Supporting Information

Ito et al. 10.1073/pnas.1707741114

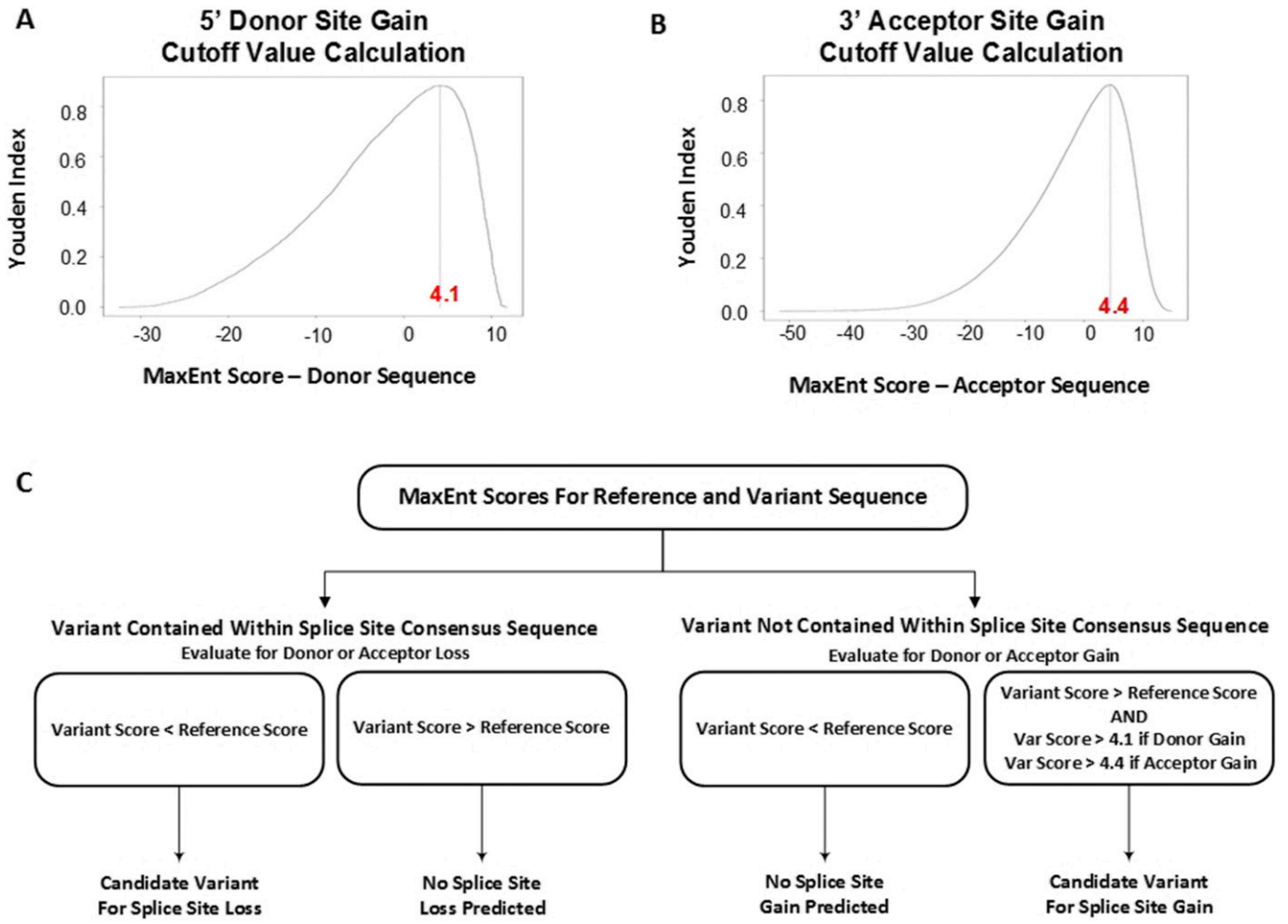

Fig. S1. Thresholds for selection of candidate splice altering variants (*Materials and Methods*). Youden's index (index = sensitivity + specificity − 1) was maximized to set a MaxEnt score value below which variants were not predicted to lead to 5′ donor or 3′ acceptor splice site gain. (*A*) At the 5′ donor splice site, Youden's index was maximized when MaxEnt score = 4.1. (*B*) At the 3′ acceptor splice site, Youden's index was maximized when MaxEnt score = 4.4. (*C*) A decision tree depicts the algorithm used for selection of candidate splice-altering variants. Variants predicted to cause loss of a donor or acceptor splice site (contained within splice site, ΔMaxEnt < 0) were selected for cell splicing assays. Variants predicted to create a gain of donor splice site (outside of splice site, ΔMaxEnt > 0) and had MaxEnt_{var} score > 4.1 were selected for the minigene splicing assay. Similarly, variants predicted to create a gain of acceptor splice site (outside of splice site, ΔMaxEnt > 0) and had MaxEnt_{var} score > 4.4 were selected for the minigene assay.

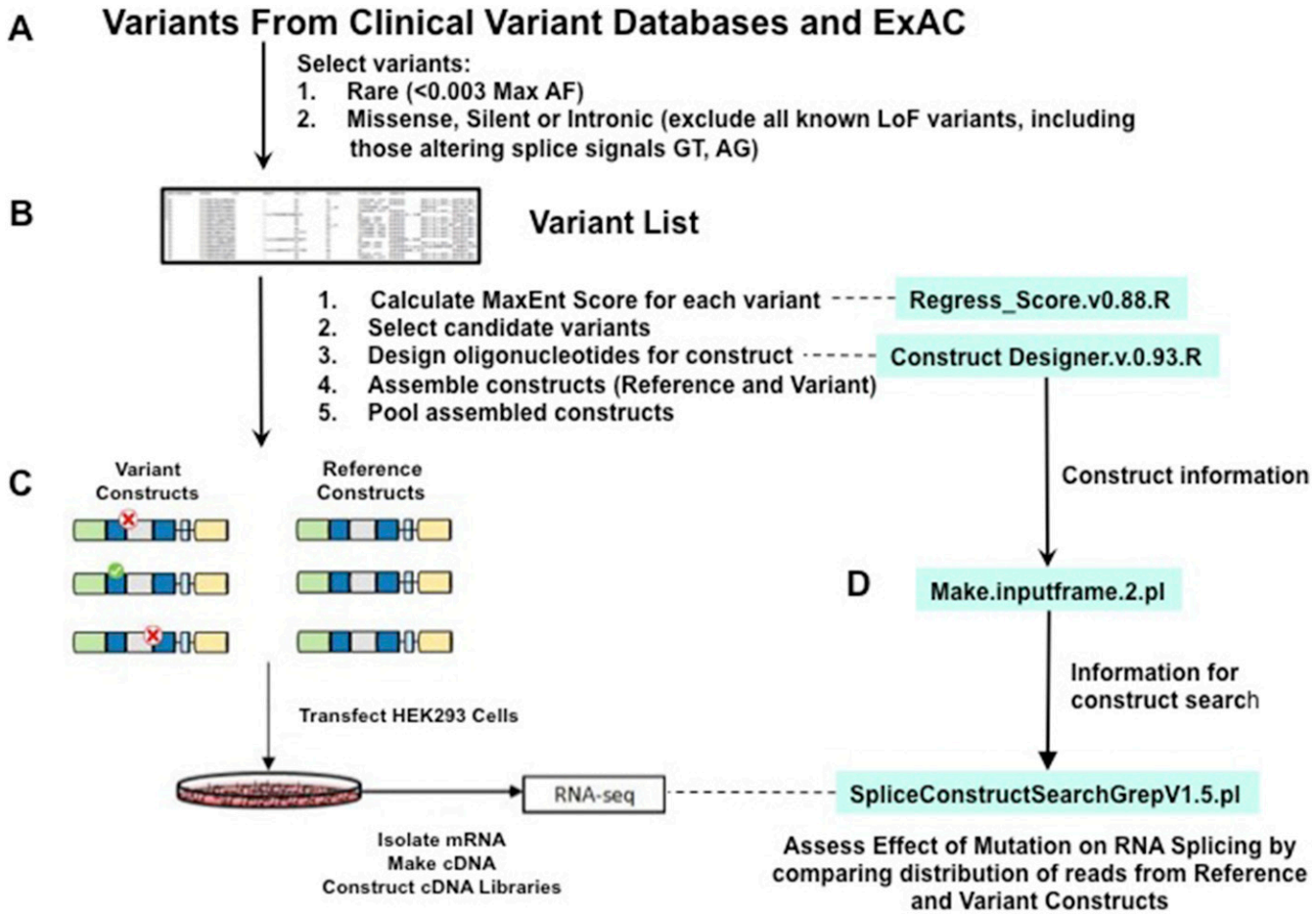

Fig. S2. The workflow of the entire minigene splicing assay. All scripts are available online at https://github.com/SplicingVariant/SplicingVariants_Beta. (*A*) Variants are extracted from the clinical diagnostic and ExAC databases. Variants were initially screened to exclude all variants with allele frequencies >0.003, as well as variants that led to nonsense, frameshift, or altered the canonical GT or AG bases in consensus splice sites. (*B*) Variant positions in a given transcript were retrieved from Biomart database and MaxEnt scores for both reference allele and variant allele were calculated. The decision tree to select candidate splice altering variants based on MaxEnt scoring is depicted in Fig. S1*C*. Candidate variants were used to design mutant and reference constructs as described in *Materials and Methods*. (*C*) Reference and variant minigene constructs are assembled using a combination of synthetic oligonucleotides and PCR. Minigene constructs are pooled and transfected into HEK293 cells. After 24 h, RNA is extracted, converted into a cDNA library, and