



Published in final edited form as:

Genet Med. 2021 July ; 23(7): 1281–1287. doi:10.1038/s41436-021-01134-9.

Computational prediction of protein subdomain stability in *MYBPC3* enables clinical risk stratification in hypertrophic cardiomyopathy and enhances variant interpretation

Andrea D. Thompson, MD, PhD¹, Adam S. Helms, MD, MS¹, Anamika Kannan¹, Jaime Yob², Neal K. Lakdawala, MD³, Samuel G. Wittekind, MD, MSc⁴, Alexandre C. Pereira, MD, PhD⁵, Daniel L. Jacoby, MD⁶, Steven D. Colan, MD⁷, Euan A. Ashley, MRCP, DPhil⁸, Sara Saberi, MD, MS¹, James S. Ware, PhD, MRCP⁹, Jodie Ingles, PhD¹⁰, Christopher Semsarian, MBBS, PhD, MPH¹¹, Michelle Michels, MD, PhD¹², Francesco Mazzarotto, PhD^{9,13,14}, Iacopo Olivetto, MD^{13,14}, Carolyn Y. Ho, MD³, Sharlene M. Day, MD², SHaRe investigators.

¹Cardiovascular Medicine, University of Michigan, Ann Arbor, MI ²Cardiovascular Medicine, University of Pennsylvania, Philadelphia, PA ³Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA ⁴Cincinnati Children's Hospital Medical Center, Heart Institute, Cincinnati, OH, USA ⁵Heart Institute (InCor), University of Sao Paulo Medical School, Sao Paulo, Brazil ⁶Cardiovascular Medicine, Yale University, New Haven, CT ⁷Department of Cardiology, Boston Children's Hospital, Boston, MA, USA ⁸Center for Inherited Heart Disease, Stanford University ⁹National Heart & Lung Institute & Royal Brompton Cardiovascular Research Centre, Imperial College London, London, United Kingdom ¹⁰Cardio Genomics Program at Centenary Institute, The University of Sydney, Sydney, Australia ¹¹Agnes Ginges Centre for Molecular Cardiology at Centenary Institute, The University of Sydney, Australia ¹²Department of Cardiology, Thoraxcenter, Erasmus MC Rotterdam, The Netherlands ¹³Cardiomyopathy Unit, Careggi University Hospital, Florence, Italy ¹⁴Department of Clinical and Experimental Medicine, University of Florence, Italy

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Authors: Andrea D. Thompson, MD/PhD Cardiovascular Division, University of Michigan, 1150 W. Medical Center Dr., Ann Arbor, MI 48109. adooley@med.umich.edu. Telephone: 734-763-7838, Fax: 734-615-3326., Sharlene M. Day, MD, Cardiovascular Division, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, PA 19104. sharlene.day@pennmedicine.upenn.edu.

Author Information

Conceptualization – A.D.T., S.M.D.; Data curation- A.S.H., N.K.L., S.G.W., A.C.P., D.L.J., S.D.C., E.A.A., S.S., J.S.W., J.I., C.S., M.M., F.M., I.O., C.Y.H.; Formal Analysis, Investigation, Methodology- A.D.T., A.K., J.Y., A.S.H.; Writing-original draft – A.D.T., S.M.D.; Writing- review & editing- A.S.H.; N.K.L., S.G.W., A.C.P., D.L.J., S.D.C., E.A.A., S.S., J.S.W., J.I., C.S., M.M., F.M., I.O., C.Y.H.

Ethics Declaration

Exported data from SHaRe was de-identified. This study complies with the Declaration of Helsinki, Institutional review board and ethics approval was obtained in accordance with policies applicable to each SHaRe site and informed consent was obtained from all participants as required. ShaRe sites are; Brigham and Women's Hospital (Boston, MA, USA), Boston Children's Hospital (Boston, MA, USA), Careggi University (Florence, Italy), Centenary Institute (Sydney, Australia), Children's Hospital of Philadelphia (Philadelphia, PA, USA), Cincinnati Children's Hospital (Cincinnati, Ohio, USA), Erasmus University (Rotterdam, Netherlands), Laboratory of Genetics and Molecular Cardiology (Sao Paulo, Brazil), Royal Brompton Hospital (London, UK), Royal Prince Alfred Hospital (Sydney, Australia), Stanford University (Palo Alto, CA, USA), University of Michigan (Ann Arbor, MI, USA), University of Pennsylvania (Philadelphia, PA, USA), University of Sydney (Sydney, Australia), Yale-New Haven Hospital (New Haven, Connecticut, USA), Akureyri Hospital Iceland (Akureyri, Iceland).

Abstract

Purpose: Variants in *MYBPC3* causing loss-of-function are the most common cause of HCM. However, a substantial number of patients carry missense variants of uncertain significance (VUS) in *MYBPC3*. We hypothesize that a structural-based algorithm, STRUM, which estimates the effect of missense variants on protein folding, will identify a subgroup of HCM patients with a *MYBPC3* VUS associated with increased clinical risk.

Methods: Among 7,963 patients in the multi-center Sarcomeric Human Cardiomyopathy Registry, 120 unique missense VUSs in *MYBPC3* were identified. Variants were evaluated for their effect on subdomain folding and a stratified time-to-event analysis for an overall composite endpoint (first occurrence of ventricular arrhythmia, heart failure, all-cause mortality, atrial fibrillation, and stroke) was performed for patients with HCM and a *MYBPC3* missense VUS.

Results: We demonstrated that patients carrying a *MYBPC3* VUS predicted to cause subdomain misfolding (STRUM +, $\Delta G = -1.2$ kcal/mol) exhibited a higher rate of adverse events compared to those with a STRUM- VUS (Hazard Ratio=2.29, P=0.0282). *In silico* saturation mutagenesis of *MYBPC3* identified 4,943/23,427 (21%) missense variants that were predicted to cause subdomain misfolding.

Conclusions: STRUM identifies patients with HCM and a *MYBPC3* VUS who may be at higher clinical risk and provides supportive evidence for pathogenicity.

Introduction:

Genetic variant interpretation is an ongoing challenge in clinical medicine, particularly when the gene of interest lacks robust functional assays^{1,2}. A variety of computational algorithms have been developed to predict variant pathogenicity, but their sensitivity and specificity are often poor, particularly when applied broadly across different diseases and different genes^{1,3}. Loss-of-function (LoF) pathogenic variants are common^{1,4,5}, resulting from either frameshift or nonsense variants creating a premature stop codon, splice errors, disruption of enzymatic activity, alteration of protein-protein interactions, or protein misfolding^{1,6,7}. Recognizing a common mechanism by which variants in a particular gene lead to LoF can inform the development of gene-specific computational algorithms to more accurately predict pathogenicity among variants that cannot be confidently classified based on clinical and family data alone^{6,7}.

