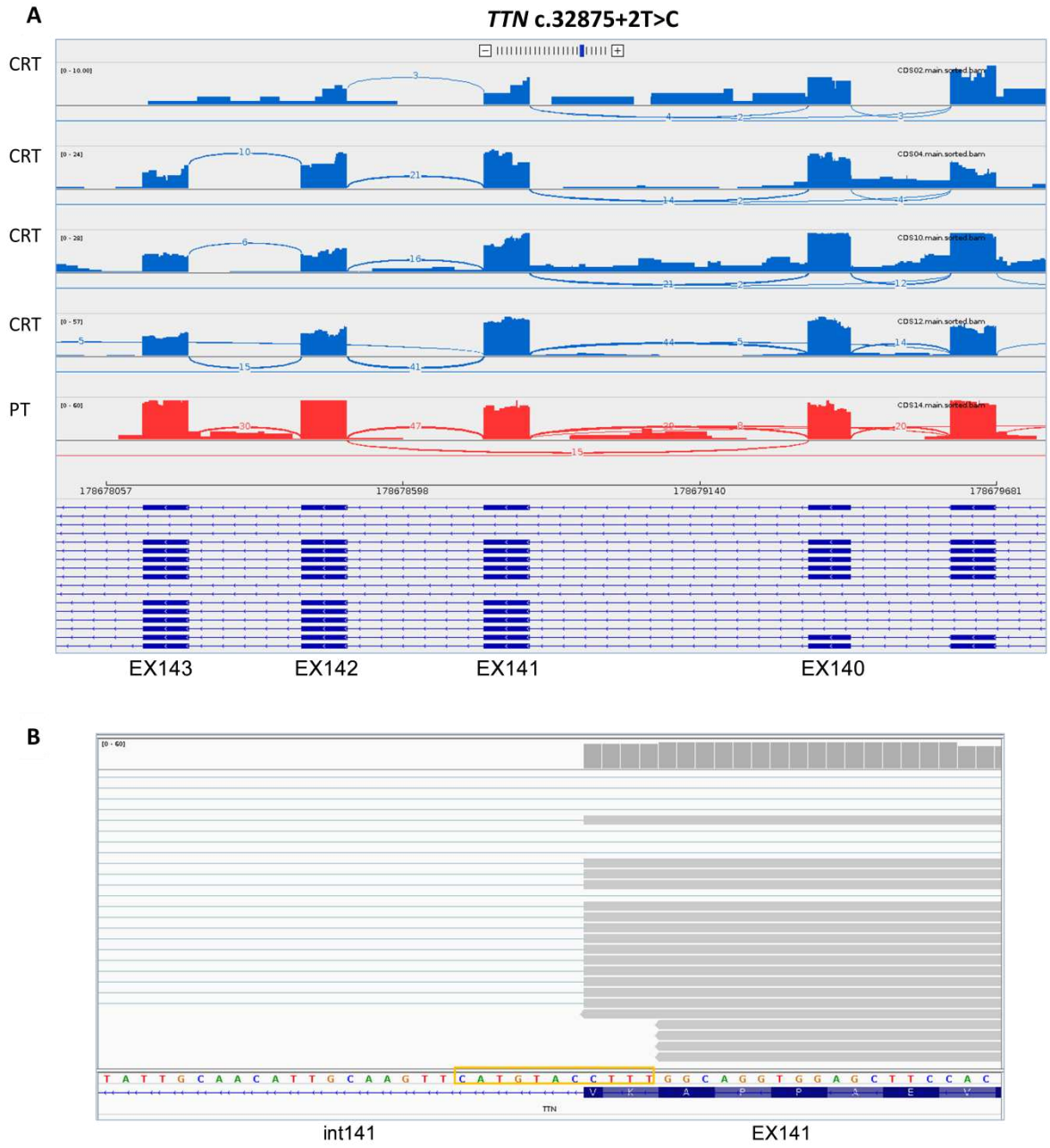


Figure 77. Visualisation of *TTN* c.32875+2 T>C gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *TTN* c.32875+2T>C. (**B**) The orange box highlights intron-exon boundaries of interest.

In the case of c.41609-1 G>A variant, previously characterised as spliceogenic, creates an exonic acceptor site in the first nucleotides of exon 227 leading to deletion of 3 bp. In agreement with previous data, RNA-seq from patient's blood confirmed the disruption of the normal acceptor site and usage of the cryptic 3' ss (Figure 78A). However, the read coverage mapped

the region of the transcripts carrying the mutated allele (Figure 78B).

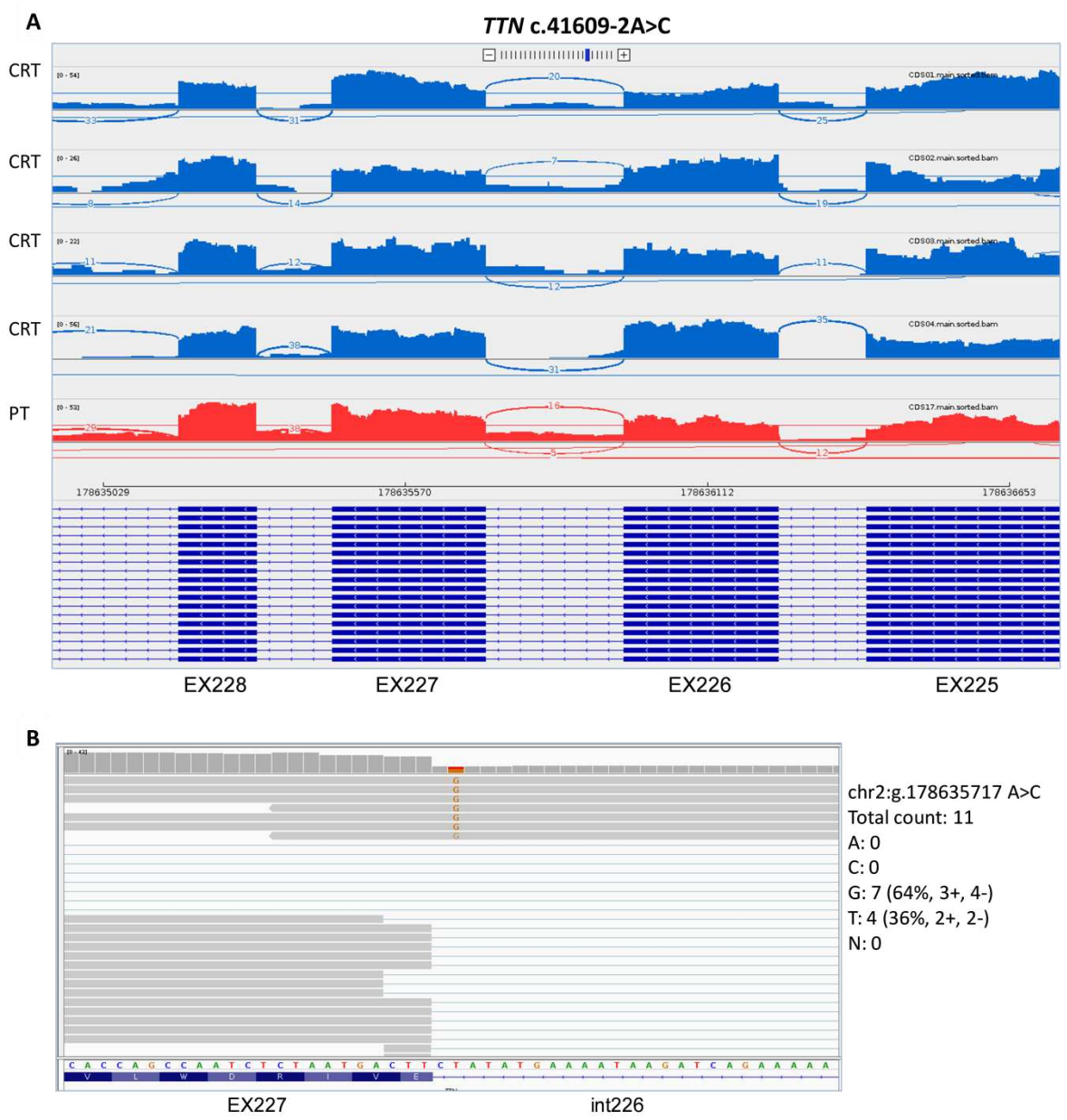


Figure 78. Visualisation of *TTN* c.41609-2 A>C gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *TTN* c.41609-2A>C. (**B**) The orange box highlights intron-exon boundaries of interest.

The *TTN* c.57847+4delGTAA variant affects the consensus donor site, which was observed to induce mis-splicing events with the inclusion of intron into the transcript by minigene assays. Whereas no splicing aberrations were identified after performing RT-PCR from blood samples. Interestingly, for this case, RNA-seq identified high coverage in intron region resulting in a frameshift due to the deletion of the 4nt which were detected by read mapping (Figure 79A-B).

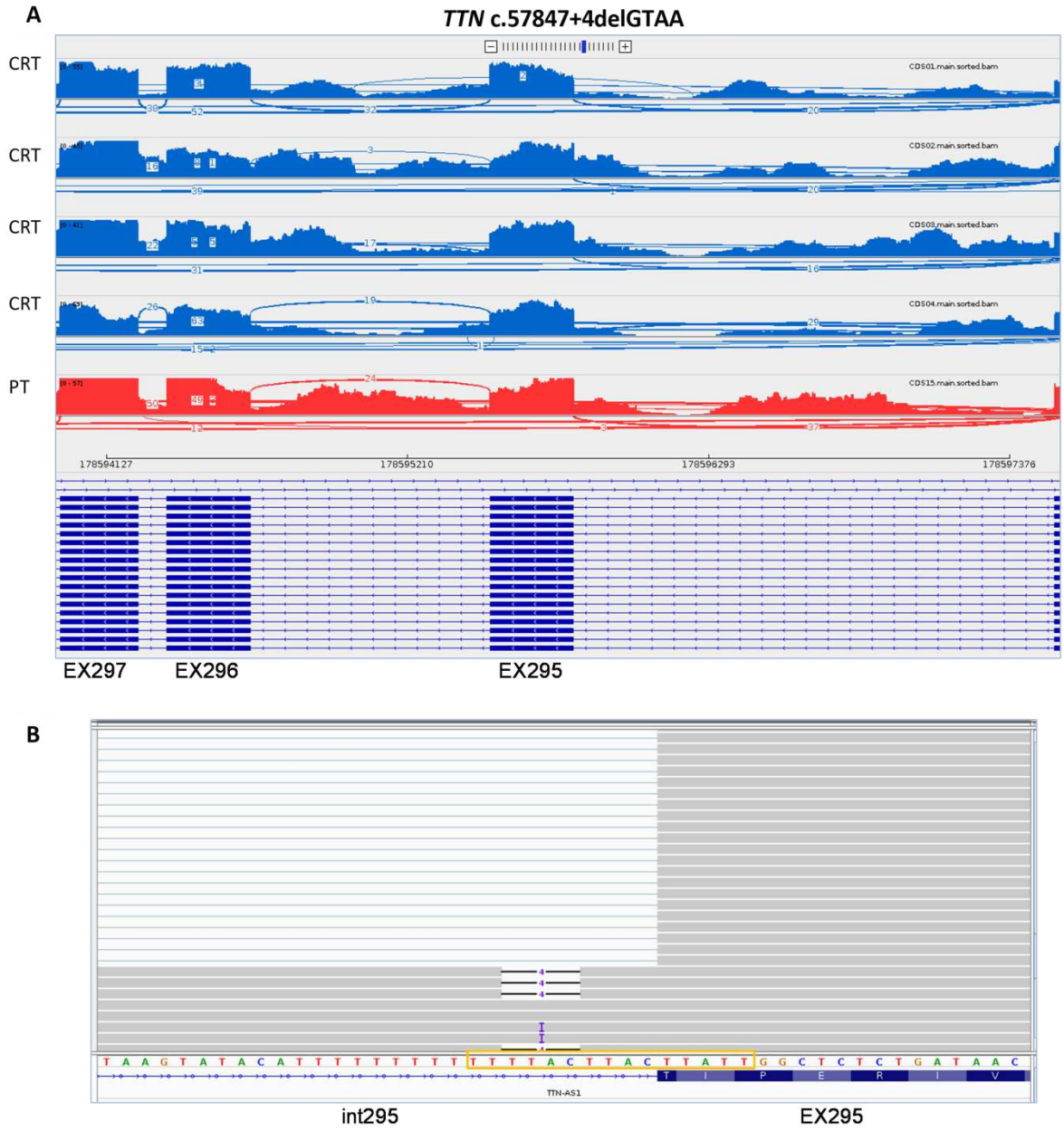


Figure 79. Visualisation of *TTN* c.57847+4delGTAA gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *TTN* c.57847+4delGTAA. (**B**) The orange box highlights intron-exon boundaries where the 4nt deletion was mapped.

RNA-seq data from *VCL* established that there was no significant alternative splicing in the region of the gene containing the variant c.622+4 C>G which the mutant allele was not identified by read mapping (Figure 80A-B). In agreement with previous data, *VCL* c.622+4C>G was reported to use the canonical junction with no effect on mRNA processing.

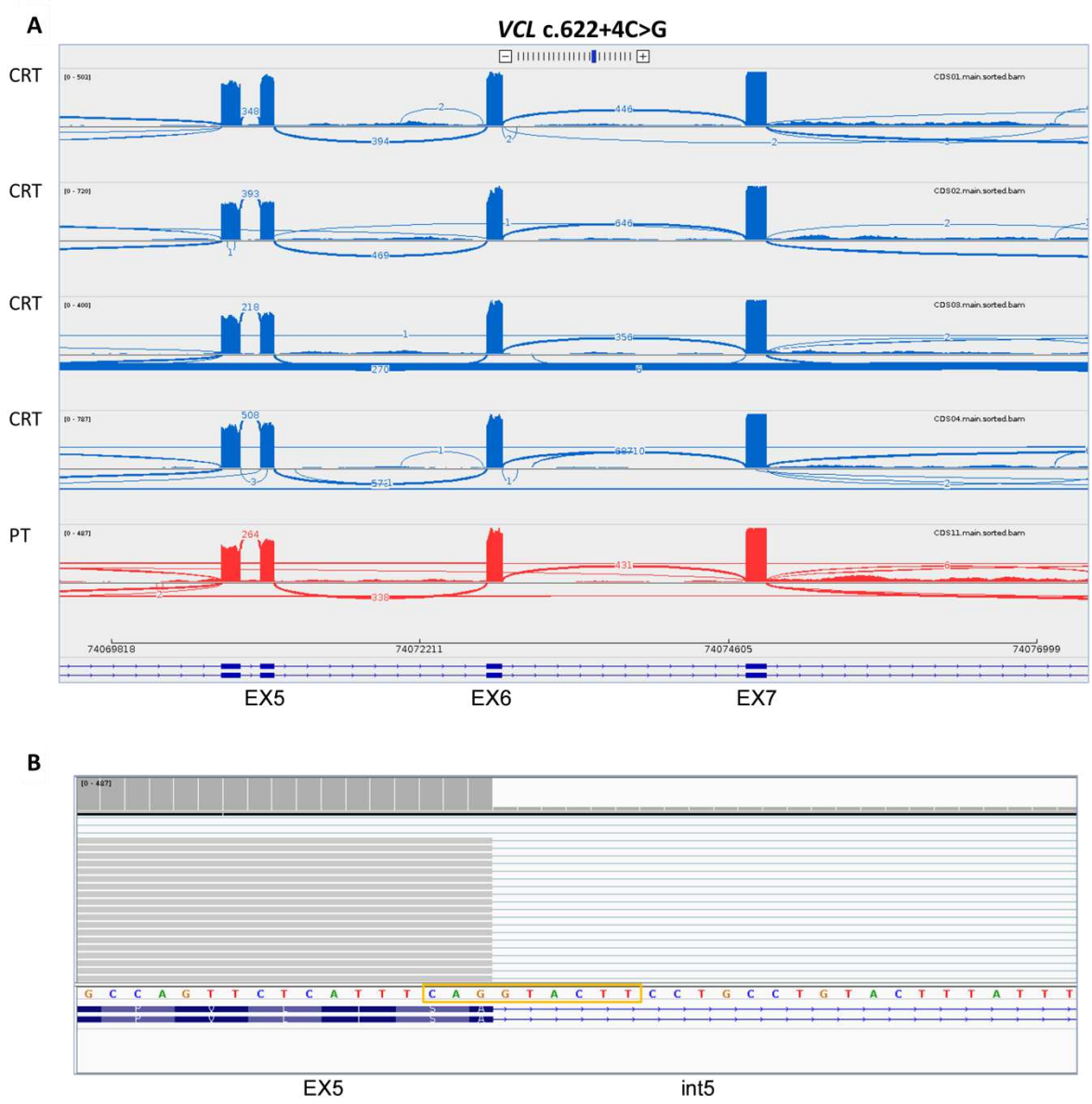


Figure 80. Visualisation of *VCL* c.622+4 C>G gene by IGV. IGV-Sashimi plot of patient (PT, red) and controls (n=4, CRT, blue) is shown which represents the read coverage and the transcript isoforms of the respective genes. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. (**A**) Sashimi plot from *VCL* c.622+4 C>G. (**B**) The orange box highlights intron-exon boundaries where the variant was mapped.

3.7 Testing of small molecule splicing modulators in HEK293T cells

Effective therapeutic strategies to overcome the consequences of aberrant splicing events are a major challenge. As stated in the introduction (section 1.6) a variety of approaches have been explored as possible promising avenues for enhancement of correct mRNA splicing.

In this study, we used small molecules to reverse the consequences of RNA mis-processing. The criteria used to select compounds that have been previously used to alter mRNA splicing

outcomes were (i) availability and (ii) which step of the splicing process they have been previously observed to influence.

Kinetin, Zeatin, and 6-Benzyladenine are part of the adenine-type cytokinin (CK) family and have been shown to affect the interaction between the donor splice site and U1 snRNP. Valproic acid (VPA) and Sodium butyrate (NaBu) are members of the histone deacetylase inhibitor family that are involved in modulating transacting factors such as SR proteins and hnRNPs. Finally, Branaplam, which is utilised for treatment of Spinal Muscular Atrophy (SMA) has been shown to enhance the interaction between the SMN2 pre-mRNA and the U1 snRNA, thereby turning a weak splice site into a strong one. The structure of the compounds is represented in figure 81.

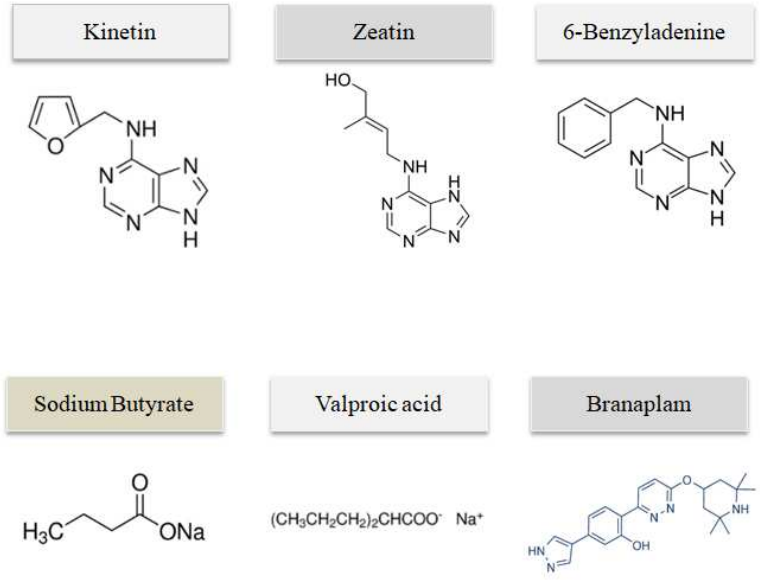


Figure 81. Chemical structures of the small molecules tested in this study. Cytokinin family: kinetin (6-furfurylaminopurine), 6-benzyladenine, zeatin (6-(4-hydroxy-3-methylbut-2-enylamino) purine). Histone deacetylase inhibitors family: valproic acid (VPA), sodium butyrate (NaBu). Branaplam.

Prior to initiating the screen, a resazurin assay was performed in order to analyse the toxicity of the selected small molecules. The assay is based on the reduction by living cells of the oxidized blue dye to a pink fluorescent resorufin product (see section 2.16). Treatment of Zeatin, VPA, NaBu, 6-Benzyladenine had no effect on cell viability as reported in figure 82. Kinetin treatment, however, showed toxicity at concentrations above 300μM, whereas Branaplam was more toxic, but did not affect viability in the range of 6.25μM to 12.5μM (Figure 70). DMSO was confirmed to be not toxic to the cells at the test concentration of 1%.

