

ARTICLE OPEN



The burden of splice-disrupting variants in inherited heart disease and unexplained sudden cardiac death

Emma S. Singer^{1,2}, Joshua Crowe^{1,2}, Mira Holliday^{1,2}, Julia C. Isbister^{1,2,3}, Sean Lal², Natalie Nowak^{1,3}, Laura Yeates^{1,2,3,4,5,6}, Charlotte Burns^{1,3}, Sulekha Rajagopalan⁷, Ivan Macciocca^{8,9,10}, Ingrid King⁸, Julie Wacker¹¹, Jodie Ingles^{1,2,3,4,5,6}, Robert G. Weintraub^{8,10,11}, Christopher Semsarian^{1,2,3} and Richard D. Bagnall^{1,2,3}✉

There is an incomplete understanding of the burden of splice-disrupting variants in definitively associated inherited heart disease genes and whether these genes can amplify from blood RNA to support functional confirmation of splicing outcomes. We performed burden testing of rare splice-disrupting variants in people with inherited heart disease and sudden unexplained death compared to 125,748 population controls. ClinGen definitively disease-associated inherited heart disease genes were amplified using RNA extracted from fresh blood, derived cardiomyocytes, and myectomy tissue. Variants were functionally assessed and classified for pathogenicity. We found 88 in silico-predicted splice-disrupting variants in 128 out of 1242 (10.3%) unrelated participants. There was an excess burden of splice-disrupting variants in *PKP2* (5.9%), *FLNC* (2.7%), *TTN* (2.8%), *MYBPC3* (8.2%) and *MYH7* (1.3%), in distinct cardiomyopathy subtypes, and *KCNQ1* (3.6%) in long QT syndrome. Blood RNA supported the amplification of 21 out of 31 definitive disease-associated inherited heart disease genes. Our functional studies confirmed altered splicing in six variants. Eleven variants of uncertain significance were reclassified as likely pathogenic based on functional studies and six were used for cascade genetic testing in 12 family members. Our study highlights that splice-disrupting variants are a significant cause of inherited heart disease, and that analysis of blood RNA confirms splicing outcomes and supports variant pathogenicity classification.

npj Genomic Medicine (2023)8:29; <https://doi.org/10.1038/s41525-023-00373-w>

INTRODUCTION

Inherited cardiomyopathies and arrhythmia syndromes are important causes of heart failure and sudden cardiac death^{1,2}. People with suspected inherited heart diseases should undergo genetic testing as this can clarify uncertain clinical diagnoses, guide treatment options, inform prognosis, and help stratify risk in family members³. While genetic testing typically focuses on rare variants in protein-coding regions and the essential splice sites of genes, there is an increasing appreciation of the range and manner by which variants can disrupt RNA splicing^{4–6}. This includes variants previously annotated as missense, nonsense, or synonymous changes and intronic variants located far away from protein-coding regions. Assigning the correct consequence of variants at the protein level is essential for the accurate classification of pathogenicity and for understanding disease mechanisms.

RNA splicing requires the recognition of the splice donor and acceptor site at the beginning and end of introns (Supplementary Fig. 1a). The donor site is defined as the last three nucleotides of an exon and the first six nucleotides of the intron, with an almost invariant 'GT' dinucleotide at the beginning of the intron. The acceptor site comprises a polypyrimidine tract around the last 20 nucleotides of the intron and includes the first three nucleotides of the exon with an almost invariant 'AG' dinucleotide at the end of the intron^{7,8}. Whilst in silico tools can predict whether a variant alters splicing with modest positive predicted values⁹, the

outcomes can only be confirmed with functional studies involving amplification and sequencing of the mRNA (Supplementary Fig. 1b).

Rare variants at the essential splice dinucleotides of genes where loss-of-function is an established disease mechanism are usually assigned a pathogenic very strong (PVS1) criterion by the American College of Medical Genetics and Association for Molecular Pathology (ACMG/AMP) variant classification guidelines¹⁰. Consequently, these variants often classify as likely pathogenic or pathogenic. In contrast, most variants in the extended splice site regions, or those predicted to create new splice sites, classify as variants of uncertain significance (VUS) due to the uncertainty about if and how they disrupt splicing. Previous work has shown that functional studies confirming aberrant splicing can provide supportive evidence of pathogenicity and can aid the classification of splicing variants as likely pathogenic^{5,11}. These functional studies have begun to reveal the types of splice-disrupting variants in selected genes; however, there is an incomplete knowledge of the range and burden of putative splice-disrupting variants in definitively disease-associated inherited heart disease genes across the different diseases, and it is unclear if these genes can be amplified from blood RNA to facilitate functional studies.

Here, we sought to describe the burden of in silico predicted, splice-disrupting variants in definitively disease-associated genes across a large clinical cohort of people with inherited heart disease or unexplained sudden cardiac death (SUD) compared to

¹Agnes Ginges Centre for Molecular Cardiology at Centenary Institute, The University of Sydney, Sydney, NSW, Australia. ²Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. ³Department of Cardiology, Royal Prince Alfred Hospital, Sydney, NSW, Australia. ⁴Cardio Genomics Program at Centenary Institute, The University of Sydney, Sydney, NSW, Australia. ⁵Centre for Population Genomics, Garvan Institute of Medical Research, and UNSW, Sydney, NSW, Australia. ⁶Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia. ⁷Department of Clinical Genetics, Liverpool Hospital, Sydney, NSW, Australia. ⁸Murdoch Children's Research Institute, University of Melbourne, Melbourne, VIC, Australia. ⁹Victorian Clinical Genetics Services, Melbourne, VIC, Australia. ¹⁰University of Melbourne, Melbourne, VIC, Australia. ¹¹Department of Cardiology, Royal Children's Hospital, Melbourne, VIC, Australia. ✉email: r.bagnall@centenary.org.au

