Protein haploinsufficiency drivers identify MYBPC3 variants that cause hypertrophic cardiomyopathy

Received for publication, April 5, 2021, and in revised form, May 21, 2021 Published, Papers in Press, June 5, 2021, https://doi.org/10.1016/j.jbc.2021.100854

Carmen Suay-Corredera1, Maria Rosaria Pricolo1,2, Elias Herrero-Galán1, Diana Velázquez-Carreras1, David Sánchez-Ortiz1, Diego García-Giustiniani1, Javier Delgado4, Juan José Galano-Frutos2,6, Helena García-Cebollada2, Silvia Vilches7,8, Fernando Domínguez7,8,9, María Sabater Molina1, Roberto Barriales-Villa1,2,2, Giulia Friso21, Javier Sancho7,8,13, Luis Serrano7,8,15, Pablo García-Pavía7,8,9,16, Lorenzo Monserrat1, and Jorge Alegre-Cebollada1,2

From the1Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain; 2Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università di Napoli Federico II, Naples, Italy; 3Health in Code, A Coruña, Spain; 4EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; 5Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain; 6Biocomputation and Complex Systems Physics Institute (BIFI). Joint Units BIFI-IQFR (CSIC) and GBs-CSIC, Universidad de Zaragoza, Zaragoza, Spain; 7Heart Failure and Inherited Cardiac Diseases Unit. Department of Cardiology. Hospital Universitario Puerta de Hierro, Madrid, Spain; 8European Reference Network for Rare and Low Prevalence Complex Diseases of the Heart (ERN GUARD-HEART), Madrid, Spain; 9Centro de Investigación Biomédica en Red en Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain; 10Hospital C. Universitario Virgen de la Arrixaca, El Palmar, Murcia, Spain; 11Unidad de Cardiopatías Familiares, Instituto de Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña, Servizo Galego de Saúde (SERGAS), Universidade da Coruña, A Coruña, Spain; 12CEINGE Biotecnologie Avanzate, scarl, Naples, Italy; 13Aragon Health Research Institute (IIS Aragón), Zaragoza, Spain; 14Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain; 15Universitat Pompeu Fabra (UPF), Barcelona, Spain; and 16Universidad Francisco de Vitoria (UFV), Pozuelo de Alarcón, Madrid, Spain

Edited by Enrique De La Cruz

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disease. Variants in MYBPC3, the gene encoding cardiac myosin-binding protein C (cMyBP-C), are the leading cause of HCM. However, the pathogenicity status of hundreds of MYBPC3 variants found in patients remains unknown, as a consequence of our incomplete understanding of the pathomechanisms triggered by HCM-causing variants. Here, we examined 44 nontruncating MYBPC3 variants that we classified as HCM-linked or nonpathogenic according to cosegregation and population genetics criteria. We found that around half of the HCM-linked variants showed alterations in RNA splicing or protein stability, both of which can lead to cMyBP-C haploinsufficiency. These protein haploinsufficiency drivers associated with HCM pathogenicity with 100% and 94% specificity, respectively. Furthermore, we uncovered that 11% of nontruncating MYBPC3 variants currently classified as of uncertain significance in ClinVar induced one of these molecular phenotypes. Our strategy, which can be applied to other conditions induced by protein loss of function, supports the idea that cMyBP-C haploinsufficiency is a fundamental pathomechanism in HCM.

Hypertrophic cardiomyopathy (HCM) is the most frequent inherited cardiac muscle disease, with an estimated prevalence of at least 0.5% (1–3). HCM is a frequent cause of sudden cardiac death in the young and can result in several cardiovascular complications including heart failure and thromboembolism (1). Identification of HCM-causing variants has transformed clinical management of HCM families. Today, genetic testing can confirm a clinical suspicion, help differential diagnosis, and forms the basis of family cascade screening to allow reproductive and professional counseling. However, with many other genetic conditions, testing is limited by the rarity of HCM genetic variants (4, 5). Because the number of individuals affected by some of these variants is so low, detailed segregation studies are seldom possible, leaving numerous variants classified as variants of uncertain significance (VUS). This situation has motivated efforts to assess pathogenicity of VUS using functional and population genetics approaches (6–10).

Variants of MYBPC3, the gene encoding cardiac myosin-binding protein C (cMyBP-C), are a leading cause of HCM (Fig. 1A) (4, 11, 12). Most well-established pathogenic variants in MYBPC3 are frameshift, nonsense, or conserved RNA splice site mutations that result in truncated polypeptides, which are more prone to degradation leading to lower total cMyBP-C protein levels (haploinsufficiency) (13–16). Indeed, cMyBP-C haploinsufficiency results by itself in the development of HCM (Fig. 1A, middle) (17–19). Intriguingly, a number of familial HCM cases are caused by variants in exon regions of MYBPC3 that do not lead to truncations (Fig. 1A, right) (17, 19, 20), including the most common pathogenic variant in

4 These authors contributed equally to this work.
* For correspondence: Jorge Alegre-Cebollada, jalegre@cnic.es.
HCM (c.1504C>T, p.R502W) (13). As of February 2021, ClinVar lists 919 nontruncating MYBPC3 variants potentially linked to HCM.