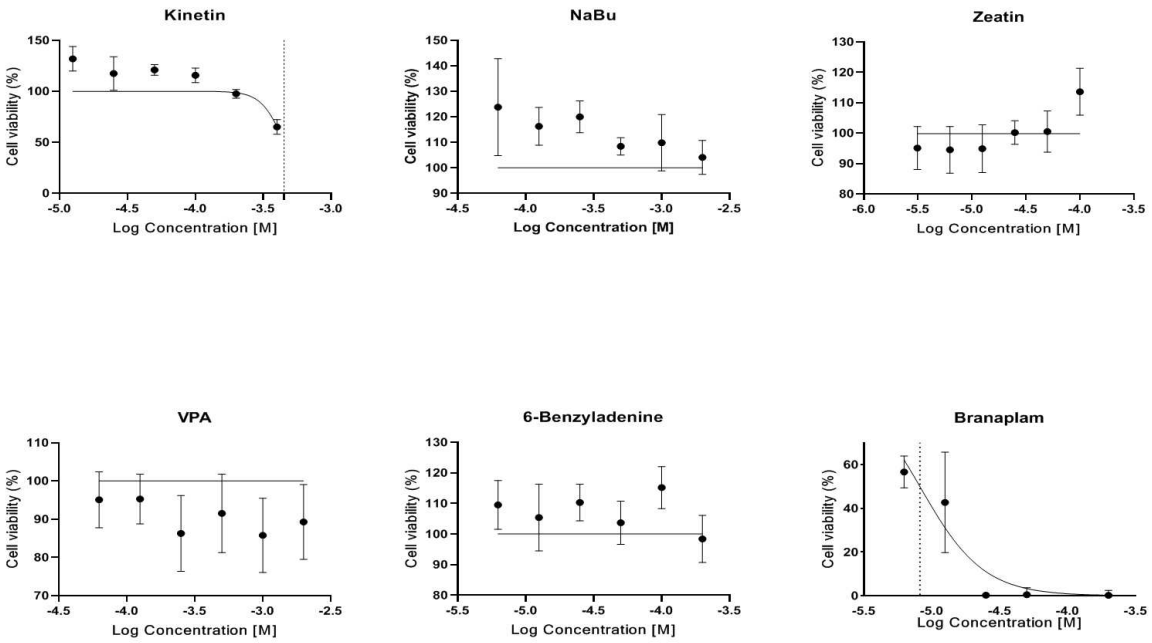


Figure 82. Cell viability assay results from HEK293T cells treated with splicing modulator compounds. Y axis, cell viability (%) and X axis, concentration in log M. Kinetin from concentration 12μM and 1:10 serial dilutions were performed up to 400μM. Sodium Butyrate (NaBu) from concentration 62.5μM to 2000μM. Zeatin from concentration 3.12μM to 100μM. Valproic acid (VPA) from concentration 62.5μM to 2000μM. 6-Benzyladenine from concentration 6.25μM to 200μM. Branaplam from concentration 6.25μM to 200μM. The curves were performed in triplicate for each concentration. Error bars represent the standard error of the mean.

3.8 Testing of small molecule compounds known to affect splicing to correct mutation deregulated splicing

After assessing cytotoxicity, compounds were tested for their effect on splicing of the wild-type and mutated minigenes previously analysed. This analysis was performed by the addition of the compound 4h after transfection of the wt and mutated minigenes and analyses of the mRNA isoforms and RT-PCR 48h after the addition of the compound. The concentration of the compounds was based on the surveyed literature and cell viability assay. 6-Benzyladenine and Zeatin have been previously used in concentrations up to 100μM. Whereas 6-Benzyladenine was observed to correct splicing in the *IKBKAP* gene carrying a splice-donor site mutation (IVS20+6 T>C) and to increase IKAP protein levels at a tested concentration of 100μM, while no response was reported for Zeatin (Slaugenhaupt *et al.*, 2004; Hims *et al.*, 2007). To identify the optimal concentration an initial experiment was performed of three independent dose-response experiments in HEK293T cells transfected with wild type and mutated pTB *DSP* c.273+5 G>A minigenes and treated with 6-Benzyladenine for 48h using three different concentrations,

ranging from 50 to 200μM. Whereas Zeatin was used at a higher dose of 200μM, at which no toxic effects were observed in HEK293T cell line.

In the study of Slaugenhaupt and colleagues (Slaugenhaupt *et al.*, 2004), Kinetin was demonstrated to correct the aberrant splicing of *IKBKAP* gene at doses higher than 50μM. As a result, concentrations up to 400μM were tested, and the results were consistent with this study's findings in terms of cellular toxicity. This data suggested that 300μM was an appropriate concentration for the initial analysis of kinetin.

VPA and NaBu are known to act as histone deacetylase (HDAC) inhibitors enhancing transcription of some genes (Göttlicher *et al.*, 2001). VPA has been observed to enhance inclusion of the *SMN2* exon 7 aberrantly splicing due to a silent variant within an exonic splicing enhancer (ESE; c.280 C>T), possibly via Htra-β1 (SR splicing factor). In this study concentrations up to 1000μM were used in fibroblast cultures obtained from patients with effects on exon 7 inclusion visible even at0.5μM. NaBu has a similar mode of action to VPA, and in β-Thalassemia, it has been reported to increase the level of normal *β-globin* transcript in primary erythroid cells derived from IVS1-110 patients at 0.5mM. Whereas higher doses have been observed to reduce cell proliferation and viability (Khaniani *et al.*, 2016). Considering the above data, VPA and NaBu were initially tested as a single concentration at 500μM. As outlined in section 3.5, these compounds showed no cellular toxicity at these concentrations in HEK293T cells.

Finally, Branaplam (NVS-SM1) is a chemical molecule that has been shown to enhance splicing at some non-canonical, mutated splice-donor sites in *SMN* gene. The study was performed using a reporter gene containing the *SMN2* exon6-exon7 splice site, and demonstrated that the molecule acts specifically at the 5' ss of exon 7 to enhance the binding affinity of the U1 snRNA by stabilising the RNA helix structure formed by SMN2 pre-mRNA and the U1 snRNP-complex, with a response at 10μM (Palacino *et al.*, 2015). After performing the cytotoxicity assay, Branaplam showed toxic effects, with logIC50 estimated at -5.088. Therefore, for preliminary analysis, this compound was used at 5μM. As a negative control, cells were treated with the same amount of solution in which the drug was resuspended, either DMSO or ddH_2O.

3.8.1 pTB *DSP* c.273+5 G>A in vitro treatment with splicing compounds

Following the methodology described above and estimating the potential response for 6-Benzyladenine to influence pTB *DSP* exon 2 splicing, HEK293T cells were transfected with wild-type and mutated minigenes that were assayed after 48h in the presence of 50, 100 and 200μM 6-Benzyladenine and DMSO as a negative control. Exon 2 splicing was determined by RT-PCR using specific primers and PCR products resolved on agarose 2% gel. Figure 83A illustrates how treatment with 6-Bensyladenine enhanced the correct splicing process of exon 2 in the

presence of the mutation and decreased aberrant splicing caused by the nucleotide substitution c.275 +5 G>A. HEK293T cells treated with 100µM demonstrated a significant increase of exon 2 inclusion with improvement at 200µM (Figure 83B). For further experiments, 200µM concentration was selected as the final concentration with a 48h treatment time for inducing a response.

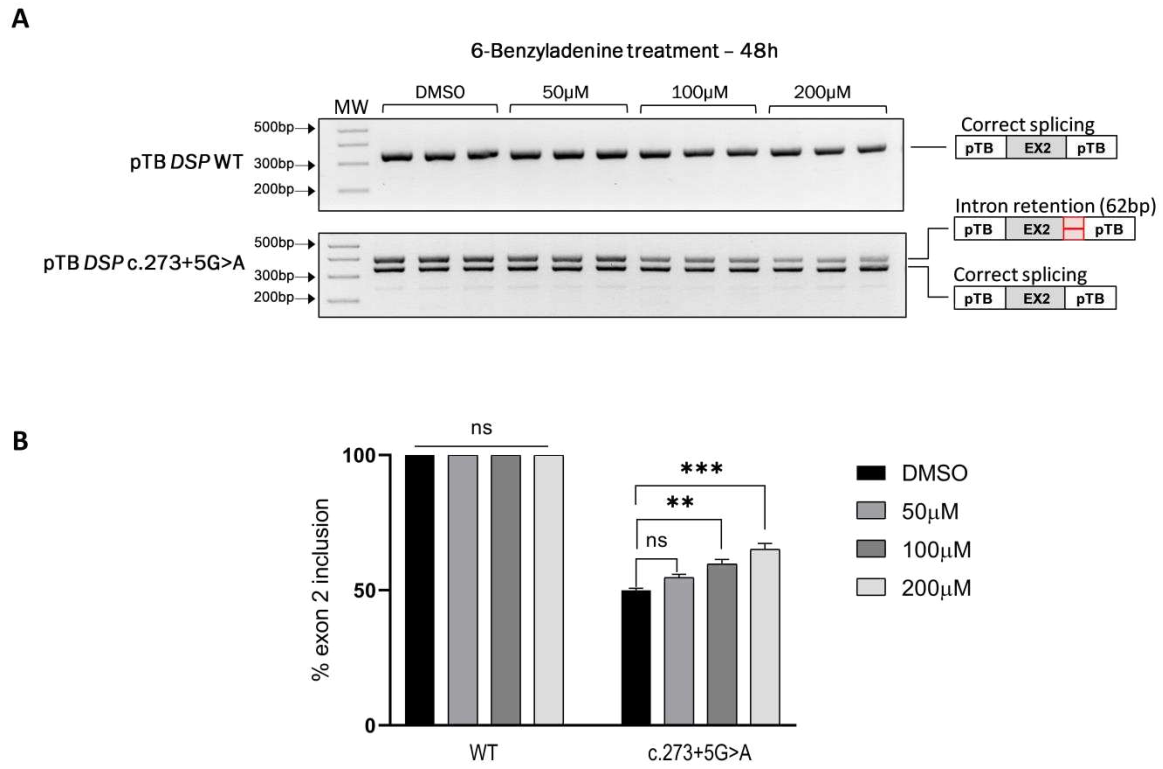


Figure 83. pTB *DSP* c.275+5 G>A minigene splicing assay in HEK293T cells treated with 6-Benzyladenine compound. (**A**) RT-PCR analysis of HEK293T cells, tested in triplicate, following treatment with 50µM, 100µM, and 200µM6-Benzyladenine in 1% DMSO or 1% DMSO alone for 48h. The top agarose gel represents the PCR products from wt minigene, while the bottom agarose gel shows PCR products from mutant minigene. M indicates the DNA marker. (**B**) Effect of DMSO and 50µM, 100µM, and 200µM 6-Benzyladenine on splicing from the minigene carrying the variant c.275+5 G>A. Quantification analysis was carried out with ImageJ of percentage (%) of normal and aberrant *DSP* transcripts expressed as means ± SD of triplicate experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

In the presence of 300µM Kinetin, 200µM Zeatin and 5µM Branaplam, the level of exon 2 canonical processed increased relative to cells treated with DMSO or ddH_2O. In figure 84A and B, representative graphs of the cells treated with 300µM Kinetin and 200µM Zeatin show that the wild-type transcript increased by approximately 10%, while cells treated with 5µM of Branaplam showed significant rescue, at ~30%, of pTB *DSP* c.275+5 G>A compared to the control (Figure

84C). Treatment with VPA at 500μM upon transfection with wild-type and mutated minigenes, no decrease in the amount of aberrantly spliced transcript was observed compared to the untreated cells (Figure 84D). In contrast, 500μM NaBu treatment showed an unexpected outcome, as illustrated in Figure 84E. This treatment resulted in a decrease of the correctly spliced transcript and a corresponding increase of the aberrantly spliced transcript.

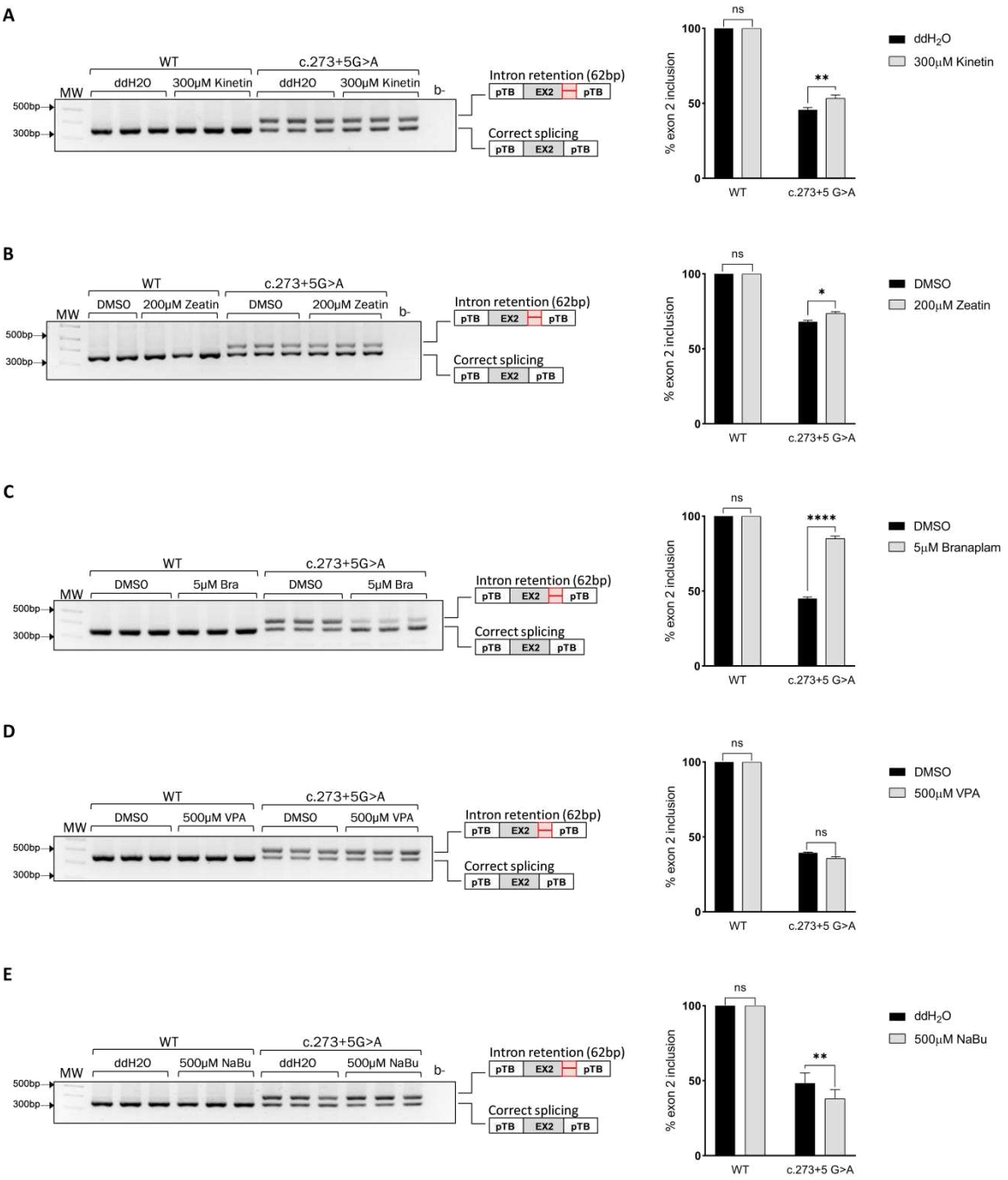


Figure 84. The effect of splicing compounds on pTB *DSP* wild type and c.275+5 G>A minigenes in HEK293T cells. (**A**) RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment with 300μM kinetin resuspended in ddH_2O or ddH_2O alone, as control, for 48h. On the right, the graph represents the effect of ddH_2O and 300μM Kinetin doses on minigene carrying the variant c.275+5 G>A. (**B**) RT-PCR analysis of HEK293T cell line, tested in

triplicate, following treatment with 200µM Zeatin resuspended in 1% DMSO and 1% DMSO alone for 48h. On the right, the graph represents the effect of DMSO and 200µM Zeatin on minigene carrying the variant c.275+5 G>A. (**C**) RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment with 5µM Branaplam (Bra) resuspended in 1% DMSO and 1% DMSO alone for 48h. On the right, the graph represents the effect of DMSO and 5µM Branaplam on minigene carrying the variant c.275+5 G>A. (**D**) RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment with 500µM VPA resuspended in 1% DMSO and 1% DMSO alone, as control, for 48h. On the right, the graph represents the effect of DMSO and 500µM VPA on minigene carrying the variant c.275+5 G>A. (**E**) RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment with 500µM NaBu resuspended in ddH_2O and ddH_2O alone, as control, for 48h. Densitometry was performed using ImageJ. The data are represented as percentages of wild-type and aberrant transcripts expressed as means ± SD of triplicate experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

3.8.2 pTB *DSP* c.273+3A>T in vitro treatment with splicing modulators

Small splicing compounds were screened on the pTB *DSP* c.233 +3A>T minigene construct, which carries the +3A>T variant leading to use a cryptic 5'ss and complete skipping of exon 2 in the transcripts.

HEK293T cells were treated with 6-Benzyladenine, a cytokine that has previously reported to affect *DSP* splicing. As expected, in the untreated cells (DMSO) exon 2 inclusion is around 14%. Whereas after 200µM 6-Benzyladenine treatment, exon 2 inclusion increases to 27% (Figure 85A).In addition, 300µM Kinetin and 200µM Zeatin were tested, and splicing profile was evaluated using RT-PCR. The results shown in Figure 85B-C confirmed the same splicing pattern as the untreated cells, indicating no rescue of splicing.

Next, HEK293T cells were assessed after 48h in the presence of VPA and NaBu at 500µM and DMSO and ddH_2O as negative control, respectively. Exon 2 inclusion was determined by RT-PCR which is shown in Figure 85D-E, confirming that the compounds do not influence the ratio of the three isoforms between the treated and untreated cells. Finally, wild-type and mutated minigenes were transfected in HEK293T cells and treated for 48h with 5µM Branaplam and 1% DMSO as a negative control. As shown in figure 85F, the percentage of exon 2 inclusion transcript increased from~10.7 to 30% in treated cells compared to untreated.

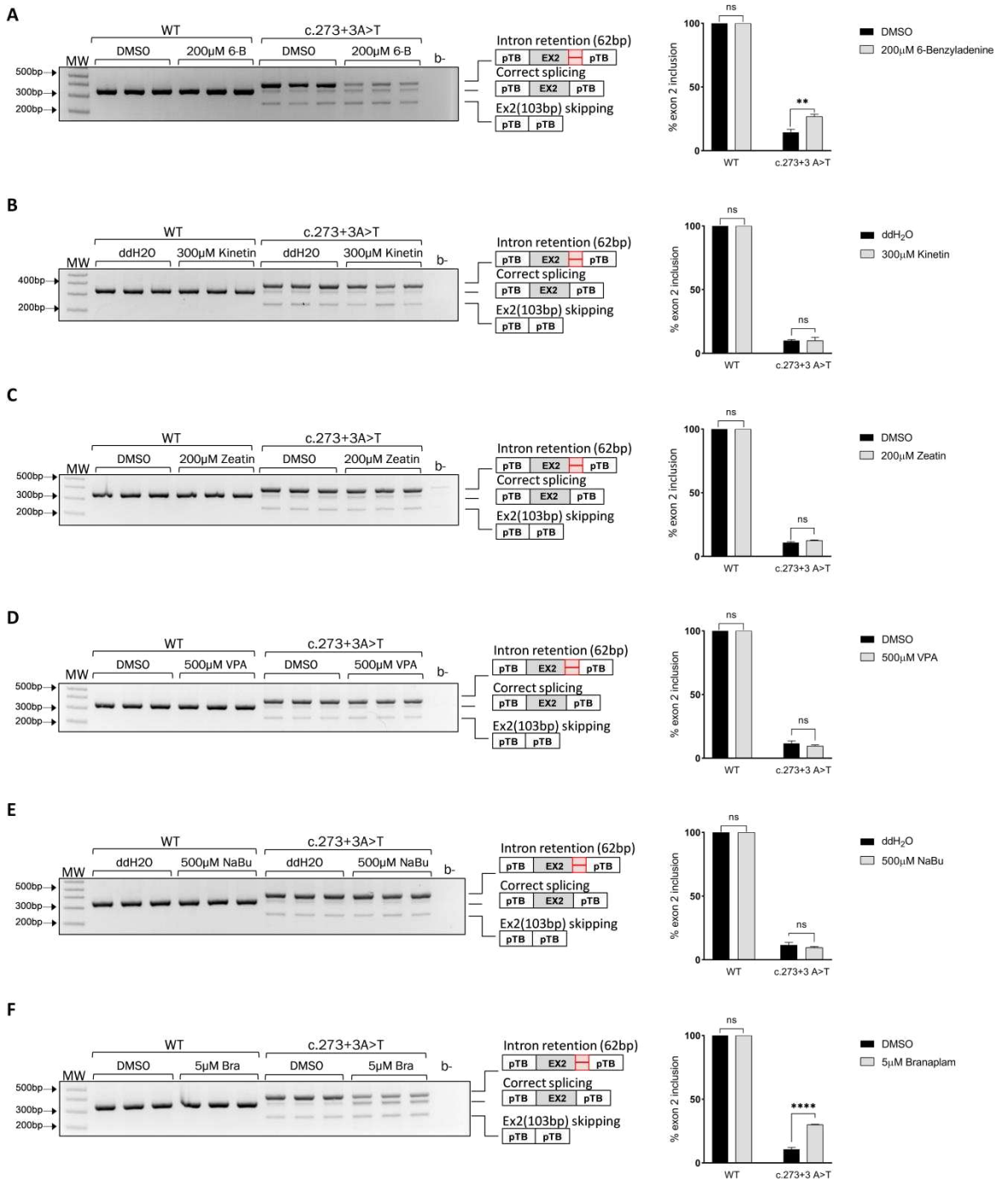
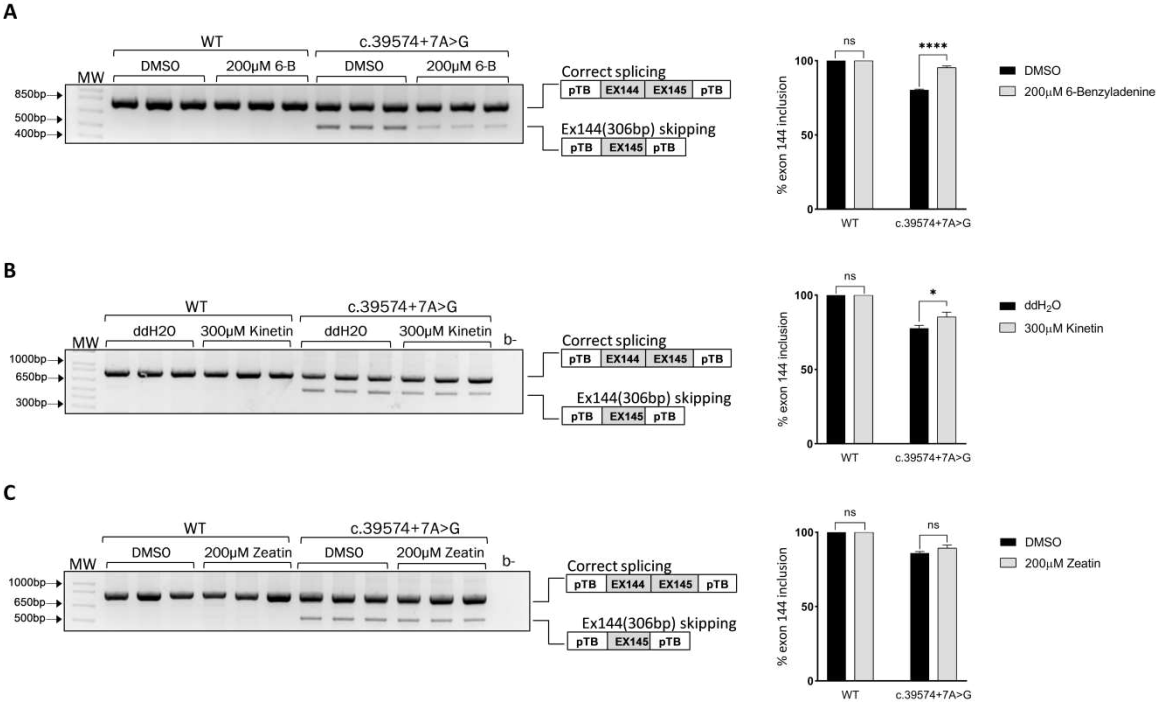