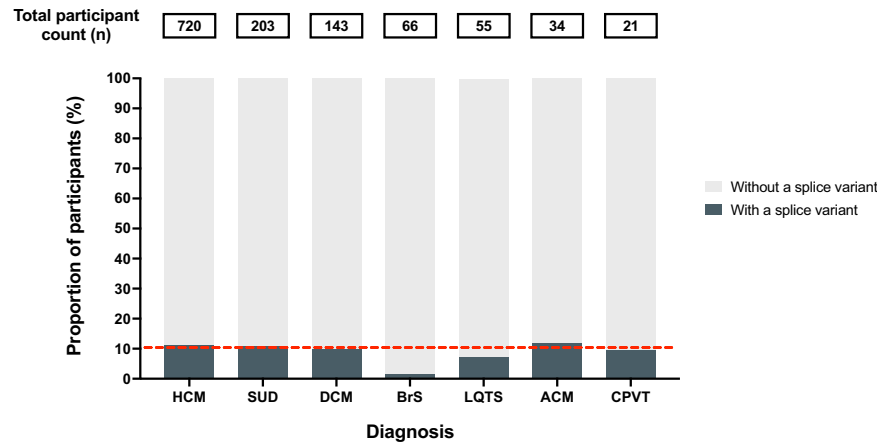


Fig. 1 The proportion of participants with an in silico-predicted splice-disrupting variant. Total participant counts within each disease group are shown in the boxes above each bar plot. The overall average of participants with an in silico-predicted splice-disrupting variant (10.3%) is highlighted by the red dotted line. HCM hypertrophic cardiomyopathy, SUD sudden unexplained death, DCM dilated cardiomyopathy, BrS Brugada syndrome, LQTS long QT syndrome, ACM arrhythmogenic cardiomyopathy, and CPVT catecholaminergic polymorphic ventricular tachycardia.

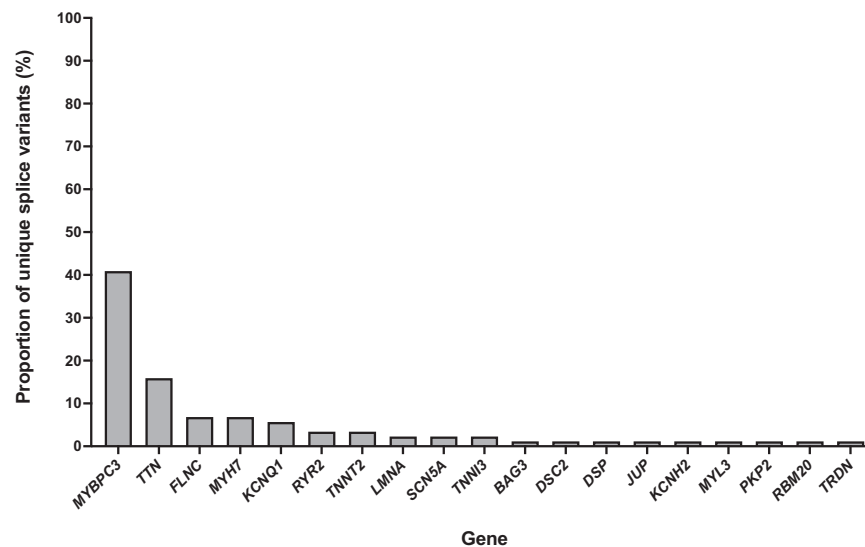


Fig. 2 The percentage of unique in silico-predicted splice-disrupting variants in each gene. Percentage of unique in silico-predicted splice-disrupting variants found in a list of 32 genes prioritised based on phenotypically concordant genes established by the ClinGen Curation Expert groups for hypertrophic cardiomyopathy, dilated cardiomyopathy, arrhythmogenic cardiomyopathy, long QT syndrome, Brugada syndrome and catecholaminergic polymorphic ventricular tachycardia.

population controls. We also show which of the 32 definitively disease-associated Clinical Genome Resource (ClinGen) curated cardiac genes can be amplified from blood RNA to support functional studies, and we confirm the outcomes of splicing variants in the participant's blood RNA, where available.

RESULTS

The burden of in silico predicted, splice-disrupting variants in definitively associated and phenotypically concordant disease genes was evaluated in 1242 unrelated participants with inherited heart disease or SUD. Hypertrophic cardiomyopathy (HCM) was the most common diagnosis in the cohort ($n = 720$), followed by SUD ($n = 203$), dilated cardiomyopathy (DCM) ($n = 143$), Brugada syndrome (BrS) ($n = 66$), long QT syndrome (LQTS) ($n = 55$), arrhythmogenic cardiomyopathy (ACM) ($n = 34$) and catecholaminergic polymorphic ventricular tachycardia (CPVT) ($n = 21$). We found 88 rare in silico predicted, splice-disrupting variants in 128

out of 1242 (10.3%) participants, and only nine participants had an alternative genetic cause for their disease (Supplementary Table 1). Within each disease group studied, approximately 10% of participants carried an in silico-predicted splice-disrupting variant, except for BrS, where only 1 out of 66 participants (1.5%) did (Fig. 1). The genes with the most splice-disrupting variants were *MYBPC3* (36/88, 41%), followed by *TTN* (14/88, 16%), *FLNC* (6/88, 7%), *MYH7* (6/88, 7%) and *KCNQ1* (5/88, 6%), with few variants, or none, in the remaining genes (Fig. 2). Self-reported ethnicities were available for 122 participants with splice-disrupting variants, with 95 European, 12 Asian, 6 North African, 8 Oceanian, and 1 'Peoples of the Americas' (Supplementary Table 1).

