

Combined use of *in silico* and *in vitro* splicing assays for interpretation of genomic variants of unknown significance in cardiomyopathies and channelopathies

Hervé Crehalet,¹ Gilles Millat,^{1,2}
Juliette Albuissou,³ Véronique Bonnet,¹
Isabelle Rouvet,⁴ Robert Rousson,¹
Dominique Bozon¹

¹Laboratoire de Cardiogénétique moléculaire, CBPE– Hospices Civils de Lyon, Bron, France; ²EA4612, Université Claude Bernard Lyon 1, Lyon, France;

³Laboratoire de Génétique Moléculaire, CHU, Hôtel-Dieu, Nantes, France;

⁴Centre de Biotechnologie Cellulaire, CBPE– Hospices Civils de Lyon, Bron, France

Abstract

The identification of molecular anomalies in patients with inherited arrhythmias or cardiomyopathies is a *multi challenge* due to: i) the number of genes involved; ii) the number of polymorphisms and the fact that most mutations are private; and iii) the variable degree of penetrance which complicates family segregation study. Consequently, a number of unclassified variants (UV) are found in patients' DNA and some (outside the canonical GT/AG) may affect splicing. Mutational screening on the most prevalent genes involved in arrhythmias syndromes or in cardiomyopathies was performed on a cohort made up of 740 unrelated French index probands. To identify splice variants among the identified UVs, a combination of available *in silico* and *in vitro* tools were used since transcript is not available. Using this approach, 10 previously identified UVs were reclassified as disease-causing mutations and, more precisely, as haploinsufficiency mutations rather than dominant-negative mutations. Most of them (7 of 10) were observed in *MYBPC3*.

Our study highlighted the importance of the combined use of *in silico* and *in vitro* splicing assays to improve the prediction of the functional impact of identified genetic variants. The primary challenge now, with new sequencing technologies, is to distinguish between background polymorphisms and pathogenic mutations. Since splice site mutations can account for almost 50% of disease-causing mutations, recognizing them from among other variations is essential.

Introduction

Cardiomyopathies and arrhythmia syndromes represent major genetic causes of sudden cardiac death.¹ Most of them are characterized by an autosomal-dominant mode of inheritance. Identification of disease-causing mutations may help to confirm a clinical diagnosis, to allow an appropriate clinical follow up and treatment in family members carrying the mutation(s) and to exclude the disease in non-carriers. However, their identification remains a challenge as molecular testing identifies nucleotide variations that may represent either known disease-causing mutations or polymorphisms, but also a large number of sequence variants of unknown clinical significance: so-called Unclassified Variants or UVs.²

Mutational screening on the most prevalent genes involved in arrhythmias syndromes or in cardiomyopathies was performed on a cohort made up of 740 unrelated French index probands [250 with long QT syndrome (LQTS)], 100 with Brugada syndrome, 280 patients with hypertrophic cardiomyopathy (HCM), and 110 patients with dilated cardiomyopathy (DCM).³⁻⁵ Based on the nature of the mutation, segregation studies and previously reported experimental data, genomic identified variants were classified either as disease-causing mutations (mutations previously reported and characterized), or as likely disease-causing mutations (frameshift mutations, nonsense mutations, severe missense mutations or mutations affecting the canonical dinucleotides of splice sites), or as UVs.

Among these UVs, some may affect normal splicing either by decreasing the efficiency of the natural donor and acceptor splice sites, by creation of new splice sites, or by alteration of sequence recognition for splicing regulatory elements such as exonic/intronic splicing enhancers or silencers (ESE, ISE, ESS, ISS). Splicing defects induced by substitutions located outside the canonical gt/ag may have different functional impact, ranging from severe (complete abolition of normal splicing) to mild (decrease in the number of normally spliced products). The aim of our study was to identify splice site mutations among nucleotide substitutions from our patients' cohort. Obviously, if the transcript is available, RNA based assay is the method of choice to visualize splicing anomalies. If not (most of the genes involved in cardiac disease are not transcribed in lymphocytes), *in silico* based predictions and *in vitro* splicing assay are valuable tools for the evaluation of splicing outcome of genomic variants.^{6,7} Given the large number of variants identified, a 3-step strategy was applied. The first round of selection was the elimination of variations with a known pathogenic effect. Then, an *in silico* screening was

Correspondence: Dominique Bozon, Laboratoire de Cardiogénétique Moléculaire, CBPE, Groupement Hospitalier Est, Hospices Civils de Lyon; 59 Bd Pinel 69677 Bron Cedex, France. Tel: +33.472129654 - Fax: 33+42785900. E-mail: dominique.bozon@chu-lyon.fr

Key words: cardiomyopathies, arrhythmia syndromes, splice mutations, molecular diagnosis.

Acknowledgments: the authors would like to thank Dr Raponi M and Dr Baralle D for providing the pTB minigene. This work was supported by the Hospices Civils de Lyon and by the Ministère de la Santé.