Nontruncating MYBPC3 variants currently pose a major challenge to genetic diagnosis of HCM. Despite recent encouraging developments, for most variants, it remains impossible to assign pathogenicity just from their location in the gene or the nature of specific amino acid changes (10, 21, 22). Interestingly, the clinical manifestations of truncating and nontruncating pathogenic variants are similar, suggesting that pathogenicity in some nontruncating variants could also result from cMyBP-C haploinsufficiency (15, 21, 23, 24). The two most frequent mechanisms of protein haploinsufficiency induced by putative nontruncating variants in monogenic diseases are RNA splicing defects that result in the appearance of premature stop codons (25) and protein destabilization (26). Both these haploinsufficiency drivers have been reported in MYBPC3 variants linked to HCM (7, 8, 19, 20, 27–30). However, how these molecular features cause disease remains unknown because of the lack of systematic comparison with nonpathogenic variants. For instance, mild changes in splicing may be better tolerated than full native splicing abrogation by a conserved splice site mutation (20). Ito et al. (7) showed that putative nontruncating MYBPC3 variants appearing in cardiomyopathy gene databases more frequently lead to RNA splicing alterations than variants in control databases; however, interpretation of results is complicated by the presence of HCM variant carriers in the general population (31). In a recent report, Thompson et al. (22) showed that 32% of pathogenic and 7% of nonpathogenic missense MYBPC3 variants are predicted bioinformatically to induce protein domain destabilization, suggesting that protein destabilizing features can provide evidence supporting variant pathogenicity. These encouraging computational results call for experimental assessment of protein destabilization phenotypes induced by MYBPC3 variants.

Here, in quest of molecular features that identify pathogenicity, we examined cMyBP-C haploinsufficiency drivers

**Figure 1.** cMyBP-C haploinsufficiency drivers induced by HCM-linked and nonpathogenic MYBPC3 variants. A, *left*, scheme of the location of cMyBP-C (in yellow) in the sarcomere. *Middle*, most HCM-causing MYBPC3 variants lead to truncated polypeptides and protein haploinsufficiency. *Right*, the remaining variants are nontruncating and result in full-length mutant proteins (mutant domain is represented in red). B, workflow to identify cMyBP-C haploinsufficiency drivers in a curated database of HCM-linked and nonpathogenic MYBPC3 variants. Bioinformatics predictions of RNA splicing alteration and protein destabilization are made for the variants building up this database. Positive hits are then further assessed experimentally to identify protein haploinsufficiency drivers induced by these variants. cMyBP-C, cardiac myosin-binding protein C; HCM, hypertrophic cardiomyopathy.
in missense and synonymous MYBPC3 variants. Our strategy builds on the characterization of variants for which cosegregation and population genetics data are available, providing information about their pathogenicity. We propose that experimentally testable variant-induced RNA splicing alteration and extensive protein destabilization provide functional evidence of pathogenicity in 11% of putative nontruncating MYBPC3 variants currently classified as VUS in ClinVar.

Results

A database of HCM-linked and nonpathogenic MYBPC3 variants

We built a database containing 20 HCM-linked and 24 nonpathogenic putative synonymous and missense MYBPC3 variants covering the entire coding sequence of the gene (see Experimental procedures section, File S1). We classified as HCM-linked those variants showing minor allele frequency (MAF) <10\(^{-4}\) in the Genome Aggregation Database and showing evidence of pathogenicity based exclusively on cosegregation and population genetics criteria validated by the American College of Medical Genetics (ACMG) (32). Our approach enables agnostic investigation and quantification of molecular pathomechanisms, since pathogenicity is not assigned according to functional or bioinformatics criteria. Interestingly, the majority of HCM-linked variants in our database target domains C3, C6, and C10, as observed in the Sarcomeric Human Cardiomyopathy Registry (21). All nonpathogenic variants have MAF >10\(^{-4}\), which is incompatible with HCM prevalence (13). We next investigated whether altered RNA splicing and protein stability are specific to HCM-linked variants, under the premise that molecular features related to disease should not appear in nonpathogenic variants. We first triaged variants using bioinformatics predictors and then assessed positive hits experimentally (Fig. 1B).

Alteration of RNA splicing by HCM-linked variants

RNA splicing is the process by which noncoding introns are removed from precursor mRNA, leading to exon-only-containing mature mRNA. Splicing involves specific recognition of the strictly conserved two first (donor site) and last (acceptor site) intron nucleotides. Mutations in these sequences impair canonical splicing, resulting in alternative mRNAs that contain premature stop codons, or lead to insertion/deletions in the polypeptide (33). Indeed, mutations in MYBPC3 that affect conserved splicing sites are considered to lead to truncations and are classified as pathogenic (8, 13). However, the correct functioning of the splicing machinery also involves recognition of sequence features in the exons, particularly in regions close to exon–exon boundaries. As a consequence, variants in exonic sequences of MYBPC3 traditionally classified as missense can result in RNA splicing alterations that lead to truncated polypeptides (7, 8, 20, 34). Indeed, we found that four HCM-linked variants in our database are predicted in silico to induce splicing site losses (Fig. 2A and Table 1). These variants target the first or last nucleotide of an exon (Note S1). Predictions also identified two HCM-linked variants that can lead to activation of new splicing sites. In contrast, all nonpathogenic variants target nucleotides outside exon–exon boundaries, and none of them is predicted to cause loss of native splicing sites. For two nonpathogenic variants, the appearance of a new splicing site is suggested (Fig. 2A, Table 1, and Note S1).

To validate predicted RNA splicing alterations, we mined data available in the literature and/or examined RNA splicing experimentally. RNA splicing of the MYBPC3 transcript is best studied from myocardial samples; however, the scarcity of genotyped myocardium has limited analysis to a few variants (8, 19, 20). We employed two alternative and more accessible strategies whose results are so far in excellent agreement with experiments using myocardial samples. These strategies are analysis of mRNA from the leukocyte fraction of peripheral blood from variant carriers, which has the advantage of immediate translational potential (34–39), and in vitro experiments using minigene constructs (7, 38, 40) (Table S1). In both cases, isolated mRNA was amplified by RT-PCR using specific pairs of primers, and results were analyzed by electrophoresis and Sanger sequencing (Fig. 2 and File S2).

In Figure 2, B–H, we show results from amplification of the exon 15–exon 21 region using mRNA obtained from blood samples that carry different MYBPC3 variants. If splicing is correct, amplification results in a fragment of 692 bp, as shown for the WT individual (Fig. 2C). We observed that variant c.1624G>C (p.E542Q) leads to a higher electrophoretic mobility band at 500 bp, marking skipping of exon 17, in agreement with predictions that the mutation perturbs a native donor site (Fig. 2, C–E). In contrast, the prediction that variant c.1505G>A (p.R502Q) causes an acceptor site gain was not validated, as splicing proceeded as for WT (Fig. 2, C and F–H). Equivalent results have been obtained before using other experimental approaches (Table 1).

In Figure 2, I–L, we present the results of a minigene strategy to test whether the nonpathogenic variant c.492C>T (p.G164G) induces the appearance of a new donor site, as suggested bioinformatically. RT-PCR amplification generated two bands both in the WT and the mutant sample (Fig. 2I). The 850 bp band corresponds to the skipping of the MYBPC3 insert, a scenario that is not uncommon in minigene assays. The 1100 bp band results from the correct inclusion of MYBPC3 exons 4 and 5 in the minigene transcript, which was further verified by Sanger sequencing (Fig. 2L). No other band was amplified in the c.492C>T sample. Hence, the prediction in Figure 2K of a new splicing site for this variant was not validated experimentally.

Combining our findings with data from the literature, we were able to collect results for all predicted RNA splicing alterations (Table 1). Four HCM-linked variants were confirmed to induce aberrant splicing leading to premature stop codons (Table S2). The validation rate in this set of variants was 4/4 for splicing site losses and 0/2 for splicing site gains.

Editors’ Pick: Molecular phenotyping of MYBPC3 variants
Importantly, none of the predicted alterations of splicing in nonpathogenic variants was confirmed experimentally (Table 1). Hence, our results indicate that alteration of RNA splicing occurs in 20% of HCM-linked variants and is associated with HCM at 100% specificity (p = 0–0.016; Table 1 and Note S2).
HCM-linked variants induce extensive protein destabilization

A reduction in protein stability leads to more frequent unfolding, which can result in degradation-sensitive polypeptides and reduced total protein levels (26, 41). To analyze protein destabilization induced by the variants in our database, we first used FoldX (CRG) (42). This software can predict protein destabilization if the high-resolution 3D structures of targeted domains are known. Hence, we examined the stability of the variants affecting domains C0, C1, C2, C3, and C5 of cMyBP-C, for which high-resolution structures are available (Protein Data Bank codes: 2K1M, 3CX2, 1PD6, 2MQ0, and 1GXE, respectively). Surprisingly, the distribution of ΔΔG values for these ten HCM-linked mutants is indistinguishable from that of nonpathogenic variants. Indeed, the HCM-linked mutant with the highest destabilization is C2-F448S (ΔΔG = 2.2 kcal/mol), but nonpathogenic variants C3-G507R and C3-A522T are predicted to induce at least the same level of destabilization (Fig. S1A and File S1). These results are in apparent contradiction with those reported by Thompson et al. (22), who predicted protein destabilization in 32% of HCM-linked variants. This discrepancy probably originates from the fact that most destabilizing mutations identified by Thompson et al. target domains C6 and C10, which we did not screen because their high-resolution structures remain unknown. Hence, we extended analysis of protein destabilization to the variants in our database with no FoldX prediction by examining experimentally recombinant WT and mutant cMyBP-C domains. We also included in this analysis the three nonpathogenic variants with the highest predicted destabilization. This approach allowed us to obtain information about the extent of protein destabilization that is not expected to be associated with development of HCM. Unfortunately, variants targeting domains C6 and C8 could not be studied because the corresponding WT domains were refractory to recombinant expression (see Experimental procedures section).

Protein expression was induced in Escherichia coli, and purified domains (Fig. S1B) were analyzed by far-UV CD spectroscopy, a technique that reports protein secondary structure (43, 44). We found that three of the four HCM-linked mutants (C4-D610N, C10-G1206D, and C10-Y1251H) could not be produced in soluble form in the most favorable expression conditions, suggesting strong destabilization (see Experimental procedures section). Of the nonpathogenic variants, eight of nine preserve protein structure (Figs. 3B and S2). To characterize the impact of mutations on domain stability, we tracked CD signals at increasing temperatures. As protein domains transition between the native and the unfolded states, the CD signal varies (Figs. S2 and S3), and the temperature at the midpoint of the denaturing transition, or melting temperature (Tm), informs on the thermal stability of the domain (44). We determined that nonpathogenic variants induce maximum ΔTm (Tm [WT] – Tm [mutant]) = 4.3°C (Fig. 3, C and D and File S1), suggesting

---

<table>
<thead>
<tr>
<th>Variant (DNA)</th>
<th>Variant (protein)</th>
<th>Variant classification</th>
<th>Predicted splicing alteration</th>
<th>Experimental splicing alteration (myocardium)</th>
<th>Experimental splicing alteration (blood)</th>
<th>Experimental splicing alteration (minigene)</th>
</tr>
</thead>
<tbody>
<tr>
<td>492C&gt;T</td>
<td>492G&gt;C</td>
<td>Nonpathogenic</td>
<td>Donor gain (DG)</td>
<td>No (Fig. 2 and L)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>725G&gt;A</td>
<td>725G&gt;G</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>772G&gt;A</td>
<td>772G&gt;G</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>1505G&gt;A</td>
<td>1505G&gt;G</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>1624G&gt;C</td>
<td>1624G&gt;C</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>2308G&gt;A</td>
<td>2308G&gt;G</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
<tr>
<td>3106C&gt;T</td>
<td>3106C&gt;T</td>
<td>Nonpathogenic</td>
<td>No (DG)</td>
<td>No (DL)</td>
<td>No (DL)</td>
<td>No (DG)</td>
</tr>
</tbody>
</table>

Abbreviations: AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss.

Assessment of predictions using myocardial biopsies, blood samples, or minigene experiments is indicated. Experimental results have been obtained from the literature and/or this study, as indicated.
that limited changes in $T_m$ up to $\sim 5$ °C are generally well tolerated and cannot be linked to pathogenicity. There was only one nonpathogenic variant that could not be produced (C9-R1138H) (Fig. S1C). Using molecular dynamics simulations, we obtained evidence that mutant domains that could not be expressed in native form are indeed destabilized (Table S3).

In summary, in four (20%) of the HCM-linked variants, our protein stability analysis workflow was able to capture domain destabilization, as indicated by lack of expression of mutant domains or by alterations in CD spectra at 25 °C. These results indicate that domain destabilization associates with pathogenicity with 94% specificity ($p = 0.01–0.075$; Fig. 3E and Note S2).

**Prevalence of cMyBP-C haploinsufficiency drivers**

Our results in the previous sections show that 40% of the HCM-linked variants in our database induce altered RNA splicing or protein destabilization (Fig. 4A). The extent of these cMyBP-C molecular phenotypes among pathogenic variants may be higher since we could not obtain information on splicing for three mutants because of limited bioinformatics predictions and on protein destabilization for six mutants because of our inability to produce WT domains. Hence, we analyzed the distribution of cMyBP-C haploinsufficiency drivers in the subset of variants for which we have information for both RNA splicing and protein stability. This analysis showed that $>55\%$ of HCM-linked variants induce RNA splicing alteration or protein destabilization (Fig. 4B). Strikingly, several variants show no evidence of either of these molecular phenotypes, as observed previously for p.R502W (19, 45), the most common pathogenic variant in HCM (13).

**Molecular phenotyping of VUS**

According to our data, defects in RNA splicing and protein destabilization are associated with pathogenicity with close to 100% specificity. Prompted by this observation, we investigated the extent to which assessment of protein haploinsufficiency drivers might contribute to assigning pathogenicity to MYBPC3 variants currently classified as VUS. For this, we studied putative nontruncating VUS reported in ClinVar with MAF $<10^{-4}$. To be able to triage domain destabilization bioinformatically, we restricted our analysis to the 73 VUS targeting cMyBP-C domains whose high-resolution structures are known (Fig. 5A and File S3). RNA splicing was predicted to be altered in 14 of 68 variants in which native splice sites were detected, and a new splicing site was predicted in one of the five remaining variants (Fig. 5B and File S3). We managed to gather experimental information for ten of these, leading to confirmation of RNA splicing alterations in two of two predicted splice site losses and in three of nine predicted site gains, all of them generating premature stop codons (Fig. 5C, Table 2, Fig. S4, and Table S2). For the five untested variants, we could not recruit carriers, and the minigene assays did not
Figure 4. Landscape of molecular phenotypes induced by putative nontruncating HCM-linked and nonpathogenic variants in MYBPC3. A, identification of cMyBP-C protein haploinsufficiency drivers in a database of nontruncating MYBPC3 variants according to the workflow proposed in Figure 1B. The number of variants positive for predicted alterations in RNA splicing or protein stability is indicated, together with the outcomes of experimental assessment. Some variants could not be tested bioinformatically because of lack of identification of native splicing sites or lack of high-resolution protein structures. Experimental assessment of protein destabilization shown in brackets corresponds to variants with no available in silico predictions. Results for HCM-linked and nonpathogenic variants are indicated in pink and green, respectively. B, pie chart summarizing the proportion of HCM-linked variants in MYBPC3 inducing different types of cMyBP-C protein haploinsufficiency drivers. For this analysis, we only considered the 14 HCM-linked variants in our database (File S1) for which data on both RNA splicing and protein stability were available. Four of these variants show altered RNA splicing. Four of them lead to domain destabilization. Our analysis assumes that bioinformatics predictions are able to capture alterations with 100% sensitivity (i.e., that there are no false negatives) (7, 42). HCM, hypertrophic cardiomyopathy.