We compared the burden of rare putative splice-disrupting variants in our disease cohorts with gnomAD (v2.1.1) exomes control populations. There was a significant excess of splice-disrupting variants in *PKP2* in people with ACM (excess burden in cases = 5.9%, $P < 0.001$), *FLNC* (2.7%, $P < 0.001$) and *TTN* (2.8%, $P < 0.001$) in people with DCM, *MYBPC3* (8.2%, $P < 0.001$) and

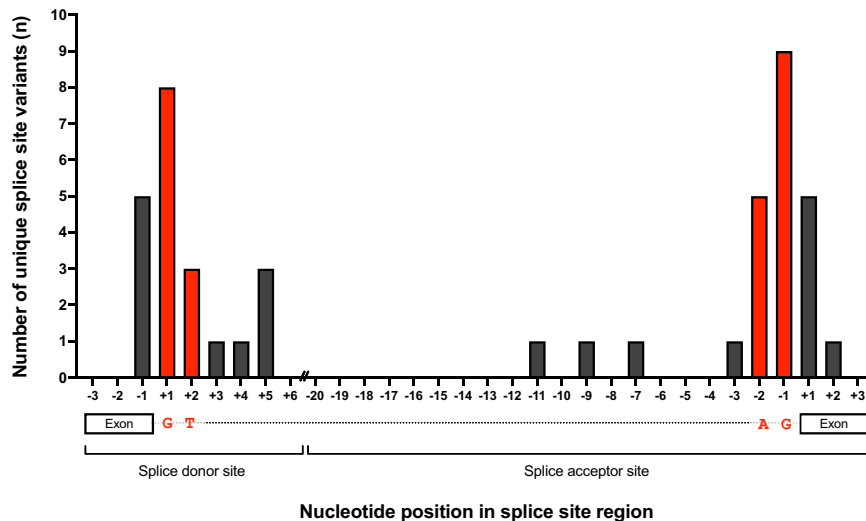


Fig. 3 Location of unique in silico-predicted splice-disrupting variants in the essential splice site region. The positions of the essential GT/AG dinucleotides are shown in red. The splice donor site consists of the last three nucleotides of the exon and the first six nucleotides of the intron. The acceptor site consists of the last 20 nucleotides of the intron and the first three nucleotides of the exon. The intervening intron is truncated '//'. Three small deletions within the splice site region are not shown.

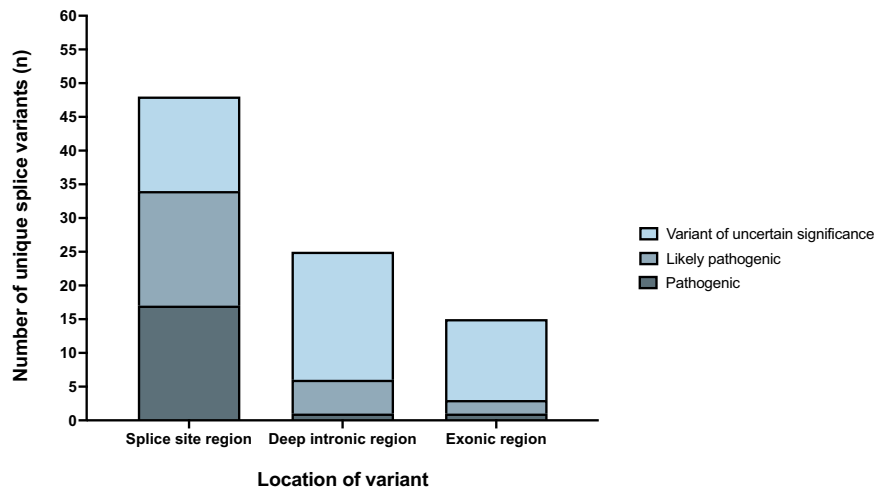


Fig. 4 Classification of in silico-predicted splice-disrupting variants. Variants located in the -3 to $+6$ region of the donor site or the -20 to $+3$ region of the acceptor site are classified as 'Splice site region' variants. All other intronic variants were labelled as 'deep intronic region'. Remaining variants in the exons, including missense, nonsense, and synonymous variants, were categorised as 'exonic region'.

MYH7 (1.3%, $P < 0.001$) in people with HCM, and *KCNQ1* in people with LQTS (3.6%, $P < 0.001$) (Supplementary Table 2). Statistical significance of these six genes was maintained when repeating the burden test with only European cases and European Non-Finnish gnomAD controls (Supplementary Table 3). The excess of putative splice-disrupting *MYH7* variants is primarily driven by six HCM participants having a Glu849Gly missense change due to a c.2681 A>G variant at the second nucleotide of exon 23, which was predicted to disrupt splicing by the in silico tools, adapting boosting (ADA) and random forests (RF)¹².

Mapping the location of splice-disrupting variants

The location of in silico predicted, splice-disrupting variants within the donor and acceptor sites was assessed (Fig. 3, Supplementary Table 4). Twenty-four variants (27%) were in the donor site, of which 13 variants disrupted the essential 'GT' dinucleotide and three clustered at the $+5$ intronic nucleotide position. Twenty-four variants (27%) were in the acceptor site, of which 14 disrupted the essential 'AG' dinucleotide. Ten variants (11%) in the first and last

nucleotide of an exon previously annotated as missense changes were predicted to disrupt the adjacent splice site.