Contributions: DB, GM, RR, research theme definition; HC, VB DB methods and experiments design, *in silico* and *in vitro* analysis, data analysis and results interpretation; HC, GM, DB, manuscript writing; GM, molecular diagnosis of cardiomyopathies and arrhythmia syndromes (Lyon); JA, molecular diagnosis of arrhythmia syndromes (Nantes); IR, Hela cells providing.

Conflict of interests: the authors have no conflict of interests.

Received for publication: 27 March 2012.

Revision received: 22 May 2012.

Accepted for publication: 23 May 2012.

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

©Copyright M. Tariq et al., 2012
Licensee PAGEPress, Italy
Cardiogenetics 2012; 2:e6
doi:10.4081/cardiogenetics.2012.e6

applied on remaining variants in order to select those that could affect natural splice sites or create new splice sites (excluding variants that could affect splicing regulators). Finally, *in vitro* functional assays were performed to evaluate the functional impact on splicing of the genomic variants previously selected.

Materials and Methods

Patients

The study included 740 unrelated index cases with a confirmed clinical diagnosis: 350 arrhythmia patients (250 with LQTS and 100 with Brugada syndrome), 280 HCM patients and 110 DCM patients.³⁻⁵ The study was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained for all cases.

Mutation analysis

Genomic DNA was extracted from whole blood using a WIZARD Genomic DNA Purification kit (Promega, Madison, WI, USA). Mutation detection was performed using an HRM/sequencing strategy as previously reported.⁸⁻¹¹ Each exon and 50 flanking intronic bases were screened for mutations in the following genes:

- LQTS cases were screened for mutations in 6 genes coding ion channels subunits: *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1, NM_000218.2); *KCNH2* (potassium voltage-gated channel, subfamily H (eag-related), member 2, NM_000238.3); *SCN5A* (sodium channel, voltage-gated, type V, alpha subunit, NM_198056.2); *KCNE1* (potassium voltage-gated channel, Isk-related family, member 1, NM_001127669.1); *KCNE2* (potassium voltage-gated channel, Isk-related family, member 2 NM_172201.1);
- Brugada cases were screened for *SCN5A* mutations;
- HCM patients were screened for mutations in 4 sarcomeric genes: *MYBPC3* (myosin binding protein C, cardiac, NM_000256.3); *MYH7* (myosin, heavy chain 7, cardiac muscle, beta, NM_000257.2); *TNNT2* (troponin T type 2 (cardiac), NM_000364) and *TNNI3* (troponin I type 3 (cardiac), NM_000363.4).
- DCM patients were screened for mutations in the gene coding the nuclear membrane protein lamin A gene (*LMNA*, NM_170707.3), in *MYH7*, *TNNT2*, and *SCN5A* for some of the DCM patients with arrhythmia.

One variant, *KCNQ1* c.683+5G>A, was identified by Dr J Albuissou, (CHU Nantes, France).

These variations were not found in 200 normal chromosomes.

Another variant, c.1359+5G>A in the prosaposin gene (*PSAP*, NM_001042465.1) is not involved in heart disease but was included as a negative control, because this substitution has no impact on splicing [MT Vanier, G. Millat, personal communication 2007, and rs11000016 (dbSNP XML build 135) is found with a relative frequency of 0.788 for G and 0.212 for A in HapMap CEU].

Splicing predictions

Several algorithms for computational scoring of 5' and 3' splice sites based on different concepts were used, applying default parameters settings.

- Alamut (Interactive Software) is a resource integrating four different algorithms: Neural Network Splice Prediction (NNSplice),¹² MaxEntScan,¹³ Splice site Finder Like (SSF), Human Splicing Finder (HSF);¹⁴
- HBond;¹⁵

SSF, NNSplice and HSF use the Weight Matrix Model. This algorithm scores the frequency of each nucleotide for the 5' or the 3'ss.¹⁶ The MaxEntScan score for 5' and 3'ss, uses the maximum entropy principle and takes into account dependencies between neighboring and non-neighboring bases. The HBond score models the 5' splice site base pairing with 11 nucleotides (the last 3 exonic and the first 8 intronic) to U1 snRNA and provides a qualitative prediction of the probability of a mutation's aberrant splicing.

Minigene splicing reporter assay

For variations with a positive *in silico* effect, the exon with 150 flanking intronic bases was amplified from the patient DNA. Wild-type (WT) and mutated polymerase chain reaction (PCR) products were inserted in the *NdeI* restriction site of the pTB minigene vector.⁶ Transfection in HeLa cells and real time PCR (RT-PCR) procedures and analysis have been previously described.¹⁷

Mutagenesis

Site-directed mutagenesis was performed with the QuickChange® II XL Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies) to obtain the *MYBPC3* p.Val219Leu and the p.Val219Ile mutant minigenes from the p.Val219 minigene.