sequenced. (*D*) The sequence data are demultiplexed and analyzed for quantitative assessment of no splice, normal splice, and aberrant splicing. The status of splicing is determined comparing statistics between reference- and variant-containing minigenes.

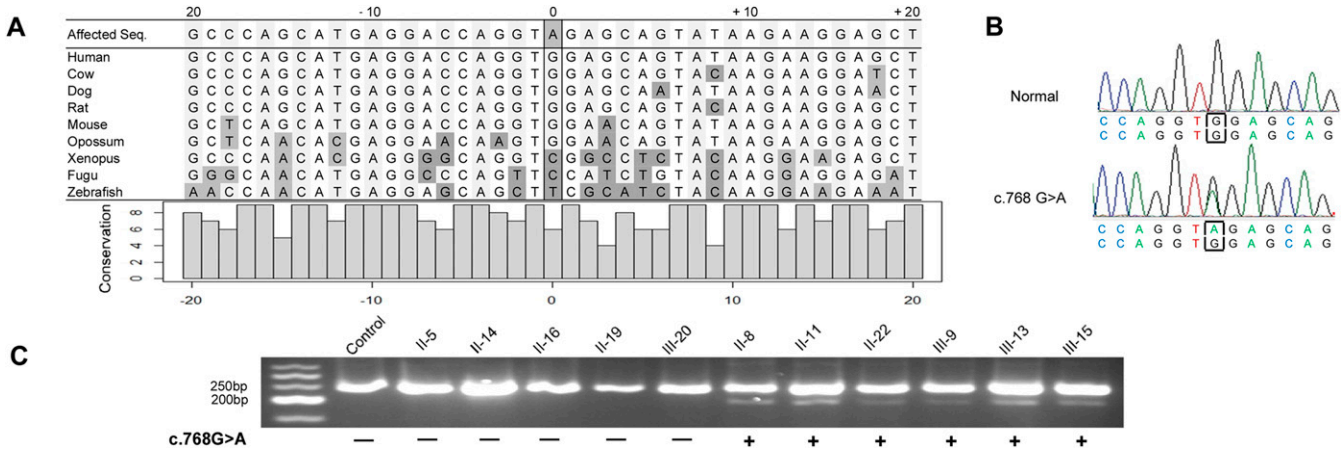

Fig. S3. *LMNA* sequence conservation and aberrant splicing caused by *LMNA* c.768G > A mutation. Related to Fig. 4. (*A*) A snapshot of *LMNA* sequence conservation. The position denotes the distance from the c.768 G > A mutation. Dark gray-filled nucleotides signify sequences differing from human reference. The bar graph represents the conservation of each individual nucleotide. Note that "9" indicates identity across all listed species). (*B*) Sanger sequencing confirmed heterozygous *LMNA* c.768 G > A mutation in the DNA of affected subject. (*C*) Gel-fractionation of RT-PCR products of lymphocyte-derived RNA from several individuals in Family MAE/MAN, showing those with the c.768G > A synonymous variant display an additional smaller species consistent with ~45-bp deletion of part of the *LMNA* transcript.