Herein we focus on *MYBPC3* (encoding the protein, cardiac myosin binding protein C, or MyBP-C). Pathogenic variants in *MYBPC3* account for ~50% of patients with sarcomeric Hypertrophic cardiomyopathy (HCM)^{8,9}, and are inherited in an autosomal dominant fashion (OMIM 115197). Patients with HCM can experience a variety of adverse clinical outcomes, including outflow tract obstruction, arrhythmias, heart failure, and sudden cardiac death⁸. Genetic variants in *MYBPC3* consist of both truncating and non-truncating types. Rarely found in healthy populations, truncating *MYBPC3* variants result in a premature stop codon and cause HCM through complete LoF and haploinsufficiency at the transcript and protein level¹⁰⁻¹³. Thus, interpretation of these truncating variants as pathogenic is straightforward¹⁴.

However, the interpretation of missense variants within *MYBPC3* presents a major challenge. Single amino acid substitutions (missense variants) are found commonly in healthy populations. Further, since missense variants do not disrupt the reading frame, protein function may be tolerant to these minor sequence changes. Thus, many missense variants lack sufficient evidence to be classified as either pathogenic or benign and are classified as variants of uncertain significance (VUS)^{14,15}. While identifying pathogenic variants allows for predictive genetic testing in at-risk relatives¹⁶, a VUS is not clinically actionable and may lead to misinterpretation by clinicians and patients¹⁷.

Identification of a pathogenic sarcomere genetic variant for HCM also has important prognostic implications. Patients with HCM and a pathogenic sarcomere variant (sarcomeric HCM) have a higher risk of adverse clinical outcomes compared to those without a sarcomere gene variant (non-sarcomeric HCM)^{8,18}. Patients carrying a sarcomere gene VUS, on average, exhibit an intermediate risk of adverse events⁸, most likely because VUSs represent a mixed pool of pathogenic and benign variants that cannot be parsed on the basis of clinical and genetic data alone.

Because LoF is an established mechanism for pathogenic variants in *MYBPC3*, we hypothesized that applying a computational approach, called STRUM^{19–21}, that incorporates both sequence-based and structure-based algorithms to missense *MYBPC3* VUSs will identify those variants that result in protein subdomain misfolding (STRUM+), thereby supporting pathogenicity and improving variant interpretation. We further predict that this approach will identify a subpopulation of patients with HCM and a STRUM+ *MYBPC3* missense VUS who are at risk for adverse clinical outcomes, at a frequency similar to patients with HCM carrying known pathogenic variants.

Methods and Materials:

SHaRe Registry Data Extraction and *MYBPC3* Variant Classification

The generation of the centralized SHaRe database has been previously described⁸. Data were exported from quarter 1 of 2019. Inclusion criteria included a site-designated diagnosis of HCM using standard diagnostic criteria⁸. SHaRe non-truncating *MYBPC3* missense variants (Tables S1,S2) were classified as previously reported¹⁴ in accordance with American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) joint guidelines, leveraging available clinical and experimental data.^{3,8,9,14,22,23} Known splice variants are classified as truncating. Since variants in *MYBPC3* present in gnomAD with allele frequencies of $> 4E-05$ and absent in SHaRe are unlikely to be independently pathogenic for HCM, these variants were included in our list of benign *MYBPC3* variants¹⁴. More details regarding variant interpretation is provided within supplemental materials.

It has previously been shown that patients carrying pathogenic non-truncating variants exhibit similar clinical outcomes to those carrying truncating *MYBPC3* variants¹⁴. Thus, a reference population including previously adjudicated truncating and non-truncating *MYBPC3* pathogenic/likely pathogenic (pathogenic) variants [*MYBPC3*-path-all] was used.

A second reference population included patients with HCM who underwent genetic testing and were negative for sarcomere variants Sarc-⁸.

Computational Structural and Protein Folding Stability Predictive Modeling

MyBP-C is made up of immunoglobulin and fibronectin subdomains (C0-C10) [NM_000256.3, NP_000247.2]. For *MYBPC3* missense variants we utilized STRUM to calculate the effect of the missense variant on the Gibbs free energy of local subdomain folding (ΔG)¹⁹ (Table S3). A negative ΔG value indicates the degree of reduced folding energy (kcal/mol) relative to the wild-type subdomain, or folding destabilization¹⁹. Previous experimental validation of this algorithm compared STRUM predictions to 3,421 experimentally tested variants from 150 proteins and demonstrated a Pearson's correlation coefficient of 0.79 and root mean square error of prediction of 1.2 kcal/mol¹⁹. Thus, a value of $\Delta G = -1.2$ kcal/mol was defined as the cut-off for destabilizing (deleterious) variants. Further details regarding STRUM analysis and structural models are provided within the Supplemental Materials (Figure S1–S3, Table S3).

Computational Sequence Based Variant Analysis (PolyPhen-2, SIFT, CardioBoost)

We compared the STRUM prediction for *MYBPC3* missense variants with a sequence-based algorithm embedded in STRUM (SIFT)^{24,25}. We also analyzed these variants with PolyPhen-2 (HumVar database), another sequence based algorithm²⁶. Finally, we compared our result to those obtained using CardioBoost which is a disease specific machine learning classifier to predict pathogenicity of rare missense variants in genes associated with cardiomyopathies and arrhythmias⁶. CardioBoost relies on minor allele frequency, whereas STRUM does not.

Clinical Outcomes Analysis

Only patients with HCM carrying a single *MYBPC3* missense VUS were included in clinical outcomes analyses to avoid confounding from cases with multiple gene variants²⁷. Comparisons using time-to-event analysis were made between variants predicted to be deleterious (STRUM +, $\Delta G = -1.2$ kcal/mol) and those predicted to be non-deleterious. The primary outcome was an overall composite previously defined as the first occurrence of any component of the ventricular arrhythmia composite, heart failure composite (without inclusions of LV ejection fraction), all-cause mortality, atrial fibrillation (AF), or stroke⁸. Results were compared to reference populations *MYBPC3*-path-all and Sarc-. A secondary analysis of a heart failure composite, ventricular arrhythmia composite, and atrial fibrillation was also performed. Finally, a secondary analysis using alternative computational algorithms (SIFT, Polyphen-2, CardioBoost) was performed. Composite outcomes are defined in more detail in the supplemental materials.

Statistical Analysis

Data presented as mean \pm standard deviation were analyzed by t-test for two groups or ANOVA for >2 groups with Tukey's post hoc test for multiple comparisons. Data presented as frequency were analyzed by a chi-square test. Odds ratio (with 95% Confidence Interval), specificity, and sensitivity were calculated to evaluate the association between computational

prediction algorithms and known pathogenic/likely pathogenic (pathogenic) or benign/likely benign (benign) variants (further details provided in supplemental materials). Primary and secondary clinical outcomes were analyzed by the Kaplan-Meier method from time of birth. Analysis from time of birth is appropriate given that the genetic variant is present from birth and variability in time to, and reason for, clinical presentation could confound the results if time from diagnosis were used. Patients who did not have the outcome of interest were censored at the time of their last recorded follow-up in SHaRe. Comparison between curves was performed using Log-rank Mantel-Cox test with p-values of < 0.05 considered statistically significant. Median event free survival and hazard ratio (mantel-Haenszel) are also reported. Statistical analyses were performed using GraphPad Prism software (San Diego, CA).