Figure 85. The effect of splicing compounds on pTB *DSP* wild type and c.273+3 A>T minigenes in HEK293T cells. Exon 2 inclusion was measured by band density analysis using ImageJ. On the left, a representative agarose gel is shown. On the right, the percentage of exon 2 inclusion is represented in the graph. The experiment was run in triplicate (n=3). (**A**) 200µM 6-Benzyladenine and 1% DMSO. (**B**) 300µM Kinetin and ddH_2O as a negative control. (**C**) 200µM Zeatin and DMSO as a negative control. (**D**) 500µM VPA and DMSO. (**E**) 500µM NaBu and ddH_2O as a negative control. (**F**) 5µM Branaplam and DMSO as a negative control. Error bars represent the SD of the three experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

3.8.3 pTB *TTN* c.39574+7A>G in vitro treatment with splicing compounds

pTB *TTN* wt and c.39574+7 minigene splicing profiles were analysed in the presence of the compounds. 6-Benzyladenine was tested at 200μM in HEK293T cell line. The cells were treated with the compounds for 48h. Total RNA isolated from cells underwent RT-PCR. Treated cells resulted in a significant increase of the wild-type transcript as compared to DMSO treated cells (Figure 86A). Similarly, 300μM Kinetin treatment also increased the percentage of exon inclusion (Figure 86B). On the other hand, the RT-PCR results (Figure 86C-D) showed that in presence of 200μM Zeatin and 500μM VPA, no significant difference was observed in the ratio isoforms between the treated and control cells. 5μM Branaplam treatment, as seen in figure 86E, affected the splicing outcome in both the wild-type minigene as well as the mutated minigene. In the wild-type minigene exon 144 skipping was observed with branaplam treatment. In the minigene carrying the genetic variant (c.39574+7 A>G) this effect was accentuated with a decrease of the canonical exon 144 splicing compared to the control and concomitantly an increase of the aberrantly spliced transcript, which corresponds to exon 144 skipping.



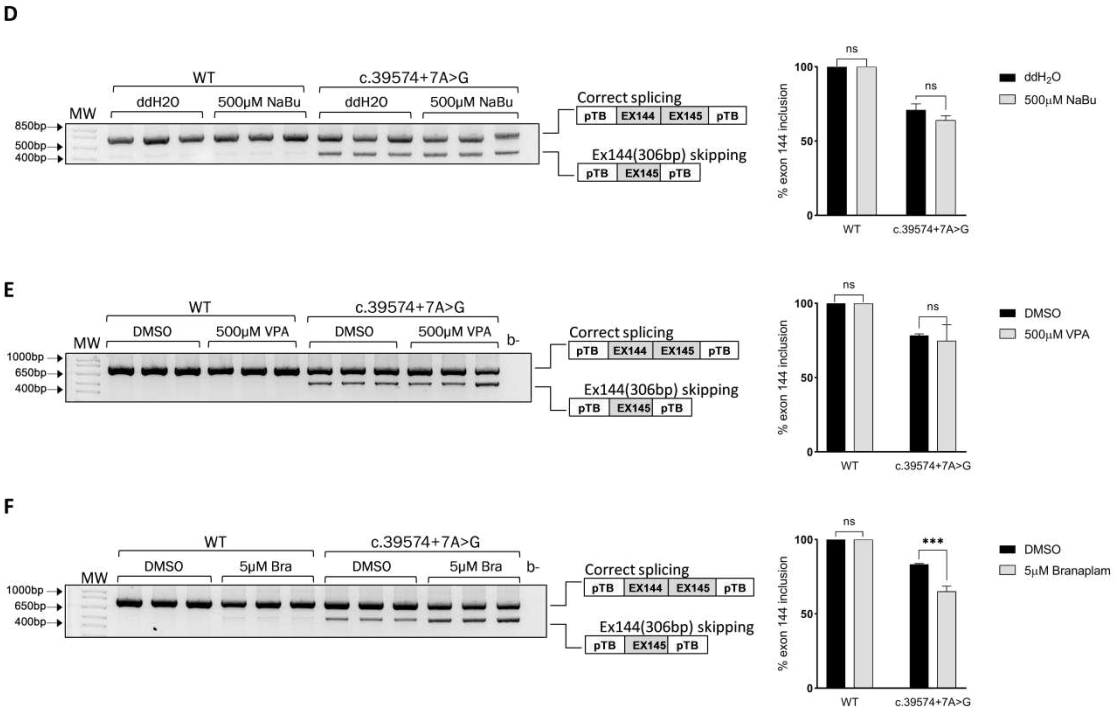


Figure 86. The effect of splicing compounds on pTB *TTN* wild type and c.39574+7 A>G minigenes in HEK293T cells. On the right, RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment. On the right, the graph represents the % exon 144 inclusion. (**A**) 200μM 6-Benzyladenine (6-B) resuspended in DMSO or DMSO alone, as control. (**B**) 300μM Kinetin resuspended in ddH_2O or ddH_2O alone, as control. (**C**) 200μM Zeatin resuspended in DMSO and DMSO alone, as control. (**D**) 500μM NaBu resuspended in ddH_2O or ddH_2O alone, as control. (**E**) 500μM VPA resuspended in DMSO and DMSO alone, as control. (**F**) 5μM Branaplam (Bra) resuspended in DMSO and DMSO alone, as control. The data are represented as the percentage of exon 144 inclusion as means ± SD of triplicate experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

3.8.4 pTB *TTN* c.57847+4delGTAA in vitro treatment with splicing compounds

pTB *TTN* wt and mutant (c.57847+4delGTAA) minigenes, where the deletion in the donor splice site induced dysfunction in splicing process (see section 3.3) were tested for changes in the relative amounts of aberrant to wild-type isoforms produced after 6-Benzyladenine and Kinetin treatment of transfected HEK293T cells. The splicing profile was investigated by RT-PCR analysis and the cells assayed for 48h in the presence of the test drug dissolved in DMSO and ddH_2O. Upon 6-Benzyladenine treatment the construct carrying GTAA deletion was responsive. As can be seen in figure 87A, pTB TTN c.57847+4del GTAA splicing changes with an increase in the correctly spliced transcript and a decrease in the aberrantly spliced transcript. Similar results were obtained with 300μM kinetin, the percent exon inclusion increased from 21% to

35% after treatment, while the average intron retention jumped from 79% to 65% (Figure 87B). In both cases, no influence on splicing was observed for the wild-type minigene (Figure 87A-B).

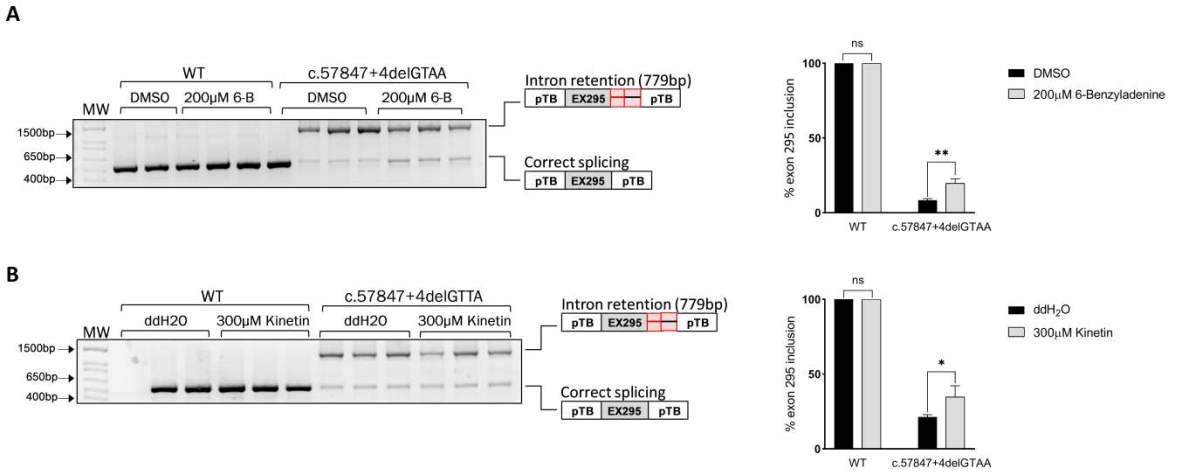


Figure 87. The effect of splicing compounds on pTB *TTN* wild type and c.57847+4delGTAA minigenes in HEK293T cells. On the right, RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment. On the right, the graph represents the % exon 295 inclusion. (**A**) 200μM 6-Benzyladenine resuspended in DMSO or DMSO alone, as control. (**B**) 300μM Kinetin resuspended in ddH_2O or ddH_2O alone, as control. The data are represented as the percentage of exon 144 inclusion as means ± SD of triplicate experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

3.8.5 pTB *TTN* c.13282+1G>A in vitro treatment with splicing compounds

Splicing small molecular compounds were tested on how act on the pTB *TTN* c.13282+1 G>A minigene. Where *in vitro* analysis reported that the minigene carrying the c.13282+1 G>A variant disrupts the canonical GT and the wt transcript (exon 46 inclusion) is not produced but induces mis-splicing events. The splicing profile was investigated through transfection of both wt and mutated minigene in HEK293T cell and treated with the compounds after 4h post-transfection. The possible rescue of exon 46 inclusion in the transcripts was determined by RT-PCR. Although each of the tested compounds showed no consistent rescue of splicing pattern compared to untreated cells. Therefore, the wt minigene showed no differences in the splicing profile after the treatment. Altogether, these experiments demonstrate that the tested compounds cannot promote the rescue of the canonical isoform and the recruitment of U1 snRNP on the end of exon 46 and force the correct splicing if there is not a low production of the canonical transcript (88A-F).

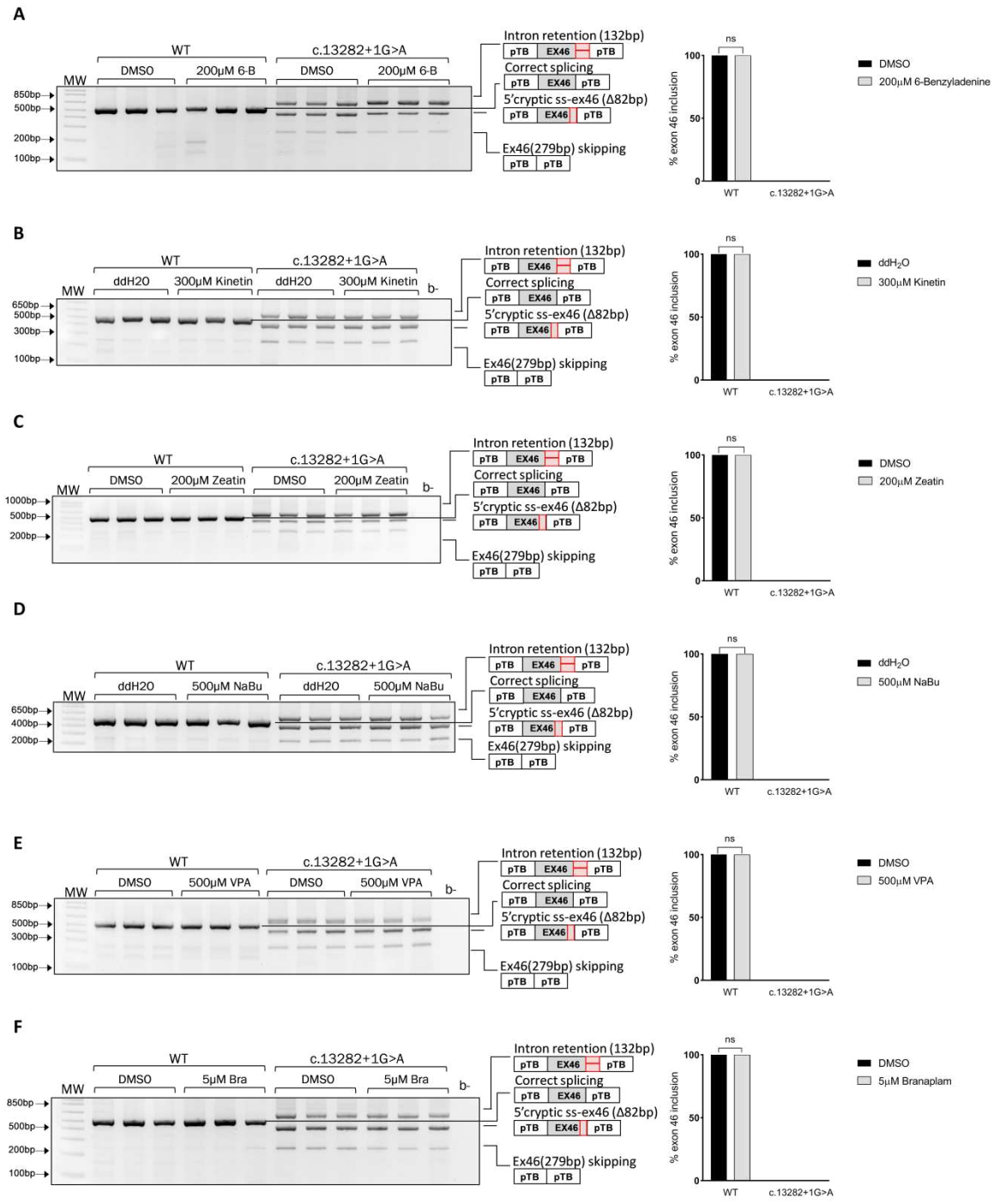


Figure 88. The effect of splicing compounds on pTB *TTN* wild type and c.13282+1 G>A minigenes in HEK293T cells. On the left, RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment. On the right, the graph represents the % exon 46 inclusion. (**A**) 200μM 6-Benzyladenine and DMSO as a negative control. (**B**) 300μM Kinetin and ddH_2O as a negative control. (**C**) 200μM Zeatin and DMSO as a negative control. (**D**) 500μM NaBu and ddH_2O as a negative control. (**E**) 500μM VPA and DMSO as a negative. (**F**) 5μM Branaplam and DMSO as a negative control. The experiment was run in triplicate (n=3). The data are represented as percentage of exon 46 inclusion. Error bar represents the SD of the three experiments.

Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

3.8.6 pTB *LMNA* c.1609-1G>A in vitro treatment with splicing compounds

In this context, splicing modulator compounds which promote the recruitment of U1 snRNP on the 5' ss and modify the activity of the *trans*- splicing factors were also investigated in minigene carrying the variant that disrupts the canonical AG splice site. The cells were transiently transfected with wt and mutated pTB *LMNA* c.1609-1G>A minigenes and treated 4h post transfection. Total RNA isolated from cells underwent RT-PCR to investigate potential differences in the splicing profile. As can be seen in figure 89A-F, HEK293T cells treated with 200μM 6-Benzyladenine, 300μM Kinetin, 200μM Zeatin, 500μM VPA, and 500μM NaBu was observed no increasing in the splicing inclusion of *LMNA* exon 10 in the mRNA transcript at the level of the minigene carrying the variant. Branaplam was the only tested compound to increase the levels of exon 10 inclusion in the transcripts (Figure 89E). The splicing modulators were also tested with the wt minigene in order to detect possible differences in the splicing mechanism.

This experiment demonstrates that the Branaplam is an essential determinant for the recruitment of the U1 snRNP on the 5'ss and improves its strength which allows recognition of the 3' ss when disrupted.

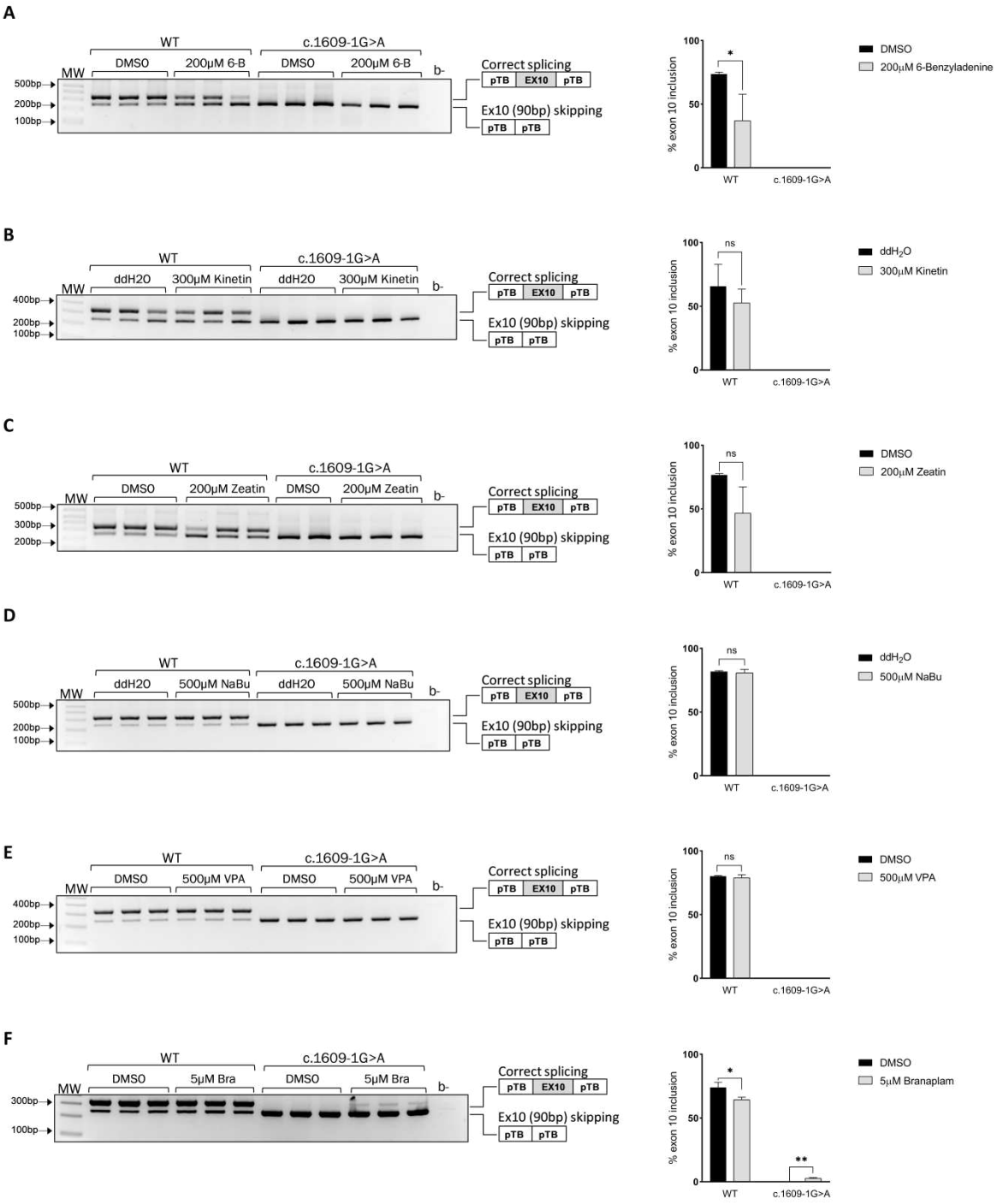


Figure 89. The effect of splicing compounds on pTB *LMNA* wild type and c.1609-1 G>A minigenes in HEK293T cells. On the left, RT-PCR analysis of HEK293T cell line, tested in triplicate, following treatment. On the right, the graph represents the % exon 10 inclusion. (**A**) 200μM 6-Benzyladenine resuspended in DMSO or DMSO alone, as control. (**B**) 300μM Kinetin resuspended in ddH_2O or ddH_2O alone, as control. (**C**) 200μM Zeatin resuspended in DMSO and DMSO alone, as control. (**D**) 500μM NaBu resuspended in ddH_2O or ddH_2O alone, as control. (**E**) 500μM VPA resuspended in DMSO and DMSO alone, as control. (**F**) 5μM Branaplam resuspended in DMSO and DMSO alone, as control. Exon 10 inclusion was measured by band density analysis using ImageJ software.