Outside of the canonical splice sites, there were 25 variants in deep intronic regions predicted to create new donor ($n = 15$) or acceptor ($n = 10$) sites, and 15 exonic variants annotated as missense, nonsense, or synonymous variant, predicted to create new donor ($n = 11$) or acceptor ($n = 4$) sites (Fig. 4, Supplementary Table 4).

RNA sources for amplification of definitively disease-associated cardiac genes

We determined which sources of mRNA would support RT-PCR amplification of the definitively associated inherited heart disease genes. In total, 21 out of 31 (68%) genes were amplified concordantly using mRNA extracted from blood, induced pluripotent stem cell-derived cardiomyocytes and myectomy tissue (Supplementary Table 5, Supplementary Fig. 2). Nine genes only amplified in induced pluripotent stem cell-derived cardiomyocytes and myectomy tissue, (*CASQ2*, *TECL1*, *FLNC*, *MYH7*, *TNNT2*, *ACTC1*,

Table 1. Functional study outcomes and sequence variant classification.

Family (diagnosis)	Variant	MES/ADA RF/SpliceAI	ClinVar ID (class)	Functional study outcome	ACMG classification without RNA analysis	ACMG classification with RNA analysis	Change in classification
014 (LQTS)	<i>KCNQ1</i> c.477+5 G > A	-5.37/0.99 0.97/0.23	53047 (P)	Exon 2 skipped (PTC)	VUS (PS4_moderate, PM2, PP1, PP3)	LP (PS3, PS4_moderate, PM2, PP1)	VUS → LP
015 (LQTS)	<i>KCNQ1</i> c.781 G > A	1.81/0.91 0.61/0.76	53103 (VUS)	Exon 6 skipped (in-frame deletion)	VUS (PM2, PS4_supporting, PP3)	VUS (PM2, PM4, PS4_supporting)	No change (VUS)
051 (HCM)	<i>MYBPC3</i> c.1224-80 G > A	7.96/NA NA/0.94	693982 (-)	Exon 14 extension (in-frame insertion)	VUS (PS4, PP3)	LP (PS4, PM2, PM4)	VUS → LP
054 (HCM)	<i>MYBPC3</i> c.1458-7 C > A	-3.93/0.99 0.45/0.99	-	Exon 17 extension (PTC)	VUS (PM2, PP3)	LP (PS3, PM2)	VUS → LP
102 (CPVT)	<i>RYR2</i> c.848+1 G > A	-8.18/1.0 0.94/0.74	201371 (Conflicting)	Exon 11 skipped (in-frame deletion)	LP (PP1_strong, PM2, PS4_supporting)	LP (PP1_strong, PM2, PM4, PS4_supporting)	No change (LP)
125 (SUD)	<i>TTN</i> c.63793 G > A	-4.57/1.0 0.99/0.74	202779 (VUS)	Intron retention (PTC)	LP (PP1_moderate, PM2, PS4_supporting, PP3)	LP (PS3, PP1_moderate, PM2, PS4_supporting)	No change (LP)

ACMG/AMP criteria: PS4 - ≥ 15 probands of concordant phenotype; PS4_moderate - ≥ 6 probands of concordant phenotype; PS4_supporting - ≥ 2 probands of concordant phenotype; PS3 - functional evidence with out-of-frame change; PM2 - rarity in the population; PM4 - functional evidence with in-frame change; PP1_strong - ≥ 7 meiosis; PP1_moderate - ≥ 5 meiosis; PP1 - ≥ 3 meiosis; and PP3 - concordant in silico tool.

LQTS long QT syndrome, HCM hypertrophic cardiomyopathy, CPVT catecholaminergic polymorphic ventricular tachycardia, SUD sudden unexplained death, MES MaxEntScan score difference, ADA adaptive boosting, RF random forest, P pathogenic, VUS variant of uncertain significance, LP likely pathogenic, PTC premature termination codon.

MYL2, *MYL3* and *TNNI3*), while *TRDN* only amplified in myectomy tissue. Seven genes amplified more than one product due to alternative splicing (*RYR2*, *TECLL*, *BAG3*, *MYH7*, *ACTC1*, *TNNI3* and *TPM1*) (Supplementary Fig. 2). *PLN*, which is definitively associated with DCM, was not included as the single coding exon in the MANE transcript, NM_002667.5, does not undergo splicing.

Functional studies of in silico-predicted splice-disrupting variants

Six in silico predicted, splice-disrupting variants without prior mRNA testing were functionally studied using blood RNA from the affected individuals and family members who carried the variant, where available (Table 1). We assessed three variants in donor splice sites. Amplification of mRNA from a female with LQTS and a *KCNQ1* c.477+5 G > A variant revealed exon 2 skipping resulting in a 91 bp deletion, leading to a frameshift and premature termination codon (Supplementary Fig. 3a). An *RYR2* c.848+1 G > A variant in a female diagnosed with familial CPVT caused exon 11 skipping, leading to an in-frame deletion of 25 amino acids (Supplementary Fig. 3b). A *TTN* c. 63793 G > A missense variant at the last nucleotide of exon 307 was found in a male with SUD, his uncle with DCM, and his clinically unaffected father, totalling six meiosis. RNA extracted from the father showed the retention of intron 307, leading to a frameshift and a premature stop codon (Supplementary Fig. 3c). We assessed two variants in acceptor splice sites. A missense variant in *KCNQ1* c.781 G > A, found in a female with LQTS, caused the skipping of exon 6, leading to an in-frame deletion of 47 amino acids (Supplementary Fig. 3d). An *MYBPC3* c.1458-7 C > A variant found in a male with HCM caused a 5 bp extension of exon 17, leading to a premature stop (Supplementary Fig. 3e). Finally, a deep intronic c.1224-80 G > A variant in *MYBPC3* found in a male with HCM created a new splice acceptor site resulting in a 78 bp extension of exon 14, leading to an in-frame inclusion of 26 amino acids (Supplementary Fig. 3f).

