SEPT–GD

Alimohamed, Mohamed Z.; Boven, Ludolf G.; van Dijk, Krista K.; Vos, Yvonne J.; Hoedemaekers, Yvonne M.; van der Zwaag, Paul A.; Sijmons, Rolf H.; Jongbloed, Jan D.H.; Sikkema-Raddatz, Birgit; Westers, Helga

Published in:
Gene

DOI:
10.1016/j.gene.2022.146984

Publication date:
2023

Gene 851 (2023) 146984

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

SEPT–GD: A decision tree to prioritise potential RNA splice variants in cardiomyopathy genes for functional splicing assays in diagnostics

Mohamed Z. Alimohamed [a,b,c,d,*], Ludolf G. Boven [a], Krista K. van Dijk [a], Yvonne J. Vos [a], Yvonne M. Hoedemaekers [a,e], Paul A. van der Zwaag [a], Rolf H. Sijmons [a], Jan D.H. Jongbloed [a,1], Birgit Sikkema-Raddatz [a,1], Helga Westers [a,*]

[a] University of Groningen, University Medical Center Groningen, Department of Genetics, The Netherlands
[b] Department of Haematology and Blood Transfusion, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania
[c] Department of Research and Training, Shree Hindu Mandal Hospital, Dar-es-salaam, Tanzania
[d] Tanzania Human Genetics Organization, Dar es salaam, Tanzania
[e] Radboud University Medical Center Nijmegen, Department of Genetics, The Netherlands

ARTICLE INFO

Edited by Claudia Montefusco

ABSTRACT

Background: Splice prediction algorithms currently used in routine DNA diagnostics have limited sensitivity and specificity, therefore many potential splice variants are classified as variants of uncertain significance (VUSs). However, functional assessment of VUSs to test splicing is labour-intensive and time-consuming. We developed a decision tree to prioritise potential splice variants for functional studies and functionally verified the outcome of the decision tree.

Materials and methods: We built the decision tree, SEPT–GD, by setting thresholds for the splice prediction programs implemented in Alamut. A set of 343 variants with known effects on splicing was used as control for sensitivity and specificity. We tested SEPT–GD using variants from a Dutch cardiomyopathy cohort of 2002 patients that were previously classified as VUS and predicted to have a splice effect according to diagnostic rules. We then selected 12 VUSs ranked by SEPT–GD to functionally verify the predicted effect on splicing using a minigene assay: 10 variants predicted to have a strong effect and 2 with a weak effect. RT-PCR was performed for nine variants. Variant classification was re-evaluated based on the functional test outcome.

Results: Compared to similar individually tested algorithms, SEPT–GD shows higher sensitivity (91 %) and comparable specificity (88 %) for both consensus (dinucleotides at the start and end of the intron, GT at the 5′ end and AG at the 3′ end) and non-consensus splice-site variants (excluding middle of exon variants). Using clinical diagnostic criteria, 1295 unique variants in our cardiomyopathy cohort had originally been classified as VUSs, with 57 predicted by Alamut to have an effect on splicing. Using SEPT–GD, we prioritised 31 variants in 40 patients. In the minigene assay, all 12 variants showed results concordant with SEPT-GD predictions. RT-PCR confirmed the minigene results for two variants, *TMEM43* c.1000 + 5G > T and *TTN* c.25922–6 T > G. Based on all outcomes, the *SGCD* c.4-1G > A and *CSRP3* c.282-5_285del variants were reclassified as likely pathogenic.

Conclusion: SEPT–GD outperforms the tools commonly used for RNA splicing prediction and improves prioritisation of variants in cardiomyopathy genes for functional splicing analysis in a diagnostic setting.

Abbreviations: VUS, Variant of unknown significance; NGS, Next generation sequencing; LP, Likely pathogenic; HGMD, Human Gene Mutation Database; P, Pathogenic; RT-PCR, Reverse transcription polymerase chain reaction; qPCR, quantitative PCR; SEPT-GD, Splice effect prediction tree, Genome diagnostics; B, Benign; LB, Likely benign; WT, Wildtype; SSF, Splicesite finder; MES, MaxEntScan; NNS, Network Splice; GS, Genesplicer; ESE, Exonic splice enhancer; TP, True positive; TN, True negative; FN, False negative; FP, False positive; HEK, Human embryonic kidney; DMEM, Dulbecco's modified eagle medium; EGFP, Enhanced green flourescent protein; LoF, Loss of function; DCM, Dilated cardiomyopathy; HCM, Hypertrophic cardiomyopathy; NMD, nonsense-mediated decay; ASSP, Alternative splice site predictor.

* Corresponding authors at: University of Groningen, Department of Genetics, HPC CB50, University Medical Center Groningen, 9700 RB, P.O.Box 30001, Groningen (Mohamed Z. Alimohamed and Helga Westers). Tel.: +31-652724566.

E-mail addresses: m.z.a.alimohamed@umcg.nl, mzahir@blood.ac.tz (M.Z. Alimohamed), h.westers@umcg.nl (H. Westers).

[1] Authors contributed equally.

https://doi.org/10.1016/j.gene.2022.146984
Received 20 February 2022; Received in revised form 9 August 2022; Accepted 13 October 2022
Available online 18 October 2022

1. Introduction

The clinical application of next-generation sequencing (NGS) using targeted, exome or whole genome sequencing approaches has resulted in a marked increase in the molecular diagnostic yield in Mendelian diseases [Shen et al, 2015]. However, before genetic variants can be classified as (likely) pathogenic ((L)P), leading to a molecular genetic diagnosis, large numbers of identified variants require prioritisation and pathogenicity assessment [Ito et al, 2017]. To predict the functional impact of variants and guide classification of variant pathogenicity, genome diagnostic centres use various decision trees, including outcomes of *in silico* prediction tools [Rhine et al, 2018]. While these computational tools are expanding [Ito et al, 2017; Rhine et al, 2018] and improving, clinical laboratory specialists still end up with long lists of variants of unknown significance (VUSs) that are clinically not actionable.

One category of VUSs consists of variants that potentially disturb normal mRNA splicing [Crehalet et al, 2012]. Approximately 9 % of all the variants in the Human Gene Mutation Database (HGMD) that are considered responsible for human inherited disease are labelled as variants with consequences for splicing (27,959/323,661; accessed on 23 December 2021) [Stenson et al., 2003]. The molecular diagnostic yield likely increases if more VUSs suspected to have an effect on splicing could be reclassified as LP or P based on functional evidence. Commonly used techniques for functional analysis of splicing are reverse transcription PCR (RT-PCR), *in vitro* minigene assays, quantitative PCR (qPCR) and protein truncation tests [Harvey and Cheng, 2016; Anna and Monika, 2018]. Ideally, all VUSs would be functionally tested for aberrant splicing. However, this goal is currently unrealistic as these tests remain very labour-intensive and time-consuming.