The experiment was run in triplicate (n=3). On the right, the percentage of exon 10 inclusion is represented in the graph. Error bars represent the SD of the three experiments. Statistical analysis was performed using unpaired Student t-test (ns: not significant; * p<0,05; ** p<0,01; *** p<0,001), comparing each lane with the control.

Chapter 4 Discussion

The introduction of NGS approaches into patient care has revolutionised human genetics. As a consequence, the number of variants identified which potentially affect genes responsible for disease is too large from manual interpretation (Adams and Eng, 2018). In this context, to integrate genome sequencing in clinical settings, specific interpretation criteria have been optimised and standardised for the classification of DNA sequence variations. For many years, variant interpretation has focused solely on the coding region of genes, without considering the possible nature of variants in non-coding DNA which can impact mRNA processing. While it is evident the possible consequences of a mutation that affects the coding DNA and results in disease via alteration in protein function, the potential influence of a variant on splicing can be difficult to predict. However, if exon skipping or intron retention occurs in the transcript it is possible to predict the consequences on protein reading frame, while accurately anticipating the activation of a cryptic splice site remains challenging.

This study focuses on investigating the potential impact of variants, classified to be either Path, LP or VUS according to ACMG/AMP guidelines, on pre-mRNA processing in patients diagnosed with CVDs and how to effectively implement variant classification. Both bioinformatics and wet lab RNA analyses have been employed to interpret variants and incorporate the results into ACMG/AMP guidelines, for variants classified as VUS.

First, bioinformatic prediction was performed on the listed variants (Table 13) to predict the potential role for the variants inducing mis-splicing events. SpliceAI, a deep learning tool, was chosen to perform the analysis as it has been established to have high accuracy in predicting spliceogenic variants. To use this tool, users require no bioinformatics expertise and default settings which display the Δscore that scanned a region +/- 500bp if intron-exon boundaries for cryptic splice site can be applied. SpliceAI predicts the disruption of the canonical splice sites with a Δscore ≥ 0.22 for 15 out of the 25variants examined in this thesis for the VUSs category, while the prediction for the LP (5/6 variants) and Path (5/5 variant) variants showed a Δscore ranging from 0.81 to 1.00, which is to clearly have an effect on splicing. Thus, SpliceAI can improve the ability to prioritise which VUS variants it is necessary study the splicing profile.

Subsequently, a total of 24 intronic variants were selected for functional studies (Table 17), whether or not an effect on splicing was predicted *in silico* to test the consistency of the approaches. Minigene assays were used to clarify the nature of variants potentially associated with splicing abnormalities. The pTB and pcDNA3 vectors were used as backbone to generate

both wt and mutated constructs. It is important to note that minigene constructs are not perfect systems and can generate technical artefacts.

Among the 24 analysed variants (6 variants in 3' ss and 18 variants in 5' ss) assessed in HEK293T cell line, only 17 (71%) variants were proven to alter splicing, 10 disrupted the canonical dinucleotides of splice sites (5 in 3'ss and 5 in 5'ss), and 7 affected the consensus splice site sequence.

Five variants were identified to disrupt 3'ss, two of which (*FLNC* c.3791-1G>A, and *LMNA* c.1609-1 G>A) caused exon skipping, whereas the rest induced the use of cryptic 3' ss (*MYBPC3* c.852-1 G>A, *TTN* c.25808-1 G>A, and *TTN* c.41609-2 A>C). The observed data was in concordance with SpliceAI prediction. Curiously, the *DMD* c.6118-3 C>A variant which is also part of the consensus 3'ss not induce mis-splicing, contrary to SpliceAI prediction. Moreover, 18 variants predicted to affect the 5' ss consensus sequence were analysed, of which only 12 variants showed spliceogenic mechanisms. Alternative 5' ss usage was observed for the variants *PKD1* c.10405+5 G>T, *RYR2* c.1961+3 A>T, and *SCN5A* c.3840+5 G>C. For three variants (*DSP* c.273+3 A>T, *DSP* c.273+5 G>A, and *TTN* c.57847+4delGTAA) alternative splice site usage and correctly spliced transcripts were observed. On the other hand, *NEXN* c.1053+1 G>A, *TTN* c.13282+1 G>A and *TTN* c.32875+2 T>C induced the disruption of the 5'ss and consequently induced an exon skipping event in parallel. In the case of *TTN* c.39574+7 A>G, exon 144 skipping was observed along with the correctly spliced transcript.

The analysis of two variants, *FLNC* c.7251+1 G>A and *TTN* c.19501+2 T>C, was more challenging. With these minigene constructs, multiple transcripts were observed in addition to those expected. From the data, it was concluded that *FLNC* c.7251+1 G>A caused exon 43 skipping, while no conclusion was possible for *TTN* c.19501+2 T>C due to the use of several cryptic donor sites within the pTB intron. The examples of these two variants, where more bands from the expected are observed in minigene assays, these might arise from technical artefacts and, suggest that redesign primers used to generate the insert, or to combine different primer sets for RT-PCR should be considered, in order to eliminate the technical artefact bands. Interestingly, in several minigene constructs, a band corresponding to the empty vector, containing only α2-3 globin joined with fibronectin exons was also observed. This is commonly considered to be a major indicator of insert misrecognition or minigene failure. Ideally, studying splicing patterns should be performed in a construct containing the entire genomic region of a gene but this is hard to clone due to the long size of introns.

For six variants classified as VUSs (*MYLK* c.3703+22 T>C, *NEXN* c.27+9 C>T, *TMEM43* c.882+8 C>T, *TTN* c.669+4 T>C, *TTN* c.50346+3 A>C, and *VCL* c.622+4 C>G) no mis-splicing events were

observed, which is in concordance with SpliceAI prediction. For these variants, the minigene assay data support the conclusion that their pathogenicity is not associated with splicing abnormalities, but rather may depend on other unidentified factors.

Furthermore, the lack of genetic context in minigene assay and the choice of the HEK293T cells can lead to non-concordant results from the affected cell type between SpliceAI and minigene assays. Thus 13 minigene constructs, which corresponded with the variants further analysed from RNA biospecimens, were selected to be tested in a more appropriate cell line (human cardiomyocyte cell line - AC16). These cells represent a closer system to the physiology and molecular mechanisms that occur in the heart. However, the main disadvantage is their proliferative capacity and can only be maintained in culture for at least 10 passages without affecting the cell marker expression and functionality (Litzkas, Jha and Ozer, 1984; Davidson *et al.*, 2005). From the 13 variants evaluated by performing minigene assay in AC16 cells, only two variants (*DMD* c.6118-3 C>A and *TTN* c.669+4 T>C) showed different outcomes from HEK293T cells.

Based on the data, a concordance rate of 92% was observed between the prediction of SpliceAI and the minigene assay. This confirmed that *in silico* analysis based on SpliceAI canprovide information regarding the nature of variants which can influence splicing, in particular improving the classification of VUSs. As well, the study demonstrates that the usage of minigene assay identified the inaccuracy of SpliceAI prediction (false negative) in the case of *DMD* c.6118-3 C>A, and *TTN* c.39574+7 A>G. The comparative analysis showed that minigene assay is more informative than SpliceAI in correctly identifying exon-skipping, cryptic splice-site activation and multiple mis-splicing events. Therefore, this study demonstrated that minigene assays performed in HEK293T cells can provide information regarding the pathogenicity of variants and splicing profile is not discordant with the outcomes using the AC16 cell line in most of the analysed variants.

GENE	VARIANT	ACMG/AMP Classification	SpliceAI Δscore loss	Minigene assay results	Observed effect in blood analysis	Protein effect
Acceptor splice site						
DMD	c.6118-3C>A	VUS	0.42	Correct splicing	No amplification for this region	No change
FLNC	c.3791-1G>A	LP	1.00	Transcript: ex22 skipping	No amplification for this region	In-frame
LMNA	c.1609-1G>A	LP	1.00	Single transcript: ex10 skipping	Transcript: correct splicing Transcript: ex10 skipping (90bp)	In-frame (p.Glu537_His566del)
MYBPC3	c.852-1G>A	LP	0.81	Transcript to be confirmed: exonic cryptic ASS	N.A.	
TTN	c.25808-1G>A	VUS	1.00	Single transcript: exonic cryptic ASS (ex105 lacks the first 2 nt)	N.A.	p.8603Glu del/out of frame and PTC 8625
TTN	c.41609-2A>C	VUS	0.99	Single transcript: exonic cryptic ASS (ex227 lacks the first 3 nt)	Transcript: correct splicing Transcript: exonic cryptic ASS	p.13870Gludel/In-frame
Donor splice site						
DSP	c.273+5G>A	VUS	0.22	Transcript: correct splicing Transcript: intronic cryptic DSS (the first 62nt of intron 2 are retained)	Correct splicing	PTC
DSP	c.273+3A>T	VUS	0.42	Transcript: correct splicing Transcript: intronic cryptic DSS (the first 62nt of intron 2 are retained) Transcript: ex2 skipping	N.A.	PTC
FLNC	c.7251+1G>A	Path	0.99	Exon 43 skipping	Correct splicing	PTC

Continued

GENE	VARIANT	ACMG/AMP Classification	SpliceAI Δscore loss	Minigene assay results	Observed effect in blood analysis	Protein effect
MYLK	c.3703+22T>C	VUS	0.00	Correct splicing	N.A.	No change
NEXN	c.27+9C>T	VUS	0.00	Correct splicing	NA	No change
NEXN	c.1053+1G>A	LP	0.98	Transcript: intronic cryptic DSS (the first 168nt of intron 9 are retained)	Correct splicing	In-frame - PTC

				Transcript: ex9 skipping		
PKD1	c.10405+5G>T	VUS	0.37	Transcript: intron 31 retention (87nt) - ex33 skipping Transcript: ex33 skipping	N.A.	In-frame (ins29 a.a.) - PTC
RYR2	c.1961+3A>T	VUS	0.96	Single transcript: intronic cryptic DSS (the first 31nt of intron 19 are retained)	N.A.	PTC
SCN5A	c.3840+5G>C	VUS	0.99	Transcript: intronic cryptic DSS (the first 51nt of intron 21 are retained) Transcript: intronic cryptic DSS (the first 47nt of intron 21 are retained)	N.A.	In-frame (ins17 a.a.) - PTC
TMEM43	c.882+8 C>T	VUS	0.00	Correct splicing	N.A.	No change
TTN	c.669+4T>C	VUS	0.00	Correct splicing	No aberration	No change
TTN	c.13282+1G>A	VUS	1.00	Transcript: exonic cryptic DSS (ex46 lacks the last 82 nt) Transcript: intronic cryptic DSS (the first 132 nt of intron 46 are retained)	N.A.	p.Asp4428_Val4472Ins / PTC
TTN	c.19501+2T>C	VUS	0.48	Transcript: exonic cryptic DSS (ex78 Δ175nt)	Transcript: Full-length Transcript: partial skipping of ex8	p.Gly6482_Ala6501del

Continued

GENE	VARIANT	ACMG/AMP Classification	SpliceAI Δscore loss	Minigene assay results	Observed effect in blood analysis	Protein effect
TTN	c.32875+2T>C	VUS	0.88	Transcript: ex141 skipping	Detect different transcripts	p.Val10931_Lys10958del
TTN	c.39574+7A>G	VUS	0.00	Transcript: correct splicing Minor transcript: ex144 skipping	N.A.	p.Try13090_Ile13192del
TTN	c.50346+3A>G	VUS	0.03	Correct splicing	No aberration	No change
TTN	c.57847+4delGTAA	VUS	0.98	Transcript: intronic cryptic DSS (the first 779bp of hybrid intron are retained) Transcript: correct splicing	Ongoing experiment	PTC
VCL	c.622+4C>G	VUS	0.00	Correct splicing	N.A.	p.Try13090_Ile13192del

Table 17. Summary of splicing analysis of 24 variants analysed. Classification of variants according to ACMG/AMP guidelines. SpliceAI prediction, with Δscore ranging from 0 to 1, with a Δscore value ≥ 0.22 which indicates an association to affect splicing. Observed effects in minigenes, blood and protein are determined by cDNA analysis. Abbreviation: Path, Pathogenic, LP, Likely Pathogenic, VUS, variants of uncertain significance, PTC, premature termination codon; p., protein; N.A., not available.

4.1 Splicing analysis directly from patient blood

The straightforward approach to validate splicing dysregulation is based on direct mRNA analysis in patient-derived cells, but in the context of CVDs, relevant tissue is rarely accessible. Therefore, blood was used as surrogate tissue for transcript analysis in 13 cases. Conventional RT-PCR followed by agarose gel electrophoresis and Sanger sequencing has been traditionally applied for determining the variant pathogenicity (Lord and Baralle, 2021).

In good agreement with the minigene assay results (HEK293T and AC16 cell lines), in three blood samples of mutation carriers, mis-splicing events were detected. For *LMNA* c.1609-1 G>A, the presence of exon 10 skipping transcript as well as the correct splicing was observed in the patient sample. While for the c.41609-2 A>C variant in *TTN,* which causes the activation of a cryptic 3' ss in minigene assay (both HEK293T and AC16 cell lines), the event was observed in patient blood sample after sub-cloning the PCR product in pGME-T easy vector due to the agarose gel not resolving 3nt differences. The *TTN* c.19501 +1 G>A variant results in the activation of a cryptic 5'ss in the transcript which was not identified in minigene assay due to the presence of abnormal transcripts that can depend on the minigene construction. However, partial deletion of exon 8 was identified in analysis of patient blood. Further analysis is required to clarify the effect of the variant on splicing and predict the possible consequences on the reading frame.

No conclusion was reported for the c.32875+2 T>C variant in *TTN (*which is still under investigation in blood), but minigene assay showed that the variant disrupts the canonical 5'ss which causes exon 141 skipping. In two cases, *DMD* c.6118-3 C>A and *FLNC* c.3791 G>A, RNA splicing analysis was not possible to investigate due to the low gene expression level in blood. However, four variants, *DSP* c.273+5 G > A, *FLNC* c.7251+1 G > A, and *NEXN* c.1053+1 G>A, *TTN* c.57847+4delGTAA showed the correctly spliced transcript. In contrast, minigene assay tested in both cell line showed that the variant leads the usage of a cryptic 5'ss which introduces a PTC in the transcripts. This observation highlights that direct RT-PCR has limitations in detecting aberrant transcripts when the variant causes out-of frame effects in the encoded transcript, resulting in targeting by NMD. NDM represents a limitation of aberrant splicing detection in blood and is known to contribute to uncertainty in quantification of usage of aberrant splice events (Miller and Pearce, 2014; Nguyen, Wilkinson and Gecz, 2014). This should be taken into consideration when an assay is assessed to confirm variant pathogenicity.

In order to investigate the activation of NMD pathway in blood and possibly predict that the variants perturbed mRNA processing, gene expression levels were analysed by RT-qPCR for

three variants (*DSP* c.273+5 G>A, *LMNA* c.1609-1 G>A, and *NEXN* c.1053+1 G>A).*LMNA* c.1609-1 G>A variant, which showed aberrant splicing in both blood sample and minigene assay was further analysed. However, no mis-splicing events were observed in *DSP* c.273+5 G>A and *NEXN* c.1053+1 G>A patient's blood samples compared to the control samples. This was in contrast to the outcomes from minigene assays, which detected that the variants induced the disruption of the canonical GT ss and the activation of a cryptic donor site immediately downstream (Table 17). After performing gene expression analysis, a reduction of transcript levels compared to the control group was observed for all analysed variants. In the case of *DSP* c.273+5 G>A the reduction was similar to the lowest expression transcript in the healthy control group. Therefore, further investigation is necessary to determine whether the different isoforms produced by the variant influence the function of the wild-type protein or whether the low expression level is more relevant to pathology. The gene expression level of *NEXN* c.1053+1 G>A was lower than the control group, indicating that the variant causes an aberrant splicing event that results in NMD. *LMNA* c.1609-1 G>A, which induces the skipping of exon 10 with an in-frame transcript, showed a large difference in expression level of transcripts, which was unexpected. This may be due to low stability of the mRNA, due to the absence of the last portion of the transcript and a shift in position of the PTC. Therefore, this data suggests that RT-qPCR should be a valid approach to detect if a variant is associated with splicing abnormalities based on gene expression levels.

RNA derived from blood was also subject to RNA-seq which is integrated into some clinical diagnostic laboratories as an approach for characterised alternative splicing patterns (Wai *et al.*, 2022). RNA-seq analysis was performed by generating 70 million, 150bp paired-end reads per sample. It should be noted that short-read sequencing may have limited coverage which is crucial for detection of mis-splicing events expressed at lower fractions, but this can be achieved by applying long-read sequencing as an alternative. However, the application of long-read sequencing would mean (at least currently), that the cost of sequencing increases and this may not be cost effective for the healthcare system (Dragoš *et al.*, 2021). In this study, evidence of mis-splicing events was detected in 3 samples (*LMNA* c.1609-1 G>A, *NEXN* c.1053+1 G>A, and *TTN* c.41609-2 A>C), in accordance with the data observed by minigene assay and direct RT-PCR in blood. While in 4 samples (*DMD* c.6118-3 C>A, *DSP* c.273+5 G>A, *FLNC* c.3791-1G>A, and *FLNC* c.7251+1 G>A) RNA-seq failed due to low gene expression. However, RNA-seq is a reliable method for detecting miss-splicing events in clinical samples, with the limitation that it is not informative in cases where gene has a low TPM value in blood. For these cases, application of an *in vitro* approach is indicated in order to confirm the pathogenicity of variants.

4.2 Variant classification following functional studies

In this study, 17 variants have been reported to alter splicing via several mechanisms. While some variants caused the total absence of canonical transcripts, others resulted in differently spliced isoform transcripts. After performing functional analysis, the variants were reclassified considering the PVS1 criterion, which was defined as “null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss-of-function (LoF) is a known mechanism of disease” (Richards *et al.*, 2015). BS3 criterion for variants that do not influence splicing and the PS3 criterion for variants that cause aberrant splicing. The results from minigene assays upgraded 7 variants (*DSP* c.273+3 A>T, *DSP* c.275+5, *RYR2* c.1961+3 A>T, *SCN5A* c.3840+5 G>C, *TTN* c.13282+1 G>A, *TTN* c.25808-1 G>A, and *TTN* c.57847+4delGTAA) from VUS to Path, with all the transcripts generated by these variants being frameshift transcripts. While for 12 VUS variants, the classification remained unchanged. The *NEXN* c.1053+1 G>A variant did not produce the canonical transcript, but the variant produces a frameshift transcript. According to the ACMG/AMP criteria, this variant should be classified from LP to Path. On the other hand, for the remaining LP as well as Path variants, the initial ACMG/AMP classification was confirmed (Table 18). However, the minigene assay was proven to be an appropriate strategy to characterise the effect of putative splice variants. In particular, this system enabled the re-classification of 37% of the VUS into Pathogenic, thereby improving genetic diagnosis.