Classification of sequence variants

We classified the pathogenicity of all 88 in silico predicted, splice-disrupting variants; 43 were classified as pathogenic or likely

pathogenic and 45 as VUS (Fig. 4, Supplementary Table 4). Most variants in the splice site regions were pathogenic or likely pathogenic, whereas most variants outside these regions were VUS. The results of our RNA-based functional studies, and previously published functional studies, were available for 29 out of 88 variants (33%). They confirmed that 19 variants caused a frameshift leading to a premature termination codon, six caused an in-frame insertion or deletion in the mRNA, one disrupted splicing with unreported consequences, one did not alter splicing, and one resulted in impaired protein function. One variant showed inconsistent results across multiple studies. These mRNA studies supported the reclassification of 11 VUS as likely pathogenic, and 3 likely pathogenic variants were upgraded to pathogenic (Supplementary Table 4). The reclassification of 11 VUS as likely pathogenic allowed these variants to be used for cascade genetic testing. Of these 11 VUS, six variants co-segregated with a concordant phenotype in 10 family members, one heterozygous family member with an unknown clinical status, and one family member who was genotype positive; phenotype negative.

DISCUSSION

We assessed the burden of splice-disrupting variants in people with inherited heart disease and SUD. We found that 10% of 1242 individuals had an in silico-predicted splice-disrupting variant in a phenotypically concordant disease gene. Most of these people had no alternative genetic explanation for their disease. There was an enrichment of splice-disrupting variants in six genes compared to controls, highlighting the importance of this class of variants. Of all splice variants found, half were clinically actionable, and the rest are prime candidates for functional studies to provide additional evidence for-or-against their clinical relevance. We determined that transcripts of 21 definitive inherited heart disease genes can be amplified using blood RNA despite very low expression in blood. Functional analysis of six splice-disrupting variants confirmed that they caused frameshifts or in-frame alterations to the protein-coding sequence. Our findings highlight the significant contribution that splice-disrupting variants make to

inherited heart diseases and the value of functional studies in achieving a genetic diagnosis.

Understanding how variants affect the protein sequence is essential for interpreting pathogenicity. Splice-disrupting variants in genes where a loss-of-function is the primary disease mechanism typically caused frameshifts and were enriched in *MYBPC3* in HCM, *TTN* and *FLNC* in DCM, *PKP2* in ACM and *KCNQ1* in LQTS. Loss-of-function variants in these genes would generally classify as likely pathogenic or pathogenic. In contrast, splicing defects in genes where deleterious missense variants prevail tended to cause in-frame deletions and insertions, such as the c.848+1 G > A variant in *RYR2* causing an in-frame exon skipping. The clinical relevance of these in-frame protein length changes is often uncertain. Enrichment of predicted splice-disrupting variants in *MYH7* in people with HCM was unexpected since *MYH7* loss-of-function variants are not usually pathogenic in HCM. There is a reported enrichment of *MYH7* truncating variants in people with left ventricular noncompaction, primarily driven by a c.732+1 G > A splice donor variant¹³. The enrichment in our study was largely driven by a c.2681 A > G variant, which could also be annotated as a Glu849Gly missense change. *MYH7* transcripts could not be amplified from blood RNA and whether this variant alters splicing is yet to be determined; however, the missense change is currently the most likely cause of HCM. While variants at the essential splice dinucleotides are expected to disrupt canonical splicing, variants outside of these dinucleotides have a more uncertain effect, which is challenging for variant interpretation. This variable effect on splicing may explain why the c.341 C > T *TNNT2* variant has shown different outcomes with functional studies using in vitro assays and primary tissue, or this may be due to differences between the assays. Variants that cause multiple outcomes on protein sequence might partly explain phenotypic variability between people harbouring them. A further consideration with variant interpretation is that missense, nonsense, and synonymous variants are often overlooked as causing a splice defect, but they have drastically different effects on the protein sequence. For example, we have previously demonstrated that an assumed benign synonymous *MYBPC3* c.2274 C > T, Gly758Gly variant in a family with HCM creates a new splice site causing truncation of exon 23 in the mRNA⁶.

Deep intronic splice-gain variants are an increasingly recognised cause of HCM, with an *MYBPC3* c.1224-52 G > A variant, alone, responsible for 1% of HCM^{6,14,15}. We found three deep intronic splice-gain variants in *MYBPC3* intron 13 in five unrelated people, of whom three had the c.1224-52 G > A variant. Altogether, we found 25 deep intronic in silico-predicted splice variants, 12 of which were found within 100 nucleotides of flanking intron sequence with exome sequencing. The contribution of deep intronic splice variants in our study is likely underestimated, as only 81 people had all intronic regions sequenced with genome sequencing.