Given the current limitations of functional test capacity, diagnostics laboratories turn first to predictive *in silico* testing and then follow up with functional *in vitro* analysis of selected variants prioritised based on the prediction outcome. These *in silico* prediction tools are based on nucleotide frequency matrices and algorithms that measure the interdependence of adjacent (Markov model) and distant (Maximum entropy model) positions of core splicing consensus sequences [Vorechovsky, 2006; Anna and Monika, 2018]. Although these algorithms perform well for the canonical splice sites (most commonly found dinucleotides GT and AG for donor and acceptor sites, respectively) [Jian et al., 2014], they work less well outside these consensus regions, often leading to false positive results [Jian et al., 2014]. To demonstrate the clinical value of routinely used *in silico* predictions of effects on splicing, one would ideally perform functional mRNA analysis in a larger series of gene variants across the complete range of predicted non-splice-affecting and splice-affecting cases. As the algorithm should preferably be applicable to all genes tested clinically, the variants in an algorithm study should have been identified in a wide range of genes, or at least in those most frequently tested in the clinic. Moreover, having a reliable decision tree to help prioritise RNA splicing variants for functional verification would be advantageous. So far, studies on these topics have been limited and there are no recognised thresholds for distinguishing between positive and negative effects on splicing for a particular variant at a particular site [Jian et al., 2014]. In this study, we set out to develop a workflow to better prioritise potential splice-affecting VUSs for follow-up functional analysis. For this purpose, we designed a decision tree, Splice Effect Prediction Tree – Genome Diagnostics (SEPT–GD), by setting thresholds for parameters in the prediction tools that are integrated in the widely used Alamut® variant interpretation software package. SEPT-GD aims to be a more stringent, structured and quantifiable method to prioritize potential splice affecting VUSs for functional follow up using the same algorithms present in Alamut®, after initial selection of such VUSs using our routine diagnostic splice prediction criteria. To test the robustness of SEPT–GD, we used variants with known effects on splicing in cardiomyopathy genes, one of the most frequently tested group of genes in diagnostics. We then used SEPT–GD to predict the effect on splicing of VUSs previously identified in a cohort of 2002 cardiomyopathy patients [Alimohamed et al, 2021] and functionally tested a selection of these variants in an *in vitro* minigene assay and in RNA isolated from blood.

2. Materials and methods

2.1. Patient samples and variants

We previously reported the yield of targeted NGS data in a cohort of 2002 cardiomyopathy patients [Alimohamed et al, 2021]. Patients included in this study were referred to our clinical genetics laboratory for genetic testing for various types of cardiomyopathies. Variant interpretation was based on guidelines recommended by the American College of Medical Genetics and Genomics [Richards et al, 2015]. Variants were classified as benign (B), likely benign (LB), VUS, LP or P. The study was performed in accordance with UMCG and Dutch national ethical guidelines. Informed consent was obtained for all patients.

2.2. Splice vacriant prediction – Routine diagnostic analysis

Alamut® software version 2.11 (Interactive Bio software, Rouen, France) was used for *in silico* prediction of splice-affecting nucleotide variants. Within Alamut, *in silico* scores comparing wild type (WT) and mutant alleles for all genetic variants were obtained using four splicing prediction tools: SpliceSiteFinder (SSF)-like, MaxEntScan (MES), Neural Network Splice (NNS) and GeneSplicer (GS). A variant is considered potentially splice-altering when 3 out of 4 of the prediction tools show a significant score difference between the WT and mutant allele as manually scrutinized by the responsible laboratory specialist clinical genetics.

2.3. Variant datasets

LP/P variants and VUSs from our cardiomyopathy cohort that were predicted to be splice-altering using Alamut® software, known splice variants and true negatives confirmed from literature were analysed in two datasets:

A. Reference set

The reference set to optimise the analysis and interpretation procedure, i.e., setting the thresholds in the SEPT–GD decision tree, consisted of the following list of variants:

1) (L)P splice-altering variants at canonical splice sites in cardiomyopathy-related genes detected in our cardiomyopathy cohort (Supplemental Table 1),
2) HGMD-listed proven exonic splice variants of all genes in our cardiomyopathy gene panel and.
3) Variants identified by a systematic PubMed search (5 March 2019) using the search items: *splicing/splice mutations/variants in cardiomyopathy minigene/RNA analysis, duration of 10 years, sorted of best match with medical subject headings (MeSH) terms; cardiomyopathies, RNA splicing, mutation, RNA, mutation; sub heading: analysis*. A variant was included for analysis when all the following criteria were met: a) the variant (gene) was relevant to cardiomyopathy, b) it was implicated to alter splicing and, c) there was functional evidence available (RNA analysis and/or minigene splicing assay).

B. Test set - cardiomyopathy cohort VUSs

The VUS test set consisted of potential RNA splice variants classified as VUS for cardiomyopathy-related genes in our previously described cohort [Alimohamed et al, 2021].

2.4. Decision tree for splice variant selection

In Alamut®, the following prediction algorithms were used: SSF, MES, NNS, GS. In addition, we incorporated the RESCUE-ESE that identifies candidate exonic splicing enhancers in vertebrate exons [Fairbrother et al, 2004] and EX-SKIP, a tool that quickly estimates which allele is more susceptible to exon skipping [Raponi et al, 2011]. For interpretation using SEPT-GD, if the EX-SKIP icon indicates for a variant a higher probability that a mutant will undergo skipping compared to WT, we considered this as one of the evidence criteria needed towards variant prioritisation. Similarly, for Rescue-ESE, under the ESE prediction icon, if a hexanucleotide sequence as candidate ESE is indicated under a specific variant or mutated sequence then we considered this as one of the evidence criteria needed towards variant prioritisation'. Variants from the control and test sets were further evaluated using our decision tree, as shown in Fig. 1.

Based on the genomic position of a variant within the gene of interest, the variants were split into four main categories: (1) consensus splice sites, (2) intronic variants, (3) near-consensus exonic variants and (4) middle-of-exon variants. For each of the main categories, criteria based on scores from the prediction algorithms in Alamut® were provided to prioritise variants for functional follow-up. In addition, a grey zone category was introduced to indicate variants with inconclusive *in silico* predictions.

1. Variants ± 2 base pairs intronic from the start or end of an exon (consensus splice sites) should always be prioritised for follow-up.
2. Intronic VUSs that do not meet criterium 1 must meet two of the following criteria to be prioritised:

i) two out of four algorithms predict a score difference ≥ 50 % between the mutated and WT sequence at the original splice site,
ii) the end of the exon is visible (within approx. 180 bp Alamut® window) and the score difference between WT and intronic variant is clearly seen and recorded whilst showing the investigated variant in the same window and
iii) an alternative splice site is clearly present (defined as presence of score difference between mutant and WT sequence above 50 % in a minimum of two out of four scores (SSF range 0–100, MES range 0–16, NNS range 0–1 and GS range 0–21).

3. Exonic VUS present from the start of the exon up to 5 bp in the exon for acceptor sites and from 5 bp before the end of the exon for donor sites must meet two of the criteria below to be prioritised:

i) two out of four algorithms predict a score difference ≥ 50 % between the mutated and WT sequence at the original splice site,
ii) the start or end of the exon should be visible (within approx. 180 bp Alamut® window) and there is a clear score difference between WT and the exonic splice variant whilst showing the investigated variant in the same window and
iii) clear presence of an alternative splice site (see 2 (iii)).

4. VUS present in the middle of an exon (a predicted donor or acceptor site) > 5 bp from the start or the end of the exon have to meet three of the following criteria to be prioritised:

i) two out of four algorithms predict a score difference ≥ 50 % between the mutated and WT sequence of the original splice site,
ii) the start or end of exon is visible (within approx. 180 bp Alamut® window) and there is a clear score difference between the WT and mutated exon whilst showing the investigated variant in the same window,
iii) there is clear presence of an alternative splice site location (see 2 (iii)),
iv) Ex-skip predicts a higher probability for the mutant to undergo skipping compared to WT,
v) Rescue-ESE, under the ESE Predictions icon, indicates the variant at WT or mutated sequence to be a hexanucleotide sequence as candidate ESE (exonic splicing enhancer) and/or a branch point difference ≥ 50 % between the mutated and WT sequence variant,
vi) Note: in the case of a variant where both the donor and acceptor sites indicate a difference between the mutated and WT sequence, follow the relevant closest-site scores depending on the start and end of an exon (i.e., if the start-site of the exon is closer to the variant position than the end-site, focus on the acceptor site, otherwise focus on the donor site) and
vii) for a variant whose exon ends are not visible for scoring comparison, three criteria from i, iii, iv and v must be met.