			ACMG CLASSIFICATION	
	GENE	**VARIANT**	**Prior functional validation by minigene assay**	**After functional validation by minigene assay**
Acceptor donor site	*DMD*	c.6118-3C>A	**VUS** (out-of-frame), PP3_Yes, PM2_No, PS4_Moderate	**VUS** (out-of-frame), PP3_Yes, PM2_No, PS4_Moderate
	FLNC	c.3791-1G>A	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Moderate	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Moderate
	LMNA	c.1609-1G>A	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Supporting	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Supporting
	MYBPC3	c.852-1G>A	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Moderate	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Moderate
	TTN	c.25808-1G>A	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No	**Pathogenic** Out-of-frame - PVS1, PP3_Yes, PM2_No
	TTN	c.41609-2A>C	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No
Donor splice site	*DSP*	c.273+3A>T	**VUS** (out-of-frame), PP3_Yes, PM2_No	**Pathogenic** Out-of-frame_PVS1, PP3_Yes, PM2_No
	DSP	c.273+5G>A	**VUS** (out-of-frame), PP3_Yes, PM2_NFE-0.05%, PS4_Moderate	**Pathogenic** Out-of-frame_PVS1, PP3_Yes, PM2_NFE-0.05%, PS4_Moderate
	FLNC	c.7251+1G>A	**Pathogenic** PVS1_Out-of-frame, PP3_Yes, PM2_No, PS4_Moderate, PS3_Moderate	**Pathogenic** PVS1_Out-of-frame, PP3_Yes, PM2_No, PS4_Moderate, PS3_Moderate
	MYLK	c.3703+22T>C	**VUS** (In-frame), PP3_No, PM2_No	**VUS** (In-frame), PP3_No, PM2_No
	NEXN	c.27+9C>T	**VUS** (In-frame), PP3_No, PM3_No	**VUS** (In-frame), PP3_No, PM3_No
	NEXN	c.1053+1G>A	**LP** In-frame - PVS1_Moderate, PP3_Yes, PM2_No, PS4_Moderate	**Pathogenic** Out-of-frame - PVS1, PP3_Yes, PM2_No, PS4_Moderate
	PKD1	c.10405+5G>T	**VUS** (Out-of-frame), PP3_Yes, PM2_No, PS4_Supporting	**VUS** (Out-of-frame), PP3_Yes, PM2_No, PS4_Supporting
	Continued			

	GENE	VARIANT	Prior functional validation by minigene assay	After functional validation by minigene assay
Donor splice site	*RYR2*	c.1961+3A>T	**VUS** (out-of-frame), PP3_Yes, PM2_No	**Pathogenic** Out-of-frame_PVS1, PP3_Yes, PM2_No
	SCN5A	c.3840+5G>C	**VUS** (In-frame), PP3_Yes, PM2_No	**Pathogenic** PVS1_Out-of-frame, PP3_Yes, PM2_No
	TMEM43	c.882+8 C>T	**VUS** (In-frame), PP3_No, PM3_No	**VUS** (In-frame), PP3_No, PM3_No
	TTN	c.669+4T>C	**VUS** (out-of-frame), PP3_No, PM3_No	**VUS** (out-of-frame), PP3_No, PM3_No
	TTN	c.13282+1G>A	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No	**Pathogenic** Out-of-frame - PVS1, PP3_Yes, PM2_No
	TTN	c.19501+2T>C	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No
	TTN	c.32875+2T>C	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No	**VUS** In-frame - PVS1_Moderate, PP3_Yes, PM2_No
	TTN	c.39574+7A>G	**VUS** (In-frame), PP3_No, PM3_No	**VUS** In-frame - PVS1_Moderate, PP3_No, PM3_No
	TTN	c.50346+3A>G	**VUS** (out-of-frame), PP3_Yes, PM2_No	**VUS** (out-of-frame), PP3_Yes, PM2_No
	TTN	c.57847+4delGTAA	**VUS** (In-frame), PP3_yes, PM2_No	**Pathogenic** Out-of-frame_PVS1, PP3_yes, PM2_No
	VCL	c.622+4C>G	**VUS** (In-frame), PP3_Yes, NFE - 0.009%	**VUS** (In-frame), PP3_Yes, NFE - 0.009%

Table 18. Variant classification based on ACMG/AMP criteria. Variant classification was performed using the ACMG/AMP criteria. In the case of eight variants, variant classification changed after combining the results from functional RNA studies, leading to the application of PVS1. For a description of all evidence criteria please refer to ACMG/AMP guidelines (Richards *et al.*, 2015). Abbreviation: LP, likely Pathogenic; VUS, variant with unknown significance; PVS1, Pathogenic, very strong; PP3, Pathogenic, supporting; PM2, Pathogenic, moderate; PS3, Pathogenic, strong; PS4, Pathogenic, strong; NFE, European - non-Finnish.

In summary, incorporating RNA-related data from functional splicing studies in the ACMG/AMP guidelines improved the ability to classify and study splicing variants with accuracy and completeness, thus avoiding the non-actionable diagnostic endpoint of a VUS. This also reinforces clinical benefits of being able to reliably predict probable mis-splicing events to improve accuracy and completeness of conclusions drawn from RNA diagnostics. The minigene assay provides important information about the nature of variants since it allows functional investigation of almost any change in DNA sequence and identification of all generated transcripts. Furthermore, genes have tissue-specific expression and therefore may be difficult or impossible to amplify directly from a blood sample with more optimal tissue often unavailable. In the context of this study, cardiac disease tissue sample availability is limited (Gonorazky *et al.*, 2019). Despite these considerations, approximately 80% of human genes have been demonstrated to have a basal level of transcription in blood (Liew *et al.*, 2006). Moreover, it appears that analysing splicing events from blood by RT-PCR is certainly sufficient to classify a variant in a proportion of cases. While it cannot be used for variants creating a PTC, where the combination of SpliceAI and minigene assay is recommended for this category of variant. It is also important to highlight that RT-qPCR from blood can be a useful method to identify the possible alteration of spliceosome activity due to the recognition of a low gene expression in patients carrying the variant. However, for cardiac genes which are lowly expressed in blood, variants cannot be analysed using this surrogate tissue and minigene assays for predicting potential pathogenicity are recommended. However, determining exactly what the variant is doing requires a minigene system, which also allows investigation of heterozygous variants. The combination of several analytical approaches is required to avoid false negative or positive predictions, in order to better clarify if the variant is associated with splicing abnormalities and also provide the correct information to categorise the variants.

Given the number of variants analysed in this study, it should be considered whether it is possible to directly infer the potential effect of a variant based solely on its position relative to the canonical intron-exon boundary. However, as can be seen in figure 90, currently the number of variants analysed is too small to draw conclusions, However, as the number of variants for which functional assays are undertaken increases (both from this study and future work), incorporation of this data into training data sets for tools such as SpliceAI, may allow even more accurate estimation of where intronic mutations are more harmless or more deleterious (Figure 90A). The genetic variants were shown to induce different aberrant splicing events, including intronic nucleotide retention, exonic nucleotide deletion, and exon skipping (Figure 90B).

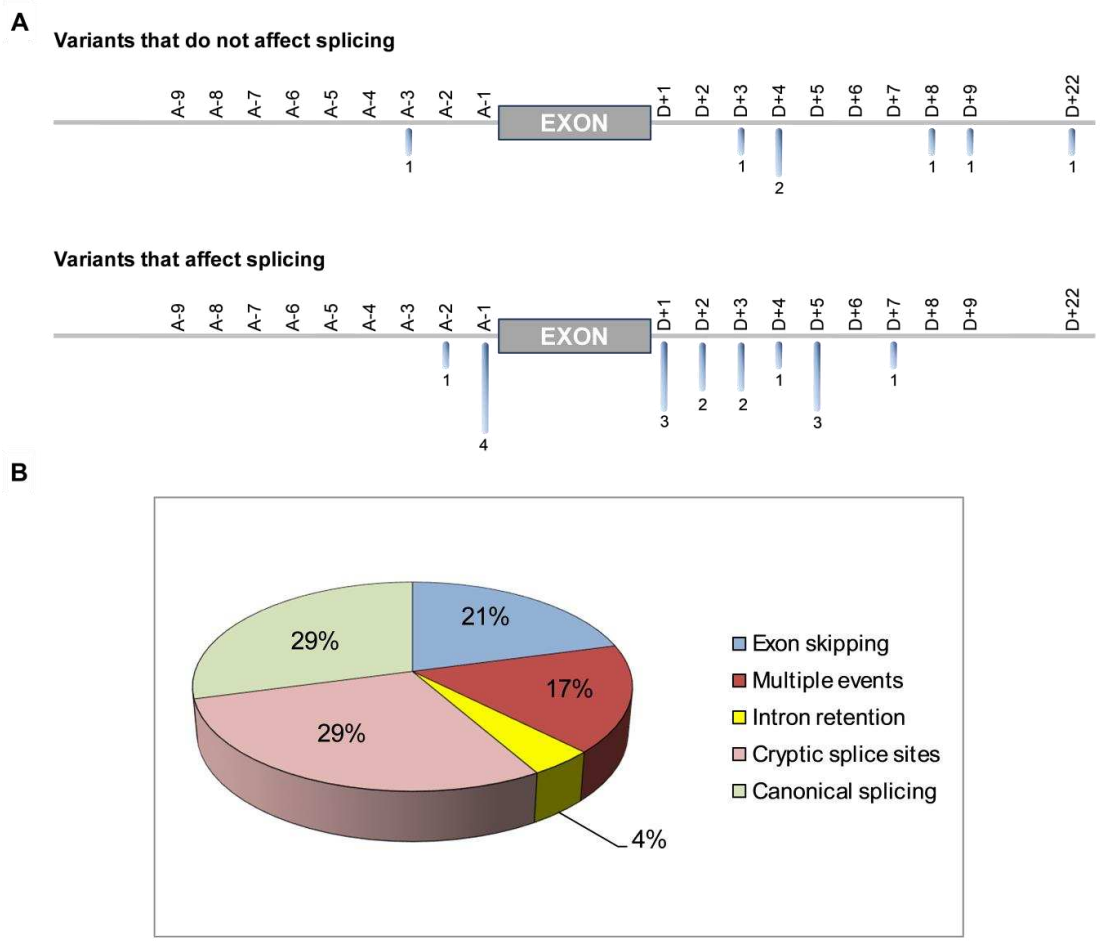


Figure 90. Schematic representation of aberrant splicing. (**A)** Location of the analysed variants. In the upper panel of the cartoon are reported the number of variants that do not affect splicing with the relative position. In the lower panel of the cartoon is reported the number of variants that affect splicing with relative numbers indicated. (**B**) Representation of all the splicing events observed by RNA analysis.

4.3 Analysis of the ability of splicing correcting compounds on restoration of correct splicing profile

As described in the Introduction (section 1.9.6), small-molecule compounds represent a promising strategy for correction of splicing abnormalities and have been applied to recover splicing in different *in vitro* and *in vivo* models of human disease. Their structural differences can result in variations in binding and subsequent different biological effects. However, despite their therapeutic potential, some aspects related to their mechanism of action and interaction with their target are still uncertain. Additionally, it is important to ensure that these compounds are effective and safe, since off targets can impact the crucial mechanisms involved in gene expression regulation.

In this thesis, using six different minigenes carrying variants in both 5' and 3' splice sites, the effect of six compounds and their ability to correct aberrantly spliced transcripts observed in the minigene assay as well as in patient blood samples was investigated. The results demonstrated that kinetin increased correct splicing of the three minigenes carrying the intronic variants, *DSP* c.273+5 G>A, *TTN* c.39574+7 A>G, and *TTN* c.57847+4delGTAA. These results suggest that the effect of kinetin on splicing is not correlated with the nature of the mutation. To date, the suggested mechanism of action of kinetin has been to enhance the interaction between the donor splice site and U1 snRNP (Slaugenhaupt *et al.*, 2004). It will be interesting to see if this compound also affects naturally occurring alternative splicing events and to what extent. A comparison of the variants analysed thus far shows that kinetin can modulate the binding of U1 snRNP at the canonical donor splice site, especially in the cases of *DSP* c.273+5 G>A, *TTN* c.39574+7 A>G, and *TTN* c.57847+4delGTAA variant.

6-Benzyladenine increased correct mRNA processing in four minigenes carrying the mutations tested, albeit to varying degrees. Comparing the results, a significant response was observed in pTB *DSP* c.273+5 G>A, pTB *DSP* c.273+3 A>T, pTB *TTN* c.39574+7 A>G, pTB *TTN* c.57847+4delGTAA. These findings suggest that 6-benzyladenine likely influences the activation or cellular localisation of specific spliceosome proteins that are dependent on the specific make up of the *cis*-acting elements defining the splice site. Future studies will explore the mechanism of action, which may improve the efficacy of the compound. The effect of Zeatin on the aberrant transcripts in all the analysed minigenes was not efficient, except in pTB *DSP* c.273+5 G>A where a low rescue of exon 2 inclusion in the transcript was observed. Overall, these data suggest that the compounds that belong to the family of N6-substituted adenine derivates known as cytokinin, or plant growth factors, may play a role in splicing modulation. Further investigation will be performed in order to understand the mechanism and why the different effects that they produce among the class of compounds and also for specific mutation types.

Sodium butyrate and valproic acid both have been reported to promote the inclusion of exon 7 in the *SMN2* transcript(Chang *et al.*, 2001; Sumner *et al.*, 2003). In the variants in this study, VPA had no effect, although work is still ongoing for the pTB *TTN* c.57847+4delGTAA mutation. NaBu on the other hand, was observed to produce contrary responses. For pTB *DSP* c.273+3 A>T, pTB *LMNA* c.1609-1 G>A, and pTB *TTN* c.39574+7 A>G no rescue of aberrant splicing was observed, whilst in pTB *DSP* c.273+5 G>A NaBu induced an increase of the aberrantly spliced transcript (Figure 84E). Although the splicing pattern is still to be analysed for both wild-type and mutated pTB *TTN* c.57847+4delGTAA minigenes. This further highlights the potential critical role of the *trans*- splicing factor and the sequence context in modulating splicing rescue activity.

Finally, a compound involved in the rescue of *SMN* gene Branaplam was tested. This compound has been reported to modulate binding between the 5' splice site and U1 snRNP protein (a component of spliceosome). The pTB *DSP* minigene carrying the mutation located at +5 of the donor site of exon 2 was observed to induce intron retention, which was dramatically decreased after Branaplam treatment (Figure 84C). The same response was detected in the minigene carrying the variant c.273+3 A>T in *DSP*. This may be due to modulation of the U1 snRNP binding at the canonical donor splice site, increasing efficient recognition of the cryptic 5' splice site. The pTB *TTN* minigene carrying the mutation at +7 of exon 144 shows an increase in the exon 144 skipping event, as had been seen in the wt minigene. Branaplam was also tested in minigene carrying a single point mutation in 3' ss, pTB *LMNA* c.1609-1 G>A, which was observed a slight rescue of exon 10 inclusion in the transcript. This data demonstrates that helping U1 snRNP to land on the 5' ss is a major avenue to help rescue of splicing impairments.

The results obtained from these preliminary experiments show that further studies into the mechanism of action of these compounds will no doubt lead to better understanding of fundamental regulatory processes controlling tissue-specific gene expression.

Chapter 5 Conclusion

This project has established the value of applying functional splicing analysis to clarify the pathological effects of genetic variations linked to mis-splicing events and improve the accuracy of variant classification.

This data supports the notion that only a thorough investigation of genetic variants by employing both in silico and experimental approaches (*in vitro* and *in vivo*) provides beneficial findings to correctly interpret genetic variants, particularly in patients with cardiovascular disorders. Therefore, future research should be focused on integrating splicing analysis with the current ACMG/AMP guidelines. The ability to accurately predict the effect of a variant is particularly crucial for genetic variants classified as VUS.

RNA analysis from accessible clinical samples, such as blood, can be a valuable tool to investigate variants linked with splicing abnormalities. Both RT-PCR and RNA-seq can identify mis-splicing events, however, these approaches are not applicable for splicing analysis in the cases of insufficient gene expression, and when variants introduce a premature termination codon. In the case of transcripts targeted by NMD, RT-qPCR provides substantial evidence in detecting low gene expression levels, confirming that variants influence canonical splicing and that aberrant transcripts are targeted by NMD. Despite the importance of performing a direct RNA analysis from blood, significant issues with detecting aberrant transcripts or tissue-gene expression can make mRNA analysis difficult to assess. Therefore, minigene assays are recommended to provide a reliable conclusion on the pathogenicity of variants. For initial analysis, minigene assay was assessed in HEK293T cells, and no discrepancy was observed compared with the splicing outcomes from AC16 cardiac cell line.

Each technical approach to assessing potential effects of variants on splicing has different strengths and weaknesses, but nonetheless applying RNA analysis is promising for characterising splicing variants, and its integration is crucial to create a robust guideline for the interpretation of different variants involved in genetic disorders.

The final part of this study was focused on possible approaches to rescue mis-splicing events, by application of small molecule splicing modulators. Minigenes carrying variants at the donor and acceptor splice sites were used to investigate the potential effects of six compounds on rescuing correct exon inclusion of different genes. The analysis of splicing profile after treatment, revealed that cytokin-based compounds and Branaplam showed high efficiency in rescue of mis-splicing events, while valproic acid and sodium butyrate did not induce any notable response. This approach, already tested for the correction of splicing mutations in

previous studies, shows great promise for future treatment of disease caused by splicing mutation. Further development of these compounds will extend in the therapy of human genetic diseases.

Appendix A FastQC raw RNA-seq data

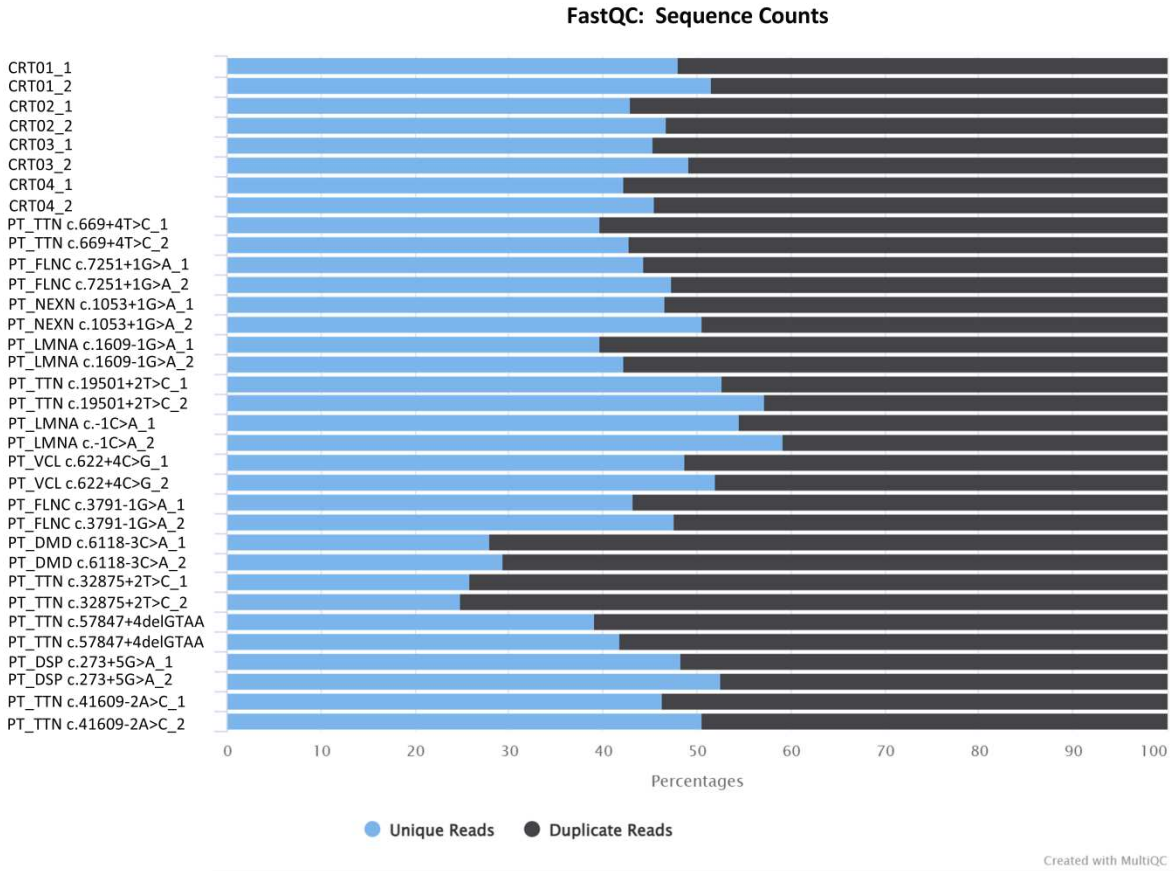


Figure 91. FastQC analysis of the RNA-seq data from each sample. Results from FastQC analysis display the percentage of unique reads, and duplicate reads for each sample in the raw RNA-seq data.