We used stringent in silico parameters to predict variants causing a splice defect. RF and ADA tools are limited to evaluating variants at canonical splice sites, whereas MaxEntScan (MES)¹⁶, and SpliceAI¹⁷ can be applied genome-wide. For variants at canonical splice sites, all four tools predicted a splicing defect for 25 variants, three tools predicted 11 variants, two tools predicted eight variants, and one tool predicted one variant (Supplementary Table 6). We, or others, have confirmed that variants in each category disrupt splicing. Only MES and SpliceAI tools could evaluate variants outside of canonical splice sites. Thirty variants were predicted to disrupt splicing by MES alone, 8 variants were predicted to disrupt splicing by both tools, and 2 variants were predicted to disrupt splicing by SpliceAI alone. Variants within each category have been confirmed by us, or others, using functional studies. The variability in splicing predictions is likely based on the underlying design and algorithms that each tool uses. Most of these tools are designed to predict whether variants

alter splicing at canonical GT/AG splice sites and perform less well with the rare minor introns that begin and end with different dinucleotides, such as AT/AC. We found an *SCN5A* c.4434+5 G > A variant in one such AT/AC minor intron that was predicted to disrupt splicing by SpliceAI, whereas the other tools did not recognise the wild-type splice sequence. Given the prediction variability, multiple in silico tools should be used to identify putative splicing variants.

Our findings have important clinical implications for people and families with inherited heart diseases. In almost all inherited heart diseases, genetic testing currently fails to identify a cause in at least 50% of cases. The recognition of splice-disrupting variants in the current study begins to explain some of these gene-elusive people and families, with very important clinical implications including a more precise diagnosis, guiding therapies, screening at-risk family members, and in providing options for reproductive decisions.

Limitations of our study include that we did not use tools to evaluate variants that disrupt splice branchpoints, enhancers and silencer sequences or in the 5' and 3' untranslated regions. Our study focused on definitively disease-associated genes and did not include genes with more recently reported disease associations, such as *ALPK3*¹⁸, and *FHOD3*^{19,20}. We were underpowered to confirm a significant enrichment of splice variants in *SCN5A* in people with BrS. Only 81 people had genome sequencing, so the contribution of deep intronic splice-disrupting variants may be underestimated.

In conclusion, splice-disrupting variants are a significant cause of inherited heart diseases and SUD. Missense, nonsense and synonymous should be evaluated for possible effects on splicing as this can impact the interpretation of pathogenicity. Blood RNA is an accessible surrogate for primary heart tissue for RNA testing of many inherited heart disease genes, which increases the diagnostic yield of genetic testing. Confirmation of a deleterious splicing outcome in families with inherited heart disease and SUD has immediate clinical benefits, including guiding treatment options and facilitating risk stratification in family members.

METHODS

Study cohort

People with inherited heart disease or sudden unexplained death (SUD) were recruited from the Genetic Heart Disease and Hypertrophic Cardiomyopathy Clinics, Royal Prince Alfred Hospital, Sydney, Australia and the Royal Children's Hospital, Melbourne, Australia. We included people diagnosed with ACM (MIM 607450), DCM (MIM 604145), HCM (MIM 613426), BrS (MIM 601144), CPVT (MIM 604772), LQTS (MIM 613688), or SUD³. We excluded people with restrictive cardiomyopathy and isolated left ventricular non-compaction as ClinGen curation of disease-gene associations had not been performed. All participants were enrolled in protocol X20-0157/ETH00776, approved by the Sydney Local Health District Ethics Review Committee, Australia, or protocol #32092, approved by The Royal Children's Hospital Melbourne Research Ethics Committee. Written informed consent was provided by participants, parents for enrolled children or by next of kin for enrolled deceased participants.

In silico prioritisation of splice-altering variants

All participants had undergone prior genetic testing with cardiac gene panel (36%), exome (57%), or genome sequencing (7%), as previously described^{5,21}. Variants were retrieved in phenotypically concordant and definitively disease-associated genes, as established by the ClinGen Gene Expert panels²²⁻²⁷, except for SUD, in which we retrieved variants in all 32 definitively disease-associated genes (Supplementary Table 7). Candidate splice-disrupting variants were required to have an allele frequency <0.0001 in

the Genome Aggregation Database (gnomAD)^{28,29}, equating to an allele count ≤ 15 in gnomAD v2.1.1 and v3.1.2. In silico splice prediction tools MES¹⁶, ADA, RF¹², and SpliceAI¹⁷ were used to identify putative splicing variants as they are readily available with Ensembl's Variant Predictor and have been shown to offer sensitivity and specificity for splice site variants in cardiac genes³⁰. Variants were selected if they were located in a canonical splice site and reduced the MES score by >4 , or created a new splice site with an MES score >4 and showed an increase of >4 from wild-type sequence³¹, or if they scored >0.6 with ADA or RF¹², or >0.5 with SpliceAI¹⁷, as recommended by the developers of the tools.

Rare splice variant burden testing

Rare splice variant burden testing between case cohorts and gnomAD exomes, v2.1.1 was performed as per Mazzarotto et al. (2021)¹³. Rare variants were defined as having an African, East Asian, Latino, non-Finnish European and South Asian sub-population allele frequency and overall allele frequency in gnomAD exomes <0.0001 . Variants must also be in protein-coding regions or within the first and last 75 bp of intron sequence to account for exome sequencing target regions. We excluded the *TNNT2* NM_001276345.2: c.601-1 G > A variant, found in 5 participants with Oceanian ancestry, as this is a rare Oceanian polymorphism. The case denominator was the number of individuals within each disease group, and for gnomAD, the denominator was the average number of people sequenced for each gene to account for variable coverage in exome-sequenced samples. Statistical significance for the enrichment of variants in cases was assessed with a one-sided Fisher's exact test, with Bonferroni correction applied for testing the number of definitively disease-associated genes for each disease. The case excess was defined as the difference in rare variant frequencies between cases and gnomAD. Power calculations were performed using the `ES.h()` and `pwr.2p2n.test()` functions of the R package `pwr`.