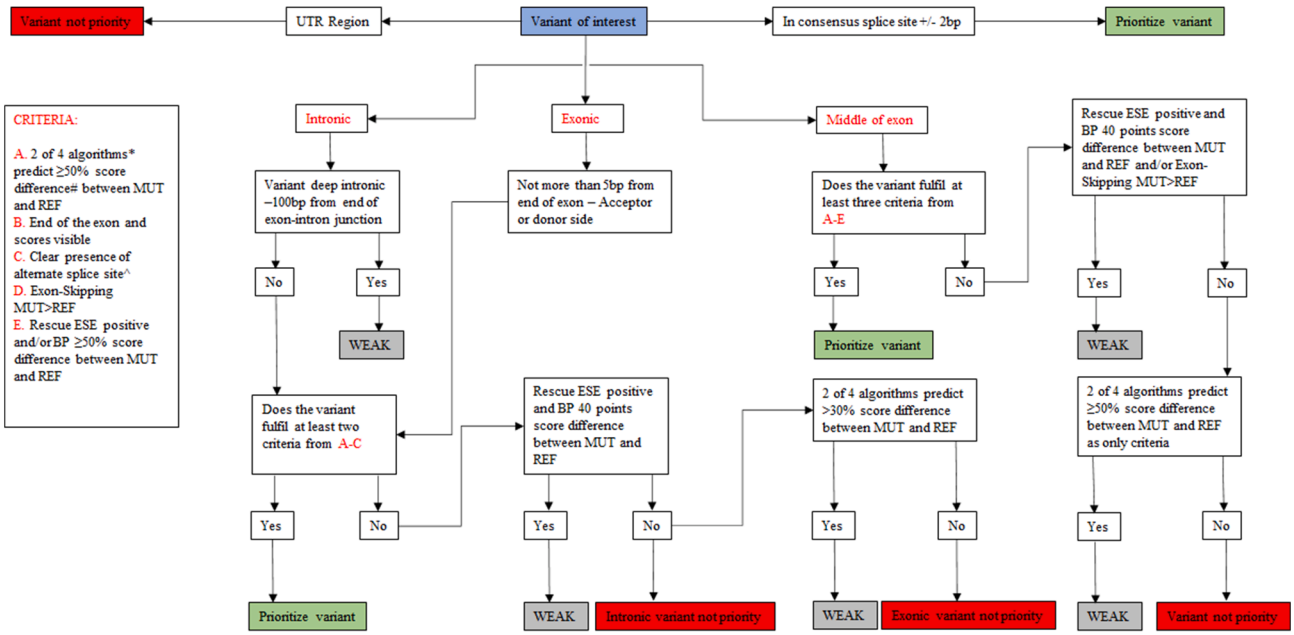

Fig. 1. Schematic representation of Splice Effect Prediction Tree – Genome Diagnostics (SEPT–GD).

2.5. Criteria for weak splice prediction

To balance specificity and sensitivity, we introduced weak splice effect criteria to show indeterminate prioritisation due to inconclusive *in silico* evidence. These include:

i) intronic variants that do not present with any scores but are predicted by Rescue-ESE to be a hexanucleotide sequence and by SSF to have a branch point difference of 40 points between the mutated and WT sequence variant,
ii) deep intronic variants (100 bp away from exon-intron junction) that have met the intronic VUS criteria to be prioritised for functional follow-up,
iii) exon variants that have not met the criteria to be prioritised for functional follow-up but are labelled as hexanucleotide sequence using Rescue-ESE and show a branch point difference of 40 points between the mutated and WT sequence variant,
iv) exon variants for which two or more algorithms present a score difference > 30 % (one of which is completely abolished) between WT and variant,
v) middle-of-exon variants with not enough score difference that present with predictions from Rescue-ESE and a branch point difference of 40 points or skipping predicted by Ex-skip and
vi) middle-of-exon variant calls with enough score difference but no additional indications.

The outcome of the decision tree, i.e., whether to prioritise a variant as a potential splice variant for functional follow-up or as variant with no priority, was considered true positive (TP) if the outcome was concordant with literature results based on functional proof. The outcome was considered true negative (TN) if the variant was not considered splice-altering by the decision tree and the results from functional studies. The outcome was considered false negative (FN) if the decision tree indicated a variant to be non-splice-altering when it is a splicing variant according to literature reports based on functional experiments. The outcome was considered false positive (FP) when the decision tree analysis indicated a variant to be splice-altering when it was not shown to be a splice variant according to functional analysis in literature reports. The sensitivity and specificity of the prediction of splicing affecting variants using SEPT–GD was calculated as: Sensitivity = TP/ (TP + FN), Specificity = TN/ (TN + FP).

2.6. Constructs for ex vivo splicing assay (minigene assay)

To functionally verify the variant prioritisation results from the decision tree, we tested selected variants with an *ex vivo* splicing assay, the minigene assay [Gaildrat et al, 2010]. Variants predicted by the decision tree to influence splicing were selected for testing based on availability of samples and consent of patients.

Genomic WT and mutant fragments containing an exon or exons in the region of interest and up to 250 bp of 5′and 3′ flanking intronic sequences were PCR-amplified (primer sequences in Supplemental Table 2). The products were subcloned into the pJET cloning vector, following manufacturer's instructions (Thermo Fisher Scientific, MA, United States). The inserts were verified with Sanger sequencing and correct inserts were cloned into the pSPL3 exon-trapping vector (Invitrogen Corporation, Carlsbad, CA, United States).

2.7. Transfection of HEK293 cells and RT-PCR

Human embryonic kidney (HEK) 293 cells were plated in 6-well plates containing 6×10^5 cells/well and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with glutamine, 10 % foetal bovine serum, 1 % penicillin/streptomycin (penicillin 10,000 U/ml, streptomycin 10000 μg/ml) and incubated at 37 °C, 5 % CO_2. After 24 h, the cells were transfected with 1 μg plasmid DNA using polyethylenimine according to the manufacturer's instructions (Polyscience Inc, Warrington, PA, USA). As positive control, we used the pSPL3 plasmid containing WT *KIAA* exon 28 or the *KIAA* exon 28c.4862G > A p. (Arg1621Glu) sequence known to generate a new splice site and previously confirmed in the minigene assay (loss of 54 nucleotides). The empty pSPL3 vector was used as negative control. Transfection with an Enhanced Green Flourescent Protein (EGFP)-containing vector was performed to check the transfection efficiency. After 48 h, the cells were lysed and RNA was isolated according to the manufacturer's instruction (Qiagen, Hilden, Germany). 5 μg total RNA was used as a template to synthesise cDNA (RevertAid H Minus First Strand cDNA Synthesis Kit, Thermo Fisher Scientific) using the cDNA random hexamer primers pd(N)6 and/or oligo (dT)18 primers. PCR was performed using the primers (SD6) 5′-CTGAGTCACCTGGACAACC-3′ and (SA2) 5′-ATCTCAGTGGTATTTGTGAGC-3′, of which the sequences are complementary to sequences of the exons standardly available in pSL3, and Amplitaq Gold Fast PCR mix (Thermo Fisher Scientific, MA, United States) and the following amplification programme: 5 min at 96 °C, followed by 35 cycles of 1 min at 94 °C, 1 min at 58 °C, 1 min at 72 °C (depending on insert size) and a final elongation time of 10 mins at 72 °C. PCR products were analysed by agarose gel electrophoresis and Sanger sequencing (Supplemental Table 2). Splice assay minigene experiments were performed in duplicates. To predict the functional consequences of the cloned sequence on the minigene assay and the effect of the splice variant on the transcript, we used the Human Splicing Finder 3.1 programme, as previously described [Desmet et al, 2009]. The HSF was used to seek the consensus values comparing WT and mutant sequences which is not available on the Alamut platform. The program generated consensus values (CV) in a range from 0 to 100 for each nucleotide input. WT and mutant sequences were uploaded in the program and difference between the CV were analyzed.