Discussion

The clinical management of HCM families has benefited from the discovery of causative genes over the last two decades (1). Currently, genetic testing is a class I recommendation in both the European and American HCM clinical practice guidelines. Thanks to the advent of next-generation DNA sequencing techniques, the number of genetic variants found in HCM patients has increased considerably, creating new challenges in variant interpretation (4, 37, 47–49). Many genetic variants are present in only a few patients, limiting the power of cosegregation analyses to assess pathogenicity. Alternatively, functional deficits linked to disease progression can enable classification of rare pathogenic variants, as implemented in the ACMG guidelines (32, 47, 50–52).

Here, we have phenotyped MYBPC3 variants in the search for protein haploinsufficiency drivers that can sustain pathogenicity. Importantly, we compared HCM-linked and nonpathogenic variants, which allowed us to link molecular properties to pathogenicity. Although our approach leads to small sample sizes, our results demonstrate that around 50% of putative nontruncating HCM-linked variants in MYBPC3 significantly and specifically induce altered RNA processing or extensive protein destabilization (Fig. 4B). Furthermore, we have identified these pathogenicity drivers in 11% of MYBPC3 variants currently classified as VUS in ClinVar (Fig. 5). Considering these results, we estimate that close to 20% of missense and synonymous variants in MYBPC3 currently classified as VUS in ClinVar may show functional evidence of pathogenicity based on the induction of protein haploinsufficiency drivers.
The first step of our workflow to assess protein haploinsufficiency drivers in MYBPC3 variants uses in silico tools to predict alterations in RNA splicing and protein stability, which are then validated experimentally (Fig. 5). Hence, it is important to consider the sensitivity and specificity of these bioinformatics tools. For example, experimental analysis following negative predictions is discouraged because the sensitivity of predictions is generally high (7, 42). Direct experimental testing would only be indicated if algorithms fail to capture native splicing sites, or when there is no high-resolution protein structural information. Regarding specificity of predictions of RNA splicing alterations, we observed 0% (0/6) false positives for the prediction of site losses, contrasting with a 77% (10/13) false-positive rate for the prediction of new sites, in agreement with previous observations (7, 37) (Tables 1 and 2). Hence, experimental validation of alterations of RNA splicing, particularly for prediction of new sites, is a strict requirement for the identification of true positives, which in the clinical setting can be implemented easily by analyzing the peripheral blood of carriers. We also recommend validation of protein destabilization predictions, since we could only verify extensive protein destabilization in 40% of FoldX-triaged VUS variants (Fig. 5, D and E).

Our study supports the emerging view that genetic variants that affect RNA splicing are pathogenic (7, 8, 53, 54). Interestingly, these RNA-splicing–altering mutants can be found in

Figure 5. Assessment of cMyBP-C haploinsufficiency drivers in MYBPC3 VUS. A, 73 VUS in ClinVar were screened for alterations to RNA splicing and protein stability. B, results from predictions of RNA splicing. Blue triangles indicate exon–exon boundaries. Each bar corresponds to a single variant and is colored according to the predicted effect on RNA splicing. C, experimental assessment of predicted changes. Variants whose effects on splicing could not be tested experimentally are colored light pink (see Table 2). D, predicted protein destabilization of the 73 VUS. Each bar corresponds to a single variant and is colored according to the predicted protein destabilization. The dotted line marks the highest destabilizing change in ΔΔG detected for a nonpathogenic variant (Fig. S1A). E, experimental determination of changes in thermal stability for the 10 VUS with predicted ΔΔG > 3 kcal/mol. The green reference line at ΔTm = 5 °C marks destabilization values that can be found in nonpathogenic variants (Fig. 3D), whereas we consider ΔTm > 10 °C (pink reference line) to be a signature of HCM-linked variants (corresponding bars are colored red, whereas variants below the 10 °C threshold are shown in gray). Figure S6 shows the CD data. cMyBP-C, cardiac myosin-binding protein C; VUS, variants of uncertain significance.
both exonic and intronic regions that are far from splice sites if the consequence of the mutation is the activation of a new splice site (Table 2) (7, 37). Interestingly, some nonpathogenic variants are found in regions close to splice sites; however, they are not predicted to alter splicing to the same extent as pathogenic variants (8). Our data are consistent with this observation (Note S1). To date, there have been no reports of nonpathogenic MYBPC3 variants causing altered RNA splicing.

To the best of our knowledge, our study is the first to experimentally characterize the frequency of protein destabilization induced by HCM-linked missense MYBPC3 variants (Fig. 4B). Our results are in good agreement with previous estimates suggesting that 32% of pathogenic variants could induce protein destabilization (22). However, we acknowledge that there are challenges associated with the functional assessment of MYBPC3 variants based on protein stability. First, there is no high-resolution structural information for many cMyBP-C domains, which may hamper effective triaging of potentially disruptive variants.

A second challenge stems from experimental validation of protein destabilization phenotypes using recombinant domains, which in the case of highly destabilizing mutations can be difficult or even impossible. Our interpretation that strong destabilization hampers mutant domain production is supported by molecular dynamics simulations (Table S3) and is consistent with similar observations in similar settings (55); however, recombinant production of mutant domains may fail for reasons other than reduced stability, including toxicity to host or codon bias (56). In our set of experiments, we could not produce the nonpathogenic C9 domain carrying variant R1138H. The WT C9 domain has the lowest $T_m$ value among all domains assayed (42.2 °C; Fig. 3C). However, C9 is a fibronectin-III type domain, a protein fold whose typical $T_m$ values are well above 50 °C (57). We speculate that our recombinant C9 domain may not recapitulate native stability and that the impact of minor, nonpathogenic changes on stability may be higher in this context. It is also possible that the native environment of the sarcomere limits the destabilizing effect of the mutation via post-translational modifications or protein–protein interactions.

The fact that protein stability is not binary poses a third challenge to protein destabilization–guided functional classification of variants. We observed that nonpathogenic variants can cause slight protein destabilization ($\Delta T_m < 5 ^\circ C$; Fig. 3D). Based on the typical distribution of $\Delta T_m$ in unselected single amino-acid polymorphisms across different proteins (46), we considered that $\Delta T_m > 10 ^\circ C$ is a signature of pathogenic protein destabilization. The confidence of this threshold could be increased by measuring the stability of additional nonpathogenic variants. In addition, more refined thresholds could be obtained by assessing the stability of multidomain cMyBP-C constructs, although the highly similar thermal stability of consecutive cMyBP-C domains (Fig. 3C) can complicate interpretation of results.

From a mechanistic point of view, both the alteration of RNA splicing and the destabilization of cMyBP-C domains can lead to reduced cMyBP-C levels, similar to the situation induced by truncating variants (18, 28, 58, 59). It is remarkable though that many HCM-linked MYBPC3 variants do not alter RNA splicing or protein stability (Fig. 4B). The pathogenicity triggers in those cases, which include well-established HCM variants such as c.1504C $>$ T, p.R502W, remain unknown. A tempting hypothesis is that some of them can induce protein haploinsufficiency by alternative mechanisms, including decreased rates of transcription and translation (60–62), increased recognition by mRNA and protein degradation machineries (21, 60, 63, 64), and defective incorporation of cMyBP-C in the sarcomere (65). Alternatively, in the absence of protein haploinsufficiency, variants can lead to perturbed binding to protein partners, resulting in altered sarcomere function (66–70). Tantalizingly, both protein haploinsufficiency and altered binding could result in converging misregulation of the super-relaxed state of myosin, inducing sarcomere hypercontractility typical of HCM (71).

In summary, we propose that identification of protein haploinsufficiency drivers in MYBPC3 variants provides functional evidence of pathogenicity (PS3 criterion in the

Table 2

<table>
<thead>
<tr>
<th>VUS (cDNA)</th>
<th>VUS (protein)</th>
<th>Predicted splicing alteration</th>
<th>Experimental splicing alteration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>39C&gt;T</td>
<td>S13S</td>
<td>DG</td>
<td>—</td>
<td>File S2</td>
</tr>
<tr>
<td>194C&gt;T</td>
<td>T65M</td>
<td>AG</td>
<td>—</td>
<td>File S2</td>
</tr>
<tr>
<td>479G&gt;A</td>
<td>R160Q</td>
<td>AG</td>
<td>No</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>505G&gt;A</td>
<td>G169S</td>
<td>DL, DG</td>
<td>Yes (DL, AG)</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>659A&gt;G</td>
<td>Y220C</td>
<td>DG</td>
<td>No</td>
<td>(7)</td>
</tr>
<tr>
<td>1213A&gt;G</td>
<td>M405V</td>
<td>DG</td>
<td>Yes</td>
<td>(7, 49)</td>
</tr>
<tr>
<td>1287G&gt;A</td>
<td>A429A</td>
<td>AG</td>
<td>—</td>
<td>File S2</td>
</tr>
<tr>
<td>1309G&gt;A</td>
<td>V437M</td>
<td>DG</td>
<td>—</td>
<td>File S2</td>
</tr>
<tr>
<td>1466A&gt;T</td>
<td>D489V</td>
<td>DG</td>
<td>—</td>
<td>File S2</td>
</tr>
<tr>
<td>2155T&gt;C</td>
<td>C719R</td>
<td>DG</td>
<td>No</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>2217G&gt;A</td>
<td>E739E</td>
<td>AG</td>
<td>No</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>2234A&gt;G</td>
<td>D745G</td>
<td>DG</td>
<td>No</td>
<td>(7)</td>
</tr>
<tr>
<td>2249C&gt;T</td>
<td>T750M</td>
<td>AG</td>
<td>No</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>2288A&gt;G</td>
<td>N763S</td>
<td>DG</td>
<td>Yes</td>
<td>Fig. 4A, B and C</td>
</tr>
<tr>
<td>2308G&gt;C</td>
<td>D770H</td>
<td>DL</td>
<td>Yes</td>
<td>Fig. 5-B</td>
</tr>
</tbody>
</table>

Abbreviations: AG, acceptor gain; DG, donor gain; DL, donor loss; VUS, variants of uncertain significance.

Five variants could not be studied. We experimentally studied the variants for which no information was available in the literature. Experimental results were obtained from the literature and/or this study, as indicated.
EDITORS’ PICK: Molecular phenotyping of MYBPC3 variants

ACMG guidelines (32)) and can contribute to the assessment of pathogenicity of putative nontruncating variants in the MYBPC3 gene. RNA splicing alterations and protein destabilization can be validated using readily available laboratory assays, addressing the urgent need for methods to assign pathogenicity of genetic variants associated with HCM (51, 72). Our results increase the number of actionable variants in MYBPC3, thus having the potential to improve clinical management of HCM families, and can be extended to other diseases caused by protein loss of function.