RNA extraction

RNA was extracted from 2.5 mL of peripheral blood collected in PAXGene® Blood RNA Tubes (PreAnalytiX, Hombrechtikon, Switzerland) using the PAXGene® blood RNA kit (PreAnalytiX) according to manufacturer's instructions, except for a final elution in 40 μ L of elution buffer. RNA was extracted from induced pluripotent stem cell-derived cardiomyocytes from two unrelated people, as previously described³². RNA was extracted from 20 μ g of snap-frozen myectomy tissue homogenised in 1 mL of TRIzol® reagent using the phenol-chloroform method according to the manufacturer's instructions, except for a final resuspension in 30 μ L of diethyl pyrocarbonate (DEPC) treated water. RNA samples were quantified using the Nanodrop ND1000 (Thermo Fisher Scientific, Massachusetts, United States) and assessed with 2% agarose gel electrophoresis stained with 1% Gel Red using the GelDoc Go System (BioRad, California, United States).

RT-PCR amplification of definitively disease-associated cardiac genes

Primers were designed to amplify at least three consecutive exons of MANE (matched annotation between NCBI and EBI) transcripts, with an RT-PCR product size between 340 to 500 bp, except for *BAG3*, *KCNH2* and *TTN*, which amplified larger products due to large exon sizes. One microgram of RNA was reverse transcribed using 200 U Superscript™ III (Invitrogen, Massachusetts, United States), 4 μ L Superscript™ III™ 5X first-strand buffer, 5 mM DTT, 40 U RNaseOUT™ (Thermo-Fisher), 0.25 mM dNTP (Roche, Basel, Switzerland) and 0.01 mM random hexamer primer (Thermo-Fisher) in a final volume of 20 μ L. Complementary DNA products were diluted 1:2 in DEPC-treated water and amplified as previously described⁵, using an annealing temperature of 60 °C unless

otherwise specified (Supplementary Table 5). PCR products were resolved on an agarose gel and all gel images were derived from the same experiment. PCR products were purified using 5 U Antarctic Phosphatase (New England BioLabs, Massachusetts, United States) and 20 U Exonuclease I (New England BioLabs), and Sanger sequenced at the Macrogen Sequencing Facility (Seoul, Republic of Korea). Chromatograms were analysed using Sequencher™ v.5.4.6 (Gene Codes Corporation, Michigan, United States).

Functional studies of splice-altering variants

Functional studies of putative splicing variants in genes that successfully amplified in blood RNA was performed on 500 ng RNA extracted from the participant's fresh blood using the PAXGene® blood RNA kit. Functional studies were performed as previously described⁵. PCR primer sequences are shown in Supplementary Table 8.

Classification of variants using the ACMG/AMP framework

All putative splice variants were classified for pathogenicity using the ACMG/AMP classification framework³³. PVS1 was assigned to variants causing a frameshift in genes where loss-of-function is an established disease mechanism. We also applied PVS1 to loss-of-function variants in genes with a ClinGen gene dosage haploinsufficiency score of 3, which represents the highest rating given to genes with sufficient evidence supporting a dosage sensitivity (<https://search.clinicalgenome.org/kb/gene-dosage>). PVS1_strong was given to genes with a haploinsufficiency score of 2 and to *titin* (*TTN*), representing genes with emerging evidence of dosage sensitivity. PVS1_moderate was applied to genes with a haploinsufficiency score of 1, representing little evidence of dosage sensitivity. PVS1_supporting was applied to genes that have not yet been curated for dosage sensitivity but have a ratio of <0.35 for the observed to expected (o/e) number of loss-of-function variants in that gene, as provided by gnomAD.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Data used and/or analysed during the current study is provided in the Supplementary information, original gel images and Sanger sequencing files are available from the corresponding author. Gene panel testing was performed by a clinically accredited testing facility, Victorian Clinical Genetics Services (Melbourne, Australia), and therefore unavailable. Whole exome and genome sequencing data are subject to conditions of Ethics agreement X20-0157/ETH00776 under which the data was generated, and therefore unavailable unless a data sharing agreement has been obtained from the Sydney Local Health District Ethics Review Committee, Australia. Control dataset was obtained from gnomAD (<https://gnomad.broadinstitute.org/>).

Received: 11 May 2023; Accepted: 29 September 2023;
Published online: 11 October 2023

REFERENCES

1. Al-Khatib, S. M. et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: Executive Summary. *Circulation* **138**, e272–e391 (2018).
2. McKenna, W. J., Maron, B. J. & Thiene, G. Classification, Epidemiology, and Global Burden of Cardiomyopathies. *Circ. Res.* **121**, 722–730 (2017).
3. Wilde, A. A. M. et al. European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) Expert Consensus Statement on the state of genetic testing for cardiac diseases. *EP Eur.* **24**, 1307–1367 (2022).