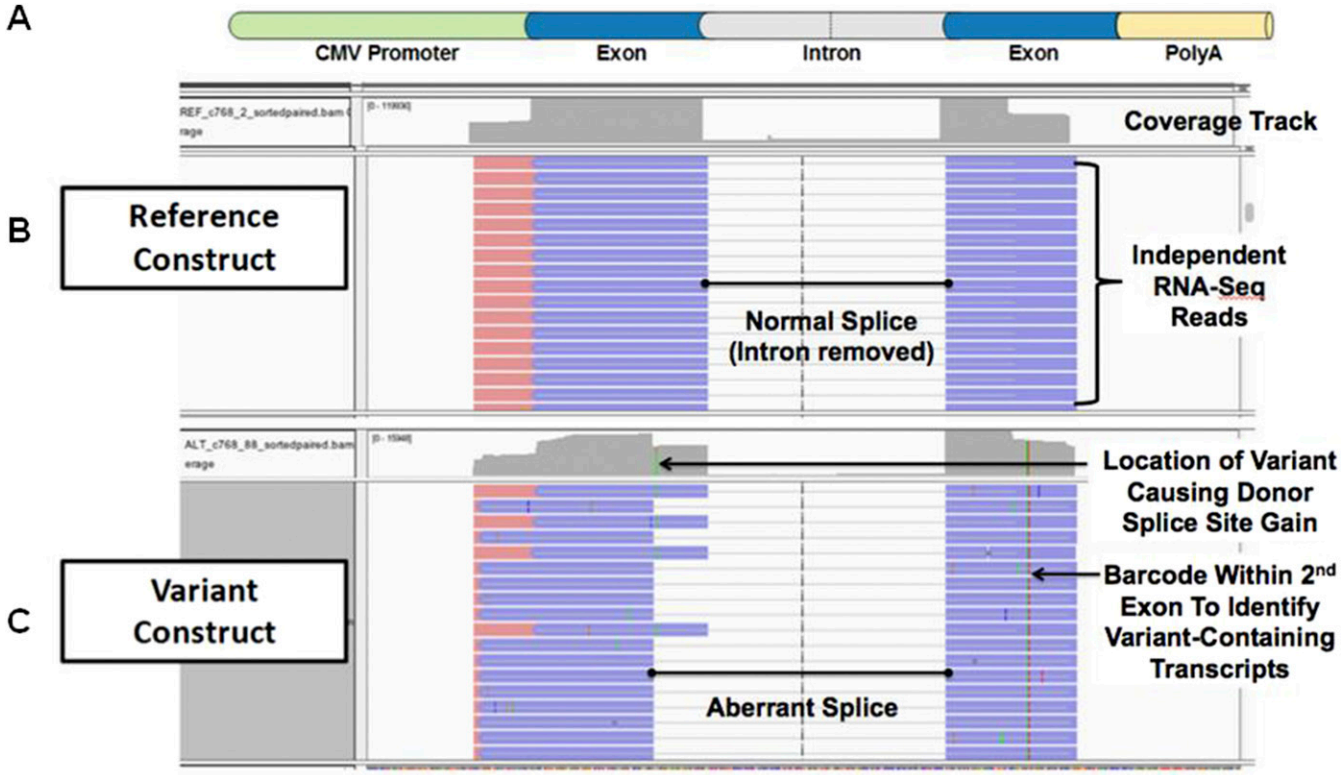

Fig. S4. *LMNA* c.768 G > A confirmed to lead to donor site gain in the minigene splice assay. In addition to quantifying the number of reads with no splicing, normal splicing, and aberrant splicing for each reference- and variant-containing minigene, all reads were also aligned to the original minigene sequence using STAR, and splice patterns for each construct were directly visualized using the IGV to confirm computations. (*A*) A visual schematic of a standard minigene construct used in the splicing assay containing the CMV promoter, exon–intron–exon sequence, and the SV40PA signal. (*B*) Reads derived from the reference construct are depicted. Independent RNA sequencing reads uniformly demonstrate normal splicing out of the intron. (*C*) Reads from the variant construct demonstrate both the normal splice and aberrant splice due to a novel donor site. Aberrant splicing deleted 45 bp from the end of the first exon.