Results

Due to the large number of genomic variants identified, a first round of selection was applied with elimination of both common reported polymorphisms and variations with a known well-characterized pathogenic effect. After the first selection, the number of variations to be submitted to *in silico* analysis was 366. Use of several algorithms for computational scoring of 5' and 3' splice sites selected 15 genomic variations that could putatively affect natural splice sites or create new splice sites. Each variant with a potential positive or unclear *in silico* effect was investigated using an *in vitro* minigene splicing assay. Tables 1 and 2 showed a detailed overview of the results.

Donor splice site (5'ss)

Canonical donor splice sites of the human protein-coding genes has well-defined, more or less conserved, additional positions to the invariable intronic dinucleotide **gt**: **MAG | gtragt** (M: A or C; R: purine) for pairing with U1 snRNA.¹⁸ After *in silico* analysis, a 10% change between wild-type and mutant score values in at least two different algorithms,¹⁹ or unclear predictions (such as an increase in

the score of the mutant splice site for one program while another program lowered the score for the same mutant) were criteria for performing an *in vitro* functional assay for the variations concerned.

Using these criteria, 9 genomic variations were selected and further investigated using minigene splicing assay: 4 of them affected the highly conserved G (80%) at intronic position +5. These 4 variations were analyzed using the *PSAP* c.1359+5G>A variation as a negative control. If we scan this variation with the splicing prediction softwares, 4 of them predicted an effect ranging from -11.1% (NNSplice) to -27.9% (MaxEntScan) (Table 1). HBond predicted that it would not interfere with splicing and, indeed, transcript analysis in the patient's lymphoblasts, skin fibroblasts and minigene experiments confirmed that this change does not affect splicing. The donor splice site of the *PSAP* exon 12 (CAG|gtacgc) has a high score and the number of adjacent bonds (14), as defined by HBond, is identical in the WT and the mutant. The 5 other variants were all exonic and 4 of them affected the last nucleotide of exon.

Detailed characteristics of each of these 9 nucleotide variations are presented below.

KCNQ1 c.477+5G>A (intron 2)

All 5 algorithms predicted a severe impact on splicing with a strong decrease in the natural donor splice site strength and the potential use of a donor cryptic site located at intronic position c.477+80 (Table 1). These predictions were confirmed by minigene assay as, in HeLa cells transfected with the mutant minigene, the donor site at position c.477+80 was exclusively used (Supplementary Figures 1-1). The corresponding transcript would encode the first 159 amino acids (AA) followed by 4 aberrant AA and a premature termination codon (PTC). This aberrant mRNA is likely to be degraded by nonsense-mediated decay (NMD) causing a complete loss of function from this allele. This mutation was identified in a 56-year old LQTS patient.

KCNQ1 c.683+5G>A (intron 4)

This variant was also predicted to have a severe impact on splicing (Table 1). *In vitro* assay showed exon 4 to have been completely skipped and the resulting transcript would encode the first 202 AA followed by 8 aberrant AA and a PTC (Supplementary Figures 1-2). This transcript is probably degraded by NMD and must be considered as a *null* allele. c.683+5G>A was identified in a young compound heterozygous patient with the *KCNQ1* missense p.Arg190Leu (c.569G>T). The patient suffered from syncope triggered by emotion or swimming from the age of three years while his father, carrier of the c.683+5G>A, is currently asymptomatic.

KCNQ1 c.1032+5G>A (intron 7)

Both *in silico* and *in vitro* analysis showed a severe impact of this substitution on splicing. In HeLa cells transfected with the mutant minigene, 2 aberrantly spliced products were found: one skipped exon 7 and the other used a cryptic donor splice site at intronic position c.1032+16 as suggested by SSFL, MaxEnt Scan and HSF (Table 1, Supplementary Figures 1-3). Exon 7 skipping maintains the reading frame but deletes 37 AA, while the use of the c.1032+16 as a donor splice site adds 5 aberrant amino acids and also maintains the reading frame. This amino acid deletion/insertion would alter the 6th transmembrane segment of the KvLQT1 channel. *KCNQ1* c.1032+5G>A was identified in 3 LQTS patients from the same family.

MYBPC3 c.3190+5G>A (intron 29)

Three out of the 5 algorithms tested predicted an impact on splicing by complete abolition of the donor splice site. This donor splice site does not have a strong consensus as NNSplice did not detect the physiological donor site. This is illustrated by the minigene assay as, in HeLa cells transfected with the wild-type sequence, some of the transcripts are lacking exon 29 (248 bp fragment, Supplementary Figures 1-4 b). Mutant minigene transfection showed complete exon 29 skipping and the resulting transcript would encode a truncated protein containing the 1063 amino acids, followed by 10 aberrant amino acids and a PTC, probably inducing transcript degradation via NMD (Supplementary Figures 1-4). HBond did not make a reliable prediction, because the WT donor splice site score of intron 29 is too weak for this software: only 8 bonds in a continuous stretch of complementary nucleotides were found, which is below the threshold of 12 adjacent bonds required for a more accurate prediction by HBond. This splice site mutation was identified in a 48-year old HCM patient.