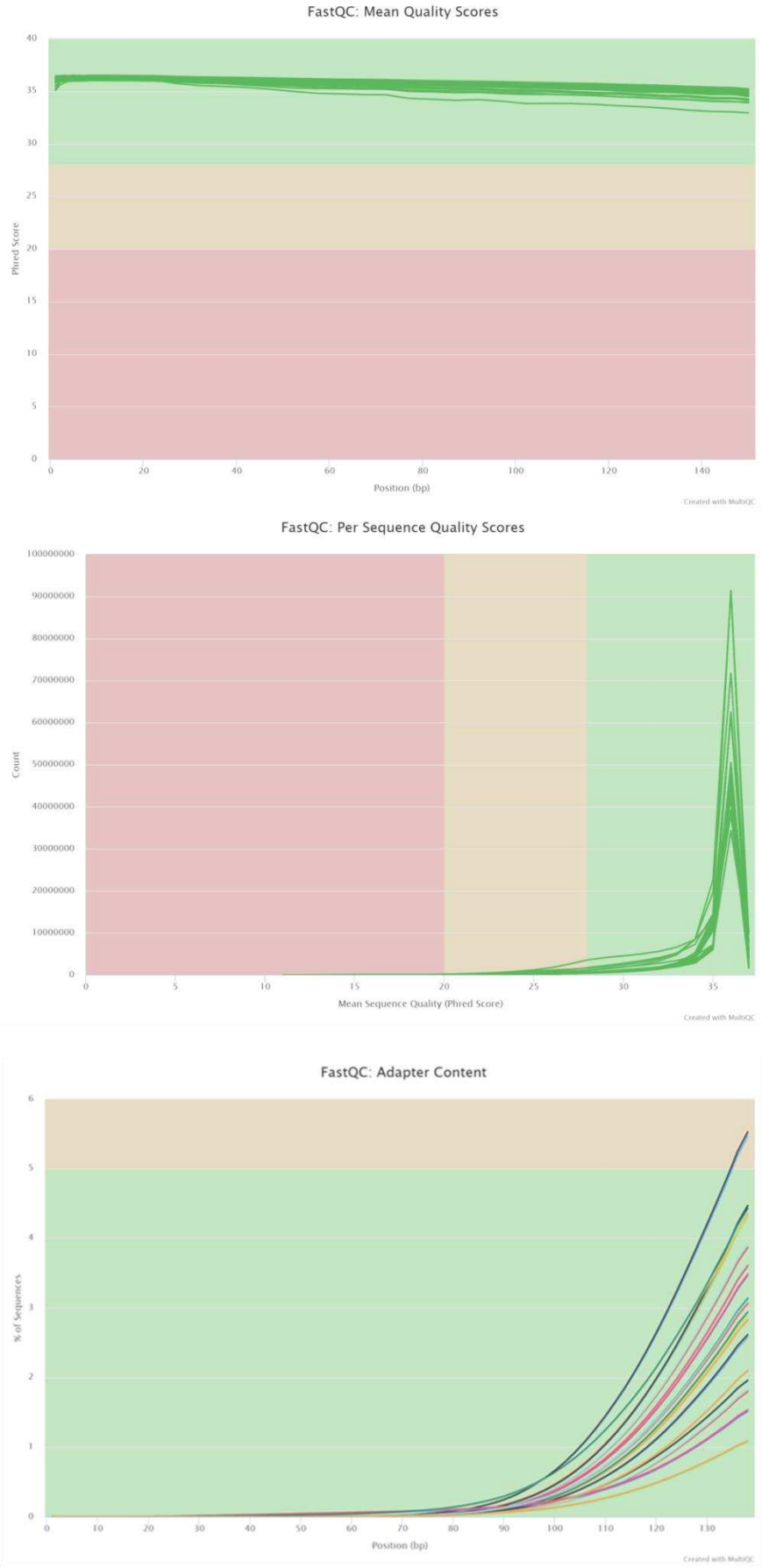


Figure 92. Read quality scores and adapter content. Results after FastQC analysis on the raw RNA-seq data. Figures are divided into three colour sections; red colour low quality, orange colour medium quality, and green colour good quality.

Appendix B Alignment scores of RNA-seq data

Sample Name	% Aligned	M Aligned
CRT01	83.1%	58.7
CRT02	85.3%	60.9
CRT03	86.2%	61.0
CRT04	84.6%	60.4
PT_TTN c.669+4T>C	85.2%	61.3
PT_FLNC c.7251+1G>A	84.8%	59.5
PT_NEXN c.1053+1G>A	84.0%	59.7
PT_LMNA c.1609-1G>A	84.3%	76.5
PT_TTN c.19501+2T>C	85.6%	60.4
PT_LMNA c-1C>A	85.3%	60.2
PT_VCL c.622+4C>G	85.5%	63.0
PT_FLNC c.3791-1G>A	82.2%	57.8
PT_DMD c.6118-3C>A	60.1%	40.5
PT_TTN c.32875+2T>C	57.0%	73.4
PT_TTN c.57847+4delGTAA	81.5%	118.5
PT_DSP c.273+5G>A	87.0%	62.2
PT_TTN c.41609-2A>C	86.9%	62.5

Table 19. STAR. %Alignment (Description: % Uniquely mapped reads); M Aligned (Uniquely mapped reads (millions)).

Appendix C Genomic features with HTseq

Sample Name	% Assigned	M Assigned
CRT01_intron	31.7%	54.5
CRT02_intron	30.7%	53.7
CRT03_intron	35.8%	59.6
CRT04_intron	29.8%	53.0
PT_TTN c.669+4T>C_intron	30.1%	54.4
PT_FLNC c.7251+1G>A_intron	31.3%	55.0
PT_NEXN c.1053+1G>A_intron	35.0%	60.3
PT_LMNA c.1609-1G>A_intron	34.9%	79.6
PT_TTN c.19501+2T>C_intron	39.3%	66.9
PT_LMNA c-1C>A_intron	40.4%	67.8
PT_VCL c.622+4C>G_intron	37.3%	64.9
PT_FLNC c.3791-1G>A_intron	33.5%	65.9
PT_DMD c.6118-3C>A_intron	14.6%	36.6
PT_TTN c.32875+2T>C_intron	17.7%	77.0
PT_TTN c.57847+4delGTAA_intron	32.0%	124.3
PT_DSP c.273+5G>A_intron	42.2%	70.9
PT_TTN c.41609-2A>C_intron	34.2%	58.5

Table 20. HTSeq Count. % Assigned (description: % Assigned reads), M Assigned (description: Assigned Reads (millions))

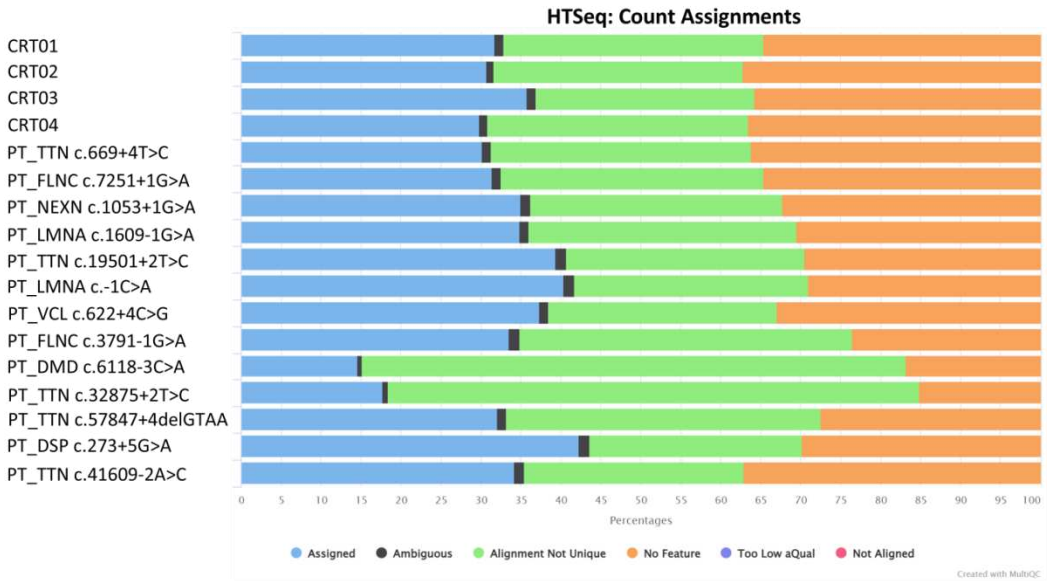


Figure 93. Percentage of reads mapped to genomic features with HTseq.

Bibliography

Abou Tayoun, A. N. *et al.* (2018) 'Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion', *Human Mutation*, 39(11). doi: 10.1002/humu.23626.

Ackerman, M. J. (2015) 'Genetic purgatory and the cardiac channelopathies: Exposing the variants of uncertain/unknown significance issue', *Heart Rhythm*, 12(11). doi: 10.1016/j.hrthm.2015.07.002.

Adams, D. R. and Eng, C. M. (2018) 'Next-Generation Sequencing to Diagnose Suspected Genetic Disorders', *New England Journal of Medicine*, 379(14). doi: 10.1056/nejmra1711801.

Adams, M. D., Rudner, D. Z. and Rio, D. C. (1996) 'Biochemistry and regulation of pre-mRNA splicing', *Current Opinion in Cell Biology*. doi: 10.1016/S0955-0674(96)80006-8.

Aicher, J. K. *et al.* (2020) 'Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq', *Genetics in Medicine*, 22(7). doi: 10.1038/s41436-020-0780-y.

Alanis, E. F. *et al.* (2012) 'An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects', *Human Molecular Genetics*, 21(11). doi: 10.1093/hmg/dds045.

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-A Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31(2). doi: 10.1093/bioinformatics/btu638.

Anna, A. and Monika, G. (2018) 'Splicing mutations in human genetic disorders: examples, detection, and confirmation', *Journal of Applied Genetics*. doi: 10.1007/s13353-018-0444-7.

Au, E. D. and Farkas, M. H. (2019) 'RNA sequencing and transcriptome analysis', in *Genetics and Genomics of Eye Disease: Advancing to Precision Medicine*. doi: 10.1016/B978-0-12-816222-4.00004-6.

Bang, M. L. *et al.* (2001) 'The complete gene sequence of titin, expression of an unusual ≈700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system', *Circulation Research*. doi: 10.1161/hh2301.100981.

Baralle, D. and Baralle, M. (2005) 'Splicing in action: Assessing disease causing sequence changes', *Journal of Medical Genetics*. doi: 10.1136/jmg.2004.029538.

Baralle, D. and Buratti, E. (2017) 'RNA splicing in human disease and in the clinic', *Clinical Science*, 131(5), pp. 355–368. doi: 10.1042/CS20160211.

Baralle, D., Lucassen, A. and Buratti, E. (2009) 'Missed threads: The impact of pre-mRNA splicing defects on clinical practice', *EMBO Reports*, 10(8). doi: 10.1038/embor.2009.170.

Baralle, M. and Baralle, F. E. (2018) 'The splicing code', *BioSystems*. doi: 10.1016/j.biosystems.2017.11.002.

Barbitoff, Y. A. *et al.* (2020) 'Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage', *Scientific Reports*, 10(1). doi: 10.1038/s41598-020-59026-y.

Barbosa-Morais, N. L. *et al.* (2012) 'The evolutionary landscape of alternative splicing in vertebrate species', *Science*, 338(6114). doi: 10.1126/science.1230612.

Baskin, B. *et al.* (2013) 'TMEM43 mutations associated with arrhythmogenic right ventricular cardiomyopathy in non-Newfoundland populations', *Human Genetics*, 132(11). doi: 10.1007/s00439-013-1323-2.

Begay, R. L. *et al.* (2016) 'FLNC Gene Splice Mutations Cause Dilated Cardiomyopathy', *JACC: Basic to Translational Science*. doi: 10.1016/j.jacbts.2016.05.004.

Behjati, S. and Tarpey, P. S. (2013) 'What is next generation sequencing?', *Archives of Disease in Childhood: Education and Practice Edition*, 98(6). doi: 10.1136/archdischild-2013-304340.

Belkin, A. M. *et al.* (1988) 'Diversity of vinculin/meta-vinculin in human tissues and cultivated cells. Expression of muscle specific variants of vinculin in human aorta smooth muscle cells', *Journal of Biological Chemistry*, 263(14). doi: 10.1016/s0021-9258(18)68688-0.

Bengtsson, L. and Otto, H. (2008) 'LUMA interacts with emerin and influences its distribution at the inner nuclear membrane', *Journal of Cell Science*, 121(4). doi: 10.1242/jcs.019281.

Beqqali, A. (2018) 'Alternative splicing in cardiomyopathy', *Biophysical ReviewsAlternative splicing is an important mechanism used by the cell to generate greater transcriptomic and proteomic diversity from the genome. In the heart, alternative splicing is increasingly being recognised as an important layer of pos*. doi: 10.1007/s12551-018-0439-y.

Berget, S. M. (1995) 'Exon recognition in vertebrate splicing', *Journal of Biological Chemistry*. doi: 10.1074/jbc.270.6.2411.

Bergsma, A. J. *et al.* (2018) 'Alternative Splicing in Genetic Diseases: Improved Diagnosis and Novel Treatment Options', in *International Review of Cell and Molecular Biology*. doi: 10.1016/bs.ircmb.2017.07.008.

Bick, D. *et al.* (2019) 'Case for genome sequencing in infants and children with rare,

undiagnosed or genetic diseases', *Journal of Medical Genetics*. doi: 10.1136/jmedgenet-2019-106111.

Black, D. L. (2003) 'Mechanisms of Alternative Pre-Messenger RNA Splicing', *Annual Review of Biochemistry*. doi: 10.1146/annurev.biochem.72.121801.161720.

Blencowe, B. J. (2000) 'Exonic splicing enhancers: Mechanism of action, diversity and role in human genetic diseases', *Trends in Biochemical Sciences*. doi: 10.1016/S0968-0004(00)01549-8.

Brauch, K. M. *et al.* (2009) 'Mutations in Ribonucleic Acid Binding Protein Gene Cause Familial Dilated Cardiomyopathy', *Journal of the American College of Cardiology*, 54(10). doi: 10.1016/j.jacc.2009.05.038.

Breathnach, R. *et al.* (1978) 'Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries', *Proceedings of the National Academy of Sciences of the United States of America*, 75(10). doi: 10.1073/pnas.75.10.4853.

Brnich, S. E. *et al.* (2019) 'Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework', *Genome Medicine*. doi: 10.1186/s13073-019-0690-2.

Brow, D. A. (2002) 'Allosteric Cascade of Spliceosome Activation', *Annual Review of Genetics*. doi: 10.1146/annurev.genet.36.043002.091635.

Bryen, S. J. *et al.* (2022) 'Prevalence, parameters, and pathogenic mechanisms for splice-altering acceptor variants that disrupt the AG exclusion zone', *Human Genetics and Genomics Advances*, 3(4). doi: 10.1016/j.xhgg.2022.100125.

Buratti, E. (2016) ' The minor spliceosome could be the major key for FUS / TLS mutants in ALS ', *The EMBO Journal*, 35(14). doi: 10.15252/embj.201694763.

Burd, C. G. and Dreyfuss, G. (1994) 'Conserved structures and diversity of functions of RNA-binding proteins', *Science*, 265(5172). doi: 10.1126/science.8036511.

Burge, C. and Karlin, S. (1997) 'Prediction of complete gene structures in human genomic DNA', *Journal of Molecular Biology*, 268(1). doi: 10.1006/jmbi.1997.0951.

Burset, M., Seledtsov, I. A. and Solovyev, V. V. (2000) 'Analysis of canonical and non-canonical splice sites in mammalian genomes', *Nucleic Acids Research*, 28(21). doi: 10.1093/nar/28.21.4364.

Byron, S. A. *et al.* (2016) 'Translating RNA sequencing into clinical diagnostics: Opportunities

and challenges', *Nature Reviews Genetics*. doi: 10.1038/nrg.2016.10.

Cabral, R. M. *et al.* (2010) 'Identification and characterization of DSPIa, a novel isoform of human desmoplakin', *Cell and Tissue Research*, 341(1). doi: 10.1007/s00441-010-0989-1.

Cáceres, J. F., Screaton, G. R. and Krainer, A. R. (1998) 'A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm', *Genes and Development*, 12(1). doi: 10.1101/gad.12.1.55.

Campuzano, O. *et al.* (2015) 'Determining the Pathogenicity of Genetic Variants Associated with Cardiac Channelopathies', *Scientific Reports*, 5. doi: 10.1038/srep07953.

Carrier, L. *et al.* (1997) 'Organization and sequence of human cardiac myosin binding protein C gene (MYBPC3) and identification of mutations predicted to produce truncated proteins in familial hypertrophic cardiomyopathy', *Circulation Research*, 80(3). doi: 10.1161/01.res.0000435859.24609.b3.

Carson, J. H., Cui, H. and Barbarese, E. (2001) 'The balance of power in RNA trafficking', *Current Opinion in Neurobiology*. doi: 10.1016/S0959-4388(00)00249-X.

Cartegni, L. *et al.* (2003) 'ESEfinder: A web resource to identify exonic splicing enhancers', *Nucleic Acids Research*. doi: 10.1093/nar/gkg616.

Cartegni, L., Chew, S. L. and Krainer, A. R. (2002) 'Listening to silence and understanding nonsense: Exonic mutations that affect splicing', *Nature Reviews Genetics*. doi: 10.1038/nrg775.

Chabot, B. and Shkreta, L. (2016) 'Defective control of pre-messenger RNA splicing in human disease', *Journal of Cell Biology*. doi: 10.1083/jcb.201510032.

Chan, R. C. and Black, D. L. (1997) 'The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream', *Molecular and Cellular Biology*, 17(8). doi: 10.1128/mcb.17.8.4667.

Chang, J. G. *et al.* (2001) 'Treatment of spinal muscular atrophy by sodium butyrate', *Proceedings of the National Academy of Sciences of the United States of America*, 98(17). doi: 10.1073/pnas.171105098.

Chaudhury, A., Chander, P. and Howe, P. H. (2010) 'Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles', *RNA*. doi: 10.1261/rna.2254110.

Chen, H., Choudhury, D. M. and Craig, S. W. (2006) 'Coincidence of actin filaments and talin is

required to activate vinculin', *Journal of Biological Chemistry*, 281(52). doi: 10.1074/jbc.M607324200.

Chen, M. and Manley, J. L. (2009) 'Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm2777.

Cheng, J. *et al.* (2019) 'MMSplice: Modular modeling improves the predictions of genetic variant effects on splicing', *Genome Biology*, 20(1). doi: 10.1186/s13059-019-1653-z.

Chou, M.-Y. *et al.* (1999) ' hnRNP H Is a Component of a Splicing Enhancer Complex That Activates a c- src Alternative Exon in Neuronal Cells ', *Molecular and Cellular Biology*, 19(1). doi: 10.1128/mcb.19.1.69.

Christiaans, I. *et al.* (2019) 'Large next-generation sequencing gene panels in genetic heart disease: challenges in clinical practice', *Netherlands Heart Journal*. doi: 10.1007/s12471-019-1251-4.

Chung, C. T., Niemela, S. L. and Miller, R. H. (1989) 'One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.86.7.2172.

Cirino, A. L. and Ho, C. (1993) *Hypertrophic Cardiomyopathy Overview*, *GeneReviews®*.

Conesa, A. *et al.* (2016) 'A survey of best practices for RNA-seq data analysis', *Genome Biology*. doi: 10.1186/s13059-016-0881-8.

De Conti, L., Baralle, M. and Buratti, E. (2013) 'Exon and intron definition in pre-mRNA splicing', *Wiley Interdisciplinary Reviews: RNA*. doi: 10.1002/wrna.1140.

Cooper, T. A. (2005) 'Use of minigene systems to dissect alternative splicing elements', *Methods*. doi: 10.1016/j.ymeth.2005.07.015.

Corvelo, A. *et al.* (2010) 'Genome-wide association between branch point properties and alternative splicing', *PLoS Computational Biology*, 6(11). doi: 10.1371/journal.pcbi.1001016.

Cretu, C. *et al.* (2018) 'Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds', *Molecular Cell*, 70(2). doi: 10.1016/j.molcel.2018.03.011.

Crick, F. (1970) 'Central dogma of molecular biology', *Nature*, 227(5258). doi: 10.1038/227561a0.

Crooke, S. T. (2017) 'Molecular Mechanisms of Antisense Oligonucleotides', *Nucleic Acid Therapeutics*, 27(2). doi: 10.1089/nat.2016.0656.

Cullinane, A. R. *et al.* (2011) 'Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia', *Journal of Investigative Dermatology*, 131(10). doi: 10.1038/jid.2011.157.

Cummings, B. B. *et al.* (2017) 'Improving genetic diagnosis in Mendelian disease with transcriptome sequencing', *Science Translational Medicine*, 9(386). doi: 10.1126/scitranslmed.aal5209.

Daguenet, E., Dujardin, G. and Valcárcel, J. (2015) ' The pathogenicity of splicing defects: mechanistic insights into pre- mRNA processing inform novel therapeutic approaches ', *EMBO reports*, 16(12). doi: 10.15252/embr.201541116.

Dal Mas, A. *et al.* (2015) 'Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA', *American Journal of Human Genetics*, 96(1). doi: 10.1016/j.ajhg.2014.12.009.

Dana, H. *et al.* (2017) 'Molecular Mechanisms and Biological Functions of siRNA.', *International journal of biomedical science : IJBS*, 13(2).

Davidson, M. M. *et al.* (2005) 'Novel cell lines derived from adult human ventricular cardiomyocytes', *Journal of Molecular and Cellular Cardiology*, 39(1). doi: 10.1016/j.yjmcc.2005.03.003.

Dehainault, C. *et al.* (2007) 'A deep intronic mutation in the RB1 gene leads to intronic sequence exonisation', *European Journal of Human Genetics*, 15(4). doi: 10.1038/sj.ejhg.5201787.

Desmet, F. O. and Béroud, C. (2012) 'Bioinformatics and mutations leading to exon skipping', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-767-5_2.

Dharshini, S. A. P., Taguchi, Y. H. and Gromiha, M. M. (2020) 'Identifying suitable tools for variant detection and differential gene expression using RNA-seq data', *Genomics*, 112(3). doi: 10.1016/j.ygeno.2019.12.011.

Dheda, K. *et al.* (2004) 'Validation of housekeeping genes for normalizing RNA expression in real-time PCR', *BioTechniques*, 37(1). doi: 10.2144/04371rr03.

Dickson, A., Osman, E. and Lorson, C. L. (2008) 'A negatively acting bifunctional RNA increases survival motor neuron both in vitro and in vivo', *Human Gene Therapy*, 19(11). doi: 10.1089/hum.2008.067.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1). doi: 10.1093/bioinformatics/bts635.

Dogan, R. I. *et al.* (2007) 'SplicePort-An interactive splice-site analysis tool', *Nucleic Acids Research*. doi: 10.1093/nar/gkm407.

Dolzhenko, E. *et al.* (2017) 'Detection of long repeat expansions from PCR-free whole-genome sequence data', *Genome Research*, 27(11). doi: 10.1101/gr.225672.117.

Donadon, I. *et al.* (2018) 'Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model', *Human Molecular Genetics*, 27(14). doi: 10.1093/hmg/ddy151.