2.8. RT-PCR on patient RNA

An additional consent was obtained from patients with the variants tested using minigene (primary test) to obtain a separate blood sample for RNA isolation and RT-PCR analysis. RNA was isolated from whole blood collected in PAXgene® Tubes using the Maxwell® 16 Instrument and the Maxwell® 16 LEV simply RNA Blood Kits (Promega Corporation, Madison, WI, United States). To investigate the effect of a potential splice-site variant at RNA level, equal amounts of RNA were synthesised to first strand cDNA using the RevertAid H Minus First Strand cDNA synthesis kit (Thermo Fisher Scientific). RT-PCR was performed using gene-specific primers designed to amplify the exon expected to be affected by the variant and flanking region sequences. The resulting PCR products were analysed by agarose gel electrophoresis and Sanger sequencing. A result was considered positive on RT-PCR when the expected splice effect was observed as a specific-sized band on agarose gel and considered negative when the expected band was not observed. Minigene and RT-PCR results were independently interpreted.

3. Results

In our cohort of 2002 cardiomyopathy patients, we detected 1904 variants that were classified as VUS, (L)P or P. Of these 1904 variants, 485 were unique variants classified as (L)P. Forty-one of those variants, present in 59 patients (3 % of the cohort), have an (known) effect on splicing. The prevalences of (L)P splicing variants per gene, cardiomyopathy subtype and gender are provided in Supplemental Fig. 1. Using the routine diagnostic criteria for predicting splicing, 57 of the 1295 unique variants classified as VUS were predicted to alter splicing (Fig. 2).

3.1. Validation of SEPT–GD for splice prediction using the reference set

To test the validity and set thresholds within the Alamut®-based

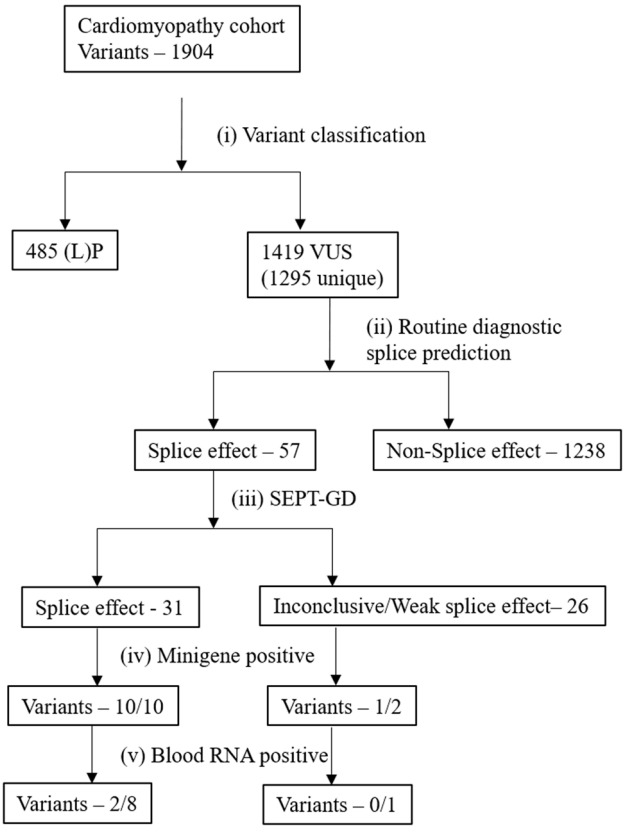

Fig. 2. Schematic representation of prioritised splice variants for *in vitro* testing, following SEPT–GD based on *in silico* splicing prediction tools. (i) Cardiomyopathy cohort containing 2002 patients and 1904 variants separated into (L)P and VUS following routine diagnostic criteria. (ii) VUSs from cardiomyopathy cohort were tested using the routine diagnostic decision criteria in Alamut® software for splicing prediction. (iii) Variants underwent additional analysis with SEPT–GD, and we selected VUSs predicted to affect splicing selected for minigene analysis. (iv) Variants tested splice-affecting with the minigene assay. (v) Variants tested splice-affecting with RT-PCR.

SEPT–GD decision tree, we used a reference set, built as described in the Methods section, which included:

i) 41 unique (L)P splice-altering variants for cardiomyopathy-related genes that we detected in our cardiomyopathy cohort: 39 intronic and 2 exonic,
ii) an additional 36 exonic splice variants from the HGMD database that were present in genes also screened using our cardiomyopathy-targeted panel genes as external control variants set to equalise the number of intronic and exonic splice variants and
iii) variants identified by a systematic PubMed literature search. This provided 50 papers matching the initial criteria. After careful reading, 34 were rejected because the data presented was not relevant to cardiomyopathy or humans, the variants described affected the splicing machinery (factors), or the papers lacked functional analysis using RNA or minigene assay. In total, 16 papers met the inclusion criteria and contributed 266 unique variants to the external control variant set.

The total control variant set comprised of 343 unique variants. Of these, 183 were splicing and 160 were non-splicing variants [Table 1], while 161 variants were intronic and 182 variants were exonic, including 111 middle-of-exon variants (102 from literature 1–50 and 9 from HGMD), i.e., variants not present in the first or last 5 bp at the beginning or end of the exon, respectively. Using our SEPT–GD decision tree resulted in 140 TP variants (including all the (L)P splicing variants detected in our cardiomyopathy patient cohort), 80 TN variants, 27 FP variants and 14 FN variants, leading to a sensitivity of 91 % and specificity of 75 % in predicting a splice effect. None of the 14 FN variants showed a difference in splice predictions between the mutant and WT using our splice prediction tree. Of the 27 FP variants, 20 were located in the middle of the exon, 5 in the intron and 2 in the first or last 5 bp of the exon. For these, the splicing prediction algorithms used showed a significant score difference between the mutant and WT sequence.

SEPT–GD labelled 82 variants out of the 343 as having a 'weak splice effect', of which 33 % (27/82) were reported as splice variants and 67 % (55/82) as non-splicing variants. Of the variants predicted to have a weak splice effect, 56 % (46/82) were located in the middle of an exon or were deep intronic.

To assess the performance of SEPT–GD in comparison to other *in*

Table 1
Overview of literature-reported splicing variants and performance of the splice effect decision tree. Control numbers refer to literature files used to extract variants, as shown in Supplemental Table 4.

	Literature variants			Decision tree results				Inconclusive		
Control No.	Splicing	Non-splicing	Total	TP	TN	FP	FN	Variants	T	F
1	9	0	9	7	0	1	0	1	0	1
2	3	2	5	3	2	0	0	0	0	0
3	54	143	197	33	64	25	8	67	13	54
4	7	0	7	3	0	0	0	4	4	0
5	8	0	8	8	0	0	0	0	0	0
6	5	0	5	3	0	0	2	0	0	0
11	6	3	9	6	2	1	0	0	0	0
14	1	8	9	1	8	0	0	0	0	0
15	1	0	1	1	0	0	0	0	0	0
21	1	2	3	1	2	0	0	0	0	0
22	1	0	1	1	0	0	0	0	0	0
26	1	0	1	1	0	0	0	0	0	0
30	1	0	1	1	0	0	0	0	0	0
37	5	0	5	3	0	0	0	2	2	0
46	4	0	4	2	0	0	0	2	2	0
50	1	0	1	1	0	0	0	0	0	0
(L)P	41	0	41	41	0	0	0	0	0	0
HGMD	34	2	36	24	2	0	4	6	6	0
Excl. MOE*				126	50	7	13			
Total	183	160	343	140	80	27	14	82	27	55

*MOE – Middle of exon variants, defined as variants not present in the first or last 5 bp at the beginning or end of the exon – not included in total row.

silico splice prediction programmes, the middle-of-exon variants were removed as they are known to be difficult to predict for splicing effects by existing programmes. re-calculating the specificity and sensitivity of SEPT–GD on literature-reported splice variants excluding middle-of-exon variants (N = 196) led to an increase of specificity to 88 %, while the sensitivity remained 91 %. Of these, 36 variants were labelled as having a weak predicted effect (SEPT–GD decision inconclusive), of which 59 % (21/36) were true splice variants and 41 % (15/36) were non-splicing variants.