MYBPC3 c.1090G>A; p.Ala364Thr (exon 12)

This variation affects the last nucleotide of exon 12 and changes the highly conserved Ala residue at position 364 to a Thr. The last exonic nucleotide is also highly conserved (80% G) and is part of the pairing sequence with U1 snRNA (G at this position and at the first and at the fifth intronic positions makes 3 GC pairing). All 5 algorithms predicted an impact on splicing and, indeed, mutant minigene assay showed complete exon 12 skipping (Supplementary Figures 1-5). HeLa cells transfected with the wild-type sequence also showed partial exon 12 skipping because of

the very weak acceptor splice site of intron 11 (Table 2, *MYBPC3* c.927-9G>A). The cryptic donor splice site at position c.1090+74, detected by SSF and HSF, was not used in the presence of the mutation in the minigene assay. Exon 12 skipping produces a PTC and the abnormal transcript would be degraded by NMD.

MYBPC3 c.1213A>G; p.Met405Val (exon 13)

This variation changes the Met residue at position 405 to a Val; Met 405 is not highly conserved and Met and Val are closely related amino acids. This exonic variant, located at 10 nucleotides before the end of exon 13, was also predicted by the 4 algorithms tested to create a new donor splice site inside the exon. This mutant donor splice is stronger than the WT site which has a poor consensus, not detected by SSFLike and NNSplice. Minigene experiments confirmed these *in silico* data: transfection of the normal minigene showed both a normal spliced product (379 bp) and exon skipping (248 bp) because of the weak consensus of the normal donor splice site at c.1223+1 (Supplementary Figures 1-6). Transfection of the mutant minigene showed only a shorter 368 bp product corresponding to the exclusive use of the strong c.1213 mutant splice donor site. Furthermore, exon skipping disappeared in cells transfected with the mutant minigene because of the strong donor splice site created by the mutation. The mutant transcript missed the last 11 bases of exon 13 and, if not completely degraded by NMD, would encode a polypeptide containing 404 amino acids followed by 4 aberrant amino acids and a PTC. *MYBPC3* c.1213G>A is a splice mutation identified in a 32-year old HCM patient.

MYBPC3 c.1624G>C; p.Glu542Gln (exon 17)

This nucleotide change affects the last nucleotide of exon 17. All algorithms predicted a severe impact on splicing. For testing this variation in a minigene assay, a 917 bp genomic fragment (from intron 14 to intron 17) was inserted in the minigene because of the small size of introns 15 and 16 (86 bp). This was complicated by the weak acceptor splice sites of introns 15 and 16. Results from the normal minigene transfection showed several spliced products including the normally spliced fragment and two others products retaining introns 15 and 16 (Supplementary Figure 1-7). Transfection of the mutant minigene showed a complete absence of the normal spliced product and presence of only mis-spliced products, confirming the *in silico* predictions. This splice mutation was identified in a patient with a

hypertrophic obstructive cardiomyopathy diagnosed at 21 years of age.

MYBPC3 c.2067G>T; p.Gln689His (exon 21)

This variation affects the last nucleotide of exon 21 and changes the highly conserved Gln residue at position 689 to a His. *In silico* and *in vitro* analyses showed both abolition of the normal donor site and complete exon 21 skipping, creating a PTC (Supplementary Figures 1-8). This mutation was identified in a 7-month old HCM patient.

LMNA c.1488G>A; p.Thr496Thr (exon 8)

This variation also affects the last exonic nucleotide. All software used predicted a potential impact on splicing with a cryptic donor site at position c.1489-26 becoming much stronger than the mutated site. A minigene assay was performed, with a genomic fragment from intron 7 to intron 9 because of the small size of intron 8 (84 bp). Minigene experiments showed only normally spliced products even in the presence of the mutation (Supplementary Figures 1-9). The cryptic donor splice site at c.1489-26 is not used in transfection experiments because it is probably repressed; even in the absence of the mutation, this cryptic donor splice site has a higher score than the WT donor splice site for SSFL, NNSplice and HSF. The study of this mutant highlights the requirement of functional tests to validate *in silico* predictions.