Donadon, I. *et al.* (2019) 'Rescue of spinal muscular atrophy mouse models with AAV9-Exon-specific U1 snRNA', *Nucleic acids research*, 47(14). doi: 10.1093/nar/gkz469.

Dong, Y., Siegwart, D. J. and Anderson, D. G. (2019) 'Strategies, design, and chemistry in siRNA delivery systems', *Advanced Drug Delivery Reviews*. doi: 10.1016/j.addr.2019.05.004.

Douglas, A. G. L. and Baralle, D. (2021) 'Translating RNA splicing analysis into diagnosis and therapy', *OBM Genetics*. doi: 10.21926/obm.genet.2101125.

Dragoš, V. Š. *et al.* (2021) 'New approach for detection of normal alternative splicing events and aberrant spliceogenic transcripts with long-range PCR and deep RNA sequencing', *Biology*, 10(8). doi: 10.3390/biology10080706.

Dreyfuss, G. *et al.* (1993) 'hnRNP proteins and the biogenesis of mRNA', *Annual Review of Biochemistry*. doi: 10.1146/annurev.bi.62.070193.001445.

Dreyfuss, G., Philipson, L. and Mattaj, I. W. (1988) 'Ribonucleoprotein particles in cellular processes', *Journal of Cell Biology*. doi: 10.1083/jcb.106.5.1419.

Eckstein, F. (2014) 'Phosphorothioates, essential components of therapeutic oligonucleotides', *Nucleic Acid Therapeutics*. doi: 10.1089/nat.2014.0506.

Eraslan, G. *et al.* (2019) 'Deep learning: new computational modelling techniques for genomics', *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0122-6.

Estigoy, C. B. *et al.* (2009) 'Intercalated discs: Multiple proteins perform multiple functions in non-failing and failing human hearts', *Biophysical Reviews*. doi: 10.1007/s12551-008-0007-y.

Ewels, P. *et al.* (2016) 'MultiQC: Summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19). doi: 10.1093/bioinformatics/btw354.

Fairbrother, W. G. *et al.* (2002) 'Predictive identification of exonic splicing enhancers in human genes', *Science*, 297(5583). doi: 10.1126/science.1073774.

Fernández, A., Josa, S. and Montoliu, L. (2017) 'A history of genome editing in mammals', *Mammalian Genome*, 28(7–8). doi: 10.1007/s00335-017-9699-2.

Fill, M. and Copello, J. A. (2002) 'Ryanodine receptor calcium release channels', *Physiological Reviews*. doi: 10.1152/physrev.00013.2002.

Fiset, P.-O. and Gounni, A. S. (2001) 'Antisense Oligonucleotides: problems with use and solutions', *Reviews in Biology and Biotechnology*, 1(2).

Van der Flier, A. and Sonnenberg, A. (2001) 'Structural and functional aspects of filamins', *Biochimica et Biophysica Acta - Molecular Cell Research*. doi: 10.1016/S0167-4889(01)00072-6.

Ford, L. P., Wright, W. E. and Shay, J. W. (2002) 'A model for heterogeneous nuclear ribonucleoproteins in telomere and telomerase regulation', *Oncogene*, 21(4). doi: 10.1038/sj.onc.1205086.

Freiburg, A. *et al.* (2000) 'Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity', *Circulation Research*, 86(11). doi: 10.1161/01.RES.86.11.1114.

French, J. D. and Edwards, S. L. (2020) 'The Role of Noncoding Variants in Heritable Disease', *Trends in Genetics*. doi: 10.1016/j.tig.2020.07.004.

Fu, X. D. (1995) 'The superfamily of arginine/serine-rich splicing factors.', *RNA (New York, N.Y.)*.

Fu, X. D. and Ares, M. (2014) 'Context-dependent control of alternative splicing by RNA-binding proteins', *Nature Reviews Genetics*. doi: 10.1038/nrg3778.

George, C. H., Chang, C. Y. and Lai, F. A. (2005) 'Toward a molecular understanding of the structure-function of ryanodine receptor Ca2+ release channels: Perspectives from recombinant expression systems', *Cell Biochemistry and Biophysics*. doi: 10.1385/CBB:42:2:197.

Geuens, T., Bouhy, D. and Timmerman, V. (2016) 'The hnRNP family: insights into their role in health and disease', *Human Genetics*. doi: 10.1007/s00439-016-1683-5.

Giudicessi, J. R. and Ackerman, M. J. (2013) 'Genetic testing in heritable cardiac arrhythmia syndromes: Differentiating pathogenic mutations from background genetic noise', *Current Opinion in Cardiology*. doi: 10.1097/HCO.0b013e32835b0a41.

Gloss, B. S. and Dinger, M. E. (2018) 'Realizing the significance of noncoding functionality in clinical genomics', *Experimental and Molecular Medicine*. doi: 10.1038/s12276-018-0087-0.

Gonorazky, H. D. *et al.* (2019) 'Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease', *American Journal of Human Genetics*, 104(3). doi: 10.1016/j.ajhg.2019.01.012.

Göttlicher, M. *et al.* (2001) 'Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells', *EMBO Journal*, 20(24). doi: 10.1093/emboj/20.24.6969.

Graveley, B. R. (2000) 'Sorting out the complexity of SR protein functions', *RNA*. doi: 10.1017/S1355838200000960.

Guo, W. *et al.* (2012) 'RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing', *Nature Medicine*. doi: 10.1038/nm.2693.

Habelhah, H. *et al.* (2001) 'ERK phosphorylation drives cytoplasmic accumulation of hnRNP-K and inhibition of mRNA translation', *Nature Cell Biology*, 3(3). doi: 10.1038/35060131.

Han, S. P., Tang, Y. H. and Smith, R. (2010) 'Functional diversity of the hnRNPs: Past, present and perspectives', *Biochemical Journal*. doi: 10.1042/BJ20100396.

Harrington, C. A. *et al.* (2020) 'RNA-Seq of human whole blood: Evaluation of globin RNA depletion on Ribo-Zero library method', *Scientific Reports*, 10(1). doi: 10.1038/s41598-020-62801-6.

Harrow, J. *et al.* (2012) 'GENCODE: The reference human genome annotation for the ENCODE project', *Genome Research*, 22(9). doi: 10.1101/gr.135350.111.

Hasegawa, M. *et al.* (2011) 'Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product', *ACS Chemical Biology*, 6(3). doi: 10.1021/cb100248e.

Havens, M. A., Duelli, D. M. and Hastings, M. L. (2013) 'Targeting RNA splicing for disease therapy', *Wiley Interdisciplinary Reviews: RNA*. doi: 10.1002/wrna.1158.

Havens, M. A. and Hastings, M. L. (2016) 'Splice-switching antisense oligonucleotides as therapeutic drugs', *Nucleic Acids Research*. doi: 10.1093/nar/gkw533.

Hebsgaard, S. M. *et al.* (1996) 'Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information', *Nucleic Acids Research*, 24(17), pp. 3439–3452. doi: 10.1093/nar/24.17.3439.

Hellemans, J. *et al.* (2008) 'qBase relative quantification framework and software for

management and automated analysis of real-time quantitative PCR data', *Genome Biology*, 8(2). doi: 10.1186/gb-2007-8-2-r19.

Helms, A. S. *et al.* (2020) 'Spatial and Functional Distribution of MYBPC3 Pathogenic Variants and Clinical Outcomes in Patients With Hypertrophic Cardiomyopathy', *Circulation: Genomic and Precision Medicine*, 13(5). doi: 10.1161/CIRCGEN.120.002929.

Henrie, A. *et al.* (2018) 'ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data', *Human Mutation*, 39(8). doi: 10.1002/humu.23555.

Herman, D. S. *et al.* (2012) 'Truncations of Titin Causing Dilated Cardiomyopathy', *New England Journal of Medicine*, 366(7). doi: 10.1056/nejmoa1110186.

Hershberger, R. E., Morales, A. and Siegfried, J. D. (2010) 'Clinical and genetic issues in dilated cardiomyopathy: A review for genetics professionals', *Genetics in Medicine*. doi: 10.1097/GIM.0b013e3181f2481f.

Hershberger, R. E. and Siegfried, J. D. (2011) 'Update 2011: Clinical and genetic issues in familial dilated cardiomyopathy', *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2011.01.015.

Hims, M. M. *et al.* (2007) 'Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia', *Journal of Molecular Medicine*, 85(2). doi: 10.1007/s00109-006-0137-2.

Hinson, J. T. *et al.* (2015) 'Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy', *Science*, 349(6251). doi: 10.1126/science.aaa5458.

Van Den Hoogenhof, M. M. G., Pinto, Y. M. and Creemers, E. E. (2016) 'RNA Splicing regulation and dysregulation in the heart', *Circulation Research*. doi: 10.1161/CIRCRESAHA.115.307872.

Islam, S. (2021) 'Artificial Intelligence in Healthcare', *International Journal of Engineering Materials and Manufacture*, 6(4). doi: 10.26776/ijemm.06.04.2021.08.

Jaganathan, K. *et al.* (2019) 'Predicting Splicing from Primary Sequence with Deep Learning', *Cell*, 176(3). doi: 10.1016/j.cell.2018.12.015.

Jiménez-García, L. F. and Spector, D. L. (1993) 'In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism', *Cell*, 73(1). doi: 10.1016/0092-8674(93)90159-N.

Kaida, D. *et al.* (2007) 'Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA', *Nature Chemical Biology*, 3(9). doi: 10.1038/nchembio.2007.18.

Kalsotra, A. *et al.* (2008) 'A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart', *Proceedings of the National Academy of Sciences of the United States of America*, 105(51). doi: 10.1073/pnas.0809045105.

Kamdar, F. and Garry, D. J. (2016) 'Dystrophin-Deficient Cardiomyopathy', *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2016.02.081.

Karczewski, K. *et al.* (2019) 'ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020', *bioRxiv*.

Kasprowicz-Maluśki, A. *et al.* (2016) 'Journey from the Center of the Cell - the intra- and intercellular transport of mRNA', *Acta Biochimica Polonica*. doi: 10.18388/abp.2016_1359.

Katsuragi, S. and Ikeda, T. (2018) 'Dilated Cardiomyopathy', in *Maternal and Fetal Cardiovascular Disease*. doi: 10.1007/978-981-10-1993-7_9.

Ke, S. *et al.* (2018) 'Saturation mutagenesis reveals manifold determinants of exon definition', *Genome Research*, 28(1). doi: 10.1101/gr.219683.116.

Keren, H., Lev-Maor, G. and Ast, G. (2010) 'Alternative splicing and evolution: Diversification, exon definition and function', *Nature Reviews Genetics*. doi: 10.1038/nrg2776.

Kessler, M. M. *et al.* (1997) 'Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast', *Genes and Development*, 11(19). doi: 10.1101/gad.11.19.2545.

Khaniani, M. S. *et al.* (2016) 'Sodium butyrate and valproic acid as splicing restoring agents in erythroid cells of β-Thalassemic patients', *Iranian Journal of Biotechnology*, 14(1). doi: 10.15171/ijb.1169.

Kiledjian, M. and Dreyfuss, G. (1992) 'Primary structure and binding activity of the hnRNP U protein: Binding RNA through RGG box', *EMBO Journal*, 11(7). doi: 10.1002/j.1460-2075.1992.tb05331.x.

Kimura, T. *et al.* (2005) 'Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1', *Human Molecular Genetics*, 14(15). doi: 10.1093/hmg/ddi223.

Knott, G. J. and Doudna, J. A. (2018) 'CRISPR-Cas guides the future of genetic engineering', *Science*. doi: 10.1126/science.aat5011.

Koenig, M., Monaco, A. P. and Kunkel, L. M. (1988) 'The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein', *Cell*, 53(2). doi: 10.1016/0092-8674(88)90383-2.

Kolovos, P. *et al.* (2012) 'Enhancers and silencers: An integrated and simple model for their function', *Epigenetics and Chromatin*. doi: 10.1186/1756-8935-5-1.

Konarska, M. M. (1998) 'Recognition of the 5' splice site by the spliceosome', *Acta Biochimica Polonica*. doi: 10.18388/abp.1998_4346.

Konarska, M. M., Vilardell, J. and Query, C. C. (2006) 'Repositioning of the reaction intermediate within the catalytic center of the spliceosome', *Molecular Cell*. doi: 10.1016/j.molcel.2006.01.017.

Kong, S. W. *et al.* (2010a) 'Heart Failure-Associated Changes in RNA Splicing of Sarcomere Genes', *Circulation: Cardiovascular Genetics*, 3(2). doi: 10.1161/CIRCGENETICS.109.904698.

Kong, S. W. *et al.* (2010b) 'Heart Failure-Associated Changes in RNA Splicing of Sarcomere Genes', *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.109.904698.

Kotake, Y. *et al.* (2007) 'Splicing factor SF3b as a target of the antitumor natural product pladienolide', *Nature Chemical Biology*, 3(9). doi: 10.1038/nchembio.2007.16.

KOTELIANSKY, V. E. *et al.* (1992) 'An additional exon in the human vinculin gene specifically encodes meta-vinculin-specific difference peptide: Cross-species comparison reveals variable and conserved motifs in the meta-vinculin insert', *European Journal of Biochemistry*, 204(2). doi: 10.1111/j.1432-1033.1992.tb16692.x.

Ku, C. S. *et al.* (2013) 'A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease', *Molecular Psychiatry*. doi: 10.1038/mp.2012.58.

Kukurba, K. R. and Montgomery, S. B. (2015) 'RNA sequencing and analysis', *Cold Spring Harbor Protocols*, 2015(11). doi: 10.1101/pdb.top084970.

Kuramoto, E. *et al.* (1992) 'Oligonucleotide Sequences Required for Natural Killer Cell Activation', *Japanese Journal of Cancer Research*, 83(11). doi: 10.1111/j.1349-7006.1992.tb02734.x.

Lahmers, S. *et al.* (2004) 'Developmental Control of Titin Isoform Expression and Passive Stiffness in Fetal and Neonatal Myocardium', *Circulation Research*, 94(4). doi: 10.1161/01.RES.0000115522.52554.86.

Le, K. Q. *et al.* (2015) 'Alternative splicing as a biomarker and potential target for drug discovery', *Acta Pharmacologica Sinica*. doi: 10.1038/aps.2015.43.

LeCun, Y., Bengio, Y., Hinton, G. (2015) 'Deep learning. nature 521 (7553): 436', *Nature*, 521.

Lee, M. S., Henry, M. and Silver, P. A. (1996) 'A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export', *Genes and Development*, 10(10). doi: 10.1101/gad.10.10.1233.

Lek, M. *et al.* (2016) 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature*, 536(7616). doi: 10.1038/nature19057.

Liew, C. C. *et al.* (2006) 'The peripheral blood transcriptome dynamically reflects system wide biology: A potential diagnostic tool', *Journal of Laboratory and Clinical Medicine*, 147(3). doi: 10.1016/j.lab.2005.10.005.

Lin, S. *et al.* (2014) 'Comparison of the transcriptional landscapes between human and mouse tissues', *Proceedings of the National Academy of Sciences of the United States of America*, 111(48). doi: 10.1073/pnas.1413624111.

Litzkas, P., Jha, K. K. and Ozer, H. L. (1984) 'Efficient transfer of cloned DNA into human diploid cells: protoplast fusion in suspension.', *Molecular and Cellular Biology*, 4(11). doi: 10.1128/mcb.4.11.2549.

López-Bigas, N. *et al.* (2005) 'Are splicing mutations the most frequent cause of hereditary disease?', *FEBS Letters*. doi: 10.1016/j.febslet.2005.02.047.

Lord, J. and Baralle, D. (2021) 'Splicing in the Diagnosis of Rare Disease: Advances and Challenges', *Frontiers in Genetics*. doi: 10.3389/fgene.2021.689892.

Maddirevula, S. *et al.* (2020) 'Analysis of transcript-deleterious variants in Mendelian disorders: Implications for RNA-based diagnostics', *Genome Biology*, 21(1). doi: 10.1186/s13059-020-02053-9.

Majewski, J. *et al.* (2011) 'A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: Novel use of genomic diagnostics and exome sequencing', *Journal of Medical Genetics*, 48(9). doi: 10.1136/jmedgenet-2011-100288.

Makarenko, I. *et al.* (2004) 'Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts', *Circulation Research*, 95(7). doi: 10.1161/01.RES.0000143901.37063.2f.

Man, E. *et al.* (2013) 'NGS identifies TAZ mutation in a family with X-linked dilated cardiomyopathy.', *BMJ case reports*, 2013. doi: 10.1136/bcr-2012-007529.

Manley, J. L. and Krainer, A. R. (2010) 'A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)', *Genes and Development*. doi: 10.1101/gad.1934910.

Mantere, T., Kersten, S. and Hoischen, A. (2019) 'Long-read sequencing emerging in medical genetics', *Frontiers in Genetics*. doi: 10.3389/fgene.2019.00426.

Marco-Puche, G. *et al.* (2019) 'RNA-Seq Perspectives to Improve Clinical Diagnosis', *Frontiers in Genetics*. doi: 10.3389/fgene.2019.01152.

Martínez-Sánchez, M. *et al.* (2019) 'Analysis of housekeeping genes in the peripheral blood of retinoblastoma patients', *bioRxiv*. doi: 10.1101/693101.

Matera, A. G. and Wang, Z. (2014) 'A day in the life of the spliceosome', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3742.

Mattick, J. S. *et al.* (2018) 'Whole genome sequencing provides better diagnostic yield and future value than whole exome sequencing', *Medical Journal of Australia*, 209(5). doi: 10.5694/mja17.01176.

Mattioli, C. *et al.* (2014) 'Unusual splice site mutations disrupt FANCA exon 8 definition', *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1842(7). doi: 10.1016/j.bbadis.2014.03.014.

McNally, E. M. and Mestroni, L. (2017) 'Dilated cardiomyopathy: Genetic determinants and mechanisms', *Circulation Research*. doi: 10.1161/CIRCRESAHA.116.309396.

Meissner, G. (2004) 'Molecular regulation of cardiac ryanodine receptor ion channel', *Cell Calcium*, 35(6). doi: 10.1016/j.ceca.2004.01.015.

Meister, G. and Tuschl, T. (2004) 'Mechanisms of gene silencing by double-stranded RNA', *Nature*. doi: 10.1038/nature02873.

Meynert, A. M. *et al.* (2014) 'Variant detection sensitivity and biases in whole genome and exome sequencing', *BMC Bioinformatics*, 15(1). doi: 10.1186/1471-2105-15-247.

Miau, L. H. *et al.* (1998) 'Identification of heterogeneous nuclear ribonucleoprotein K (hnRNP K) as a repressor of C/EBPβ-mediated gene activation', *Journal of Biological Chemistry*, 273(17). doi: 10.1074/jbc.273.17.10784.

Miller, J. N. and Pearce, D. A. (2014) 'Nonsense-mediated decay in genetic disease: Friend or foe?', *Mutation Research - Reviews in Mutation Research*. doi: 10.1016/j.mrrev.2014.05.001.

Mogensen, J. *et al.* (2015) 'The current role of next-generation DNA sequencing in routine care of patients with hereditary cardiovascular conditions: A viewpoint paper of the European Society of Cardiology working group on myocardial and pericardial diseases and members of the Europ', *European Heart Journal*. doi: 10.1093/eurheartj/ehv122.

Moles-Fernández, A. *et al.* (2021) 'Role of splicing regulatory elements and in silico tools usage in the identification of deep intronic splicing variants in hereditary breast/ovarian cancer genes', *Cancers*, 13(13). doi: 10.3390/cancers13133341.

Motulsky, H. J. (2021) 'GraphPad Prism 9 Statistics Guide - Tukey and Dunnett methods.', in *GraphPad Statistics Guide*.

Mourelatos, Z. *et al.* (2001) 'SMN interacts with a novel family of hnRNP and spliceosomal proteins', *EMBO Journal*, 20(19). doi: 10.1093/emboj/20.19.5443.

Muntoni, F., Torelli, S. and Ferlini, A. (2003) 'Dystrophin and mutations: One gene, several proteins, multiple phenotypes', *Lancet Neurology*. doi: 10.1016/S1474-4422(03)00585-4.

Murdock, D. R. (2020) 'Enhancing Diagnosis Through RNA Sequencing', *Clinics in Laboratory Medicine*. doi: 10.1016/j.cll.2020.02.001.

Murdock, D. R. *et al.* (2021) 'Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing', *Journal of Clinical Investigation*, 131(1). doi: 10.1172/JCI141500.

Na, D. E. C. and Hipertensiva, C. (no date) 'No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title'. Available at: https://cshprotocols.cshlp.org/content/2006/1/pdb.rec390.full?text_only=true.

Nachtergaele, S. and He, C. (2017) 'The emerging biology of RNA post-transcriptional modifications', *RNA Biology*. doi: 10.1080/15476286.2016.1267096.

Naito, T. (2019) 'Predicting the impact of single nucleotide variants on splicing via sequence-based deep neural networks and genomic features', *Human Mutation*, 40(9). doi: 10.1002/humu.23794.

Neagoe, C. *et al.* (2003) 'Gigantic variety: Expression patterns of titin isoforms in striated muscles and consequences for myofibrillar passive stiffness', in *Journal of Muscle Research and Cell Motility*. doi: 10.1023/A:1026053530766.