Test set of variants: In our cardiomyopathy cohort, 1419 variants (1295 unique) were classified as VUS. Using the routine diagnostic criteria, 57 of these unique variants, which were detected in 71 patients (including one variant seen in three patients), were predicted to have a splicing effect (Supplemental Table 3). We then used SEPT–GD to prioritise these variants. A total of 26 variants were labelled as having a weak effect and therefore not prioritised for follow-up. The remaining 31 strongly predicted splicing variants were detected in 40 patients (2 % of the total cohort) and labelled as priority variants. Supplemental Fig. 2 lists the prevalence of predicted splice variants classified as VUS per gene, cardiomyopathy subtype and gender.

3.2. Ex vivo splicing reporter assay (minigene testing)

We selected 12 variants for minigene testing based on availability of consent for follow up as well as DNA stored at the diagnostics section of the department. Of these, 10 were predicted by SEPT–GD to be strong splice-altering variants and two were predicted to be weak variants (one of which was detected in three families). Using this assay, we detected splice alterations for all 10 variants strongly predicted to affect splicing and for one variant with a weak prediction (*ABCC9* c.2424 + 6C > G) [Table 2]. For the remaining variant (*DES* c.79G > A), predicted to be weak, the minigene experiments were inconclusive, with no differences observed between transfected minigene constructs containing WT, mutant or no insert. Detailed results of the minigene assay and sequencing results demonstrating the functional consequences of splicing variants are provided in Supplemental Fig. 3 (1–12).

Details on the 11 variants with observed splicing alteration were as follows. For five variants, at least two aberrantly spliced products were found: *SGCD* c.4-1G > A (partial intron 2 retention, partial exon 3 skipping), *TXNRD2* c.591 + 1G > C (partial exon 6 and total exon 7 skipping, at least 170 bp intron 7 retention), *DSP* c.273 + 5G > A (partial intron 2 retention, over 250 bp intron 2 retention), *TTN* c.25922–6 T > G (partial retention of intron 90, exon 91 skipping) and *TMEM43* c.1000 + 5G > T (exon 11 skipping, at least 138 bp intron 11 retention). Partial intron retention only was detected in four variants: *RYR2* c.1477-8C > A (partial intron 15 retention), *LAMA4* c.814 + 17A > G (partial intron 7 retention), *CSRP3* c.282-5_285del (partial intron 3 retention) and *ABCC9* c.2424 + 6C > G (partial intron 19 retention). Total exon skipping only was detected for two variants: *ILK* c.1210-2A > G (exon 12 skipping) and *TTN* c.31514-3A > G (exon 120 skipping). Results comparing the *in silico* predictions and *ex vivo* assay outcomes for the variants tested are shown in Table 2.

3.3. RT-PCR on patient RNA

From the 12 variants selected for RT-PCR, for only 9 variants RNA samples for testing were available. Eight of those variants were strongly predicted to affect splicing by SEPT–GD and positive for splice-altering on minigene assay and one variant (*DES* gene) was inconclusive on minigene assay. For strong-predicted splice variants, RT-PCR showed a splice effect for two variants, *TMEM43* c.1000 + 5G > T (exon 91 skipping) and *TTN* c.25922–6 T > G (exon 11 skipping) (Table 2), concordant with the minigene assay results. For the remaining six variants, RT-PCR did not show a splice effect. Based on the RT-PCR results, the *DES* c.79G > A variant that was inconclusive in the minigene assay did not affect splicing (data not shown), underscoring its weak splicing prediction.

3.4. Variant reclassification

The data obtained with the minigene and RT-PCR assays were used to reclassify the variants tested. Two variants, *SGCD* c.4-1G > A and *CSRP3* c.282-5_285del, were reclassified to LP as their effect, now proven via our functional assays, results in haploinsufficiency and loss of function (LoF) is a known disease mechanism for these genes within the respective cardiomyopathy subtype (i.e., DCM and HCM respectively). For four variants, *ABCC9* c.2424 + 6C > G, *ILK* c.1210-2A > G, *TTN* c.31514-3A > G and *TTN* c.25922–6 T > G, our results provide additional evidence for pathogenicity, but these were not reclassified to LP because the association of LoF variants (while the result of the splicing effect in the respective exons/genes) has not clearly been established as a disease mechanism for cardiomyopathy. In addition, the results for *DSP* c.273 + 5G > A also provided more proof for pathogenicity.However, although LoF is a known mechanism for disease for *DSP*, we did not reclassify this variant to LP because of its relatively high frequency in the general population (0.05 % in non_Finnish Europeans). This variant was also previously reported to alter the donor splice site on intron 2 in the *DSP* gene [Basso et al, 2006]. We also considered the *TMEM43* c.1000 + 5G > T variant a "VUS towards LP". Skipping of exon 11 will result in a frameshift and a premature stop codon in exon 12, and the variant allele is therefore expected to escape nonsense-mediated mRNA decay (NMD) and thus the production of a truncated protein, but the association of such a variant in this gene with disease is currently unknown. The remaining four variants, *TXRND* c.591 + 1G > C, *DES c.79G > A*, *RYR2* c.1477-8C > A, and *LAMA4* c.814 + 17A > G were not considered for reclassification. For the *TXRND2* and *DES* variants this was because both were predicted to result in a frame-shift and thus LoF and the association of that type of variant in these genes with cardiomyopathy is not yet established. For the *RYR2* and *LAMA4* variant reclassification was not considered because the association of these genes with the cardiomyopathy subtypes (DCM) found in the respective patients is not yet established. All four are still classified as VUS, without considering these splice results as additional proof of pathogenicity.

4. Discussion

In this study, we developed a decision tree, SEPT–GD, based on *in silico* predictions within the widely used commercial software Alamut®. SEPT–GD supports prioritisation of variants for functional analysis of splicing by setting thresholds for reliable predictions based on a reference variant set with known effects on splicing.This allows the prioritisation of potential splice variants with a high probability of being splice-altering that have been classified as VUS in routine diagnostics, which was confirmed with *in vitro* functional assessment of selected splice variants using minigene reporter assays with 100 % concordance.

Our decision tree showed higher sensitivity (91 %) and comparable specificity (88 %) for consensus and non-consensus splice-site variants when compared to similar individually tested algorithms on consensus splice sites. In a study comparing bioinformatic programmes (HSF, MES, NNS and ASSP) for analysis of variants within splice-site consensus regions that used a collection of 222 pathogenic variants and 50 benign polymorphisms, 75.9 %–83.6 % sensitivity and 72.3 %–81.3 % specificity ranges were reported [Tang et al, 2016]. *In silico* algorithms are thus more accurate in predicting the splicing effects of variants located closer to the intron–exon boundaries [Tosi et al, 2010]. The high occurances of inconclusive evidence and weak calls for middle-of-exon and deep intronic variants in our cohort, which lowered the specificity of SEPT–GD to 75 %, highlights the on-going challenge in predicting these categories of splice alterations using current software. Notably, in our cohort, we used variants in genes implicated in cardiomyopathies, and this may not necessarily be representative of other disease types. Testing the performance of SEPT–GD for other genes in daily practice

Table 2
Table showing *in silico* predictions, minigene splice reporter assay and RT-PCR results for variants tested.