Acceptor Splice Site (3'ss)

The 3'splice site is defined by three separate elements: the branch site, the polypyrimidine tract and the 3'canonical ag: yyyyyyincag|G.²⁰ Except for the ag, other signals are degenerated and accurate identification of 3'ss by bioinformatics may sometimes be difficult. The 6 selected mutants for *in vitro* minigene splicing assay were genomic variations which either showed variable predictions (*SCN5A* c.-52-4C>T (Supplementary Figures 2-1), *TNNT2* c.68-5_68 -3delinsTT (Supplementary Figures 2-6), or affected non canonical (AT/AC) splicing sites [*SCN5A* c.4438-3C>T (Supplementary Figures 2-2)], or decreased the WT score [*KCNH2* c.308G>T (Supplementary Figures 2-3), *MYBPC3* c.655G>T (Supplementary Figures 2-4)], or affected weak acceptor sites (not found by at least one software) [*MYBPC3* c.927-9G>A (Supplementary Figures 2-5)]. Detailed characteristics of 3 of the variations are presented below.

KCNH2 c.308G>T (p.Gly103Val) and MYBPC3 c.655G>T (p.Val219Phe)

These 2 substitutions affected the first

Table 1. *In silico* predictions and *in vitro* results of the effect on splicing of sequence variations at donor splice site.

Gene	Variation	<i>In silico</i> studies WT 5'ss > MUT 5'ss			Human splicing finder	HBond	<i>In vitro</i> splicing studies
		Splice site finder like	MaxEntScan	NNSplice			
<i>KCNQ1</i>	c.477+5G>A	80.40 > below threshold cryp 5'ss (+80): 80.23 84.85 > 72.7 (-14.3%)	9.89 > 4.52 (-54.3%) cryp 5'ss (+80): 5.87 9.73 > 4.89 (-49%)	0.97 > below threshold cryp 5'ss (+80): 0.7 0.99 > 0.79 (-20.2%)	85.49 > 73.33 (-14.2%) cryp 5'ss (+80): 87.61 90.04 > 77.87 (-13.5%)	HC2 → LC3 HP HC1 → LC4 HP	Use of a cryptic 5'ss c.477+80 Exon 4 skipped
	c.688+5G>A						
	c.1032+5G>A	77.95 > below threshold cryp 5'ss (+16): 71.76	9.00 > 3.46 (-61.6%) cryp 5'ss (+16): 5.46	0.97 > below threshold	85.15 > 72.98 (-14.3%) cryp 5'ss (+16): 83.2	HC1 → LC3 HP	Use of a cryptic 5'ss c.1032+16; exon 7 skipped
	<i>MYBPC3</i>	c.1090G>A (p.Ala364Thr)	80.35 > 72.21 (-14.4%) cryp 5'ss (+74): 80.22 physiological: not detected mutant: 89.86	4.90 > below threshold physiological: 4.95 mutant: 9.60	0.98 > below threshold physiological: not detected mutant: 0.99	92.80 > 82.22 (-11.4%) cryp 5'ss (+74): 90.82 physiological: 75.94 mutant: 96.67	HC2 → LC2 HP NT
	c.1213A>G (p.Met405Val)	80.59 > below threshold cryp 5'ss (+82): 79.15	7.75 > 4.09 (-47.2%)	0.90 > 0.45 (-49.7%)	90.86 > 79.84 (-12.1%)	HC3 → HC1 HP	Exon 21 skipped
	c.1624G>C (p.Glu542Gln)	71.80 > below threshold	9.04 > below threshold	0.91 > below threshold	76.95 > 66.08 (-14.1%)	HC2 → LC2 HP	Exon 29 skipped
	c.2067G>T (p.Gln689His)	72.21 > below threshold	6.18 > below threshold	physiological: not detected	83.31 > 71.14 (-14.6%)	LC3 → LC2 LP	
	c.3190+5G>A						
<i>LMNA</i>	c.1488G>A (p.Thr496Thr)	86.42 > 74.28 (-14%) cryp 5'ss (c.1489-26: 87.98)	11.45 > 8.34 (-27.2%) cryp 5'ss (c.1489-26: 9.65)	0.99 > 0.83 (-15.7%) cryp 5'ss (c.1489-26: 1.00)	93.14 > 82.56 (-11.4%) cryp 5'ss (c.1489-26: 93.38)	HC3 → HC1 HP	No effect
	<i>PSAP</i>	c.1359+5G>A	83.44 > 71.28 (-14.6%)	10.12 > 7.29 (-27.9%)	0.99 > 0.88 (-11.1%)	89.02 > 76.86 (-13.6%)	HC2 → HC2 LP

Scores indicated are assigned by the web-based tools following analysis of the input sequence. Default thresholds were used for all bioinformatics tools. WT, wild type; MUT, mutant; cryp 5'ss, cryptic donor splice site; HC, high complementarity; LC, low complementarity; LP, high probability for aberrant splicing; LP, low probability for aberrant splicing; NT, not tested.

Table 2. *In silico* predictions and *in vitro* results of the effect on splicing of sequence variations at acceptor splice site.