Results

Patients with HCM and a *MYBPC3* missense VUS predicted to disrupt subdomain folding (STRUM+) exhibit a higher incidence of adverse clinical outcomes

We began by evaluating all *MYBPC3* missense VUS within SHaRe using STRUM. *MYBPC3* VUSs exhibited a mean ΔG of -0.73 ± 1.06 kcal/mol (Figure S4). Of 120 unique *MYBPC3* missense VUSs, 34 (28%) were predicted to cause subdomain misfolding with ΔG values < -1.2 kcal/mol (deleterious) (Table S2). Next, we evaluated clinical characteristics and outcomes in patients with HCM and a single missense *MYBPC3* VUS predicted to disrupt subdomain folding (STRUM+) compared to patients carrying a VUS not predicted to disrupt subdomain folding (STRUM-). For this analysis, we included only patients who carried a single VUS within *MYBPC3*, and excluded patients who carried a second pathogenic variant or variant of uncertain significance ($N = 105$). Patients with a STRUM+ vs STRUM- *MYBPC3* VUS exhibited similar clinical characteristics including BMI, gender, ancestry, age at diagnosis, wall thickness, ejection fraction, left ventricular outflow tract obstruction (Table 1). We observed that patients carrying a STRUM+ VUS experienced higher rates of adverse events compared to patients carrying a STRUM- VUS (Figure 1, hazard ratio 2.3, $p = 0.03$). Furthermore, patients carrying a STRUM+ VUS exhibited a similar rate of adverse clinical events compared to patients carrying a pathogenic variant (*MYBPC3*-Path-all). Conversely, patients carrying STRUM- VUSs exhibited a lower frequency of outcomes, similar to Sarc- patients (Figure 2). There were no statistically significant differences between groups for the individual component outcomes, including ventricular arrhythmias, heart failure, or atrial fibrillation (Figure S5).

STRUM exhibits improved specificity over established sequence-based prediction algorithms and improved sensitivity when combined with CardioBoost.

To determine the sensitivity and specificity of STRUM to differentiate pathogenic from benign variants within *MYBPC3* we performed STRUM analysis on all known pathogenic missense variants within ShaRe ($n = 19$) and known missense benign variants within SHaRe and gnomAD ($n = 110$, Table S1, Figure 3A). These variants were present in 412 patients with HCM within the SHaRe registry. *MYBPC3* benign variants exhibited a mean ΔG of -0.31 ± 0.60 kcal/mol which was significantly higher than *MYBPC3* VUS (ΔG of -0.73 ± 1.06 kcal/mol, $p = 0.005$) (Figure S4) and *MYBPC3* pathogenic variants (mean ΔG of

-1.00 +/- 1.08 kcal/mol, $p = 0.016$) (Figure 3A). We found that variants predicted to be deleterious by STRUM were more likely to be pathogenic variants (OR 5.9, 95% CI 1.8–19.6) (Figure 3C). Only 9 additional unique non-truncating *MYBPC3* variants were designated as pathogenic and/or likely pathogenic within ClinVar. However, all of these variants had a single submission and a review status of 0–1/4 criteria provided. By modern standards, these variants would be reclassified as VUSs and were therefore not included in our analysis.

Algorithms that were purely sequence-based achieved greater sensitivity but performed inferiorly to STRUM in regard to specificity. STRUM exhibited a 93% specificity for benign variants and PolyPhen-2 and SIFT exhibited a specificity of 62% (OR 4.5, 95% CI 1.5–13.5) and 54% (OR 1.3, 95% CI 0.5–3.4) respectively (Figure 3C, Figure S6). Additionally, variant interpretation by SIFT or PolyPhen-2 did not stratify patients carrying a *MYBPC3* VUS for clinical adverse outcomes (Figure S6).

In comparison, CardioBoost, demonstrated a specificity of 98% (OR 42.3, CI 8.0–223.6) (Figure 3, Table S1). For pathogenic variants, CardioBoost demonstrated a sensitivity of 47%. Interestingly, there was limited overlap among known pathogenic variants predicted to be deleterious by STRUM and those predicted to be deleterious by CardioBoost, making the 2 algorithms complementary (Table S1). Combining these algorithms to classify any variant predicted to be deleterious by CardioBoost or STRUM as pathogenic, maintained a high specificity of 93% and improved sensitivity to 63% (Figure 3C).

When examining patients with HCM and a *MYBPC3* missense VUS, STRUM identified a larger number of *MYBPC3* VUSs as deleterious. Only sixteen of thirty-nine (41%) patients with a STRUM+ *MYBPC3* VUS were also identified as CardioBoost+. Just 3 additional patients were uniquely identified as CardioBoost+ (Table S2). While there is a trend toward a higher rate of adverse clinical events in patients with HCM and a CardioBoost+ *MYBPC3* VUS, this difference was not statistically significant (Figure 3D).

STRUM predictions within pathogenic variants are consistent with experimental modeling

Prior experimental characterization of *MYBPC3* pathogenic missense variants within the C10 domain, Leu1238Pro and Asn1257Lys, demonstrated that these variants failed to localize to the sarcomere and were rapidly degraded within primary cardiomyocytes¹⁴. Consistent with these experimental findings, pathogenic C10 domain variants are uniformly predicted to destabilize protein folding (ΔG of -2.89 and -1.45 kcal/mol respectively) (Figure 4).

Conversely, of the pathogenic *MYBPC3* variants not predicted to be deleterious by STRUM (Figure 3), a large number were localized within the C3 domain (Figure 3A, open circles; 7/13) and exhibited a mean ΔG -0.32 kcal/mol, (range -0.93 to 0.04). A large number of known pathogenic variants cluster within the C3 domain near a surface-exposed flexible linker (Figure 4)¹⁵. Thus, these variants would be predicted to alter electrostatic protein-protein interactions but would not be expected to disrupt subdomain folding. This result is consistent with prior experimental and structural characterization data of these C3 pathogenic variants. Arg495Gln, Arg502Trp, and Phe503Leu incorporate normally into the

sarcomere and have protein 1/2 lives that are indistinguishable from wild-type MyBP-C within primary cardiomyocytes¹⁴. Further, the NMR structure of the *MYBPC3* Arg502Trp C3 domain reveals preserved subdomain folding²⁸.

While C3 and C10 pathogenic variants have a narrow range of ΔG values, ΔG predictions for C6 pathogenic variants vary from -2.33 to 0.04 (mean $\Delta G -1.11$). We previously examined two C6 domain variants, Arg810His and Trp792Arg, and found that they incorporate normally into the sarcomere and exhibit normal protein 1/2 lives in primary cardiomyocytes¹⁴. However, both of these variants were predicted to destabilize subdomain folding by STRUM, exhibiting values near the cutoff: Arg810His ($\Delta G -1.22$ kcal/mol), Trp792Arg ($\Delta G -1.28$ kcal/mol). They are also predicted to be pathogenic by CardioBoost (Table S2). These observations suggest that a subset of pathogenic variants mildly disrupt subdomain folding without causing complete destabilization of MyBP-C. Subdomain destabilization in these cases could interfere with protein-protein interactions or MyBP-C conformational dynamics.