Nr.	Gene	Variant	Transcript	Size cDNA*	New potential splice site	HSF Consensus value (0–100)	*In silico* prediction	Size New site*	*Ex vivo* confirmation	Minigene	Variant result	RT-PCR
1	*SGCD*	c.4-1G > A	NM_000337.5	452	ctcttctctcagCG	90.46	Exon 3 42 bp longer	494	42 bp insertion	Partial intron 2 retention	In frame ins	N/A
					caaatgcctcagGA	82.86	Exon 3 9 bp shorter	443	9 bp deletion	Partial exon 3 skipping	In frame del	
2	*TXNRD2*	c.591 + 1G > C	NM_001282512.1	405	AAGgtggga	87.21	Exon6/7 75 bp shorter	330	75 bp deletion	Partial exon 6 and total exon 7 skipping	Frameshift	No variation
					N/A	N/A	Exon 7 170 bp longer	575	170 bp insertion	Intron 7 retention (at least 170 bp cloned)	Frameshift	
3	*RYR2*	c.1477-8C > A	NM_001035.2	399	ttttttttaaagTT	84.08	Exon 16 6 bp longer	405	6 bp insertion	Partial intron 15 retention	In frame ins	No variation
4	*DSP*	c.273 + 5G > A	NM_004415.2	366	CTGgttagc	88.05	Exon 2 61 bp longer	427	61 bp insertion	Partial intron 2 retention	Framsehift	No variation
					N/A	N/A	Exon 2 250 bp longer	610	250 bp insertion	Over 250 bp intron 2 retention	Frameshift	
5	*LAMA4*	c.814 + 17A > G	NM_001105206.2	359	ACAgtacgt	74.91	Exon 7 12 bp longer	371	12 bp insertion	Partial intron 7 retention	In frame ins	No variation
6	*ILK*	c.1210-2A > G	NM_001014795.2	644	N/A	N/A	Exon 12 150 bp shorter	494	150 bp deletion	Exon 12 skipping	In frame del	N/A
7	*DES*	c.79G > A	NM_001927.3	2311	cttcccactcagCT	88.43	Exon 1 71 bp shorter	2240	inconclusive result	N/A	N/A	No variation
8	*CSRP3*	c.282-5_285del	NM_003476.4	396	CAGatgagg	70.07	Exon 4 110 bp longer	506	110 bp insertion	Partial intron 3 retention	Frameshift	No variation
9	*TTN*	c.25922–6 T > G	NM_001267550.1	542	ctttttttccaagTT	82.49	Exon 91 5 bp longer	547	5 bp insertion	Partial intron 90 retention	Frameshift	Exon 91 skipping
					N/A	N/A	Exon 91 279 bp shorter	264	279 bp deletion	Exon 91 skipping	In frame del	
10	*TTN*	c.31514-3A > G	NM_001267550.1	431	N/A	N/A	Exon 120 81 bp shorter	350	81 bp deletion	Exon 120 skipping	Frameshift	No variation
11	*TMEM43*	c.1000 + 5G > T	NM_024334.2	381	N/A	N/A	Exon 11 118 bp shorter	264	118 bp deletion	Exon 11 skipping	Frameshift	Exon 11 skipping
					N/A	N/A	Exon 11 138 bp longer	519	138 bp insertion	At least 138 bp intron 11 retention	Frameshift	
12	*ABCC9*	c.2424 + 6C > G	NM_020297.2	348	GAGgtatat	77.27	Exon 19 5 bp longer	352	5 bp insertion	Partial intron 19 retention	Frameshift	N/A

*cDNA- Exons + Vector pSPL3 – 263 bp.

and functionally following up would be ideal as a validation step.

Our decision tree is based on the splice prediction tools available in the Alamut® software. This commercially available software is used in many genetic diagnostic labs and integrates several splice effect prediction tools. However, these algorithms are often used with default parameter settings [Millat et al, 2015] or by adapting variable cut-off thresholds for the same algorithm [Houdayer et al, 2012; Steffensen et al, 2014; Bonnet et al, 2008]. SEPT–GD shows promising potential for predicting splice-affected variants with high(er) accuracy. Its application adds value to routine practice in that it reduces the large burden of testing variants that can be splice-affecting by narrowing down the list to strong candidates for functional assessment, potentially reducing the resources needed and time taken.

To confirm our decision tree predictions, we performed *in vitro* DNA analysis using minigene splicing reporter assays. This showed 100 % concordant results for variants with a strong predicted splice effect with SEPT–GD. In total, we analysed 12 selected VUSs, of which 10 variants were predicted to be strongly damaging by our decision tree, while one weak splice-effecting variant, *ABCC9* c.2424 + 6C > G, also showed positive minigene results. For the remaining weak predicted *DES* variant c.79G > A, a conclusion was not achieved due to lack of evidence. Pathogenicity assessment of candidate variants resulted in reclassification of two variants to LP. Notably, reclassification cannot only rely on the results of functional data and other criteria as presented by the ACMG/AMP guidelines (Richards et al, 2015) should also be met, like criteria PM2 (absent from or rare in controls) and/or PP3 (multiple lines of computational evidence support pathogenicity). In case of the two variants that were reclassified, data from our functional splice analyses (criteria PS3; well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect) suggest these variants result in LoF and therefore also criteria PVS1 (null variant in a gene where LoF is a known mechanism of disease) being met. When the full strength of these criteria would be considered, these variants would be reclassified as pathogenic, however, like also suggested by Rofes et al., 2020 in a comparable study, these criteria should be weighed more carefully, all together justifying reclassification these as LP. Likewise, for another six variants, we provide additional evidence for pathogenicity, although this was not yet sufficient for reclassification as LP. For these variants, co-segregation data may establish the association to disease. In the majority of the cases, the minigene-based assay is considered to provide a reliable assessment of whether a variant is splice-affecting. *In vitro* results, however, must be interpreted with caution, particularly for classification of VUS. Such methods by themselves cannot prove variant pathogenicity, as the pathobiological consequence may not be the same in the tissue of interest, and thus require complementing *in vivo* analyses [Groeneweg et al, 2014].

Using SEPT–GD, 31 unique variants that seemed to have a potential effect on the splicing machinery were detected in 39 patients, making up 2 % of the total cohort and 4.8 % of the patients with a VUS. This leads to an estimate of a 40 % potential increase (31 new variants added to the initial 41 splicing (L)P reported from the cohort (31/(31 + 41)) and a 5 % increase in total potential (L)P variants (72/526) in identification of potential pathogenic splice variants seen in cardiomyopathy patients. This result is comparable to a previously reported finding where the inclusion of variants functionally validated to alter splicing yielded a 50 % increase in pathogenic splicing variants in cardiomyopathy patients and demonstrated that ~5% of VUSs from affected patients alter splicing and are undetected disease-causing variants [Ito et al, 2017].