Gene	Variation	<i>In silico</i> studies WT 3'ss > MUT 3'ss			Human splicing finder	<i>In vitro</i> splicing studies
		Splice site finder like	MaxEntScan	NNSplice		
<i>SCN5A</i>	c.-52-4C>T	87.8 (no change)	9.07 > 10.01 (+10.4%)	0.85 > 0.83 (-2.4%)	87.16 > 87.82 (+0.7%)	No effect
	c.4438-3C>T		Non canonical AT/AC intron			No effect
<i>KCNH2</i>	c.308G>T (p.Gly103Val)	84.9 > 78.7 (-7.4%)	9.5 > 7.6 (-20.0%)	1.0 > 0.8 (-16.7%)	91.33 > 87.04 (-5%)	No effect
	<i>MYBPC3</i>	c.655G>T (p.Val219Phe)	78.04 > 71.80 (-8%)	7.45 > 5.39 (-27.7%)	0.87 > 0.44 (-49.4%)	86.3 > 82 (-4.9%)
	c.927-9G>A		Physiological acceptor splice site not detected		81.91 > 81.79 new 3'ss: 87.17	Exon 11 skipped
<i>TNNT2</i>	c.68-5_68-3delinsTT	74.46 > below threshold	1.42 > 3.62 (+155.7%)	Physiological: not detected	83.17 > 74.52 (-10.4%)	No effect

Scores indicated are assigned by the web-based tools following analysis of the input sequence. Default thresholds were used for all bioinformatics tools. WT, wild type; MUT, mutant; new 3'ss, new acceptor splice site.

exonic nucleotide with a G>T change. In silico analysis predicted, for these 2 variations, a decrease in the score ranging from 5% to 57% (Table 2). However, minigene assay showed that *KCNH2* c.308G>T had no effect on splicing (Supplementary Figures 2-3), while *MYBPC3* c.655G>T induced complete exon 6 skipping (Supplementary Figures 2-4). Another mutation affecting the same nucleotide, c.655G>C, p.Val219Leu, has been previously reported.²¹ To evaluate its impact on splicing, the corresponding mutant minigenes and the non-reported transition c.655G>A, p.Val219Ile were also generated by site directed mutagenesis. The c.655G>C transversion had the same effect as c.655G>T (complete exon 6 skipping), while the transition c.655G>A showed both exon 6 skipping and exon 6 inclusion. The *MYBPC3* c.655G>T splice mutation was identified in a 48-year old HCM patient.

MYBPC3 c.927-9G>A (intron 11)

This previously reported variation²² was identified in our cohort in 2 unrelated patients: a 12-year old HCM boy and a 30-year old man with HCM and atrial fibrillation. The weak acceptor splice site of intron 11 was detected only by HSF which predicted the creation, at position c.927-9, of a new acceptor site stronger than the WT (Table 2). HeLa cells transfected with the WT sequence showed two different transcripts (Supplementary Figures 2-5 b): a 410 bp normal product containing exon 12 and the 248 bp product corresponding to exon 12 skipping, as for the WT minigene c.1090G (Supplementary Figures 1-5 b), confirming that the weak acceptor site of intron 11 is not always recognized in HeLa cells. Mutant minigene assay showed complete exon 12 skipping and the transcript lacking exon 12 would encode a truncated protein (Supplementary Figures 2-5 b).

Discussion

The aim of our study was to detect splice site mutations among UVs previously identified in our cohort of patients with arrhythmias or cardiomyopathies. For this purpose, in the absence of an RNA based assay, a combination of *in silico* and *in vitro* tests were used to distinguish between neutral and pathological nucleotide substitutions.

Considering the donor splice site, 4 out of 9 mutants affected the +5 position with a G>A nucleotide change and 4 variants changed the last exonic base (position -1 of the donor splice site). Positions -1 and +5 are highly conserved (G in 80% and 78.1%, respectively) and 35% of the mutations affecting these positions caused aberrant splicing compared to 9% for the other positions, excluding the invariant positions.²³

Positions -1 and +5 are the most prominent positions (excluding +1 and +2) in base pairing with U1 snRNA: G at -1, +1, and +5 are involved in GC pairing, and a mispairing at one of these positions may be critical. The analysis of 45519 5'ss showed that the number of base pairs ranges from 3 to 9 but the average is around 7, and this analysis also showed a strong dependency between some and adjacent positions.^{24,25} The exonic -1 and the intronic +5 positions of the donor splice site are also involved in pairing with U5 snRNA and U6 snRNA, respectively, and this may also contribute to their high degree of conservation.