Nguyen, L. S., Wilkinson, M. F. and Gecz, J. (2014) 'Nonsense-mediated mRNA decay: Inter-individual variability and human disease', *Neuroscience and Biobehavioral Reviews*. doi: 10.1016/j.neubiorev.2013.10.016.

NHS (2021) 'Testing Criteria for Rare and Inherited Disease', *National Genomic Test Directory*, (October).

Nigro, J. M. *et al.* (1991) 'Scrambled exons', *Cell*, 64(3). doi: 10.1016/0092-8674(91)90244-S.

Nishio, H. *et al.* (1994) 'Identification of a novel first exon in the human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter', *Journal of Clinical Investigation*, 94(3). doi: 10.1172/JCI117417.

O'Brien, K. *et al.* (2008) 'The biflavonoid isoginkgetin is a general inhibitor of pre-mRNA splicing', *Journal of Biological Chemistry*, 283(48). doi: 10.1074/jbc.M805556200.

Ohe, K. and Hagiwara, M. (2015) 'Modulation of alternative splicing with chemical compounds in new therapeutics for human diseases', *ACS Chemical Biology*. doi: 10.1021/cb500697f.

Opitz, C. A. *et al.* (2004) 'Developmentally Regulated Switching of Titin Size Alters Myofibrillar Stiffness in the Perinatal Heart', *Circulation Research*, 94(7). doi: 10.1161/01.RES.0000124301.48193.E1.

Ottesen, E. W. *et al.* (2016) 'Severe impairment of male reproductive organ development in a low SMN expressing mouse model of spinal muscular atrophy', *Scientific Reports*, 6. doi: 10.1038/srep20193.

Pacini, C. and Koziol, M. J. (2018) 'Bioinformatics challenges and perspectives when studying the effect of epigenetic modifications on alternative splicing', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2017.0073.

Padgett, R. A. (2012) 'New connections between splicing and human disease', *Trends in Genetics*. doi: 10.1016/j.tig.2012.01.001.

Palacino, J. *et al.* (2015) 'SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice', *Nature Chemical Biology*, 11(7). doi: 10.1038/nchembio.1837.

Patro, R. *et al.* (2017) 'Salmon provides fast and bias-aware quantification of transcript expression', *Nature Methods*, 14(4). doi: 10.1038/nmeth.4197.

Pertea, M., Lin, X. and Salzberg, S. L. (2001) 'GeneSplicer: A new computational method for splice site prediction', *Nucleic Acids Research*, 29(5). doi: 10.1093/nar/29.5.1185.

Piñol-Roma, S. and Dreyfuss, G. (1992) 'Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm', *Nature*, 355(6362). doi: 10.1038/355730a0.

Pinotti, M. *et al.* (2008) 'U1-snRNA mediated rescue of mRNA processing in severe factor VII deficiency', *Blood*, 111(5). doi: 10.1182/blood-2007-10-117440.

Piva, F. *et al.* (2012) 'SpliceAid 2: A database of human splicing factors expression data and RNA target motifs', *Human Mutation*. doi: 10.1002/humu.21609.

Pollard, M. O. *et al.* (2018) 'Long reads: Their purpose and place', *Human Molecular Genetics*. doi: 10.1093/hmg/ddy177.

Pros, E. *et al.* (2009) 'Antisense therapeutics for neurofibromatosis type 1 caused by deep intronic mutations', *Human Mutation*, 30(3). doi: 10.1002/humu.20933.

Protonotarios, A. and Elliott, P. M. (2019) 'Arrhythmogenic cardiomyopathies (ACs): Diagnosis, risk stratification and management', *Heart*, 105(14). doi: 10.1136/heartjnl-2017-311160.

Pugh, T. J. *et al.* (2014) 'The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing', *Genetics in Medicine*, 16(8). doi: 10.1038/gim.2013.204.

Raponi, M. *et al.* (2014) 'BRCA1 exon 11 a model of long exon splicing regulation', *RNA Biology*, 11(4). doi: 10.4161/rna.28458.

Reese, M. G. (1997) 'Improved splice site detection in Genie', in *Journal of Computational Biology*. doi: 10.1089/cmb.1997.4.311.

Rentzsch, P. *et al.* (2019) 'CADD: Predicting the deleteriousness of variants throughout the human genome', *Nucleic Acids Research*, 47(D1). doi: 10.1093/nar/gky1016.

Ribeiro, M. *et al.* (2020) 'RNA splicing defects in hypertrophic cardiomyopathy: Implications for diagnosis and therapy', *International Journal of Molecular Sciences*. doi: 10.3390/ijms21041329.

Richards, S. *et al.* (2015) 'Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology', *Genetics in Medicine*, 17(5). doi: 10.1038/gim.2015.30.

Richardson, P. *et al.* (1996) 'Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the definition and classification of cardiomyopathies', *Circulation*. doi: 10.1161/01.CIR.93.5.841.

Riedmayr, L. *et al.* (2018) 'Construction and Cloning of Minigenes for in vivo Analysis of Potential Splice Mutations', *BIO-PROTOCOL*, 8(5). doi: 10.21769/bioprotoc.2760.

Riepe, T. V. *et al.* (2021) 'Benchmarking deep learning splice prediction tools using functional splice assays', *Human Mutation*, 42(7). doi: 10.1002/humu.24212.

Riolo, G., Cantara, S. and Ricci, C. (2021) 'What's wrong in a jump? Prediction and validation of splice site variants', *Methods and Protocols*. doi: 10.3390/mps4030062.

Rivas, M. A. *et al.* (2015) 'Effect of predicted protein-truncating genetic variants on the human transcriptome', *Science*, 348(6235). doi: 10.1126/science.1261877.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*. doi: 10.1038/nbt.1754.

Robinson, J. T. *et al.* (2023) 'igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV)', *Bioinformatics (Oxford, England)*, 39(1). doi: 10.1093/bioinformatics/btac830.

Roca, X. and Krainer, A. R. (2009) 'Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA', *Nature Structural and Molecular Biology*, 16(2). doi: 10.1038/nsmb.1546.

Rogozin, I. B. *et al.* (2012) 'Origin and evolution of spliceosomal introns', *Biology Direct*. doi: 10.1186/1745-6150-7-11.

Roscigno, R. F., Weiner, M. and Garcia-Blanco, M. A. (1993) 'A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing', *Journal of Biological Chemistry*, 268(15). doi: 10.1016/s0021-9258(18)82114-7.

Rosenstein, B. J. and Cutting, G. R. (1998) 'The diagnosis of cystic fibrosis: A consensus statement', *Journal of Pediatrics*, 132(4). doi: 10.1016/S0022-3476(98)70344-0.

Rossi, F. *et al.* (1996) 'Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I', *Nature*, 381(6577). doi: 10.1038/381080a0.

Rowlands, C. F., Baralle, D. and Ellingford, J. M. (2019) 'Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing', *Cells*. doi: 10.3390/cells8121513.

Sapra, A. K. *et al.* (2009) 'SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo', *Molecular Cell*, 34(2). doi: 10.1016/j.molcel.2009.02.031.

Saw, P. E. and Song, E. W. (2020) 'siRNA therapeutics: a clinical reality', *Science China Life Sciences*. doi: 10.1007/s11427-018-9438-y.

Scalzitti, N. *et al.* (2021) 'Spliceator: multi-species splice site prediction using convolutional neural networks', *BMC Bioinformatics*, 22(1). doi: 10.1186/s12859-021-04471-3.

Schneider-Poetsch, T., Chhipi-Shrestha, J. K. and Yoshida, M. (2021) 'Splicing modulators: on the way from nature to clinic', *Journal of Antibiotics*. doi: 10.1038/s41429-021-00450-1.

Schoch, K. *et al.* (2020) 'Alternative transcripts in variant interpretation: the potential for missed diagnoses and misdiagnoses', *Genetics in Medicine*, 22(7). doi: 10.1038/s41436-020-0781-x.

Schoenauer, R. *et al.* (2011) 'EH-myomesin splice isoform is a novel marker for dilated cardiomyopathy', *Basic Research in Cardiology*, 106(2). doi: 10.1007/s00395-010-0131-2.

Scotti, M. M. and Swanson, M. S. (2016) 'RNA mis-splicing in disease', *Nature Reviews Genetics*. doi: 10.1038/nrg.2015.3.

Shapiro, M. B. and Senapathy, P. (1987) 'RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression', *Nucleic Acids Research*, 15(17). doi: 10.1093/nar/15.17.7155.

Shen, S. *et al.* (2014) 'rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data', *Proceedings of the National Academy of Sciences of the United States of America*, 111(51). doi: 10.1073/pnas.1419161111.

Shibata, A. *et al.* (2016) 'IntSplice: Prediction of the splicing consequences of intronic single-nucleotide variations in the human genome', *Journal of Human Genetics*, 61(7). doi: 10.1038/jhg.2016.23.

Singh, G. and Cooper, T. A. (2006) 'Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing', *BioTechniques*. doi: 10.2144/000112208.

Singh, R. K. and Cooper, T. A. (2012) 'Pre-mRNA splicing in disease and therapeutics', *Trends in Molecular Medicine*. doi: 10.1016/j.molmed.2012.06.006.

Siomi, M. C. *et al.* (2011) 'PIWI-interacting small RNAs: The vanguard of genome defence', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3089.

Siva, K., Covello, G. and Denti, M. A. (2014) 'Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases', *Nucleic Acid Therapeutics*. doi: 10.1089/nat.2013.0461.

Slaugenhaupt, S. A. *et al.* (2004) 'Rescue of a human mRNA splicing defect by the plant cytokinin kinetin', *Human Molecular Genetics*, 13(4). doi: 10.1093/hmg/ddh046.

Smith, P. J. *et al.* (2006) 'An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers', *Human Molecular Genetics*, 15(16). doi: 10.1093/hmg/ddl171.

Solis, A. S., Shariat, N. and Patton, J. G. (2008) 'Splicing fidelity, enhancers, and disease', *Frontiers in Bioscience*. doi: 10.2741/2812.

Soukarieh, O. *et al.* (2016) 'Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools', *PLoS Genetics*. doi: 10.1371/journal.pgen.1005756.

Spinelli, L., Giugliano, G. and Esposito, G. (2021) 'Cardiac Involvement in Autosomal Dominant Polycystic Kidney Disease', *Cardiogenetics*, 11(2). doi: 10.3390/cardiogenetics11020006.

Stark, R., Grzelak, M. and Hadfield, J. (2019) 'RNA sequencing: the teenage years', *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0150-2.

Strande, N. T. *et al.* (2018) 'Navigating the nuances of clinical sequence variant interpretation in Mendelian disease', *Genetics in Medicine*. doi: 10.1038/s41436-018-0100-y.

Su, Q. *et al.* (2018) 'Structure of the human PKD1-PKD2 complex', *Science*, 361(6406). doi: 10.1126/science.aat9819.

Sumner, C. J. *et al.* (2003) 'Valproic Acid Increases SMN Levels in Spinal Muscular Atrophy Patient Cells', *Annals of Neurology*, 54(5). doi: 10.1002/ana.10743.

Suñé-Pou, M. *et al.* (2017) 'Targeting splicing in the treatment of human disease', *Genes*. doi: 10.3390/genes8030087.

Suñé-Pou, M. *et al.* (2020) 'Innovative Therapeutic and Delivery Approaches Using Nanotechnology to Correct Splicing Defects Underlying Disease', *Frontiers in Genetics*. doi: 10.3389/fgene.2020.00731.

Tacke, R. and Manley, J. L. (1999) 'Determinants of SR protein specificity', *Current Opinion in Cell Biology*. doi: 10.1016/S0955-0674(99)80050-7.

Tan, W. L. W. *et al.* (2017) 'A landscape of circular RNA expression in the human heart', *Cardiovascular Research*, 113(3). doi: 10.1093/cvr/cvw250.

Tavtigian, S. V. *et al.* (2018) 'Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework', *Genetics in Medicine*, 20(9). doi: 10.1038/gim.2017.210.

Tayal, U., Prasad, S. and Cook, S. A. (2017) 'Genetics and genomics of dilated cardiomyopathy and systolic heart failure', *Genome Medicine*. doi: 10.1186/s13073-017-0410-8.

Taylor, A., Alloub, Z. and Tayoun, A. A. (2021) 'A simple practical guide to genomic diagnostics in a pediatric setting', *Genes*. doi: 10.3390/genes12060818.

Thompson, P. M. *et al.* (2017) 'A Structural Model for Vinculin Insertion into PIP2-Containing Membranes and the Effect of Insertion on Vinculin Activation and Localization', *Structure*, 25(2).

doi: 10.1016/j.str.2016.12.002.

Trefz, F. K., Lichter-Konecki, U. and Konecki, D. (1989) 'Phenylketonuria', *Current Opinion in Pediatrics*. doi: 10.1097/00008480-198912000-00028.

Truty, R. *et al.* (2021) 'Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation', *American Journal of Human Genetics*, 108(4). doi: 10.1016/j.ajhg.2021.03.006.

Tse, H. F. *et al.* (2013) 'Patient-specific induced-pluripotent stem cells-derived cardiomyocytes recapitulate the pathogenic phenotypes of dilated cardiomyopathy due to a novel DES mutation identified by whole exome sequencing', *Human Molecular Genetics*, 22(7). doi: 10.1093/hmg/dds556.

Tudurachi, B. S. *et al.* (2023) 'An Update on MYBPC3 Gene Mutation in Hypertrophic Cardiomyopathy', *International Journal of Molecular Sciences*. doi: 10.3390/ijms241310510.

Uzumcu, A. *et al.* (2006) 'Loss of desmoplakin isoform I causes early onset cardiomyopathy and heart failure in a Naxos-like syndrome.', *Journal of medical genetics*. doi: 10.1136/jmg.2005.032904.

Vandesompele, J. *et al.* (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.', *Genome biology*, 3(7). doi: 10.1186/gb-2002-3-7-research0034.

Vanichkina, D. P. *et al.* (2018) 'Challenges in defining the role of intron retention in normal biology and disease', *Seminars in Cell and Developmental Biology*. doi: 10.1016/j.semcdb.2017.07.030.

Vasile, V. C., Ommen, S. R., *et al.* (2006) 'A missense mutation in a ubiquitously expressed protein, vinculin, confers susceptibility to hypertrophic cardiomyopathy', *Biochemical and Biophysical Research Communications*. doi: 10.1016/j.bbrc.2006.04.151.

Vasile, V. C., Edwards, W. D., *et al.* (2006) 'Obstructive hypertrophic cardiomyopathy is associated with reduced expression of vinculin in the intercalated disc', *Biochemical and Biophysical Research Communications*. doi: 10.1016/j.bbrc.2006.08.106.

Veltrop, M. and Aartsma-Rus, A. (2014) 'Antisense-mediated exon skipping: Taking advantage of a trick from Mother Nature to treat rare genetic diseases', *Experimental Cell Research*. doi: 10.1016/j.yexcr.2014.01.026.

Verma, B. *et al.* (2018) 'Minor spliceosome and disease', *Seminars in Cell and Developmental*

Biology. doi: 10.1016/j.semcdb.2017.09.036.

Wahl, M. C., Will, C. L. and Lührmann, R. (2009) 'The Spliceosome: Design Principles of a Dynamic RNP Machine', *Cell*. doi: 10.1016/j.cell.2009.02.009.

Wai, H. A. *et al.* (2020) 'Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance.', *Genetics in medicine : official journal of the American College of Medical Genetics*, 0(0), pp. 1–10. doi: 10.1038/s41436-020-0766-9.

Wai, H. A. *et al.* (2022) 'Short amplicon reverse transcription-polymerase chain reaction detects aberrant splicing in genes with low expression in blood missed by ribonucleic acid sequencing analysis for clinical diagnosis', *Human Mutation*, 43(7). doi: 10.1002/humu.24378.

Wai, H., Douglas, A. G. L. and Baralle, D. (2019) 'RNA splicing analysis in genomic medicine', *International Journal of Biochemistry and Cell Biology*. doi: 10.1016/j.biocel.2018.12.009.

Wally, V., Murauer, E. M. and Bauer, J. W. (2012) 'Spliceosome-mediated trans-splicing: The therapeutic cut and paste', *Journal of Investigative Dermatology*. doi: 10.1038/jid.2012.101.

Walsh, R. *et al.* (2017) 'Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples', *Genetics in Medicine*, 19(2). doi: 10.1038/gim.2016.90.

Wang, E. T. *et al.* (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456(7221). doi: 10.1038/nature07509.

Wang, G. S. and Cooper, T. A. (2007) 'Splicing in disease: Disruption of the splicing code and the decoding machinery', *Nature Reviews Genetics*. doi: 10.1038/nrg2164.

Wang, H. *et al.* (2010) 'Mutations in NEXN, a Z-Disc gene, are associated with hypertrophic cardiomyopathy', *American Journal of Human Genetics*, 87(5). doi: 10.1016/j.ajhg.2010.10.002.

Wang, H. *et al.* (2016) 'Genome-wide analysis of alternative splicing during human heart development', *Scientific Reports*, 6. doi: 10.1038/srep35520.

Wang, R. *et al.* (2019) 'SpliceFinder: Ab initio prediction of splice sites using convolutional neural network', *BMC Bioinformatics*, 20. doi: 10.1186/s12859-019-3306-3.

Wang, Z. *et al.* (2006) 'General and Specific Functions of Exonic Splicing Silencers in Splicing Control', *Molecular Cell*, 23(1). doi: 10.1016/j.molcel.2006.05.018.

Watanabe, T., Kimura, A. and Kuroyanagi, H. (2018) 'Alternative splicing regulator RBM20 and cardiomyopathy', *Frontiers in Molecular Biosciences*. doi: 10.3389/fmolb.2018.00105.

Weighardt, F., Biamonti, G. and Riva, S. (1995) 'Nucleo-cytoplasmic distribution of human hnRNP proteins: A search for the targeting domains in hnRNP A1', *Journal of Cell Science*, 108(2). doi: 10.1242/jcs.108.2.545.

Whiley, P. J. *et al.* (2014) 'Comparison of mRNA splicing assay protocols across multiple laboratories: Recommendations for best practice in standardized clinical testing', *Clinical Chemistry*, 60(2). doi: 10.1373/clinchem.2013.210658.

Whitehead, K. A., Langer, R. and Anderson, D. G. (2009) 'Knocking down barriers: Advances in siRNA delivery', *Nature Reviews Drug Discovery*. doi: 10.1038/nrd2742.

Will, C. L. and Lührmann, R. (2011) 'Spliceosome structure and function', *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a003707.

Wong, E. K. *et al.* (2019) 'Perceptions of genetic variant reclassification in patients with inherited cardiac disease', *European Journal of Human Genetics*. doi: 10.1038/s41431-019-0377-6.

Xie, F., Wang, J. and Zhang, B. (2023) 'RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes', *Functional and Integrative Genomics*, 23(2). doi: 10.1007/s10142-023-01055-7.

Xiong, H. Y. *et al.* (2015) 'The human splicing code reveals new insights into the genetic determinants of disease', *Science*, 347(6218). doi: 10.1126/science.1254806.

Xu, J. *et al.* (2016) 'Comprehensive assessments of RNA-seq by the SEQC consortium: FDA-led efforts advance precision medicine', *Pharmaceutics*. doi: 10.3390/pharmaceutics8010008.

Xu, N., Chen, C.-Y. A. and Shyu, A.-B. (2001) 'Versatile Role for hnRNP D Isoforms in the Differential Regulation of Cytoplasmic mRNA Turnover', *Molecular and Cellular Biology*, 21(20). doi: 10.1128/mcb.21.20.6960-6971.2001.

Xue, Y. *et al.* (2014) 'Fgfr3 mutation frequency in 324 cases from the international skeletal dysplasia registry', *Molecular Genetics and Genomic Medicine*, 2(6). doi: 10.1002/mgg3.96.

Xue, Y. *et al.* (2015) 'Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: Single-gene, gene panel, or exome/genome sequencing', *Genetics in Medicine*. doi: 10.1038/gim.2014.122.

Yeo, G. and Burge, C. B. (2004) 'Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals', in *Journal of Computational Biology*. doi: 10.1089/1066527041410418.

Yuan, Z. Y. *et al.* (2021) 'Desmoplakin and clinical manifestations of desmoplakin

cardiomyopathy', *Chinese Medical Journal*. doi: 10.1097/CM9.0000000000001581.

Zaidi, S. *et al.* (2013) 'De novo mutations in histone-modifying genes in congenital heart disease', *Nature*, 498(7453). doi: 10.1038/nature12141.

Zatkova, A. *et al.* (2004) 'Disruption of exonic splicing enhancer elements is the principal cause of exon skipping associated with seven nonsense or missense alleles of NF1', *Human Mutation*, 24(6). doi: 10.1002/humu.20103.

Zhang, S. *et al.* (2018) 'Base-specific mutational intolerance near splice sites clarifies the role of nonessential splice nucleotides', *Genome Research*, 28(7). doi: 10.1101/gr.231902.117.

Zhu, C., Chen, Z. and Guo, W. (2017) 'Pre-mRNA mis-splicing of sarcomeric genes in heart failure', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. doi: 10.1016/j.bbadis.2016.11.008.

Zhu, J., Mayeda, A. and Krainer, A. R. (2001) 'Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins', *Molecular Cell*, 8(6). doi: 10.1016/S1097-2765(01)00409-9.

Zuallaert, J. *et al.* (2018) 'Splicerover: Interpretable convolutional neural networks for improved splice site prediction', *Bioinformatics*, 34(24). doi: 10.1093/bioinformatics/bty497.