Comparing RT-PCR results for variants tested on minigene showed poor concordance (25 % for strongly predicted splice-altering variant using SEPT–GD). Although patient RNA is usually preferred for splicing analysis, several issues hamper the analysis of aberrant splicing from the variant allele, such as availability, degradation of aberrant transcripts through NMD, like would be expected for the RT-PCR experiments performed in material of carriers of the *TXNRD2, DSP, ILK, DES, CSRP3,* and *TTN* variants for which the introduction of premature stop codons is the most likely effect, and the confounding presence of normal and alternative transcripts from the WT allele in heterozygous patients. Minigene assays that display high sensitivity and specificity in the assessment of aberrant splicing caused by genetic sequence variants [Tournier et al, 2008] are thus used instead. However, occasional differences in splice patterns are observed between minigene and patient RNA analysis [Bonnet et al, 2008; Acedo et al, 2012; Steffensen et al, 2014]. For minigene assays it is important to keep in mind that the construct size might be a limitation for mimicking the natural genomic environment in the best way. Furthermore, for genes such as *DES* and *ILK*, indeed it might be difficult to assess variants in first and last exons to mimic the authentic splicing mechanism *in vivo*, requiring adapted minigenes to be designed for a splice effect to be depicted [Chen et al, 2018]. Assessing splice effects in RNA isolated from whole blood is restricted by the fact that not all genes or relevant transcripts thereof are expressed in blood. Although *SGCD, RYR2, DSP, LAMA4, CSRP3, TTN* and *ABCC9* are known to be lowly expressed in blood, we continued studying the respective genes and the other highly expressed genes (*TXNRD2, ILK, DES, TMEM43*) in blood [GTEx consortium, 2013], as other tissues were unavailable. We were able to detect the respective transcripts in blood, however only found aberrant transcripts for *TMEM43* and *TTN* variants with RT-PCR. Further studies using specific tissues is needed for conclusive results. Therefore, showing a splice effect in a functional assay is on its own not enough to classify a variant as pathogenic.

Several promising developments will improve and accelerate the evaluation of potential splice-site variants in the near future. Various approaches using next-generation RNA sequencing [Davy et al, 2017, Adamopoulos et al., 2018; Bryant et al, 2012; Park et al, 2013] are being developed that might be implemented in routine diagnostics in the coming years to ease the recognition of splice effects. Moreover, massively parallel reporter assays such as MaPSy (Massive parallel splicing assay) and Vex-seq (variant exon sequencing) [Soemedi et al, 2017; Adamson et al, 2018] have now become an increasingly popular tool to study alternative splicing and are expected to be the future of testing splicing variants. In addition, existing computational tools and online resources are designed to predict the effects of missense variants on protein products [Park et al, 2018], and work particularly well for variants in mutation hotspot regions in extensively studied genes with established association between disease and variants in those regions. Therefore, for variants in these regions computational predictions and algorithms may be sufficient for classification in the future. However, more evidence suggests that missense, nonsense and silent variants within exons and intronic variants can also disrupt splicing and cause diseases and should be the focus moving forward. Furthermore, non-SNP variants such as indels and short tandem repeats should be studied as they have been reported to modify *cis* splicing regulatory elements and to affect splicing [Gymrek et al, 2016; Zhang et al, 2014].

5. Conclusion

Our data show that SEPT–GD is a reliable tool to prioritise RNA splicing variants for functional follow-up, as exemplified by the variants identified in cardiomyopathy genes. Moreover, when confirmed by functional assays, this also supports classifying more VUSs as LP or P. Further studies incorporating larger datasets and other disease indications using SEPT–GD are needed to help solve the more difficult weakly predicted splice variants. The larger datasets may provide the necessary numbers of variants needed to train AI-based software tools, which will allow routine diagnostics to rely solely on prediction algorithms/models for (near) consensus splice-site variants while functional tests can be focused on middle-of-exon and deep(er) intronic variants, which are more difficult to predict.

Author contribution

M.Z.A, H.W, B. S-R and J.D.H.J. conceptualized the study, methodology and wrote the original draft of the manuscript; M.Z.A., L.G.B. and K.K. v D. curated the data. M.Z.A performed the validation and visualization. M.Z.A, H.W, B.S-R, J.D.H.J, Y.J.V., Y.M.H. and P.vd.Z. formally analyzed the data; H.W, B.S-R, R.J.S and J.D.H.J. supervised the study, all authors critically reviewed and edited the manuscript for important intellectual content; and all authors gave final approval of the version to be published.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Acedo, A., Sanz, D.J., Durán, M., Infante, M., Pérez-Cabornero, L., Miner, C., Velasco, E. A., 2012. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. Breast Cancer Res. 14 (3), R87. https://doi.org/10.1186/bcr3202. PMID: 22632462; PMCID: PMC3446350.

Adamopoulos, P.G., Theodoropoulou, M.C., Scorilas, A., 2018. Alternative Splicing Detection Tool-a novel PERL algorithm for sensitive detection of splicing events, based on next-generation sequencing data analysis. Ann. Transl. Med. 6 (12), 244. https://doi.org/10.21037/atm.2018.06.32.

Adamson, S.I., Zhan, L., Graveley, B.R., 2018. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. Genome Biol.. 19, 71. https://doi.org/10.1186/s13059-018-1437-x.

Alimohamed, M.Z., Johansson, L., Posafalvi, A., Boven, L.G., van Dijk, K.K., Walters, L., Vos, Y.J., Westers, H., Hoedemaekers, Y.M., Sinke, R.J., Sijmons, R.H., Sikkema-Raddatz, B., Jongbloed, J.D.H., van der Zwaag, P.A., 2021. Diagnostic yield of targeted next generation sequencing in 2002 Dutch cardiomyopathy patients. Int. J. Cardiol. S0167-5273(21)00378-8. doi: 10.1016/j.ijcard.2021.02.069. Epub ahead of print. PMID: 33662488.

Anna, A., Monika, G., 2018. Splicing mutations in human genetic disorders: examples, deletions and confirmation. J Appl Genet. 59 (3), 253–268.

Basso, C., Czarnowska, E., Della Barbera, M., Bauce, B., Beffagna, G., Wlodarska, E.K., Pilichou, K., Ramondo, A., Lorenzon, A., Wozniek, O., Corrado, D., Daliento, L., Danieli, G.A., Valente, M., Nava, A., Thiene, G., Rampazzo, A., 2006. Ultrastructural evidence of intercalated disc remodelling in arrhythmogenic right ventricular cardiomyopathy: an electron microscopy investigation on endomyocardial biopsies. Eur. Heart J. 27 (15), 1847–1854. https://doi.org/10.1093/eurheartj/ehl095. Epub 2006 Jun 14 PMID: 16774985.

Bonnet, C., Krieger, S., Vezain, M., Rousselin, A., Tournier, I., Martins, A., Berthet, P., Chevrier, A., Dugast, C., Layet, V., Rossi, A., Lidereau, R., Frébourg, T., Hardouin, A., Tosi, M., 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. J. Med. Genet. 45 (7), 438–446. https://doi.org/10.1136/jmg.2007.056895. Epub 2008 Apr 18 PMID: 18424508.

Bryant Jr, D.W., Priest, H.D., Mockler, T.C., 2012. Detection and quantification of alternative splicing variants using RNA-seq. Methods Mol. Biol. 883, 97–110. https://doi.org/10.1007/978-1-61779-839-9_7. PMID: 22589127.

Chen, Y., Huang, L., Jiao, X., Riazuddin, S., Riazuddin, S.A., Fielding, H.J., 2018. A novel *LRAT* mutation affecting splicing in a family with early onset retinitis pigmentosa. Hum Genomics. 12, 35.

Crehalet, H., Millat, G., Albuisson, J., et al., 2012. Combined use of in silico and in vitro splicing assays for interpretation of genomic variants of unknown significance in cardiomyopathies and channelopathies. Cardiogenetics. 2 (1), 6. https://doi.org/10.4081/cardiogenetics.2012.e6.

Davy, G., Rousselin, A., Goardon, N., Castéra, L., Harter, V., Legros, A., Muller, E., Fouillet, R., Brault, B., Smirnova, A.S., Lemoine, F., de la Grange, P., Guillaud-Bataille, M., Caux-Moncoutier, V., Houdayer, C., Bonnet, F., Blanc-Fournier, C., Gaildrat, P., Frebourg, T., Martins, A., Vaur, D., Krieger, S., 2017. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. Eur. J. Hum. Genet. 25(10), 1147–1154. doi: 10.1038/ejhg.2017.116. Epub 2017 Jul 26. PMID: 28905878; PMCID: PMC5602017.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., Béroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 37(9), e67. doi: 10.1093/nar/gkp215. Epub 2009 Apr 1. PMID: 19339519; PMCID: PMC2685110.