4. Patel, P. N. et al. Contribution of Noncanonical Splice Variants to TTN Truncating Variant Cardiomyopathy. *Circ. Genom. Precis. Med.* **14**, 581–590 (2021).
5. Singer, E. S., Ingles, J., Semsarian, C. & Bagnall, R. D. Key Value of RNA Analysis of MYBPC3 Splice-Site Variants in Hypertrophic Cardiomyopathy. *Circ. Genom. Precis. Med.* **12**, 32–39 (2019).
6. Bagnall, R. D. et al. Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **72**, 419–429 (2018).
7. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
8. Zhang, M. Q. Statistical Features of Human Exons and Their Flanking Regions. *Hum. Mol. Genet.* **7**, 919–932 (1998).
9. Riepe, T. V. et al. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum. Mutat.* **42**, 799–810 (2021).
10. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–423 (2015).
11. Bournazos, A. M. et al. Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet. Med.* **24**, 130–145 (2022).
12. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
13. Mazzarotto, F. et al. Systematic large-scale assessment of the genetic architecture of left ventricular noncompaction reveals diverse etiologies. *Genet. Med.* **23**, 856–864 (2021).
14. Holliday, M. et al. Transcriptome Sequencing of Patients With Hypertrophic Cardiomyopathy Reveals Novel Splice-Altering Variants in MYBPC3. *Circ. Genom. Precis. Med.* **14**, 183–191 (2021).
15. Harper, A. R. et al. Reevaluation of the South Asian MYBPC3Δ25bp Intronic Deletion in Hypertrophic Cardiomyopathy. *Circ. Genom. Precis. Med.* **13**, 101–109 (2020).
16. Yeo, G. & Burge, C. B. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.* **11**, 377–394 (2004).
17. Jaganathan, K. et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535–548 (2019).
18. Lopes, L. R. et al. Alpha-protein kinase 3 (ALPK3) truncating variants are a cause of autosomal dominant hypertrophic cardiomyopathy. *Eur. Heart J.* **42**, 3063–3073 (2021).
19. Ochoa, J. P. et al. Formin Homology 2 Domain Containing 3 (FHOD3) Is a Genetic Basis for Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **72**, 2457–2467 (2018).
20. Semsarian, C., Ingles, J. & Bagnall, R. D. Revisiting Genome Sequencing Data in Light of Novel Disease Gene Associations. *J. Am. Coll. Cardiol.* **73**, 1365–1366 (2019).
21. Bagnall, R. D. et al. Exome-based analysis of cardiac arrhythmia, respiratory control, and epilepsy genes in sudden unexpected death in epilepsy. *Ann. Neurol.* **79**, 522–534 (2016).
22. Adler, A. et al. An International, Multicentered, Evidence-Based Reappraisal of Genes Reported to Cause Congenital Long QT Syndrome. *Circulation* **141**, 418–428 (2020).
23. Hosseini, S. M. et al. Reappraisal of Reported Genes for Sudden Arrhythmic Death. *Circulation* **138**, 1195–1205 (2018).
24. Ingles, J. et al. Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ. Genom. Precis. Med.* **12**, 57–64 (2019).
25. James, C. A. et al. International Evidence Based Reappraisal of Genes Associated With Arrhythmogenic Right Ventricular Cardiomyopathy Using the Clinical Genome Resource Framework. *Circ. Genom. Precis. Med.* **14**, 273–284 (2021).
26. Jordan, E. et al. Evidence-Based Assessment of Genes in Dilated Cardiomyopathy. *Circulation* **144**, 7–19 (2021).
27. Walsh, R. et al. Evaluation of gene validity for CPVT and short QT syndrome in sudden arrhythmic death. *Eur. Heart J.* **43**, 1500–1510 (2022).
28. Whiffin, N. et al. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* **19**, 1151–1158 (2017).
29. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
30. Rayani, K. et al. Identification and in-silico characterization of splice-site variants from a large cardiogenetic national registry. *Journal*; <https://doi.org/10.1038/s41431-022-01193-9> (2022).
31. Ito, K. et al. Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. *Proc. Natl Acad. Sci. USA.* **114**, 7689–7694 (2017).
32. Holliday, M., Ross, S. B., Lim, S. & Semsarian, C. Generation of an induced pluripotent stem cell line from a hypertrophic cardiomyopathy patient with a pathogenic myosin binding protein C (MYBPC3) p.Arg502Trp mutation. *Stem Cell Res.* **33**, 56–59 (2018).
33. Kelly, M. A. et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet. Med.* **20**, 351–359 (2018).

ACKNOWLEDGEMENTS

The study was funded by an NSW Health Cardiovascular Disease Senior Scientist Grant, an Australian Government Research Training Program Scholarship, a co-funded National Heart Foundation of Australia/National Health and Medical Research Council (NHMRC) PhD scholarship (#102568/ #191351), an NHMRC Career Development Fellowship (#1162929), an NHMRC Investigator Grant (#2016822) and an NSW Health Cardiovascular Disease Clinician Scientist Grant. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this paper.

AUTHOR CONTRIBUTIONS

Conceptualisation: R.D.B., E.S.S.; Data curation: R.D.B., E.S.S.; Formal analysis: R.D.B., E.S.S., J.C.I.; Funding acquisition: R.D.B., J.I., R.G.W., C.S.; Investigation: R.D.B., E.S.S., J.C., M.H., J.C.I., N.N., L.Y., C.B., S.R., I.M., I.K., J.W., J.I., R.G.W., C.S.; Methodology: R.D.B., E.S.S., M.H., J.I., R.G.W., C.S.; Resources: M.H., S.L., S.R., I.M., J.W., R.G.W., C.S.; Writing original draft: R.D.B., E.S.S.; Writing-review, editing and final approval: All authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-023-00373-w>.

Correspondence and requests for materials should be addressed to Richard D. Bagnall.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023