***In silico* saturation mutagenesis of *MYBPC3* identified 4,943 missense variants predicted to cause subdomain misfolding**

Only a subset of amino acid substitutions has been observed in patients with HCM and are cataloged in publicly available databases, such as Clinvar. However, previously unreported variants frequently arise in probands with HCM who undergo clinical genetic testing²⁹. Thus, we performed STRUM on all possible *MYBPC3* single amino acid substitutions (*in silico* mutagenesis) to develop a compendium of STRUM+ variants that may be useful for the research and clinical community. We found that 4,943 of 24,665 (20%) amino acid substitutions were predicted to disrupt subdomain folding (Figure S6, Table S4–S5).

Discussion

Clinical risk stratification has been a cornerstone of clinical HCM management. It is well-established that patients with sarcomeric HCM have a higher rate of adverse clinical outcomes compared to non-sarcomeric HCM, enabling the incorporation of genetic data into clinical risk stratification in HCM^{8,18}. Yet, refinement of clinical risk for patients with a VUS remains an ongoing challenge for clinicians^{1,5}. We have identified a subpopulation of patients with a *MYBPC3* missense VUS that are predicted to disrupt subdomain protein folding (STRUM+) who exhibit clinical outcomes indistinguishable from patients with a pathogenic *MYBPC3* variant. Conversely, patients carrying a *MYBPC3* VUS not predicted to affect subdomain folding (STRUM-), exhibit a lower prevalence of adverse clinical outcomes similar to patients with non-sarcomeric HCM. Although the methodology of parsing these variants is different for *MYBPC3* because of differing underlying mechanisms, these findings are analogous to a recent study in *MYH7* in which patients with HCM carrying VUSs that were located within the interacting-heads motif had a higher rate of adverse clinical outcomes compared to patients carrying VUSs that were outside of this motif³⁰. These studies together suggest that VUSs in sarcomere genes are primarily an admixture of pathogenic and benign variants. So, while patients with HCM carrying sarcomere gene VUSs as a whole exhibit a prevalence of clinical outcomes that are

intermediate between patients with or without pathogenic sarcomere variants⁸, a computational approach specifically leveraging the pathogenic mechanism of *MYBPC3* has enabled the identification of higher risk subpopulation that exhibit clinical outcomes similar to sarcomeric HCM and a lower risk subpopulation that exhibit clinical outcomes similar to non-sarcomeric HCM.

While computational prediction should not be exclusively relied on to assign pathogenicity of a variant or risk stratify an individual patient, STRUM could be incorporated in an additive manner with other methods for variant adjudication to prioritize variants that warrant further investigation. Given that novel *MYBPC3* variants are frequently identified by genetic testing of probands with HCM²⁹, we completed an *in silico* “saturation mutagenesis” of *MYBPC3* compiling a complete list of STRUM+ variants. Excluding known pathogenic or benign variants, we estimate that ~ 0.097% (1/1033) individuals within gnomAD carry a *MYBPC3* variant predicted to cause subdomain misfolding by STRUM. STRUM+ *MYBPC3* VUSs identified in patients with HCM should be prioritized for additional clinical and experimental investigation. Specifically, functional experimental studies to evaluate the direct effects of *MYBPC3* VUSs on protein stability, folding, and localization, as we have done previously for a subset of pathogenic variants¹⁴, will be important. Familial co-segregation analysis on patients carrying a *MYBPC3* STRUM + VUS would add complementary information to these types of experimental studies.

When known benign missense variants were evaluated by STRUM, 102 out of 110 variants were correctly predicted, with an overall specificity of 93%. However, for known pathogenic variants, only 7 of 19 were predicted to alter subdomain folding by STRUM, yielding a sensitivity of 32%. This, lower sensitivity was in large part explained by a known cluster of pathogenic variants within C3¹⁵. None of the 7 known pathogenic variants in C3 had a ΔG value below the threshold of -1.2 kcal/mol. This is consistent with experimental data that demonstrates C3 variants localize normally to the sarcomere and exhibit protein 1/2 lives similar to wild-type MyBP-C. Additionally, an NMR structure of Arg502Trp demonstrates that this variant does not disrupt subdomain folding but rather is more likely to alter protein-protein interactions^{14,28}. In contrast, MyBP-C pathogenic variants in C10, predicted by STRUM to cause subdomain misfolding, fail to localize to the sarcomere and are rapidly degraded¹⁴. These experimental results support the accuracy of STRUM predictions for subdomain misfolding. Further, they highlight that STRUM is only predictive of pathogenicity for variants that significantly alter protein folding as their primary mechanism. Thus, a ΔG value of > -1.2 kcal/mol does not exclude pathogenicity for variants that cause loss or gain-of-function through an alternate mechanism such as alternative splicing or altered protein-protein interactions. STRUM is best applied to VUSs after other clinical, computational, and experimental criteria for variant adjudication have been implemented. For example, *MYBPC3* pathogenic variants which lead to LoF by mechanisms other than subdomain misfolding have previously been well characterized and defined as pathogenic, including splice variants^{14,22,23} and the cluster of pathogenic variants within C3 (aa.485–503)^{15,28,31} discussed above.

STRUM performed superiorly to sequence based-algorithms alone, such as SIFT and PolyPhen-2, which each had lower specificity and were unable to clinically risk stratify

patients with HCM and a *MYBPC3* missense VUS. Compared to using each method independently, combining STRUM and CardioBoost improved sensitivity for identifying known pathogenic variants to 63% while maintaining a specificity for known benign variants of 93%. CardioBoost supported pathogenicity for 3 missense VUSs that were STRUM-, but only predicted pathogenicity for 16/39 of *MYBPC3* STRUM+ VUSs. This result highlights the added utility of STRUM to identify a subset of VUSs within *MYBPC3* that result in local subdomain misfolding leading to allelic LoF and have a high probability of being pathogenic. Because CardioBoost and STRUM are complementary and have high specificity, we would propose that the ACMG/AMP PP3 criteria, where multiple lines of computational evidence support a deleterious effect of a variant, could be applied when one or both algorithms predict pathogenicity. Conversely, because of relatively limited sensitivity for each algorithm independently, we would propose that the BP4 criteria, where multiple lines of computational evidence support no impact of the variant, be applied only if both algorithms predict that a variant is non-pathogenic.

Although this study was limited by a moderate sample size of 105 patients with HCM, the comprehensive variant adjudication in SHaRe enabled strict inclusion of patients carrying a single VUS within *MYBPC3* to clearly discriminate genetic-clinical correlates in this population. This approach enabled us to discern a difference in a composite of adverse clinical outcomes between patients with STRUM+ and STRUM- variants. However, the sample size was insufficient for detecting differences in individual outcomes, such as arrhythmias or heart failure and did not provide sufficient power to correct for other risk predictors.

