SUPPLEMENTAL MATERIAL

Supplemental Methods:

Burden Analysis of TTNtv in DCM Cases and Controls

A total of 203 patients with familial or sporadic idiopathic DCM (30% familial, 59% male, mean age at diagnosis 35 ± 18 years) were recruited from Brigham and Women's Hospital, Boston Children's Hospital, Victor Chang Cardiac Research Institute, and Royal Brompton & Harefield Hospitals. Study participants provided informed written consent and protocols were approved by the institutional human ethics committees as previously described for UK cases[7] and for all other sites.[20] Genomic DNA libraries were constructed using standard library preparation protocols. Fragments were ligated to adaptors, amplified and purified, then hybridized to custom arrays enriched for coding regions of *TTN*. Control data were obtained from a cohort of elderly subjects with no known cardiovascular disease drawn from the Alzheimer's Disease Sequencing Project (n=3329, www.niagads.org/adsp/). These cohorts had been previously sequenced and data (stored in BAM file format) was available for each subject.

We selected rare variants from both cohorts using standard computational tools (GATK Best Practices recommendations).[25,26] Single nucleotide variants and small indels were identified with GATK HaplotypeCaller and annotated using SnpEff,[27] dbSNP, and ExAC (v3). Filters included Pop-Max ExAC AF (highest population allele frequency noted in ExAC) < 0.0001, GQ (genotype quality) ≥ 40, 30 < (alt_AD) / [(ref_AD+ alt_AD)] < 70, ref_AD > 5 and alt_AD > 5 (AD = allele depth). All subjects in the DCM cohort had self-reported European ancestry. To eliminate population stratification due to potential differences in ancestry of case and control cohorts, we used principal component analysis (PCA) and subjects (cases and controls) with non-

European ancestry were excluded from the analysis. We required that 95% of cases and 95% of controls have at least 10x coverage at all sequence positions included in the analysis.

Computational Prioritization of Variants for Splicing Assay

Variants were selected if they were located in either the 9 bp splice donor site or 23 bp splice acceptor site of any of the 364 *TTN* exons. *TTN* exon coordinates were determined by Locus Reference Genomic (LRG) annotation, the standard recommended for clinical reporting. Only variants located in splice-signals flanking exons with PSI ≥ 0.90 were considered for further evaluation. Variants creating substitutions at the canonical GT and AG positions were excluded from this analysis, as these were already classified as pathogenic TTNtv.

We previously described a bioinformatic strategy that employed MaxEntScan to analyze the likelihood that the modified 9 bp sequence could act as a splice donor site or the modified 23 bp sequence could function as a splice acceptor site (Supplemental Figure I).[13,19,28] That is, if the change in MaxEnt (ΔMaxEnt = $MaxEnt_{var}$ - $MaxEnt_{ref}$) < 0, a variant was identified as a candidate splice-altering variant. Our previous analyses of putative splice-altering variants in other genes (*MYBPC3* and *LMNA*) demonstrated that variants which increase MaxEnt (ΔMaxEnt > 0) are very unlikely to alter splicing.[13,19]

Functional Assessment of Variants Using Minigene Assay

Comprehensive details of the design of minigenes and experimental procedures have been described previously.[13,19] Briefly, minigene splicing assays tested a ~1200 bp DNA construct containing an exon-intron-exon test sequence under the expression of a CMV promoter. Test sequences were designed using a computational script that retrieved the

surrounding sequence context for a given variant from online resources and allocated these sequences to a 500 bp template. Each minigene contained approximately 215-220 bp of sequence surrounding the upstream and downstream splice junctions of interest (Supplemental Figure II). When necessary, deep intronic regions were excluded in order to fit the test sequence on a 500 bp oligonucleotide. Pairs of minigenes were designed with and without the variant, ordered as gBlock Gene Fragments (Integrated DNA Technologies), and assembled into a full-length construct using two rounds of PCR. Each minigene contained a 2 bp barcode in the 3' end of the second exon, which permitted multiplexed DNA sequencing and subsequent computational analysis.

HEK293 cells (ATCC) were cultured in DMEM (Gibco) with 10% FBS, L-glutamine and gentamycin on cell-culture plates (FALCON). Cells were seeded into a 6-well plate, and Lipofectamine 2000 (Invitrogen) was utilized for transfection when cells had reached 90% confluence. Up to 24 reference and variant constructs (100 ng of each construct) were transfected into one well, per manufacturer's protocol. A subset of minigenes underwent nucleofection (Amaxa) in human induced pluripotent stem cell-derived cardiomyocytes to confirm minigene encoded RNAs expressed in both cell lines were spliced in similar fashion.

Twenty-four hours after minigene delivery, cells were harvested, total RNA was extracted, and cDNA libraries were prepared for Illumina MiSeq Sequencing. We determined the number of sequence reads derived from unspliced RNA, normal-spliced RNA and aberrantly spliced RNA transcribed from reference and variant-containing minigene constructs (Supplemental Figure III). A minimum of 100 analyzable reads from each construct (identified by barcode) was required to assess consequences of splicing. The number of analyzable reads

was normalized to 100 and p-value was calculated by two-sided Fisher's exact test with $p<0.001$ as the threshold for significance.

Identification of Rare Splice Region TTN Variants from Clinical Variant Databases

Full *TTN* variant lists were obtained from the Laboratory of Molecular Medicine (LMM) of Partners Healthcare (December 2016) and from ClinVar of NCBI (January 2017), two large databases of human genetic variation. Only single nucleotide variants were included in this analysis. We selected for variants located in either the 9 bp splice donor site or 23 bp splice acceptor site of any of the 364 *TTN* exons. A total of 640 unique variants were found to be located within *TTN* splice donor or acceptor sites.

Among the 640 variants remaining within the splice region, we next excluded variants in low PSI exons (PSI < 0.90, n=250), variants already known to disrupt the canonical GT and AG splice signals (n=75), and variants with allele frequency > $1x10^{-4}$ (n=114). PSI values for each exon were determined from previously published data utilizing samples from the Genotype-Tissue Expression (GTEx) project.[8] Allele frequencies were annotated using the Exome Aggregation Consortium (ExAC). After all filtering, 201 variants were identified as rare candidate variants.

Identification of Rare Splice Region Variants in MYBPC3 and *LMNA from Clinical Testing Databases*

We similarly identified rare variants reported in the splice regions of *MYBPC3* and *LMNA*, two additional autosomal dominant cardiomyopathy genes. Loss-of-function variants in *MYBPC3* are prevalent causes of hypertrophic cardiomyopathy, whereas haploinsufficiency of

LMNA is a substantial contributor to DCM with associated conduction system defects.[29-31] A total of 150 single nucleotide variants were reported within a splice donor site or splice acceptor site, did not disrupt the canonical dinucleotides, and were rare (ExAC AF < $1x10^{-4}$). Our previously published analysis, on variants submitted prior to 2015, reported on the functional consequences of 94 of these rare variants.[13] The remaining 56 splice region variants were assessed in this study for the first time

Computational Prioritization of Variants using SpliceAI

We used SpliceAI, a deep neural network-based method of splice prediction,[32] to prioritize the same set of VUS that were evaluated through the *in vitro* assay to assess the performance of this *in-silico* tool with regard to splice-altering variant prediction. Positive predictive values (PPV) and negative predictive values (NPV) were analyzed under two conditions: 1) using the highest SpliceAI delta score for the variant (inclusive of donor loss, donor gain, acceptor loss, and acceptor gain), and 2) limiting the analysis to the highest SpliceAI score for splice site loss (donor loss and acceptor loss only).

PPVs and NPVs were calculated at score thresholds of 0.2 and 0.8.[32-34] Variants were designated true positives if SpliceAI predicted a splice-altering variant and the variant disrupted splicing in the minigene assay, or false positives if SpliceAI predicted a splice-altering variant and the variant was either determined by MaxEnt or the minigene assay to have no effect on splicing. In the case where a variant did not have a sufficiently high SpliceAI score, variants were categorized as false negative if the minigene assay determined the variant altered splicing. Variants were labeled as true negatives if either MaxEnt scoring or the *in vitro* assay determined that the variant had no effect on splicing.

PSI Calculation from Genotype-Tissue Expression (GTEx) Project

RNA-seq BAM files aligned to human genome reference hg19 for skeletal muscle (main site: gastrocnemius muscle), left ventricle, and right atrial tissue were downloaded from the GTEx Portal on 03/17/2018, 03/15/2018, and 04/11/2018, respectively (dbGaP accession number phs000424.v7.p2).[35,36] Reads aligning to the *TTN* gene in the GTEx bam files were extracted using the SamToFastq tool from the Picard Toolkit (https://github.com/broadinstitute/picard) and realigned to the hg19 genome using the Tophat aligner in order to generate Tophat junctions files.[37,38] Each fastq read was aligned a single time (option "--max-multihits 1") in order to avoid read pile-up in repeat regions of *TTN*. PSI was calculated using a protocol published previously.[12,39] PSI < 0.7 exons were chosen for comparison against PSI ≥ 0.9 exons because these thresholds created subsets that optimized differences in sequences.

Statistical Analysis

Statistical analyses were performed with R (version 3.4.0; http://www.R-project.org/). Comparisons between groups were performed by two-sided tests. A level of $\alpha=0.05$ was used for significance unless otherwise specified. Corrections made for multiple testing are noted.

Supplemental Tables: **[SEE SEPARATE EXCEL FILE FOR TABLES I-IV]**

Supplemental Table I: List of TTNtv identified in DCM and control cohorts. Target capture sequencing of 203 cases with idiopathic DCM identified 21 nonsense, 16 frameshift, and 10 GT/AG splice disrupting TTNtv in high PSI exons. In contrast, within a control cohort of 3329 individuals without DCM, a total of 7 nonsense, 6 frameshift, and 3 splice disrupting TTNtv were identified in high PSI exons. Identifying features of the TTNtv identified are listed. ExAC.AF = Allele Frequency in the ExAC Database; PSI.GTEX = Proportion Spliced In for affected exon in Genotype-Tissue Expression (GTEx) project.

Supplemental Table II: Assay results for all 43 *TTN* splice region variants identified in the DCM and control cohorts. After excluding known TTNtv in the two cohorts, 3.4% of cases (n=7) and 1.1% of controls (n=36) possessed a rare *TTN* VUS near a known splice site. The MaxEnt scores for the reference sequence (Ref score), alternate sequence (Alt score), and Delta MaxEnt are shown. The computational result column identifies variants that are computationally "Selected" for the minigene assay (i.e. ΔMaxEnt < 0) or computationally "Excluded" (i.e. ΔMaxEnt > 0). Minigene assay results, if applicable, are shown in the final columns of the table. Four splice-altering VUS were identified in cases whereas 6 were found in the control cohort. ExAC.AF = Allele Frequency in the ExAC Database.

Supplemental Table III: Assay results for 201 TTN variants listed in LMM and ClinVar assessed for splicing effects. A total of 45 splice-altering variants were identified, whereas 156 splice region variants had no effect on mRNA processing and are likely benign. The MaxEnt scores for the reference sequence (Ref score), alternate sequence (Alt score), and Delta MaxEnt

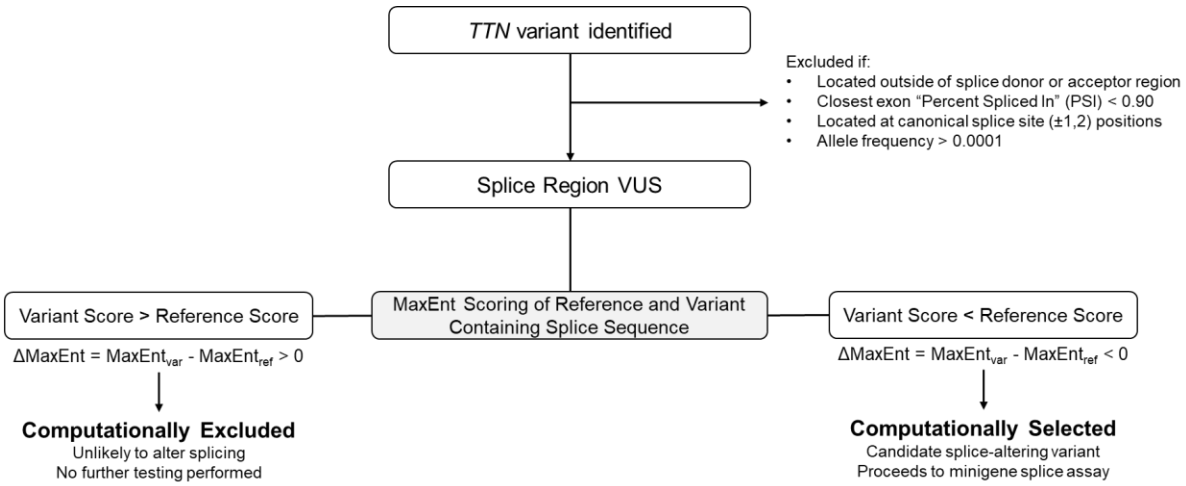
are shown. The computational result column identifies variants that are computationally "Selected" for the minigene assay (i.e. ΔMaxEnt < 0) or computationally "Excluded" (i.e. ΔMaxEnt > 0). Minigene assay results, if applicable, are shown in the final columns of the table. ExAC.AF = Allele Frequency in the ExAC Database; cDNA = position within 9 bp splice donor or 23 bp splice acceptor sequence.

Supplemental Table IV: Assay results for 150 MYBPC3 and LMNA variants listed in LMM and ClinVar assessed for splicing effects. A total of 67 splice-altering variants were identified, whereas 83 splice region variants had no effect on mRNA processing and are likely benign. The MaxEnt scores for the reference sequence (Ref score), alternate sequence (Alt score), and Delta MaxEnt are shown. The computational result column identifies variants that are computationally "Selected" for the minigene assay or computationally "Excluded". Minigene assay results, if applicable, are shown in the final columns of the table. The final column details whether the data provided for the variant is novel (n=56) or had been published previously by our group (n=94). ExAC.AF = Allele Frequency in the ExAC Database; cDNA = position within 9 bp splice donor or 23 bp splice acceptor sequence.

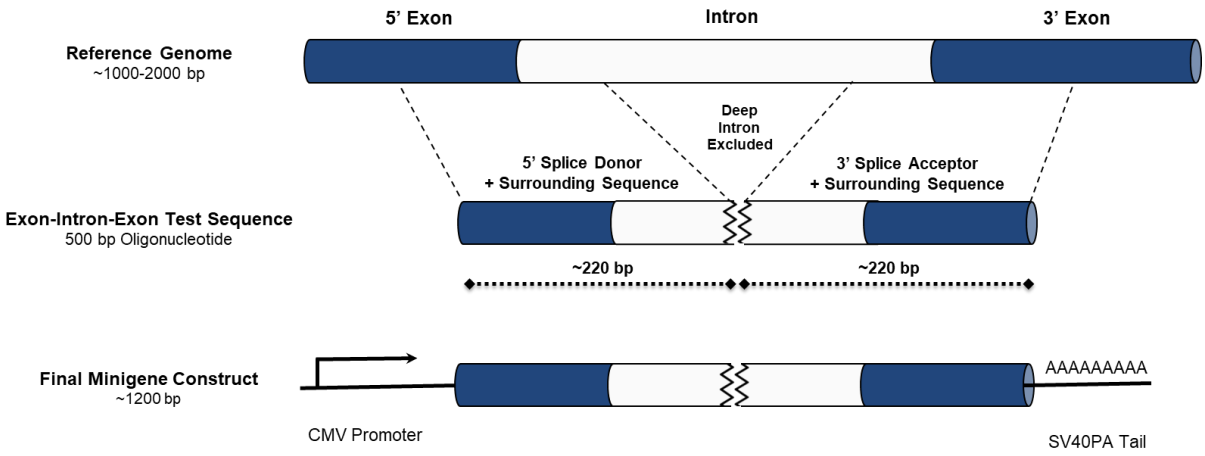
	SpliceAI ≥ 0.2		SpliceAI ≥ 0.8	
	PPV	**NPV**	**PPV**	**NPV**
All Genes	85.7%	80.9%	94.7%	72.3%
TTN	85.2%	85.3%	100.0%	79.4%
MYBPC3	82.9%	73.9%	90.0%	59.6%
LMNA	100.0%	68.4%	100.0%	60.5%

Supplemental Table V: Positive predictive values and negative predictive values of SpliceAI in each gene studied. At a SpliceAI score threshold of 0.2 for donor or acceptor site loss, the PPV across all three genes studied was 85.7%. At a more stringent score SpliceAI threshold of 0.8 for donor or acceptor site loss, the PPV increased to 94.7%. The PPVs were similar across all three genes studied. PPV=Positive Predictive Value; NPV=Negative Predictive Value.

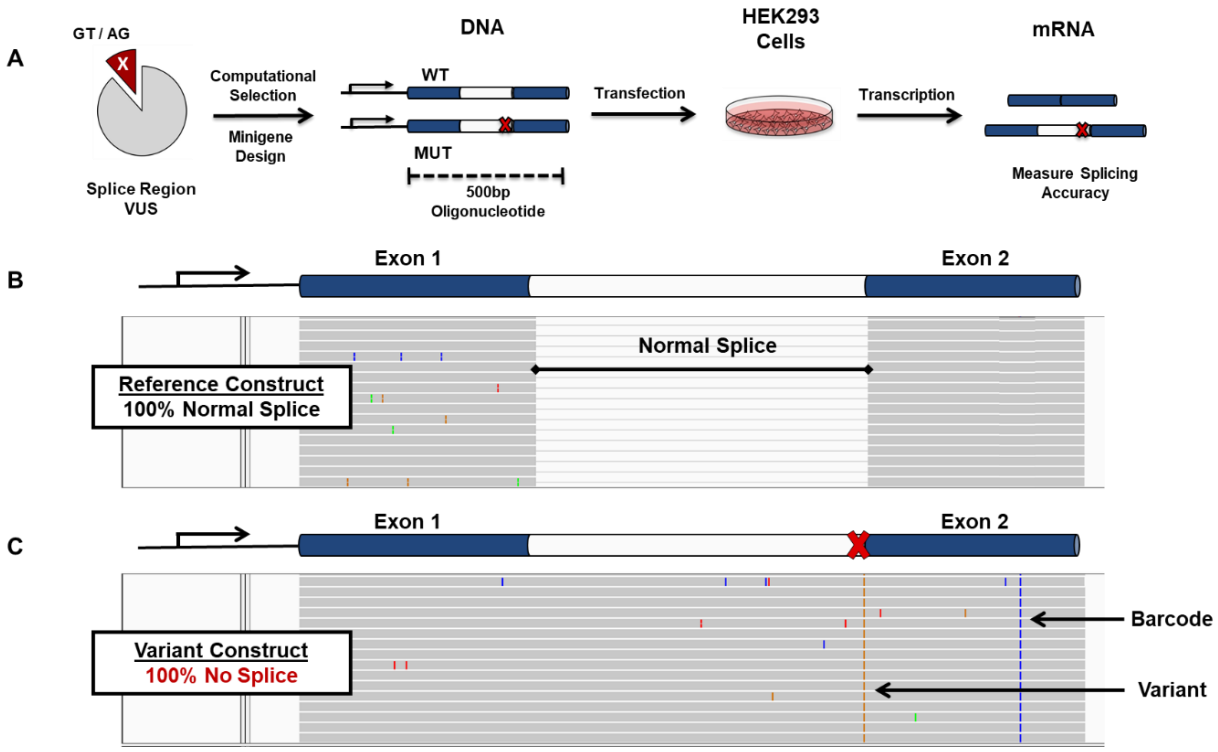
Supplemental Figures and Figure Legends:



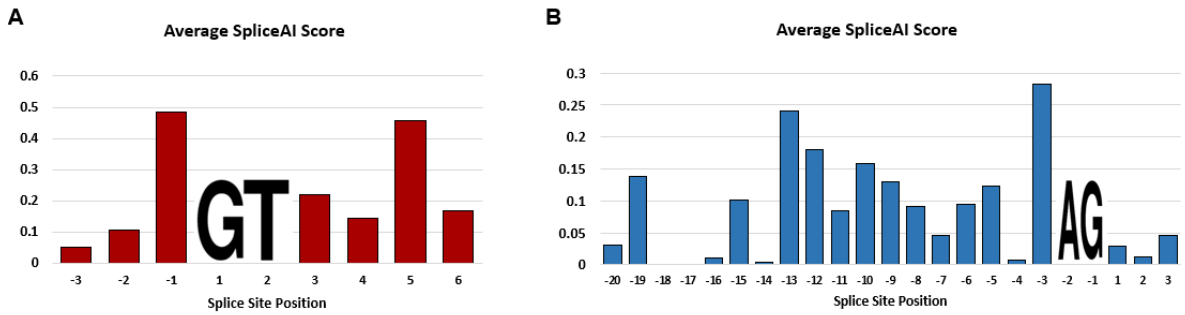
Supplemental Figure I: Computational prioritization strategy of splice region variants for minigene spice assay. The computational strategy employed MaxEntScan to analyze the likelihood that the modified 9 bp sequence could act as a splice donor site or the modified 23 bp sequence could function as a splice acceptor site. If the change in MaxEnt was less than zero (i.e. $\Delta MaxEnt < 0$) the variant of interest was deemed to be a candidate splice altering variant and tested further in the minigene assay. Variants in which the sequence change strengthened the splice site (i.e. $\Delta MaxEnt > 0$) were computationally excluded and did not undergo further testing.



Supplemental Figure II: Design of Minigenes from Reference Genomic Sequence. Test sequences were designed using a computational script that allocated the ~220 bp sequences surrounding the splice junctions of interest to a 500 bp template. When necessary, deep intronic regions were excluded in order to fit the test sequence on a 500 bp oligonucleotide. The final minigene construct contained an exon-intron-exon test sequence under the expression of a CMV promoter.



Supplemental Figure III: Overview of cellular assay to identify sequence variants that affect RNA splicing. (**A**) From a list of splice region VUS, variants are bioinformatically filtered to prioritize those most likely to affect splicing. Variants most likely to alter splicing have minigenes designed to assess for splicing effects. A pair of minigenes containing the reference and alternate allele is synthesized for each candidate variant. Minigenes are transfected into a cellular environment, RNA is extracted, and libraries are sequenced. (**B**) A schematic of a reference minigene containing an exon-intron-exon test sequence under the expression of a promoter is shown. A screenshot of the RNA sequencing result using the Integrative Genomics Viewer shows normal splicing (intron removal) in all reads. (**C**) A schematic of a variant-containing minigene with corresponding Integrative Genomics Viewer output reveals the presence of the variant, inactivation of the splice site, and inclusion of the intron. A two base pair barcode in the second exon is used to identify whether the read was transcribed from the reference or variant-containing minigene in the event the variant is spliced out.



Supplemental Figure IV: Average SpliceAI score across each position of donor and acceptor sites. Among 394 variants assessed in this study, the average maximum SpliceAI score varied with position in the splice donor and acceptor sites. **(A)** In the splice donor site, variants located at the last nucleotide of the upstream exon and the +5 position tended to have the highest SpliceAI scores. **(B)** In the splice acceptor site, the variants located at the -3 position had an average SpliceAI score higher than the remainder of the splice region.