Accuracy of *in silico* prediction for variations affecting the 5'ss (outside *gt*) will depend on the algorithm used. Nucleotide frequency based matrices must be completed by matrices that consider the influence of neighboring bases and the interdependence of nucleotides of different site positions. The number of hydrogen bonds with U1 snRNA (in our study measured by HBond) is an additional interesting parameter. For the 3 *KCNQ1* mutants, the web tools predicted a severe impact on splicing with a decrease in score ranging from 14% (HSF) to below threshold, and minigene experiments confirmed the severe effect on splicing. All the 5 tested variants affecting the intronic +5 position (Table 1) had the same impact for HSF (-13.5 to -14.6% decrease in score) showing the lack of sensitivity of this algorithm for the +5G>A change. HBond was the only bioinformatic tool to make the right prediction for the PSAP c.1359+5G>A, but it was unable to predict the effect of the *MYBPC3* c.3190+5G>A because the WT splice site score is low. The donor site of *MYBPC3* exon 29 (TTGgtgcgt) has a poor consensus (not detected by NNSplice), has a weak hydrogen bonding with U1 snRNA and changing the +5G to A dramatically decreases this interaction. Such weak donor sites may require additional factors for proper splicing such as splicing regulators or, as shown for NF1 exon 29, may have a lower U1 snRNA dependence.²⁶ The majority of +5G>A transitions reported in the literature affect splicing and induced either exon skipping and/or cryptic site activation.

The 3'splice site is defined by three separate elements: the branch site (BS), the polypyrimidine tract and the 3' *yag*/R. Except for the *ag*, the other signals are degenerated and accurate identification of 3'ss by bioinformatics may sometimes be difficult, making it almost impossible to predict sequence variations (*MYBPC3* c.927-9G>A), or making prediction unclear (*TNNT2* c.68-5_68-3delinsTT) (Table 2). A possible explanation is that the new *ag* created by *MYBPC3* c.927-9G>A recruits spliceosome components that could compete with the authentic site, preventing fully functional splicing complex assembly.²⁷ Another

possibility is that *MYBPC3* c.927-9G>A would modify a regulatory element necessary for the recognition of the weak acceptor splice site of intron 11.

Two substitutions affected the first exonic nucleotide and involved a G>T change: *KCNH2* c.308G>T and *MYBPC3* c.655G>T. The first exonic base is part of the acceptor splice site but the G is mildly conserved (about 49%), with a purine preference (77%) at this position¹⁶ which is also involved in an interaction with the U5 snRNP loop 1.²⁸ Changing the *MYBPC3* c.655G to a C showed also complete exon 6 skipping. This shows that, indeed, a purine is required at position c.655, but it is not sufficient for correct splicing since changing the c.655G to an A only partially rescued inclusion of exon 6 on the transcript. A G as the first exonic base appears to be necessary for the correct splicing of *MYBPC3* exon 6. A similar case has been described in the *DHFR* gene in CHO cells: a G>T mutation involving the first base of exon 5 induced exon 5 skipping. In a revertant of this mutant, this base was changed to A and about 25% of mRNA molecules were correctly spliced.²⁹ A regulatory sequence may also be involved in *MYBPC3* exon 6 recognition and could explain that a purine is not sufficient for proper splicing: a search for exonic splicing enhancer identified putative binding sites for SF2/ASF and SRp40 (ESE Finder, via HSF) with SRp40 scores ranging from +71% above threshold for G, +48% for C, and only +6% for T and A.

As previously reported,^{6,7,19,30} our study confirmed that the most accurate *in silico* prediction of splice site mutations required the use of a combination of different algorithms. Another important point is that, if the transcript is not available, predictions must be confirmed by a minigene assay and, ideally, all nucleotide variations should be tested *in vitro*. However, a systematic minigene experiment for each new substitution is costly in terms of both time and money and would most probably give only a few or even no more positive results.^{31,32} However, for non-canonical introns (*SCN5A* c.4438-3C>T) prediction is impossible and the minigene is mandatory. Minigene based assay provides a reliable assessment of whether a nucleotide variation is deleterious for splicing but the consequence (exon skipping and/or activation of cryptic site) may not be strictly the same in the tissue of interest. There is a difference in transcription speed between the gene promoter and the vector promoter. Intron size also differs and these parameters are known to be important for exon definition in alternative splicing.^{33,34}

In summary, our results allowed us to reclassify 10 previously identified UVs as disease-causing mutations and, more precisely, as haploinsufficiency mutations rather than dominant-negative mutations. Most of them (7 of

10) were observed in *MYBPC3*. This result was expected as 70% of *MYBPC3* mutations introduce a premature termination codon. Identification of these mutations is important as it has been shown that patients with *MYBPC3* splice mutations have a higher risk of sudden death before the age of 35 years.³⁵ Our study highlights the importance of strategies combining the use of *in silico* and *in vitro* assays to be able to identify those genetic variants of UVs that have a clinical significance. There will be an increasing need for this strategy as the more exhaustive genetic testing of cardiomyopathies and arrhythmias using next generation sequencing will identify an increasing number of genetic variations.