Experimental procedures

Human samples

Human samples were obtained from patients with informed consent according to Declaration of Helsinki guidelines. Research involving humans was authorized by the Comité de Ética de Investigación de Instituto de Salud Carlos III (ISCIII; PI 39_2017).

Selection of genetic variants

We retrieved MYBPC3 variants from the Health in Code-Mutations database, which includes information about >155,000 individuals obtained from >50,000 articles in the literature, as well as from Health in Code’s own clinical reports. Variants were included in the HCM-linked group if they show MAF < 10−4 in the Genome Aggregation Database and are enriched in HCM individuals (ACMG PS4 criterion), cosegregate with disease (ACMG PP1 criterion), or are de novo mutations in the context of no family history of disease (ACMG PS2 criterion). All nonpathogenic variants satisfy MAF > 10−4. To retrieve VUS in MYBPC3 from ClinVar, we considered all nontruncating variants associated with HCM, excluding those with conflicting interpretations of pathogenicity or showing MAF > 10−4. We restricted our VUS analysis to variants targeting domains for which there is high-resolution structural information.

Bioinformatics prediction of haploinsufficiency drivers

We used Alamut Visual (Interactive Biosoftware) to predict RNA splicing alterations induced by all variants in our HCM-linked/nonpathogenic and VUS databases. Alamut implements analyses of four RNA splicing prediction algorithms (SSF, MaxEnt, NNSPLICE, and GeneSplicer). We first determined if at least two of the algorithms identified the canonical splicing site around the mutation site. We then calculated the percent change in splicing score for native sites and potential new splicing sites. Positive hits result in at least two tools predicting >10% decrease (for native sites) or increase (for new sites) in the splicing score, in agreement with published guidelines (73). Changes in protein thermodynamic stability were estimated by FoldX. This software estimates changes in free energy (ΔΔG) upon mutation from empirically and statistically derived energy functions (42).

Analysis of RNA splicing

For experimental determination of RNA splicing, we analyzed blood samples from variant carriers or used engineered minigene constructs (File S2). In the case of blood samples, total RNA was extracted from leukocytes of carriers using Trizol (Thermo Fisher Scientific). RNA retrotranscription was done with random primers by SuperScript IV VILO Master Mix (Thermo Fisher Scientific), and the region of interest was PCR amplified using specific oligos (File S2). RT-PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and then Sanger sequenced when necessary. Correct RNA splicing generates readable electropherograms, obtained for both primers, and whose sequence matches the sequence of the canonical complementary DNA (cDNA) for cMyBP-C. For minigene experiments, the WT and mutant genomic DNA fragments were obtained from Integrated DNA Technologies, including at least the exon of interest and the 5’ and 3’ intronic flanking regions (File S2). The constructs were cloned into the β-globin’s intron 2 of the pMGene vector (74) using KpnI. Alternatively, pMGene vectors including the inserts of interest were obtained from GeneArt Gene Synthesis (from Thermo Fisher Scientific). In both cases, the resulting constructs were expressed in human embryonic kidney 293 monolayers cultured in Dulbecco’s modified Eagle’s medium (Gibco) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37 °C and 5% CO2. Cells were transiently transfected with 1 µg of WT or mutant pMGene using FuGENE HD (Promega) according to the manufacturer’s protocol. About 24 to 48 h after transfection, cells were collected and mRNA was extracted and retrotranscribed, and PCR products were purified and sequenced to compare splicing of WT and mutant constructs as aforementioned. Alternatively, purified PCR products were loaded onto 2% preparative agarose gels to isolate specific bands of interest (Fig. S4). These bands were purified using the QIAquick Gel Extraction Kit (Qiagen) prior to Sanger sequencing.

Protein expression and purification

The cDNAs encoding the cMyBP-C domains and their mutants were cloned from myocardial RNA, produced by PCR mutagenesis, or acquired commercially from Integrated DNA Technologies. Sequences are available in File S4. cDNAs were cloned into a custom-modified pQE80L expression plasmid (Qiagen) using BamHI and BglII enzymes. Final expression plasmids were verified by Sanger sequencing. Domains were expressed in E. coli BLR(DE3). Cultures at an absorbance of 0.6 to 1 at 600 nm were induced with IPTG (specific induction conditions can be found in Files S1, S3, and S4). In general, we found that expression at lower concentrations of IPTG and temperature ≤25 °C resulted in better yield of purified challenging-to-express domains, so these conditions were preferred for the expression of mutant domains. Purification of His-tagged domains was achieved by metal affinity and gel filtration chromatographies (75). Although different expression conditions were assayed, WT domains C6 and C8 could not be produced (File S4), so variants targeting these domains could not be analyzed. Proteins were eluted from the final size-exclusion chromatography in 20 mM NaPi, pH 6.5, and...
63.6 mM NaCl. Proteins were stored at 4 °C. Results of the expression and purification procedures were evaluated by SDS-PAGE and Western blotting following standard procedures (Fig. S1B).

**CD**

CD spectra were collected using a Jasco J-810 spectropolarimeter. Purified proteins were tested in 20 mM NaPi, pH 6.5, and 63.6 mM NaCl at protein concentrations ranging from 0.1 to 0.5 mg/ml in 0.1-cm-pathlength quartz cuvettes. Protein concentration was determined from absorbance at 280 nm values using theoretical extinction coefficients (Files S1, S3, and S4). Spectra were recorded at 50 nm/min scanning speed and a data pitch of 0.2 nm. Four scans were averaged to obtain the final spectra. The contribution of the buffer was subtracted, and spectra were normalized by peptide bond intensity as described (76). Homology models were obtained using SWISS-MODEL (Swiss Institute of Bioinformatics) (77). Mutations were modeled on the WT structures using Swiss-PdbViewer, version 4.1.0 (Swiss Institute of Bioinformatics) (77). For each WT and variant structure, three 1-μs long trajectories were obtained. Trajectories were analyzed using the model described in (78) to obtain estimates of ΔΔG upon mutation at 298 K.

**Molecular dynamics simulations**

Full atom molecular dynamics simulations of the WT domains and their mutants (Table S3) in dodecahedral boxes (1 nm minimal distance between protein atoms and box edges) filled with Tip3p water molecules were performed at 410 K in protonation conditions mimicking pH 7.0 using Charmm27 + CMAP force field as described (76). Control temperatures were achieved using a Peltier thermostatic system. Changes in CD signal were fitted to a sigmoidal function using IGOR Pro (Wavemetrics) to estimate $T_m$.

**Data availability**

All data described in this work is contained within this article and the supporting information. Sourcing data are available on reasonable request to the corresponding author.

**Supporting information**—This article contains supporting information (79–86).

**Acknowledgments**—We thank Natalia Vicente for excellent technical support (through grant PEJ16/MED/TI-1593 from Consejería de Educación, Junta de Comunidades de Madrid and the European Social Fund, European Union). We thank the Spectroscopy and Nuclear Magnetic Resonance Core Unit at CNIO for access to CD instrumentation. We thank Andrea Thompson and Sharlene Day for critical reading of the article. Simon Bartlett (Centro Nacional de Investigaciones Cardiovasculares [CNIC]) provided English editing. We thank all members of the Molecular Mechanics of the Cardiovascular System team for helpful discussions.


**Funding and additional information**—J. A.-C. acknowledges funding from the Ministerio de Ciencia e Innovación (MCIN) through grants BIO2014-54768-P, BIO2017-83640-P/MINECO/AEI/FEDER, UE; EIN2019-102966 and RYC-2014-16604, the European Research Area Network on Cardiovascular Diseases (ERA-CVD/ISCIII, Spain MINOTAUR, AC16/0045), and the Comunidad de Madrid (consortium TecBio-CM, S2018/NMT-4443, FEDER). The CNIC is supported by the ISCIII, MCIN, and the Pro CNIC Foundation and was a Severo Ochoa Center of Excellence (SEV-2015-0505). We acknowledge funding from ISCIII to the Centro de Investigación Biomédica en Red (CB16/11/00425). L. S. acknowledges funding from MCIN (BFU2015-63571-P). J. S. acknowledges funding from AEI (PID2019-107293GB-I00), Gobierno de Aragón (E45_17R), and ERDF-Interreg V A POCTEFA, Spain (PIREPRED-EFA086/15). C. S.-C. is the recipient of an FPI-SO predoctoral fellowship BES-2016-076638. M. R. P. was the recipient of a PhD fellowship from the Italian Ministry of Education, Universities and Research, Italy. H. G.-C. is the recipient of an FPU16/04232 doctoral contract from MCIN.

**Conflict of interest**—L. M. holds share in Health in Code. All other authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: ACMG, American College of Medical Genetics; cDNA, complementary DNA; cMyBP-C, cardiac myosin-binding protein C; HCM, hypertrophic cardiomyopathy; MAF, minor allele frequency; VUS, variants of uncertain significance.

**References**


Bonne, G., Carrier, L., Bercovici, J., Cruaud, C., Richard, P., Hainque, B., Gautel, M., Labeit, S., James, M., Beckmann, J., Weissbach, J., and Vosberg,


EDITORS’ PICK: Molecular phenotyping of MYBPC3 variants

Carmen Suay-Corredera is a PhD student at the Molecular Mechanics of the Cardiovascular System group, led by Dr. Jorge Alegre-Cebollada at Centro Nacional de Investigaciones Cardiovasculares (Madrid, Spain). She studies the molecular mechanisms by which missense mutations in cardiac myosin-binding protein c (cMyBP-C) cause hypertrophic cardiomyopathy (HCM). By unravelling the pathomechanisms involved in the development of HCM, she aims at providing key insights for the design of new potential treatments and for the classification of genetic variants in the context of disease. https://www.cnic.es/en/investigacion/1/7246/miembros.

Maria Rosaria Picolo is a postdoctoral scientist in the Molecular Mechanics of the Cardiovascular System group at Centro Nacional de Investigaciones Cardiovasculares (Madrid). She obtained a PhD in Molecular Medicine and Medical Biotechnology from University of Naples—Federico II. Her PhD project dealt with examination of molecular phenotypes of cardiac myosin-binding protein c missense hypertrophic cardiomyopathy variants, particularly mRNA splicing defects. She is currently working on the development of novel tools to study pathomechanisms in titin mutations that cause dilated cardiomyopathy and to examine protein mechanics in living cells. https://www.cnic.es/en/investigacion/1/7246/miembros.