The approach of using STRUM as an adjunctive tool for decision making may also be applicable to other genes for which LoF is a pathogenic mechanism. Approximately 50% of disease associated variants within Human Gene Mutation Database are truncating variants predicted to result in LoF¹¹. These genes, like *MYBPC3*, also have missense VUSs that may be evaluated for protein misfolding using STRUM. For example, there are several causal genes for hypertrophic, dilated and arrhythmogenic cardiomyopathies with truncating pathogenic variants, including lamin A/C, desmoplakin, and plakophilin 2, Titin, and phospholamban^{11,32}. This approach is best suited for non-enzymatic proteins where high-quality structural modeling can be performed, and for which the primary pathogenic mechanism has been established to be LoF.

Conclusions

We show that the computational algorithm STRUM, that predicts protein structure stability in response to missense variation, enables identification of patients carrying a *MYBPC3* VUS who may be at higher clinical risk of adverse events. This approach also provides supportive evidence for pathogenicity, prioritizing variants for functional experimental studies and clinical familial segregation to improve *MYBPC3* variant adjudication. Finally, STRUM may be broadly applicable to variants in other genes for which LoF is an established mechanism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding for SHaRe has been provided by an unrestricted research grant by Myokardia, Inc a startup company that is developing therapeutics that target the sarcomere. Myokardia, Inc had no role in the preparation of this manuscript or approving the content of this manuscript. The following individuals are supported by the NHLBI at the NIH; Thompson [T32 HL007853], Helms [K08HL130455], Day [R01 11572784], Ho [1P50HL112349] & [1U01HL117006]. Dr. Thompson is supported by the PFDI and M-BoCA at the University of Michigan. Dr Ware is supported by the Wellcome Trust [107469/Z/15/Z], the Medical Research Council (United Kingdom), the British Heart Foundation, NIHR, Royal Brompton Cardiovascular Biomedical Research Unit, and the NIHR Imperial College Biomedical Research Centre. Dr Ingles is a recipient of a National Health and Medical Research Council. Dr Semsarian is the recipient of a NHMRC Practitioner Fellowship [#1154992]. Dr Olivotto is supported by the Italian Ministry of Health [RF-2013-02356787] and [NET-2011-02347173] and by the Tuscany Registry of Sudden Cardiac Death (ToRSADE) project [FAS-Salute 2014, Regione Toscana].

Disclosures:

Funding for SHaRe has been provided by an unrestricted research grant by Myokardia, Inc a startup company that is developing therapeutics that target the sarcomere. Myokardia, Inc had no role in the preparation of this manuscript or approving the content of this manuscript. Drs. Helms, Ho, Day, Saberi, Olivotto, Colan, Ingles and Ashley receive research support from MyoKardia, Inc. Dr. Thompson receives compensation as an editor for Merck Manuals. Research funding for all authors is detailed within the acknowledgement sections of this manuscript. The other authors report no relevant conflicts of interest.

Data availability

De-identified data will be made available by request to the authors.

References

1. Yi S, Lin S, Li Y, Zhao W, Mills GB, Sahni N. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat Rev Genet.* 2017;18(7):395–410. [PubMed: 28344341]
2. Starita LM, Ahituv N, Dunham MJ, et al. Variant Interpretation: Functional Assays to the Rescue. *Am J Hum Genet.* 2017;101(3):315–325. [PubMed: 28886340]
3. Richards S, Aziz N, Bale 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.* 2015;17(5):405–424. [PubMed: 25741868]
4. Balasubramanian S, Fu Y, Pawashe M, et al. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun.* 2017;8(1):382. [PubMed: 28851873]
5. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science.* 2012;335(6070):823–828. [PubMed: 22344438]
6. Zhang X; Walsh R; Whiffin N; Buchan R; Midwinter W, WAGR Li N.; Ahmad M; Mazzarotto F; Roberts A; Theotokis PI; Mazaika E; Allouba M; de Marvao A; Pua CJ; Day SM; Ashley E, Colan SD, Michels M; Pereira AC; Jacoby D; Ho CY; Olivotto I; Gunnarsson GT; Jefferies JL; Semsarian C; Ingles J; O'Regan DP; Aguib Y; Yacoub MH; Cook SA; Barton PJR; Bottolo L; Ware JS Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *Genetics in Medicine* 2020;Published online: 13 Oct 2020
7. Whiffin N, Walsh R, Govind R, et al. CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation. *Genet Med.* 2018;20(10):1246–1254. [PubMed: 29369293]

8. Ho CY, Day SM, Ashley EA, et al. Genotype and Lifetime Burden of Disease in Hypertrophic Cardiomyopathy: Insights from the Sarcomeric Human Cardiomyopathy Registry (SHaRe). *Circulation*. 2018;138(14):1387–1398. [PubMed: 30297972]
9. Carrier L, Mearini G, Stathopoulou K, Cuello F. Cardiac myosin-binding protein C (MYBPC3) in cardiac pathophysiology. *Gene*. 2015;573(2):188–197. [PubMed: 26358504]
10. Marston S, Copeland O, Jacques A, et al. Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency. *Circ Res*. 2009;105(3):219–222. [PubMed: 19574547]
11. Glazier AA, Thompson A, Day SM. Allelic imbalance and haploinsufficiency in MYBPC3-linked hypertrophic cardiomyopathy. *Pflugers Archiv : European journal of physiology*. 2019;471(5):781–793. [PubMed: 30456444]
12. O’Leary TS, Snyder J, Sadayappan S, Day SM, Previs MJ. MYBPC3 truncation mutations enhance actomyosin contractile mechanics in human hypertrophic cardiomyopathy. *Journal of molecular and cellular cardiology*. 2019;127:165–173. [PubMed: 30550750]
13. Monteiro da Rocha A, Guerrero-Serna G, Helms A, et al. Deficient cMyBP-C protein expression during cardiomyocyte differentiation underlies human hypertrophic cardiomyopathy cellular phenotypes in disease specific human ES cell derived cardiomyocytes. *J Mol Cell Cardiol*. 2016;99:197–206. [PubMed: 27620334]
14. Helms AS, Thompson AD, Glazier AA, et al. Spatial and Functional Distribution of MYBPC3 Pathogenic Variants and Clinical Outcomes in Patients With Hypertrophic Cardiomyopathy. *Circ Genom Precis Med*. 2020;13(5):396–405. [PubMed: 32841044]
15. Walsh R, Mazzarotto F, Whiffin N, et al. Quantitative approaches to variant classification increase the yield and precision of genetic testing in Mendelian diseases: the case of hypertrophic cardiomyopathy. *Genome Med*. 2019;11(1):5. [PubMed: 30696458]
16. Gersh BJ, Maron BJ, Bonow RO, et al. 2011 ACCF/AHA Guideline for the Diagnosis and Treatment of Hypertrophic Cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Developed in collaboration with the American Association for Thoracic Surgery, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2011;58(25):e212–260. [PubMed: 22075469]
17. Kelly MA, Caleshu C, Morales 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*. 2018;20(3):351–359. [PubMed: 29300372]
18. Ko C, Arcsott P, Concannon M, et al. Genetic testing impacts the utility of prospective familial screening in hypertrophic cardiomyopathy through identification of a nonfamilial subgroup. *Genet Med*. 2018;20(1):69–75. [PubMed: 28640247]
19. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 2016;32(19):2936–2946. [PubMed: 27318206]
20. Kurolap A, Eshach-Adiv O, Gonzaga-Jauregui C, et al. Establishing the role of PLVAP in protein-losing enteropathy: a homozygous missense variant leads to an attenuated phenotype. *J Med Genet*. 2018;55(11):779–784. [PubMed: 29875123]
21. Amir M, Ahmad S, Ahamad S, et al. Impact of Gln94Glu mutation on the structure and function of protection of telomere 1, a cause of cutaneous familial melanoma. *J Biomol Struct Dyn*. 2020;38(5):1514–1524. [PubMed: 31014199]
22. Ito K, Patel PN, Gorham JM, et al. Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. *Proc Natl Acad Sci U S A*. 2017;114(29):7689–7694. [PubMed: 28679633]
23. Singer ES, Ingles J, Semsarian C, Bagnall RD. Key Value of RNA Analysis of MYBPC3 Splice-Site Variants in Hypertrophic Cardiomyopathy. *Circ Genom Precis Med*. 2019;12(1):e002368. [PubMed: 30645170]

24. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012;40(Web Server issue):W452–457. [PubMed: 22689647]
25. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073–1081. [PubMed: 19561590]
26. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7:Unit7 20.
27. Van Driest SL, Vasile VC, Ommen SR, et al. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol.* 2004;44(9):1903–1910. [PubMed: 15519027]
28. Zhang XL, De S, McIntosh LP, Paetzel M. Structural characterization of the C3 domain of cardiac myosin binding protein C and its hypertrophic cardiomyopathy-related R502W mutant. *Biochemistry.* 2014;53(32):5332–5342. [PubMed: 25058872]
29. Alfares AA, Kelly MA, McDermott G, et al. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet Med.* 2015;17(11):880–888. [PubMed: 25611685]
30. Toepfer CN, Garfinkel AC, Venturini G, et al. Myosin Sequestration Regulates Sarcomere Function, Cardiomyocyte Energetics, and Metabolism, Informing the Pathogenesis of Hypertrophic Cardiomyopathy. *Circulation.* 2020;141(10):828–842. [PubMed: 31983222]
31. Cohn R, Thakar K, Lowe A, et al. A Contraction Stress Model of Hypertrophic Cardiomyopathy due to Sarcomere Mutations. *Stem Cell Reports.* 2019;12(1):71–83. [PubMed: 30554920]
32. Ho CY, Charron P, Richard P, Girolami F, Van Spaendonck-Zwarts KY, Pinto Y. Genetic advances in sarcomeric cardiomyopathies: state of the art. *Cardiovasc Res.* 2015;105(4):397–408. [PubMed: 25634555]
33. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015;12(1):7–8. [PubMed: 25549265]
34. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 2015;43(W1):W174–181. [PubMed: 25883148]
35. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725–738. [PubMed: 20360767]

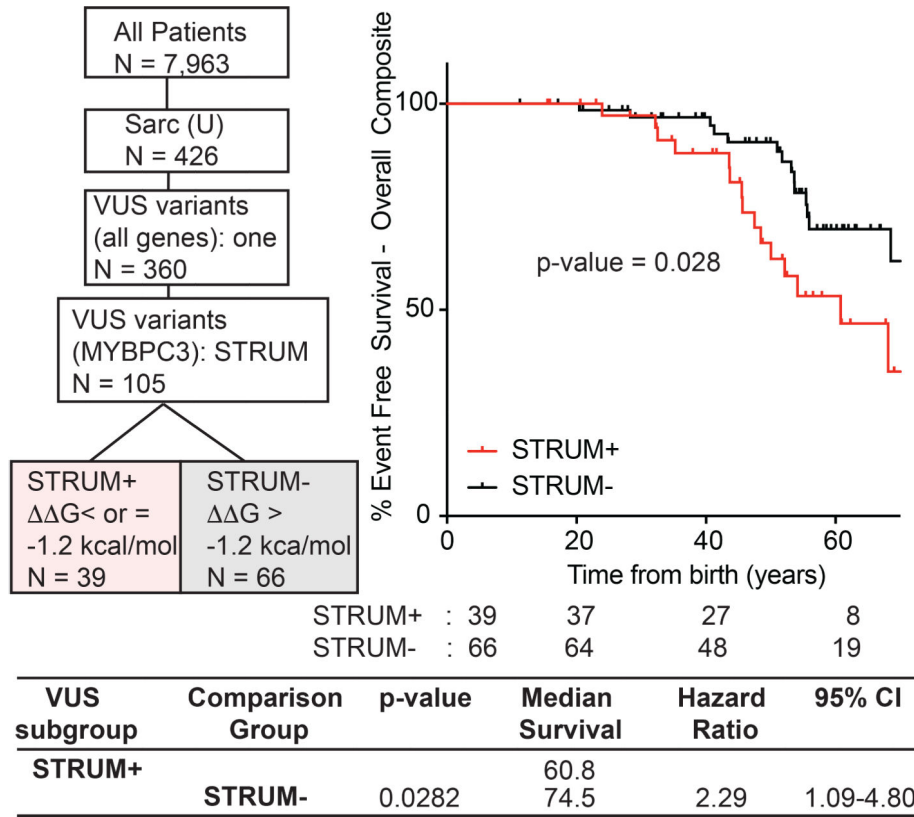


Figure 1: Patients with a *MYBPC3* VUS identified as deleterious by STRUM (STRUM+) are associated with an increased risk for adverse HCM-related outcomes. Selection within ShaRe of patients with HCM carrying a single *MYBPC3* missense VUS is shown on the left. 105 patients carry a single *MYBPC3* missense VUS, covering 77 distinct *MYBPC3* VUSs. Kaplan Meier curves, median event free survival (years), and hazard ratio with corresponding 95% CI reveal that patients carrying a STRUM+ *MYBPC3* VUS (red) exhibited a higher rate of adverse HCM-related outcomes (Overall Composite) compared to patients carrying a STRUM- variant (black).