References

- Barsheshet A, Brenyo A, Moss AJ, Goldenberg I. Genetics of sudden cardiac death. *Curr Cardiol Rep* 2011;13:364-76.
- Monserat M, Mazzanti A, Ortiz-Genga M, et al. The interpretation of cardiogenetic tests in inherited cardiovascular diseases. *Cardiogenetics* 2011;1:e8.
- Millat G, Chevalier P, Restier-Miron L, et al. Spectrum of pathogenic mutations and associated polymorphisms in a cohort of 44 unrelated patients with long QT syndrome. *Clin Genet* 2006;70:214-27.
- Millat G, Bouvagnet P, Chevalier P, et al. Prevalence and spectrum of mutations in a cohort of 192 unrelated patients with hypertrophic cardiomyopathy. *Eur J Med Genet* 2010;53:261-7.
- Millat G, Bouvagnet P, Chevalier P, et al. Clinical and mutational spectrum in a cohort of 105 unrelated patients with dilated cardiomyopathy. *Eur J Med Genet* 2011;54:e570-5.
- Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. *J Med Genet* 2005;42:737-48.
- Houdayer C, Dehainault C, Mattler C, et al. Evaluation of *in silico* splice tools for decision-making in molecular diagnosis. *Hum Mutat* 2008;29:975-82.
- Millat G, Chanavat V, Créhalet H, Rousson R. Development of a high resolution melting method for the detection of genetic variations in Long QT Syndrome. *Clin Chim Acta* 2011;412:203-7.
- Millat G, Chanavat V, Créhalet H, Rousson R. Development of a high resolution melting method for the detection of genetic variations in hypertrophic cardiomyopathy. *Clin Chim Acta* 2010;411:1983-91.
- Millat G, Chanavat V, Julia S, et al. Validation of high-resolution DNA melting analysis for mutation scanning of the LMNA gene. *Clin Biochem* 2009;42:892-8.
- Millat G, Chanavat V, Rodriguez-Lafrasse C, Rousson R. Rapid, sensitive and inexpensive detection of SCN5A genetic variations by high resolution melting analysis. *Clin Biochem* 2009;42:491-9.
- Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311-23.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377-94.
- Desmet FO, Hamroun D, Lalande M, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.
- Freund M, Asang C, Kammler S, et al. A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res* 2003;31:6963-75.
- Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987;15:7155-74.
- Crehalet H, Latour P, Bonnet V, et al. U1 snRNA mis-binding: a new cause of CMT1B. *Neurogenetics* 2010;11:13-9.
- Buratti E, Chivers M, Královicová J, et al. Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res* 2007;35:4250-63.
- Bonnet C, Krieger S, Vezain M, et al. 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* 2008;45:438-46.
- Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003;17:419-37.
- Van Driest SL, Vasile VC, Ommen SR, et al. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004;44:1903-10.
- Rodriguez-Garcia MI, Monserat L, Ortiz M, et al. Screening mutations in myosin binding protein C3 gene in a cohort of patients with hypertrophic cardiomyopathy. *BMC Med Genet* 2010;11:67.
- Nakai K, Sakamoto H. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* 1994;141:171-7.
- Carmel I, Tal S, Vig I, Ast G. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* 2004;10:828-40.
- Roca X, Olson AJ, Rao AR, et al. Features of 5'-splice-site efficiency derived from disease-causing mutations and comparative genomics. *Genome Res* 2008;18:77-87.
- Raponi M, Buratti E, Dassi E, et al. Low U1 snRNP dependence at the NF1 exon 29 donor splice site. *FEBS J* 2009;276:2060-73.
- Královicová J, Christensen MB, Vorechovský I. Biased exon/intron distribution of cryptic and *de novo* 3' splice sites. *Nucleic Acids Res* 2005;33:4882-98.
- McKeown M. The role of small nuclear RNA in RNA splicing. *Curr Opin Cell Biol* 1993;5:448-54.
- Mitchell PJ, Urlaub G, Chasin L. Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol Cell Biol* 1986;6:1926-35.
- Le Guédard-Méreuze S, Vaché C, Baux D, et al. *Ex vivo* splicing assays of mutations at noncanonical positions of splice sites in USHER genes. *Hum Mutat* 2010;31:347-55.
- Whiley PJ, Pettigrew CA, Brewster BL, et al. Effect of BRCA2 sequence variants predicted to disrupt exonic splice enhancers on BRCA2 transcripts. *BMC Med Genet* 2010;11:80.
- Pettigrew CA, Wayte N, Wronski A, et al. Colocalisation of predicted exonic splicing enhancers in BRCA2 with reported sequence variants. *Breast Cancer Res Treat* 2008;110:227-34.
- Berget SM. Exon recognition in vertebrate splicing. *J Biol Chem* 1995;270:2411-4.
- Haj Khelil A, Deguillien M, Morinière M, et al. Cryptic splicing sites are differentially utilized *in vivo*. *FEBS J* 2008;275:1150-62.
- Ehlermann P, Weichenhan D, Zehlein J, et al. Adverse events in families with hypertrophic or dilated cardiomyopathy and mutations in the MYBPC3 gene. *BMC Med Genet* 2008;9:95.