Fairbrother, W.G., Yeo, G.W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.A., Burge, C. B., 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res. 32(Web Server issue), W187–W190. doi: 10.1093/nar/gkh393. PMID: 15215377; PMCID: PMC441531.

Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frébourg, T., Tosi, M., 2010. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol. Biol. 653, 249–257. https://doi.org/10.1007/978-1-60761-759-4_15. PMID: 20721748.

Groeneweg, J.A., Ummels, A., Mulder, M., Bikker, H., van der Smagt, J.J., van Mil, A.M., Homfray, T., Post, J.G., Elvan, A., van der Heijden, J.F., Houweling, A.C., Jongbloed, J.D., Wilde, A.A., van Tintelen, J.P., Hauer, R.N., Dooijes, D., 2014. Functional assessment of potential splice site variants in arrhythmogenic right ventricular dysplasia/cardiomyopathy. Heart Rhythm. 11 (11), 2010–2017. https://doi.org/10.1016/j.hrthm.2014.07.041. Epub 2014 Jul 31 PMID: 25087486.

GTEx Consortium, 2013. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45 (6), 580–585. https://doi.org/10.1038/ng.2653. PMID: 23715323; PMCID: PMC4010069.

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M.J., Price, A.L., Pritchard, J.K., Sharp, A.J., Erlich, Y., 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat. Genet. 48, 22–29.

Harvey, S.E., Cheng, C., 2016. Methods for characterization of alternative RNA splicing. Methods Mol. Biol. 1402, 229–241.

Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., Lefol, C., Léone, M., Mazoyer, S., Muller, D., Remenieras, A., Révillion, F., Rouleau, E., Sokolowska, J., Vert, J.P., Lidereau, R., Soubrier, F., Sobol, H., Sevenet, N., Bressac-de Paillerets, B., Hardouin, A., Tosi, M., Sinilnikova, O.M., Stoppa-Lyonnet, D., 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. Hum. Mutat. 33 (8), 1228–1238. https://doi.org/10.1002/humu.22101. Epub 2012 May 11 PMID: 22505045.

Ito, K., Patel, P.N., Gorham, J.M., et al., 2017. Identification of pathogenic gene mutations in *LMNA* and *MYBPC3* that alter RNA splicing. Proc. Natl. Acad. Sci. 114 (29), 7689–7694. https://doi.org/10.1073/pnas.1707741114.

Jian, X., Boerwinkle, E., Liu, X., 2014a. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. Genet. Med. 16 (7), 497–503. https://doi.org/10.1038/gim.2013.176.

Jian, X., Boerwinkle, E., Liu, X., 2014b. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 42 (22), 13534–13544. https://doi.org/10.1093/nar/gku1206. PMID: 25416802; PMCID: PMC4267638.

Millat, G., Lafont, E., Nony, S., Rouvet, I., Bozon, D., 2015. Functional Characterization of Putative Novel Splicing Mutations in the Cardiomyopathy-Causing Genes. DNA Cell Biol. 34 (7), 489–496. https://doi.org/10.1089/dna.2015.2842.

Park, E., Pan, Z., Zhang, Z., Lin, L., Xing, Y., 2018. The expanding landscape of alternative splicing variation in human populations. Am. J. Hum. Genet. 102 (1), 11–26. https://doi.org/10.1016/j.ajhg.2017.11.002.

Park, J.W., Tokheim, C., Shen, S., Xing, Y., 2013. Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. Methods Mol. Biol. 1038, 171–179. https://doi.org/10.1007/978-1-62703-514-9_10. PMID: 23872975.

Raponi, M., Kralovicova, J., Copson, E., Divina, P., Eccles, D., Johnson, P., Baralle, D., Vorechovsky, I., 2011. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in *BRCA1* exon 6. Hum. Mutat. 32, 436–444. https://doi.org/10.1002/humu.21458.

Rhine, C.L., Cygan, K.J., Soemedi, R., et al., 2018. Hereditary cancer genes are highly susceptible to splicing mutations. PLoS Genet. 14 (3), 1–18. https://doi.org/10.1371/journal.pgen.1007231.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. ACMG Laboratory Quality Assurance Committee: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424.

Rofes, P., Menéndez, M., González, S., Tornero, E., Gómez, C., Vargas-Parra, G., Montes, E., Salinas, M., Solanes, A., Brunet, J., Teulé, A., Capellá, G., Feliubadaló, L., Del Valle, J., Pineda, M., Lázaro, C., 2020. Improving genetic testing in hereditary cancer by RNA analysis: tools to prioritize splicing studies and challenges in applying American college of medical genetics and genomics guidelines. J. Mol. Diagn. 22 (12), 1453–1468. https://doi.org/10.1016/j.jmoldx.2020.09.007. Epub 2020 Oct 1 PMID: 33011440.

Shen, T., Pajaro-Van de Stadt, S.H., Yeat, N.C., Lin, J.C.H., 2015. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Front. Genet. 215 (6), 1–9.

Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., Fairbrother, W.G., 2017. Pathogenic variants that alter protein code often disrupt splicing. Nat. Genet. 49 (6), 848–855. https://doi.org/10.1038/ng.3837. Epub 2017 Apr 17. PMID: 28416821; PMCID: PMC6679692.

Steffensen, A.Y., Dandanell, M., Jønson, L., Ejlertsen, B., Gerdes, A.M., Nielsen, F.C., Tv, H., 2014. Functional characterization of BRCA1 gene variants by minigene splicing assay. Eur. J. Hum. Genet. 22 (12), 1362–1368. https://doi.org/10.1038/ejhg.2014.40. Epub 2014 Mar 26. PMID: 24667779; PMCID: PMC4231409.

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., Abeysinghe, S., Krawczak, M., Cooper, D.N., 2003. Human Gene Mutation Database (HGMD): 2003 update. Hum. Mutat. 21 (6), 577–581. https://doi.org/10.1002/humu.10212. PMID: 12754702.

Tang, R., Prosser, D.O., Love, D.R., 2016. Evaluation of bioinformatic programmes for the analysis of variants within splice site consensus regions. Adv. Bioinformatics. 2016, 5614058. doi: 10.1155/2016/5614058. Epub 2016 May 24. PMID: 27313609; PMCID: PMC4894998.

Tosi, M., Stamm, S., Baralle, D., 2010. RNA splicing meets genetic testing: detection and interpretation of splicing defects in genetic diseases. Eur. J. Hum. Genet. 18 (6), 737–738. https://doi.org/10.1038/ejhg.2010.18. Epub 2010 Feb 24. PMID: 20179747; PMCID: PMC2987332.

Tournier, I., Vezain, M., Martins, A., Charbonnier, F., Baert-Desurmont, S., Olschwang, S., Wang, Q., Buisine, M.P., Soret, J., Tazi, J., Frébourg, T., Tosi, M., 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum. Mutat. 29 (12), 1412–1424. https://doi.org/10.1002/humu.20796. PMID: 18561205.

Vorechovsky, I., 2006. Aberrant 3'splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nuclei Acids Res. 34 (16), 4630–4641.

Zhang, X., Lin, H., Zhao, H., Hao, Y., Mort, M., Cooper, D.N., Zhou, Y., Liu, Y., 2014. Impact of human pathogenic micro-insertions and micro-deletions on post-transcriptional regulation. Hum. Mol. Genet. 23, 3024–3034.