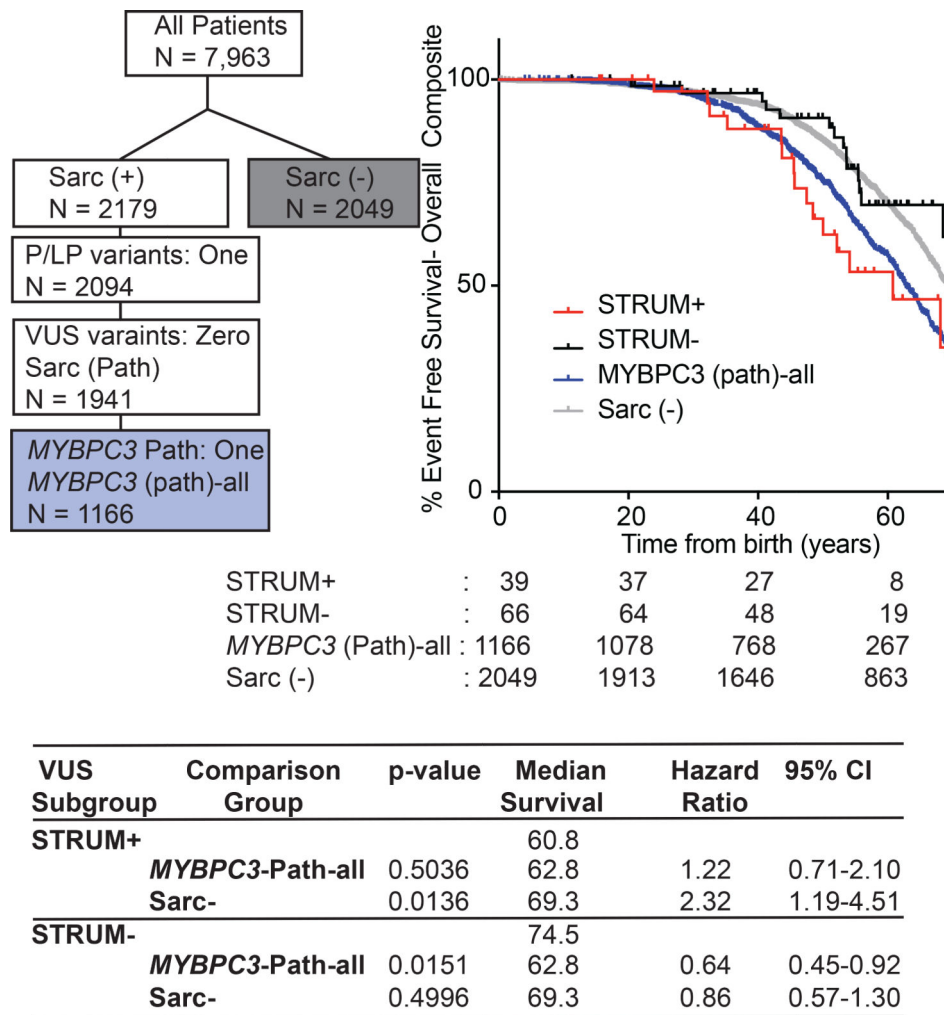


Figure 2: Patients with a *MYBPC3* VUS identified as deleterious by STRUM (STRUM +) exhibit clinical outcomes similar to patients with a *MYBPC3* pathogenic variants.

Selection within SHaRe of patients with HCM and a single *MYBPC3* pathogenic variant (*MYBPC3*-Path-all) and patients with HCM without a sarcomere gene variant after clinical genotype analysis (Sarc -) is shown on the left. Kaplan Meier curves, median event free survival (median survival), and hazard ratio with corresponding 95% confidence interval (CI) reveal patients carrying a STRUM+ *MYBPC3* VUS (red) exhibited overall composite outcomes similar to *MYBPC3*-Path-all patients (blue, p-value 0.5036). Whereas, patients carrying a STRUM- variant (black) exhibited a lower rate of adverse HCM-related outcomes (overall Composite) similar to Sarc- patients (grey).

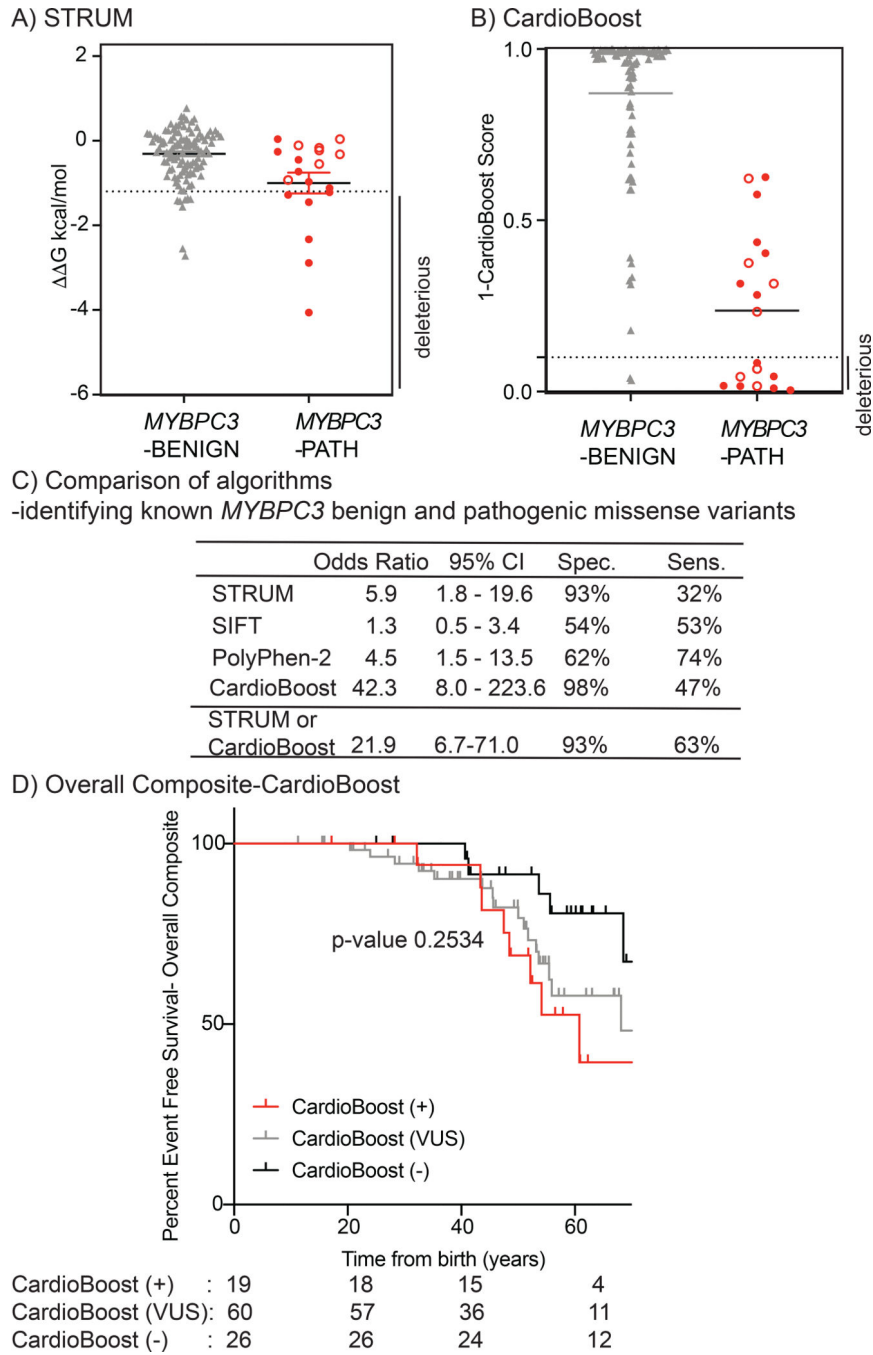


Figure 3. STRUM is complementary to CardioBoost.