Dataset S1. List of all 815 *LMNA* variants with calculated MaxEnt scores for both donor and acceptor sequences

Dataset S1

A MaxEnt score for a donor splice site and acceptor splice site were separately calculated for each variant, and ΔMaxEnt score is listed. A variant is assigned for evaluation as a loss of splice site (AL or DL) if it is located within 9 bp of a functional 5′ donor splice site or 23 bp of a functional 3′ acceptor splice site. Database of origin of variant (LMM/ClinVar or ExAC), ExAC allele frequency (AF), clinical significance (as listed in clinical variant database), and amino acid consequence of variant are shown. AG, acceptor gain; AL, acceptor loss; Ben, benign; DG, donor gain; DL, donor loss, Lik.Path, likely pathogenic; and Path, pathogenic. A final column indicates the variants that passed our computational filter and were tested in the cell splicing assay (n = 57). The variants that passed our filter but were not tested are also indicated (n = 20).

Dataset S2. List of all 1,575 MYBPC3 variants with calculated MaxEnt scores for both donor and acceptor sequences

Dataset S2

A MaxEnt score for a donor splice site and acceptor splice site were separately calculated for each variant, and ΔMaxEnt score is listed. A variant is assigned for evaluation as a loss of splice site (AL or DL) if it is located within 9 bp of a functional 5′ donor splice site or 23 bp of a functional 3′ acceptor splice site. Database of origin of variant (LMM/ClinVar or ExAC), ExAC allele frequency (AF), clinical significance (as listed in clinical variant database), and amino acid consequence of variant are shown. AG, acceptor gain; AL, acceptor loss; Ben, benign; DG, donor gain; DL, donor loss, Lik.Path, likely pathogenic; and Path, pathogenic. A final column indicates the variants that passed our computational filter and were tested in the cell splicing assay (n = 139). The variants that passed our filter but were not tested are also indicated (n = 39).

Dataset S3. *LMNA* gBlock sequences used for cell splicing assays

Datasets S3

The 500-bp constructs were designed by our in-house script from variant information files. Each sequence contains a 24-bp head (lowercase), an exon–intron–exon motif (exons represented by capital letters, introns lowercase), a 2-bp barcode, and 39-bp tail (both lowercase). Both a reference (REF) and alternate (ALT) sequence are listed for each variant.

Dataset S4. MYBPC3 gBlock sequences used for cell splicing assays

Dataset S4

The 500-bp constructs were designed by our in-house script from variant information files. Each sequence contains a 24-bp head (lowercase), an exon–intron–exon motif (exons represented by capital letters, introns lowercase), a 2-bp barcode, and 39-bp tail (both lowercase). Both a reference (REF) and alternate (ALT) sequence are listed for each variant.

Dataset S5. The splicing assay results for all *LMNA* variants that pass our splicing variant selection filter (n = 57)

Dataset S5

This table provides a variant name, the variant type, splice type, the raw number of reads in the splicing assay in the reference and the variant constructs, and the clinical phenotype linked with this variant in clinical genetic databases. Abbreviations in the second column: In, intronic; Mis, missense; and Syn, synonymous. Abbreviations for splice type: AG, acceptor site gain; AL, acceptor site loss; DG, donor site gain; and DL, donor site loss. Abbreviations for phenotype; DCM-CD, dilated cardiomyopathy with conduction system disease and LGMD, limb-girdle muscular dystrophy. For database, "Clinical" represents a variant seen in either LMM or ClinVar. "ExAC" only represents a variant only observed in the general population.

Dataset S6. The splicing assay results in all MYBPC3 variants that passed our splicing variant selection filter (n = 139)

Dataset S6

The results of the splicing assay for the pathogenic intron 32 deletion (see refs. 12 and 13) are included in the final row of the table. This table provides a variant name, the variant type, splice type, the raw number of reads in the splicing assay in the reference and the variant constructs, and the clinical phenotype linked with this variant in clinical genetic databases. Abbreviations in the second column: In, intronic; Mis, missense; and Syn, synonymous. Abbreviations for splice type: AG, acceptor site gain; AL, acceptor site loss; DG, donor site gain; and DL, donor site loss. Abbreviation for phenotype: LVNC, left ventricular noncompaction. For database, "Clinical" represents a variant seen in either LMM or ClinVar. "ExAC" only represents a variant only observed in the general population.

Dataset S7. The result of splicing assay for all 74 variants that were determined to be low-probability splice-altering candidates by our computational filter

Dataset S7

The *P* value is listed as 1 if the variant effect significantly stabilized the splice site (i.e., the percent of normal spliced constructs increased).

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)