Results of computational analysis for each unique *MYBPC3*-Benign (grey triangles, n= 110) and *MYBPC3*-Path (red circles, n =19) variant are shown for (A) STRUM and (B) CardioBoost. Mean and SEM for each group depicted. The cut-off for deleterious variants for STRUM was $\Delta\Delta G < -1.2$ kcal/mol. The cut-off for deleterious variants for CardioBoost (CardioBoost+) was a probability score > 0.90 this is graphed as $1-\text{CardioBoostScore} < 0.10$. C3 pathogenic variants are depicted in open circles in panel A and B. (C) Statistical Analysis of computational method utilized herein STRUM (Figure 3), CardioBoost (Figure 3), SIFT

(Figure S6), and Polyphen-2 (Figure S6) is shown including Odds ratio, 95% confidence interval (CI), Sensitivity, and Specificity. (D) Using the same patient selection criteria in SHaRe detailed in Figure 1, patients with HCM and a MYBPC3 missense VUS were analyzed by CardioBoost. CardioBoost (+) was a probability score > 0.90 , CardioBoost VUS 0.10 and 0.90 , and CardioBoost (-) < 0.10 . Of the 105 patients analyzed by STRUM 19 were CardioBoost(+). Kaplan Meier curves reveal that patients carrying a CardioBoost (+) *MYBPC3* VUS (red) exhibited higher rates of adverse HCM-related outcomes (Overall Composite), than patients carrying a CardioBoost (-) *MYBPC3* VUS (black), however the null hypothesis could not be excluded, p-value 0.0945. This remains true when comparing patients carrying a *MYBPC3* VUS that is CardioBoost (+) (red), CardioBoost (VUS) (grey), and CardioBoost (-) (black) (p-value 0.2534).

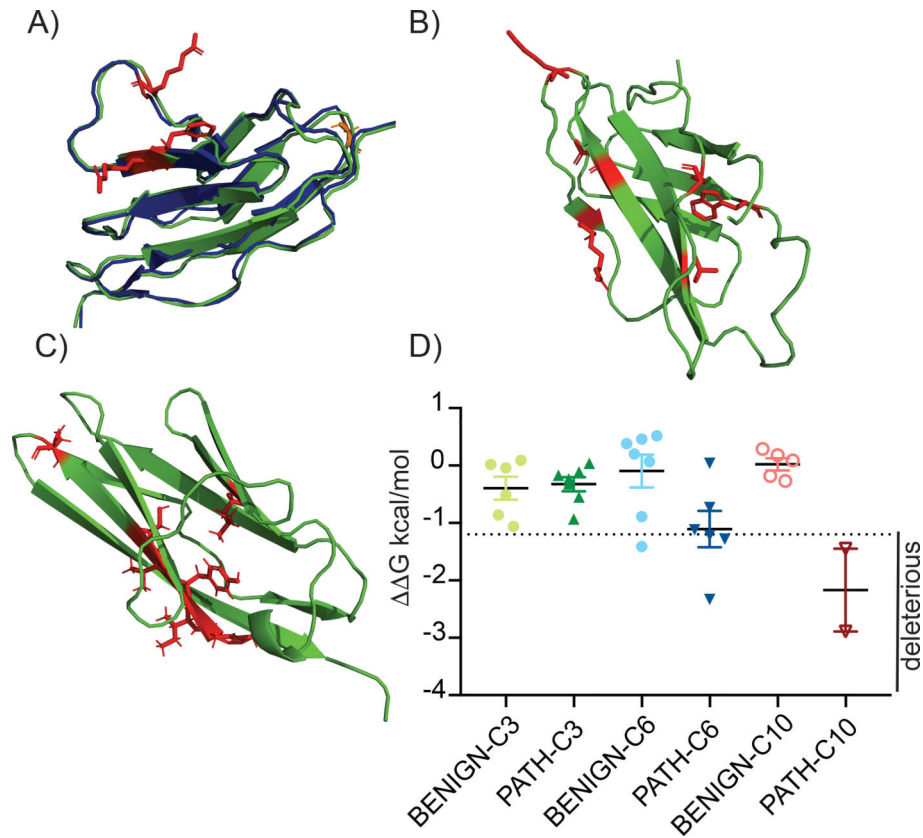


Figure 4. Structural Analysis of Pathogenic Missense *MYBPC3* variants. MyBP-C (the protein encoded by *MYBPC3*) domains C3, C6, and C10 were structurally modeled using I-TASSER³³⁻³⁵ (PyMOL, cartoon, green) Wild-type residues that are affected by missense pathogenic variants are depicted in red (PyMOL, sticks). (A) For the C3 domain, the I-TASSER model is aligned with an available NMR structure [2mq0.pdb²⁸, blue, PyMOL cartoon]. Pathogenic variants within C3 largely cluster in a surface exposed region. (B) C6 domain and (C) C10 domain pathogenic variants do not cluster within a specific region of the domain. (D) Results of STRUM¹⁹ analysis for *MYBPC3* pathogenic and benign variants within C3, C6, and C10 are shown, with mean and SEM for each group depicted. Graph is labeled to indicate variants predicted to be deleterious.

Table 1:Demographics of patients with HCM and single *MYBPC3* VUS

	STRUM (+) n = 39 (37%)	STRUM (-) n = 66 (63%)	p-value STRUM (+) vs. STRUM (-)
Baseline Characteristics			
Female, n (%)	10 (26%)	14 (21%)	0.6308
Age at diagnosis, mean (STD), year	40.31 (17.13)	41.34 (21.17)	0.9930
Follow-up time, mean (STD), year	9.82 (10.00)	12.18 (11.83)	0.7575
Maximum BMI, mean (STD) kg/m ²	28.87 (4.10)	27.96 (6.03)	0.9390
lb/ft ²	685.09 (97.29)	663.49 (143.09)	
Race, n %			0.5107
White	35 (90%)	60 (91%)	
Black	2 (5%)	1 (2%)	
Other/not reported	2 (5%)	5 (8%)	
Proband, n (%)	35 (90%)	64 (97%)	0.1232
Family History HCM, n (%)	13 (33%)	19 (29%)	0.6249
Family History SCD n, (%)	8 (21%)	9 (14%)	0.3553
Echocardiogram Data			
Maximal LVWT, mean (STD), mm	22.73 (6.97)	20.83 (7.29)	0.4287
Minimum LVEF, mean (STD), %	59.88 (8.32)	60.12 (11.34)	0.9996
LVOT peak gradient > or = 30 mmHg; n (%)	9 (23%)	21 (32%)	0.3380

BMI = body mass index, LVWT = left ventricular wall thickness, LVEF = left ventricular ejection fraction, LVOT = left ventricular outflow tract

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript