Spatial genomics of the cardiac sarcomere

Christian Michael Cadisch[1,2], Victoria Parikh[1] MD, Megan Grove[1], MS LCGC, Aditya Singh[1], DVM PhD, Dominic Abrams[3], Neal Lakdawala[4], Carolyn Ho[4], Erin Miller[5], Thomas Ryan[5], Joseph Rossano[6], Kimberly Y. Lin[6], Anna Axelsson Raja[7], Henning Bundgaard[7], Michelle Michels[8], Peter Paul Zwetsloot[8], Jodie Ingles[9], Alexandre Pereira[11], Adam Helms[12], Sara Saberi[12], Lia Crotti[13, 19], Anjali Owens[14], Sharlene M. Day[14], Belinda Gray[15], John C. Stendahl[16], Rachel Lampert[16], Iacopo Olivotto[17], Alessia Argiro[17], Niccolo Maurizi[18], George Powell[10], James Ware[10], Masataka Kawana[1], MD and Euan A. Ashley[1], FRCP DPhil**

Institutions

[1] *Stanford University*
[2] *ETH Zurich*
[3] *Boston Children's Hospital*
[4] *Brigham and Women's Hospital*
[5] *Cincinnati Children's Hospital*
[6] *Children's Hospital of Philadelphia*
[7] *University of Copenhagen; Copenhagen University Hospital Rigshospitalet*
[8] *Erasmus University*
[9] *Garvan Institute of Medical Research*
[10] *Imperial College London; Royal Brompton Hospital*
[11] *São Paulo University*
[12] *University of Michigan*
[13] *Department of Medicine and Surgery, University Milano Bicocca, Italy*
[14] *University of Pennsylvania*
[15] *Royal Prince Alfred Hospital*
[16] *Yale University*
[17] *University of Florence*
[18] *Lausanne University Hospital*
[19] *IRCCS, Istituto Auxologico Italiano, Department of Cardiology, San Luca Hospital, Cardiomyopathy Unit, Milan, Italy*

Correspondence:

Euan A. Ashley, euan@stanford.edu

Abstract

Structure-function mapping of proteins has improved our understanding of disease mechanisms and protein domains while human population genetics has provided a window into variational tolerance that can be modeled at a structural level. Here, to reveal novel insight into the cardiac sarcomere, the motor unit of the heart, we develop a novel analytical framework to integrate data from over 17,000 patients with hypertrophic cardiomyopathy (HCM) and combine it with two population-scale genomic databases that incorporate data from more than 800,000 individuals. We integrate *in silico* genetic predictions of gene variant pathogenicity with 3 dimensional integrative spatial scanning across multiple structural models of cardiac motor proteins. Results reveal both recognized and novel regions of structural variant intolerance across the critical genes of the cardiac sarcomere including novel insights into protein function. We discuss the structural relevance of variant enrichment in the context of sarcomere organization and destabilization of sequestered myosin leading to hypercontractility seen in HCM, by incorporating the recently defined high-resolution structure of human cardiac myosin filament. We extend and validate these findings through pathogenic variant class enrichment and reveal novel associations with the earlier onset of disease in a large clinical cohort. In summary, our study provides a multi-dimensional framework for integrating structural, genomic, and modeling data to reveal novel insight into the cardiac sarcomere.

Introduction

Structure-function mapping of proteins has improved our understanding of disease mechanisms and critical protein domains. Specifically, mapping the population-level disease association of variants can help to identify regions intolerant to variation that can be inferred to have disproportionate functional importance. We have previously shown that extension of these comparisons to three-dimensional space can greatly increase the information gleaned about proteins of the cardiac sarcomere, and variants that cause hypertrophic cardiomyopathy (HCM) [5, 8]. HCM is a genetic disease of the heart muscle characterized by left ventricular hypercontractility and hypertrophy, and affects ~1 in 500 individuals [1]. The clinical course of HCM is notably variable, with some patients experiencing minimal symptoms while others develop severe complications including arrhythmia, heart failure, or sudden cardiac death [1].

HCM is primarily a disease of the cardiac sarcomere. Variants in over a dozen genes encoding sarcomere-associated proteins are known to cause HCM, among these, variants in the *MYH7* gene, encoding the β-cardiac myosin heavy chain, and *MYBPC3*, encoding myosin-binding protein C, are the most prevalent causes [2]. The beta myosin heavy chain protein includes a globular head domain where ATP binding, hydrolysis, and actin interaction occur, followed by a lever arm and a long alpha-helical coiled-coil rod domain. Over 200 disease-associated variants, primarily missense, have been identified in *MYH7*, with a significant concentration in the motor domain of the head region, which ClinGen (a National Institute of Health funded central resource for variant clinical relevance) recognizes as a pathogenic hotspot [47]. Past research has identified the converter domain and the myosin mesa, a relatively flat surface on the myosin motor domain composed of several myosin subdomains that is conserved in all cardiac myosin from all species, as enriched for variants associated with HCM with a more severe prognosis and earlier disease onset [4, 5, 17]. The clinical genetics and functional analysis of these variants led to significant improvements in understanding the structure-function relationships underlying the hypercontractility observed in HCM, and informed design of several putative therapies now commercially approved or

pending Food and Drug Administration (FDA) review [16, 48, 49]. Early experience with these precision therapies shows a significant reduction in symptoms and biomarkers of disease severity [48, 49].

In contrast to *MYH7*, mapping of disease-observed variants to known or predicted 3D protein structures of other sarcomeric proteins has not been performed. While other common sarcomeric genes known to cause HCM, such as *MYBPC3*, have had several enriched regions identified by 2D mapping [10], most *MYBPC3* disease-observed variants are truncating rather than missense variants. Similarly, disease-observed variants in genes encoding the troponin complex proteins (*TNNT2*, *TNNI3*); and the myosin light chains (*MYL2*, *MYL3*) are generally less common [6]. The lack of structure-function modeling integrating clinical data for these genes may stem from the lower prevalence of disease-observed variants and the resulting limitations in statistical power. Previous attempts to identify disease-associated regions have employed various approaches, from spatial scan statistics [5] to generalized additive modeling [10], but have been limited by the available data and computational tools. As *MYH7* is a well-studied gene with prior 3D analyses and functional validation, we first re-examined it to validate our methodology and determine whether additional data would reveal previously unrecognized structural insights [16].

Recent technological advances have dramatically enhanced our ability to study structure-function relationships at the level of single amino acid substitutions in the sarcomere. Numerous large-scale biobanks with genetic and imaging data from both disease-affected and general populations have enabled the compilation of extensive datasets of human genetic variation. Furthermore, the introduction of transformer-based computational models like AlphaFold (and the related AlphaMissense predictor) has significantly improved our ability to predict variant pathogenicity [12] and protein structure [13]. In addition, advancement in structural biology and particularly cryogenic electron microscopy (cryo-EM) has enabled us to obtain high-resolution images of human sarcomere structure and individual proteins such as myosin, which would further improve the understanding of the inter- and intra-molecular interactions that are critical in defining sarcomere function [22, 24, 34, 51, 52]. Combining clinical genetics and

structural biology holds high promise for fulfilling the current unmet need for improving variant assessment in daily practice where clinicians encounter a large number of variants of unknown significance.We hypothesized that integrating these advanced computational tools with clinical and population genetic databases could reveal novel insights into the structure-function relationships within key sarcomeric genes and identify previously unknown variant enrichments. In this study, we combine exomic data representing 814,392 individuals with the Sarcomeric Human Cardiomyopathy Registry (SHaRe), a HCM clinical database of over 17,000 HCM patients, and AlphaMissense pathogenicity predictions. Using a spatial scan statistic approach, we investigated regions of disease enrichment in the 3D protein structures of *MYH7*, *MYBPC3*, *TNNT2*, *TNNI3*, *MYL2* and *MYL3*. Finally, we validated our findings by comparing predicted pathogenicity with established disease-associated variants in ClinVar, and explored phenotypic differences across the cohorts, focusing on the age of disease onset.

Results and Discussion

We first quantified the distribution of variants across major sarcomeric genes in both population and patient cohorts, complemented by *in silico* pathogenicity predictions (Table 1). Our framework compares the spatial distribution of variants found in HCM patients to those seen in population cohorts to identify regions of disease enrichment. These regions were then validated using two orthogonal approaches: enrichment of pathogenic classifications in ClinVar and differences in age of disease onset in affected individuals. This analysis integrated three large-scale genetic resources: gnomAD (730,947 exomes, with 586,403 individuals carrying variants in the genes of interest), the UK Biobank (500,000 individuals, of whom 346,977 carried variants in these genes), and the SHaRe registry (17,000 patients with cardiomyopathy, including 12,326 with HCM, of whom 4,427 carried variants in the six genes studied). Our analysis revealed strong functional implications of 15 regions enriched for variants across key sarcomeric genes, including the kinetics of myosin ATPase activity, the biophysical properties of the myosin molecule, and the stability of myosin sequestration in the sarcomere (Table 2).

MYH7 converter and transducer domains are associated with early disease onset

In the Subfragment 1 (S1) domain of the beta myosin heavy chain, we identified two key regions enriched for variants observed in disease (see Fig. 1a). The first region identified was the converter domain, centered on residue 724 with a 15 Å sphere ($p=0.004$). This region was associated with dramatically earlier disease onset (10.5 years earlier, $p=2.8\times10^{-7}$, see Fig. 2) and showed significant enrichment of pathogenic/likely pathogenic variants in ClinVar ($p=0.024$).

The converter domain enrichment aligns with findings from previous studies [5, 14, 15], and our observation of earlier diagnosis is consistent with earlier research [4, 5]. This domain was previously suggested to have a critical role in force transduction during systolic contraction [16] by coupling the biochemical events of ATPase activity and the power stroke, which could explain why variants in this region may cause HCM. However, functional analysis of the recombinant human cardiac myosin with converter domain variants showed surprisingly little change in motor function [42]. Rather, destabilizing the sequestered state of myosin seemed to result in increasing the availability of myosin heads for forming crossbridges [43] which creates the characteristic hypercontractility of HCM.

A second region, previously suggested as enriched but not statistically confirmed, was identified near the nucleotide-binding site centered on residue 257 with a 15 Å sphere ($p=0.005$). This region is known as the transducer domain, with structural elements at the heart of the motor domain comprising the seven-stranded beta sheet and connectors that control its conformation within the myosin motor [54]. It trended towards an onset of disease 4.2 years earlier onset ($p=0.071$, see supplemental Figure S7) and showed significant enrichment of pathogenic variants in ClinVar ($p=0.004$, where pathogenic refers to variants classified as pathogenic or likely pathogenic by ClinVar). Our mathematical characterization significantly strengthens the evidence for a pathogenic region in the ATP/ADP binding region moving beyond prior qualitative descriptions. By quantitatively establishing residue 257 and its surrounding region as a

region statistically significantly enriched with disease-observed variants, this analysis deepens our understanding of its clinical importance and functional relevance.

Functional analyses of several variants in the transducer domain have been reported previously. The p.Arg453Cys variant decreases the maximal ATPase activity and reduces actin filament velocity but significantly increases the intrinsic force of the myosin motor domain. Consequently, this variant leads to an overall hypercontractile state in cardiac muscle [44]. Similarly, the p.Arg249Gln variant significantly increases the opening of the interacting heads motif (IHM), releasing sequestered myosin heads, thereby enhancing ATPase activity, as demonstrated by Adhikari et al.[43]. Both variants result in a hypercontractile effect on cardiac myosin function.

Disease-observed *MYL3* variants cluster on the EF-hand domain

The examination of *MYL3* revealed one region enriched with variants observed in disease on the second EF-hand of the protein, defined by a 15 Å sphere centered on residue 182 ($p = 0.001$; see Fig. 1c). This region also shows a significantly higher number of pathogenic variants in ClinVar than expected by chance ($p = 0.006$). Given that this region corresponds structurally to the functional Ca^{2+} binding domain, the enrichment of disease-observed variants suggests significant functional relevance. However, previous studies have indicated that despite the presence of a canonical Ca^{2+} binding motif, *MYL3* has evolutionarily lost its calcium-binding capacity [37, 38], raising uncertainty about the true functional impact of variants in this region. To clarify these conflicting observations, we performed protein-ion interaction simulations to investigate the potential structural and functional implications in greater detail.

Within the enriched sphere, 16 out of the 41 patients observed with a *MYL3* variant carried the p.Glu143Lys variant. This variant was initially reported in a family with early-onset cardiomyopathy with mid-cavitary hypertrophy and restrictive physiology in which siblings carrying the homozygous p.Glu143Lys variant showed severe phenotype [58]. Another patient who was homozygous for *MYL3* p.Glu143Lys and heterozygous for *MYL2* p.Gly57Glu had severe restrictive cardiomyopathy (RCM) [59]. These reported

two cases suggest the functional importance of this domain in determining the relaxation of the myocardium that is abnormal in both HCM and RCM. A protein-ion interaction simulation using AlphaFold 3 demonstrated that residue 143 is in close proximity to the defunct divalent cation binding site (see Fig. 3a). Introducing the p.Glu143Lys variant into the amino acid sequence and re-running the simulation showed a displacement of the cation binding site from its original position (see Fig. 3b, Fig. 3c).

This evidence raises a hypothesis that substituting the negatively charged glutamic acid with the positively charged lysine influences the affinity of *MYL3* to divalent cations and thereby contributes to functional impairment. In general, EF-hand motifs act as calcium sensors, undergoing conformational changes upon calcium binding, which are essential for cellular function. [55] EF-hand motifs are also able to bind to Mg^{2+}, and the introduction of variants can abolish the binding of either cation. The simulated structural changes may suggest that protein-ion interaction is significantly altered by the p.Glu143Lys variant.

Several variants in this region have been shown to affect protein function and stability. The variant p.Glu143Lys has been shown to significantly weaken the binding of *MYL3* to the myosin lever arm [39], and a transgenic mouse model harboring this variant showed increased ATPase activity, duty ratio, and crossbridge formation that led to a significant increase in force generation and diastolic dysfunction[40]. Additionally, the p.Arg154His variant significantly weakens the binding to the myosin lever arm, exhibiting a 3-fold higher KD in protein-binding experiments [39].

Despite human *MYL3* losing its ability to bind Ca^{2+}, the enriched region identified on the protein structure matches exactly the evolutionarily lost divalent cation-binding site. Our simulation raises the hypothesis that while *MYL3* does not directly bind Ca^{2+}, the evolutionarily conserved second EF-hand still has a regulatory function, possibly by interactions through ions such as Mg^{2+} and a major structural alteration that affects its binding to myosin.

Integration of protein variant effect prediction model with real-world observations reveals nine novel enrichment regions

By incorporating *in silico* pathogenicity predictions (AlphaMissense) into our spatial analysis framework, we expanded our ability to detect variant clustering across protein structures. This integrative approach revealed statistically significantly enriched regions in six sarcomeric genes. Notably, *MYL2* and *MYBPC3*, which showed no clustering using real-world data alone, now exhibited a clear enrichment signal, underscoring the added sensitivity gained from combining computational predictions with real-world observations and adding a significant amount of data points to the analysis (see Table 1). Computational pathogenicity predictions have been shown to correlate with real-world disease associations and outcomes supporting the validity of integrating such estimates with observed variant data [61, 62].

Structural prediction model identified additional clusters on *MYH7* myosin mesa

Using a combined approach with real-world data and pathogenicity predictions, we identified two additional enriched regions and successfully re-identified the previously observed regions in the converter domain and transducer domain. Notably, the statistical significance for the transducer domain improved, with the p-value decreasing from p=0.005 to p=0.001.

The two additional regions were identified on the myosin mesa (see Fig. 1a). The first sphere was centered on residue 493 with a 10 Å radius (p=0.001) and showed significant enrichment of pathogenic variants in ClinVar (p=0.043). This region falls under the relay helix, a connector essential for allosteric communication between the motor domain and the lever arm [22, 56]. Previous structural modeling and molecular dynamics simulation suggested that the variants at residue 493 (p.Met493Leu) and its surrounding area are expected to cause destabilization of IHM [57]. The second sphere, centered on residue 696 with a 10 Å radius (p=0.001), was associated with 8.4 years later disease onset (p=0.016, see supplemental Figure S6) and also showed significant enrichment of pathogenic variants in ClinVar (p=0.041). The residue 696 resides within

the SH1 helix, another critical connector that transmits the structural rearrangements within the motor domain to the converter domain [56]. This region is also predicted to destabilize the IHM, leading to hypercontractility [57].

The myosin mesa findings both confirmed and expanded upon previous research [5, 17]. While earlier studies identified a surface area of interest, our work uncovered a three-dimensional enriched region, providing new insights into the spatial characteristics of this functional zone. The functional validation, as well as structural modeling and prediction results, provide further support for our approach.

Four variant clusters revealed in *MYBPC3* near myosin tail and IHM motor domain interfaces

Our analysis of *MYBPC3* revealed four regions of enrichment in disease-observed variants distributed across different domains of the protein (see Fig. 1b). The first region was identified on the previously reported C6 domain, centered on residue 857 with a 15 Å sphere (p=0.007). Two novel regions were found in the C8 domain, centered on residue 1030 with a 10 Å sphere (p=0.001), and in the C9 domain, centered on residue 1102 with a 10 Å sphere (p=0.001). The fourth region was identified in the previously reported C10 domain, centered on residue 1253 with a 10 Å sphere (p=0.008). Although all four regions showed a higher-than-expected proportion of pathogenic variant annotations, only the C6 domain reached statistical significance (p=0.003), and none of the regions were associated with a significant difference in disease onset.

These regions did not show statistically significant enrichment when only mapping real-world data to the 3D protein structure. This lack of significance could be due to the variability in disease severity and later onset associated with *MYBPC3* variants, and underscores the enhanced sensitivity achieved by incorporating *in silico* pathogenicity predictions into our spatial analysis.

The disease-observed enrichments in domains C6-C10 are particularly interesting due to their established interaction with the light meromyosin (LMM) region of myosin that forms the myosin tail in thick filament, which is critical for maintaining the stability of the

sequestered (inactive) state of myosin (Fig. 4a) [24]. Dutta et al. reported the cryo-EM structure of the human cardiac thick filament in a relaxed state, which revealed the packing of myosin tails in the filament backbone and their interaction with *MYBPC* and titin [24]. The fitted atomic model showed interactions between domains C6-C10 and myosin tail. In particular, domains C8 and C10 interacted with the motor domain of IHM myosin, suggesting that these two domains are involved in the stabilization of IHM. While the C6 and C9 domains do not directly interact with the IHM myosin itself, their interactions with the LMM region may affect the structure of myosin tails and affect stability of myosin head sequestration. Taken together, these results suggest that *MYBPC3* C6–C10 variants may alter sarcomeric structure in a manner that increases myosin head availability, thereby contributing to hypercontractility in HCM.

Upon examining the disease enriched regions illustrated in Fig. 1b and Fig. 4a, it is also apparent that these regions are predominantly localized within regions where the secondary protein structure consists of beta-sheets. A t-test supports this observation, indicating that the clustering of variants observed in disease on beta-sheets is highly unlikely to occur by chance ($p=1\times10^{-6}$). This result is further validated by comparing the pathogenic and benign/likely benign (B/LB) annotations from ClinVar ($p=0.049$), corroborating the finding that disease-observed variants cluster on beta-sheets (see Fig. 1b).

Previous studies analyzing disease regions on *MYBPC3* using two-dimensional sequence-based analyses rather than three-dimensional structural mapping have shown varying results. Several studies specifically compared HCM cohorts to gnomAD as a reference population. Waring et al. identified clusters in the C1, C3, C7, and C10 domains [10], while Helms et al. found variant concentration in the C3, C6, and C10 domains [8]. Walsh et al. observed clustering restricted to the C3 and C10 domains [9]. The C10 domain, in particular, has been consistently highlighted as a potential region for disease-associated variants across multiple studies [8, 9, 10]. However, the significance of other regions, such as C6, C8, and C9, remains less certain. Notably, some studies have reported no significant clustering patterns in any domain [7, 11]. These discrepancies may stem from differences in sample sizes, statistical methods, or

the later-onset and variable penetrance of *MYBPC3* variants, which are often truncating and associated with a more favorable prognosis [3].

Our discovery that both the identified disease-associated regions and individual variants classified as pathogenic in ClinVar are disproportionately located on beta-sheets represents a valuable insight into the structural basis of variant pathogenicity in *MYBPC3* (see Fig. 1b). One possible explanation for this clustering on beta-sheets could be the structural role they play in protein stability. Helms et al. identified several *MYBPC3* missense variants that markedly decreased protein stability, suggesting that haploinsufficiency due to rapid degradation could underlie their pathogenicity [8]. Beta-sheets often form the core of protein structures, and disruptions in these regions may more readily lead to misfolding or destabilization, thereby triggering pathogenic effects [18, 19]. Another hypothesis is that beta-sheets may be involved in important protein-protein interactions, such as the ones suggested from the thick filament structure described above, and variants in these regions could disrupt essential functional interfaces between myosin. Further research is needed to investigate whether the clustering of disease-observed variants in beta-sheets of *MYBPC3* could be linked to disruptions in its structural integrity or functional binding interfaces.

Variant enrichment in *MYL2* surrounding the *MYH7* lever arm

With integration of a pathogenicity prediction model, our analysis revealed a novel disease-enriched region characterized by a 20 Å sphere centered on residue 98 ($p = 0.005$), associated with 12.8 years later disease onset ($p = 0.005$, see Fig. 1d, see supplemental Figure S12). This region is in close proximity to the *MYH7* lever arm (see Fig. 3d), and encapsulates the phosphorylation site on residue Serine 15 [41].

Previous studies have hypothesized that variants within *MYL2* significantly influence cardiac contractility by altering interactions with the myosin lever arm, potentially modifying its mechanical stiffness [45][46]. Our findings, identifying a region closely surrounding the *MYH7* lever arm, support this stiffness hypothesis, suggesting that variants within this region may similarly impact lever-arm stiffness and consequently

influence mechanochemical properties. These results underscore the critical role of the *MYL2–MYH7* interaction and highlight opportunities to further investigate how increased mechanical stiffness may enhance the intrinsic force of individual myosin molecules, ultimately contributing to greater ensemble sarcomere force in the pathogenesis of HCM.

Methodological considerations and future directions

In this study, we conducted the largest-scale analysis of genetic data on HCM to date, encompassing over 814,392 population-based genome sequences alongside more than 10,000 targeted exome sequences from HCM patients. For the first time, we integrated pathogenicity predictions from a transformer-based pathogenicity prediction model (AlphaMissense) into the analysis, expanding on traditional approaches that use clinical and population data and functional validation, and introducing a framework to interpret predictive data in a structural context. While the prediction model does not distinguish between HCM and DCM, our downstream phenotype analyses focused on HCM provides disease-specific context for the findings.

Future validation strategies could include variant effect mapping using multiplexed assays [26], particularly when combined with assessment in human-induced pluripotent stem cells (hiPSCs). The measurement of relevant biomarkers, such as B-type natriuretic peptide (BNP) levels [27], could further support these findings by linking molecular alterations to functional cardiac phenotypes. BNP is released in response to increased wall stress and myocardial strain, thus monitoring BNP levels in variant carriers or hiPSC-derived cardiomyocytes could provide a functional readout of contractile imbalance or impaired relaxation. These approaches would not only help confirm the identified regions but also contribute to our understanding of disease pathways in HCM.

Our analysis identified several new disease enrichments and confirmed previously reported ones. Additionally, by examining cryo-EM structures and performing simulations with AlphaFold 3, we obtained further structural evidence supporting the proposed hypotheses on the functional impacts and disease mechanisms of these

enriched regions. Moreover, we leveraged ClinVar as a high-quality, independent dataset to validate our findings. Our approach not only highlights the utility of *in silico* models in genetic variant interpretation and classification but also underscores the importance of integrating the data with structural models of the proteins for more comprehensive disease association studies. The identification of new regions and detailed structural insights into their functional consequences could inform the development of novel therapeutic targets for treating hypertrophic cardiomyopathy. Furthermore, validating previously reported and novel disease-associated regions provides valuable supporting evidence for variant classification, particularly under the American College of Medical Genetics and Genomics (ACMG) guidelines and ClinGen framework [50, 60].

Materials and Methods

Development of protein models

Computational modelling approach

To obtain optimal structural models for each protein under study, we evaluated four computational modeling approaches and selected the most accurate models based on a set of evaluation criteria.

We first obtained pre-computed models from the AlphaFold Protein Structure Database, a collaborative effort by Google DeepMind and European Bioinformatics Institute (EMBL-EBI) [28]. These models were generated using the AlphaFold Monomer v2.0 pipeline and last updated on November 1, 2022 [13]. Secondly, we utilized ColabFold v1.5.5 to create structure predictions based on the AlphaFold2 v2.2 model weights and using the faster homology search of MMseqs2 [21, 29].

We attempted to enhance the prediction accuracy of the ColabFold model by providing experimentally obtained template structures. We utilized Stanford's shared computing cluster, Sherlock, for GPU access. ColabFold was set to use a custom template mode

with MMseqs2 browsing UniRef and environmental datasets. The number of recycles was set to 3 with early stopping tolerance on auto, and relax was limited to 200 iterations. We employed a greedy pairing strategy, and dropout was disabled.

Lastly, we used the AlphaFold Server to create structure predictions with AlphaFold 3 [20]. For all protein-ion and protein-ligand interactions, we used AlphaFold 3. For the analysis of *MYH7*, we focused exclusively on subfragment 1 (S1), encompassing residues 1-840, as this region possesses distinct functional significance and is highly enriched with variants observed in disease.

Protein template selection

For modeling the beta-myosin heavy chain (encoded by *MYH7*), we provided the model with template structures with PDB IDs 8ACT [22], 8EFH [25], 8EFI [23], 8ENC [23], and 8G4L [24]. For modeling the myosin-binding protein C (encoded by *MYBPC3*), we used template structures with PDB IDs 6CXJ [30], 6G2T [30], 7LRG [31], 7TIJ [32], and 8G4L [24]. For the essential light chain (encoded by *MYL3*) and regulatory light chain (encoded by *MYL2*), we used template structures with PDB IDs 5TBY [33], 8ACT [22], and 8G4L [24]. For modeling the cardiac troponin T (encoded by *TNNT2*) and troponin I (encoded by *TNNI3*), we used template structures with PDB IDs 6KN7 [34], 6KN8 [34], 7UTI [35], and 7UTL [35].

Assessment and comparison of predicted protein structures

To assess and compare the quality of all protein models, we analyzed predictive and experimental metrics. The structure prediction models provided two predictive scoring methods: predicted local distance difference test (pLDDT), giving a per-residue measure of local confidence, and the predicted aligned error (PAE), giving an estimate of the expected error between each pair of amino acids. We calculated the mean pLDDT and mean PAE for each protein model to understand overall accuracy and local confidence levels. Furthermore, we aligned the predicted structures with partial experimentally determined protein structures available in PDB and measured the minimal root mean square error (RMSE) between the predicted and experimental structures. Visualizations and alignments were made using PyMol 2.5.8.

Result evaluation and selection of 3D protein models

MYH7

The AlphaFold 2 structure was selected as the best structural model for *MYH7*, demonstrating superior performance across multiple evaluation metrics. With a mean pLDDT score of 73.75 and mean PAE of 23.43 Å, the AlphaFold 2 model achieved the lowest overall PAE and the second-highest pLDDT among all evaluated models (detailed PAE plots and RMSE comparisons available in the supporting information).

All models consistently showed higher accuracy in predicting the structure of subfragment 1 (S1) of the heavy meromyosin (HMM) domain, which encompasses the myosin head (residues 1-840). In contrast, the subfragment 2 (S2) (residues 841-1280) and the light meromyosin (LMM) domain (residues 1281-1935) exhibited lower confidence, with higher predicted errors in these regions.

When aligned with experimentally determined cryo-EM structures, the AlphaFold 2 model consistently produced the lowest RMSE across the tested structures. The model achieved an RMSE of 1.35 Å for the structure with PDB ID 8ACT [22], 3.73 Å for PDB ID 8ENC [23], 3.71 Å for PDB ID 8EFI [23], 6.39 Å for PDB ID 8G4L [24], and 3.25 Å for PDB ID 8EFH [25]. With an average RMSE of 3.67 Å, this model achieved the lowest overall error across the evaluated structures.

MYBPC3

In modeling *MYBPC3*, all models exhibited similar mean PAE scores, indicating comparable overall accuracy in predicting residue-residue distances across the protein. Additionally, all models, except for ColabFold v2.3 with templates, achieved similar pLDDT scores close to 80, demonstrating high confidence in local structural predictions. The AlphaFold 2 structure, with a mean pLDDT of 79.38 and a mean PAE of 25.43, was ultimately chosen as the best model due to its consistently good performance in terms of RMSE when aligned with experimentally determined cryo-EM structures (detailed PAE plots and RMSE comparisons available in the supporting information).

A detailed inspection of the PAE plots across all models revealed that while the confidence in long-distance relationships between residues far apart in the sequence is low, the local accuracy for individual domains of *MYBPC3* is high. This pattern suggests that, although the models may struggle with accurately predicting the overall topology and domain-domain interactions, they reliably capture the secondary structures and the shape of the individual domains.

The structural differences observed between our *MYBPC3* model and that reported by Thompson et al. likely arise from distinct computational modeling approaches employed [53]. Specifically, Thompson utilized the STRUM algorithm combined with I-TASSER structural predictions focused on individual protein subdomains, potentially highlighting localized folding instability, whereas our integrative approach emphasized whole-protein AlphaFold predictions and spatial clustering analyses across larger protein domains.

MYL2

For the gene *MYL2*, we selected the model created with AlphaFold 2 as the best structure. It achieved the best mean pLDDT of 83.36 and the second-best mean PAE with 13.21 Å, only surpassed by the structure created with AlphaFold 3 which achieved a mean PAE of 11.49 Å. The decision to use the AlphaFold 2 model was driven by the excellent RMSE when the structure was aligned to the cryo-EM structure with PDB ID 8ACT [22], which itself has a high resolution of 3.6 Å. An inspection of the PAE plots reveals that all structure prediction models are confident in the aligned distances in the domain of the EF-hands: the first EF-hand between residue 24 to 59, and the second and third EF-hands between residues 94-129 and 130-165.

MYL3

Upon evaluating the structural predictions of *MYL3*, the AlphaFold 2 model was identified as the most reliable, achieving a high mean pLDDT of 88.86 and a mean PAE of 11.14 Å, which, while third lowest, is comparable to the lowest PAE of 10.63Å. This model exhibited exceptional alignment with the cryo-EM structures used for comparison. When aligned with the structure having PDB ID 8ACT [22], the model achieved an

RMSE of 1.48 Å, and with the structure having PDB ID 5TBY [33], it achieved an RMSE of 1.12 Å. The PAE plots indicate that, apart from the disordered N-terminal region (amino acids 1-37), the predicted folding confidence is very high.

Datasets

We have used an extensive dataset of genetic information, combining clinical data, exonic genetic data and pathogenicity prediction scores (see Table 1).

Sarcomeric Human Cardiomyopathy Registry (SHaRe)

The SHaRe is a collaborative initiative that aggregates anonymized patient data from multiple research institutions to advance the study of cardiomyopathies. It includes information from patients diagnosed with HCM, DCM, and arrhythmogenic cardiomyopathy. The registry contains genetic, clinical, and demographic information from over 17,000 patients, of which 12,326 were diagnosed with HCM. Of these, 4,427 patients have genetic variants within the six genes analyzed in this research, totaling 1,056 unique variants.

UK Biobank (UKBB)

The UK Biobank is a comprehensive repository of biomedical data and research materials. It contains anonymized information on genetics, lifestyle habits, and health outcomes, along with biological samples, collected from 500,000 individuals across the United Kingdom. Unlike disease-focused databases such as SHaRe, UKBB is curated from the general population and not enriched for people with specific conditions. The effort results in 346,977 patients with variants in the genes of interest for our study, yielding a total of 1,479 unique variants without applying and filtering based on MAF.

Genome Aggregation Database (gnomAD)

The Genome Aggregation Database, originally established as the Exome Aggregation Consortium (ExAC), represents a global collaboration among researchers committed to sharing data from exome and genome sequencing projects. Similar to the UKBB,

gnomAD compiles data from the general population. In its latest release (v4), gnomAD has expanded its database to include 730,947 exomes of which 416,555 originate from the UK Biobank. Among these, 586,403 patients have variants in the sarcomeric genes we focused on, amounting to 4,620 unique observed variants without applying any filtering based on minor allele frequency (MAF).

AlphaMissense

AlphaMissense is a transformer-based model developed to predict the pathogenicity of missense variants in the human genome. For each gene, the model creates pathogenicity predictions for the 19 possible missense variants at each amino acid position. Within the six genes we studied, this results in a total of 77,482 pathogenicity scores. This number significantly exceeds the total number of variants observed in both the general population cohorts and the disease-enriched cohort combined.

ClinVar

ClinVar is a public archive that collects and shares information about human genetic variants and their relationships to diseases and drug responses. We used the ClinVar Miner tool to access data on our genes of interest with data current up to February 5, 2024 [36]. In total, this resulted in 5,252 annotations, of which 945 are categorized as pathogenic / likely pathogenic and 1391 are categorized as benign / likely benign. Variants with multiple or discordant submissions were counted as separate annotations. We used data from ClinVar as an independent, high-quality validation set to further substantiate the validity of the identified enriched regions.

Classification of variants

Our analysis focused on identifying disease-enriched regions using a discrete, binomial statistical framework. To accomplish this, variants were classified as either disease-observed or from the general population.

Clinical data: SHaRe vs. population datasets

For the clinical data, we adopted the same categorization strategy as proposed by Homburger et al., labeling all variants found in HCM patients from the SHaRe cohort as disease-observed [5]. Variants exclusively appearing in one of the two population datasets were considered reference variants. While the gnomAD and UKBB population datasets may include individuals with HCM, the incidence is not expected to exceed that of the general population. Thus, most rare variants observed in these cohorts are likely unrelated to HCM. Conversely, very rare or novel variants observed in the SHaRe cohort are more likely to be causative for HCM.

Integration of pathogenicity prediction model

The model AlphaMissense provides pathogenicity predictions as a continuous score ranging from 0 to 1. To incorporate these continuous scores into our discrete statistical framework, we developed a classification strategy that categorizes variants into three groups: disease-observed variants, population variants, and excluded variants.

To address the data imbalance between observed variants in clinical datasets and *in silico* predictions, we implemented a selective inclusion strategy. This approach increased the number of variants included in our analysis by 50%, while maintaining the ratio of disease-observed to reference variants constant. Before integrating AlphaMissense data, we performed a deduplication step to remove variants already present in the SHaRe cohort or population datasets.

Let $N_{Disease}$ represent the number of disease-observed variants and $N_{Population}$ the number of population variants observed in the clinical data. After deduplication, we ranked all variants based on their predicted pathogenicity score and expanded our set of disease-observed variants by adding the highest-scoring variants:

$$AddedDiseaseVariants = sorted\ Predictions[0:N_{Disease} \times 0.5]$$

Similarly, for population variants, we added the lowest-scoring variants:

$$AddedPopulationVariants = sorted\ Predictions[N_{Population} \times 0.5 : N_{Population}]$$

Spatial Scan Statistic

In our binomial spatial scan analysis, we examined the locations of unique disease-observed variants in comparison to variants found in the population datasets, using the structural models of the proteins. The aim was to identify regions where disease-observed variants occur significantly more often than can be expected by chance.

The statistical framework utilizes a sliding-window approach, using spheres with radii of 10Å, 15Å, and 20Å, which iterate across all amino acids of the protein.

For each step of the iteration, the number of disease-observed variants within (y_{in}) and outside (y_{out}) the sphere was calculated, along with the number of population variants within (z_{in}) and outside (z_{out}) the sphere. We then compared the proportions of HCM-associated variants within and outside the sphere using the following definitions:

- $p_{in}=y_{in}/(y_{in}+z_{in})$: The proportion of HCM-associated variants within a given window.
- $p_{out}=y_{out}/(y_{out} + z_{out})$: The proportion of HCM-associated variants outside the window.
- $r=(y_{in}+y_{out})/(y_{in}+z_{in}+y_{out}+z_{out})$: The overall rate of HCM-associated variants across all windows.

For each window, we compute the binomial log-likelihood ratio statistic. This statistic compares the null model, where p_{in} and p_{out} are both equal to the overall rate r, against an alternative model where p_{in} differs from p_{out}. The log-likelihood ratio statistic for each window is calculated as follows:

$$\log(\Lambda_w) = \log\left[p_{\text{in}}^{y_{\text{in}}}(1-p_{\text{in}})^{z_{\text{in}}} p_{\text{out}}^{y_{\text{out}}}(1-p_{\text{out}})^{z_{\text{out}}}\right] - \log\left[r^{y_{\text{in}}+y_{\text{out}}}(1-r)^{z_{\text{in}}+z_{\text{out}}}\right]$$

For each residue of the protein model, the likelihood ratio was calculated. We then assessed the statistical significance of the identified enriched regions using a permutation test with 1,000 permutations. During each permutation, the variant labels were shuffled randomly, and the number of variants per residue was held constant. This ensured that the significance assessment remained unaffected by any inherent non-uniform distribution of variants across the full protein.

Validation Strategies

ClinVar validation

We performed a qualitative assessment of ClinVar classifications within and outside the identified regions by examining histograms of the distribution of pathogenic and benign variants across these regions. For quantitative analysis, we compared the proportions of pathogenic and likely pathogenic variants against benign and likely benign variants using Fisher's exact test.

Disease onset age analysis

We compared the age of disease onset in patients with variants located within the identified enriched regions to those with variants outside the sphere. The statistical significance of observed differences was assessed using a Wilcoxon rank-sum test and Kaplan-Meier curves for disease onset.

External validation set

We validated our findings with an independent dataset containing variant information of patients diagnosed with HCM. The results of this validation are presented in the supporting materials.

Acknowledgements

We thank the investigators and participants in the SHaRe for their support of this project.

Figures

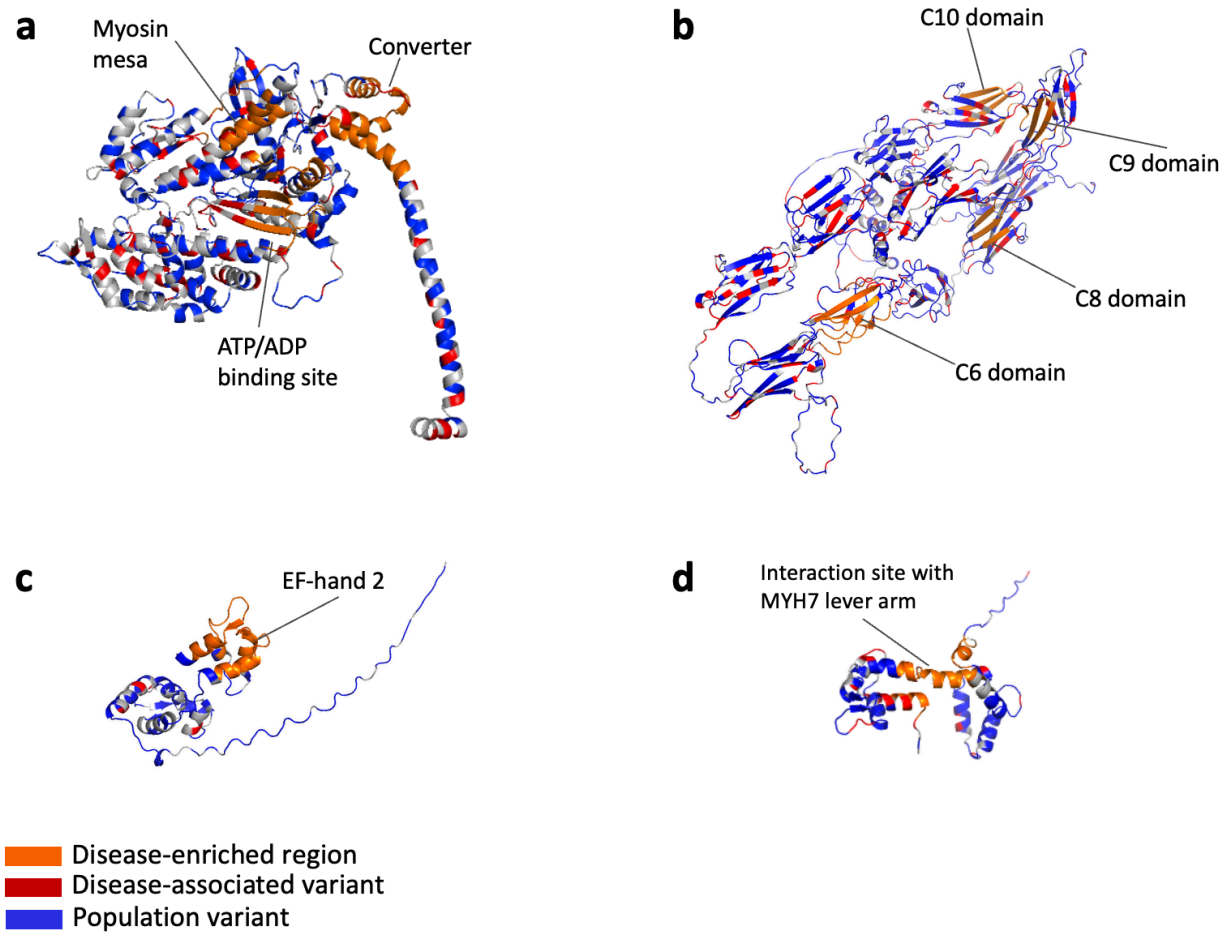

Figure 1. *Structural model of MYH7, MYBPC3, MYL2, and MYL3 highlighting regions enriched showing disease-enriched variants in orange, disease-observed variants in red and population variants in blue. (a) Structural model of MYH7 S1 (subfragment 1) region with the regions enriched with disease-observed variants highlighted in orange (myosin mesa, converter, and nucleotide binding site). (b) Structural model of MYBPC3 highlighting disease-enriched regions in the C6, C8, C9, and C10 domains. Statistical*

analysis confirmed significant clustering of disease-causing variants on beta-sheets, indicating a structural preference for disease-observed variant accumulation in these regions. (c) Structural model of MYL3 highlighting the disease-enriched region on the second EF-hand domain, centered on residue 182. (d) Structural model of MYL2 highlighting the disease-enriched region centered on residue 98 and associated with significantly later disease onset.

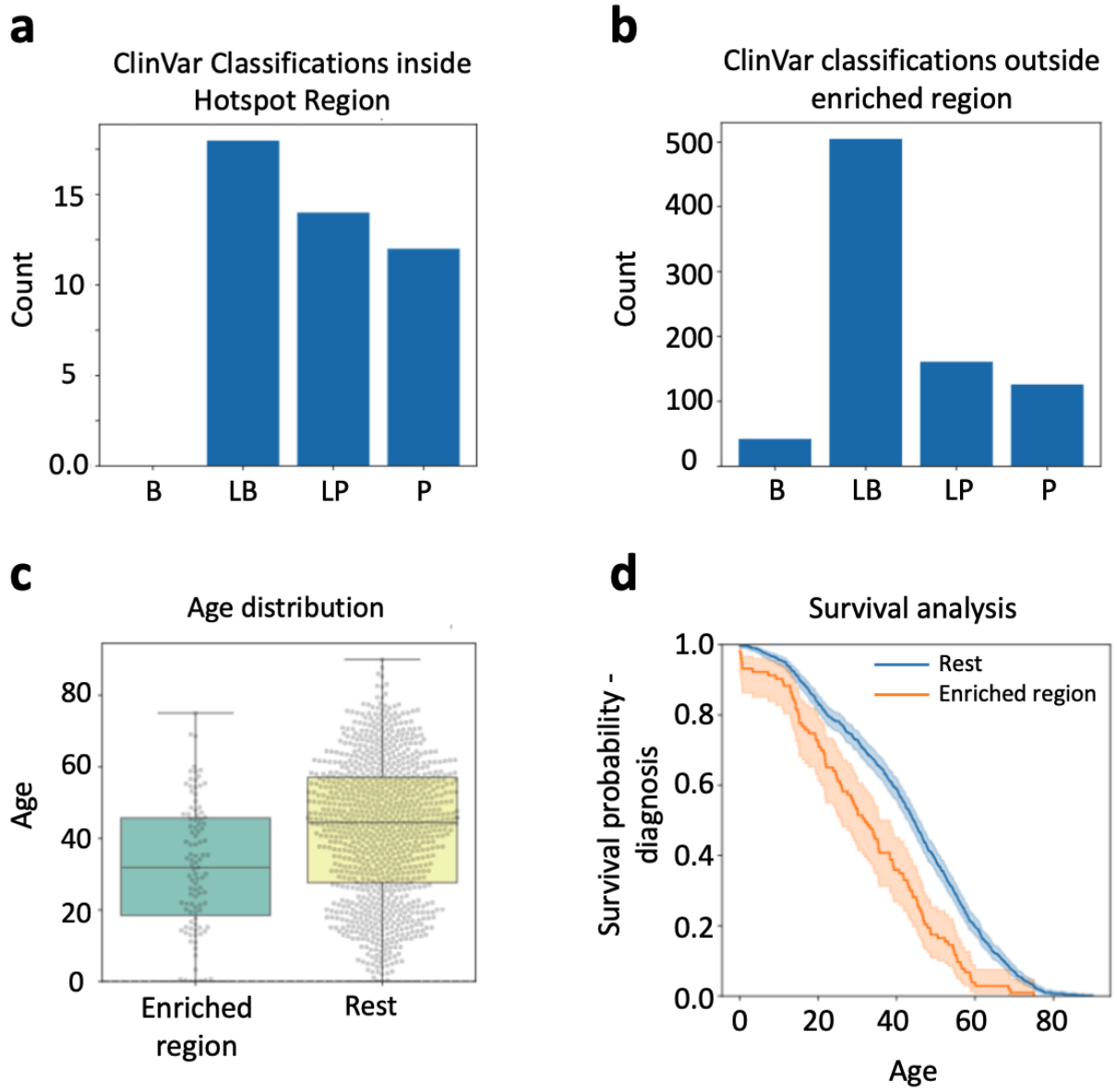

Figure 2. *As orthogonal validation of the findings, ClinVar classifications and disease onset analyses demonstrate a significant enrichment of pathogenic/likely pathogenic*

(P/LP) variants compared to benign/likely benign (B/LB) in the MYH7 converter region, as well as a significantly earlier disease onset (Δ = 10.5 years). (a) Distribution of ClinVar classifications for variants inside the enriched converter region. (b) Distribution of ClinVar classifications for variants outside the converter region. (c) Age distribution at diagnosis for diseased subjects with variants inside vs. outside the converter region. (d) Kaplan–Meier survival analysis showing earlier onset for diseased subjects with variants in the converter region.

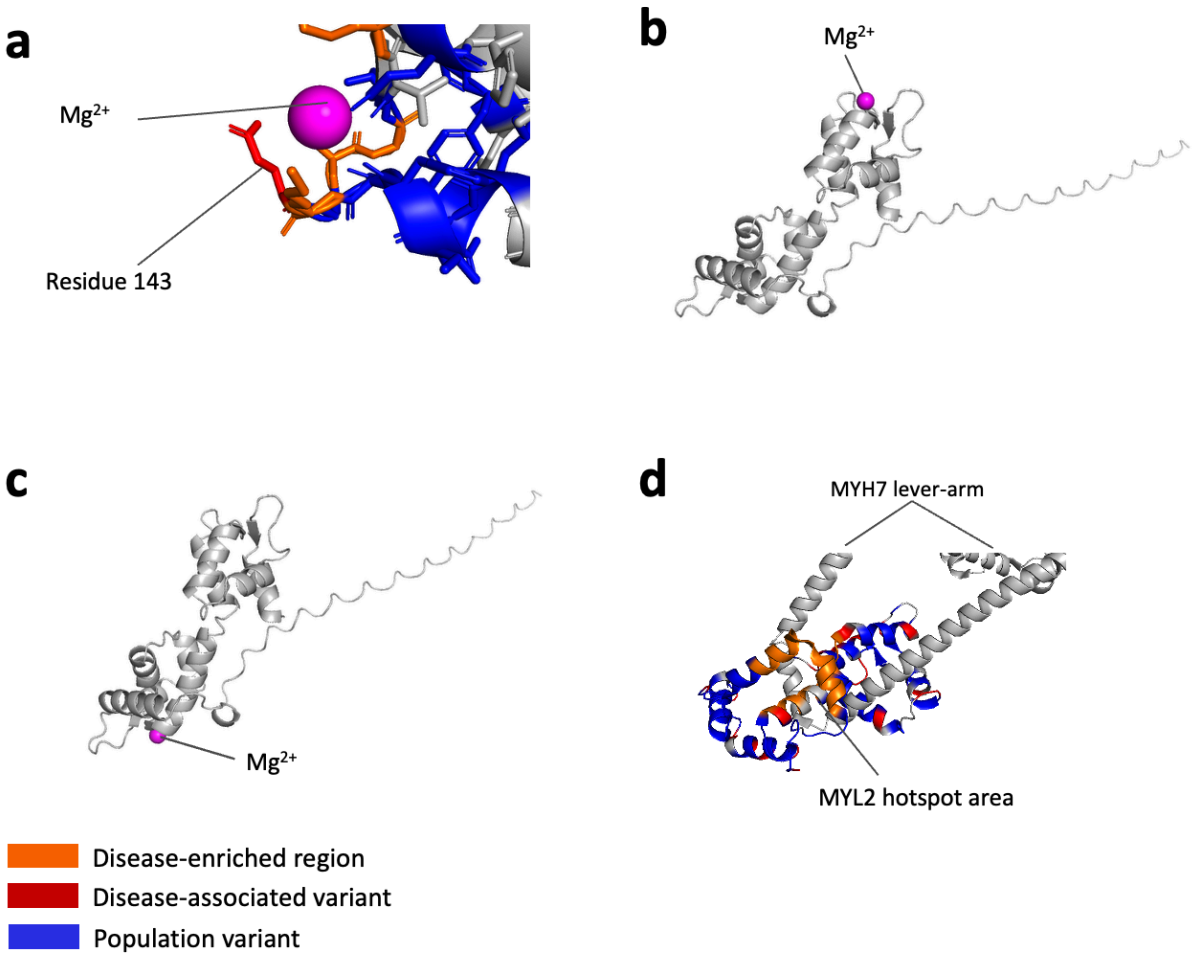

Fig. 3. *Structural models of MYL2 and MYL3 highlighting the disease-enriched region. (a) Close-up of the potential MYL3 cation interaction site, showing the proximity of residue 143 to the magnesium ion (Mg^{2+}). This variant might have its pathogenic effect by disrupting MYL3's interaction with cation. (b) For the MYL3 wild-type structure, AlphaFold 3 simulates an interaction near residue 143. (c) For the MYL3 structure with*

variant p.Glu143Lys, the interaction is moved. (d) Close-up of the MYL2 disease-associated region, showing its proximity to the MYH7 lever arm, suggesting that variants in this region may affect the interaction between MYL2 and MYH7 [24].

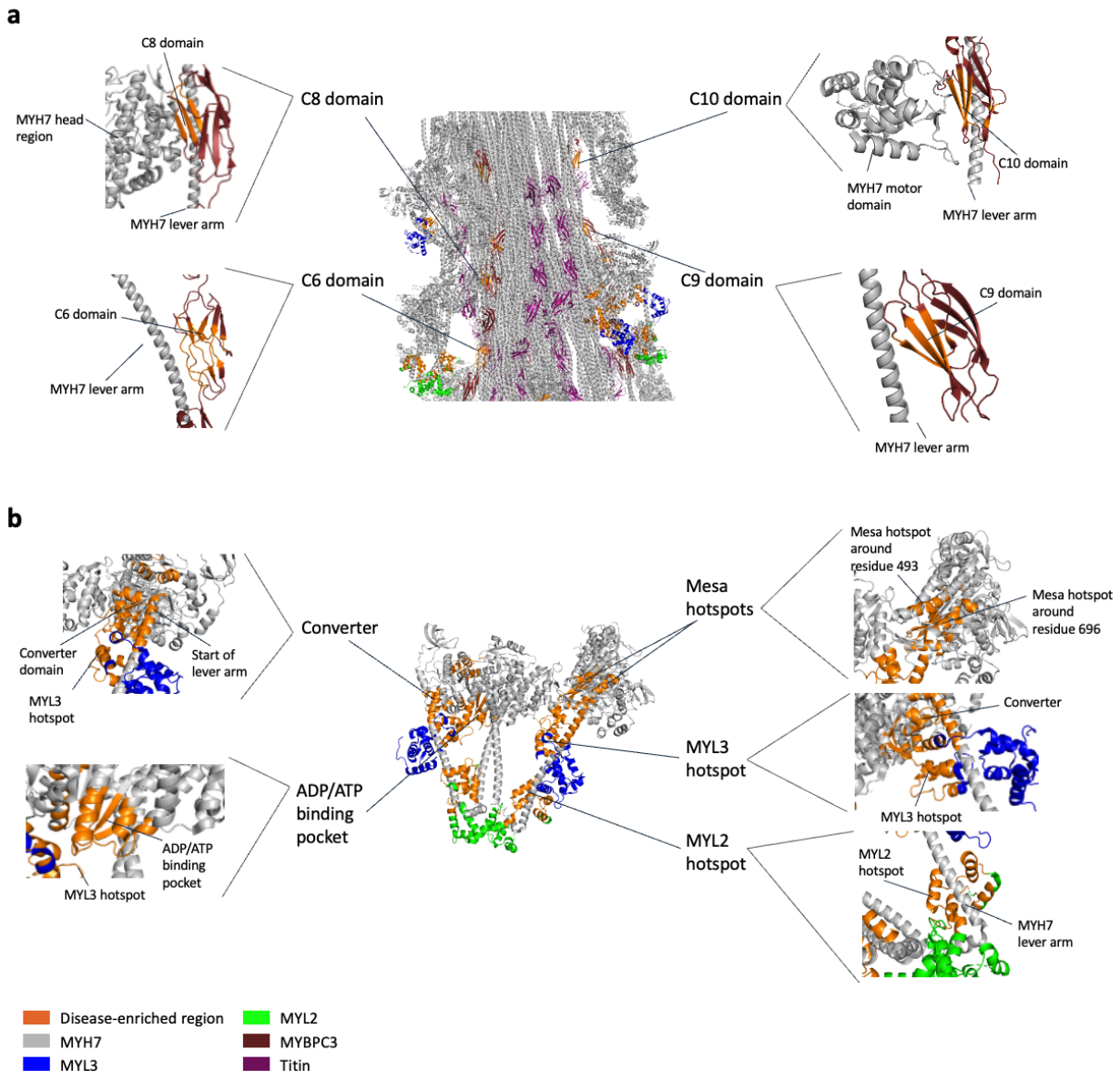

Figure 4. *(a) Cryogenic electron microscopy (Cryo-EM) structure of the human cardiac myosin filament with zoomed in visualizations of the regions identified on MYBPC3 [24]. The four disease-associated regions located in domains C6 (residue 857), C8 (residue 1030), C9 (residue 1102), and C10 (residue 1253) are highlighted, illustrating their close proximity to the myosin tails within the thick filament backbone. (b) Cryo-EM structure of*

the human beta-cardiac myosin folded-back off state displaying the regions identified on MYH7, MYL2 and MYL3 [22]. The highlighted regions indicate proximity to critical structural elements, including the converter, myosin mesa and transducer domains in MYH7, the lever-arm binding interface in MYL2, and the divalent cation-binding site in MYL3, underscoring their potential functional impact on myosin motor regulation.

Tables

Gene	No. Variants in Population Datasets	No. Patients with Variants in SHaRe	No. Variants observed in SHaRe	No. Pathogenicity Predictions
MYH7	1,893	1,332	363	38,700
MYBPC3	1,852	2,523	522	20,480
MYL2	274	103	30	3,320
MYL3	231	58	22	3,900
TNNT2	392	218	63	5,960
TNNI3	336	184	56	4,200

Table 1. *Overview of genetic variation and in silico pathogenicity predictions across six key sarcomeric genes. This table highlights the scale of the study, combining real-world clinical and population data from 800,000 individuals (gnomAD and UKBB) and 12,326 phenotyped HCM patients (SHaRe registry), alongside over 75,000 pathogenicity predictions scores to support structure-function analysis.*

Gene	Domain of center residue	P-value of enrichment
MYH7	Converter	0.004*
MYH7	Relay helix / myosin mesa (residue 493)	0.001
MYH7	SH1-SH2 helix / myosin mesa (residue 696)	0.001
MYH7	Transducer / near nucleotide-binding site	0.001*
MYBPC3	C6 - interacts with the	0.001

	myosin tail [24]	
MYBPC3	C8 - interacts with the IHM of myosin [24]	0.001
MYBPC3	C9 - interacts with the myosin tail [24]	0.001
MYBPC3	C10 - interacts with the IHM of myosin [24]	0.008
MYL2	Surrounding *MYH7* lever arm	0.005
MYL3	EF-hand 2, surrounding *MYH7* lever arm	0.001*
TNNT2	Tropomyosin-actin-binding site	0.001*
TNNT2	C-terminus	0.002*
TNNI3	Mobile domain	0.001*
TNNI3	Connecting switch to mobile domain	0.001
TNNI3	Switch domain	0.001

Table 2. *Significance assessments for identified variant enriched regions across different genes and protein domains. IHM: Interactive Head Motif*
**Region was also detectable using real-world clinical and population data only (prior to incorporating pathogenicity prediction scores into the analysis).*

References

1. Parikh, Victoria N., et al. "Advances in the study and treatment of genetic cardiomyopathies." Cell 188.4 (2025): 901-918.

2. S.J. Lehman, C. Crocini, L.A. Leinwand, Targeting the sarcomere in inherited cardiomyopathies. Nat. Rev. Cardiol. (2022).

3. S. García-Hernández, L. de la Higuera Romero, J.P. Ochoa, W.J. McKenna, Emerging themes in genetics of hypertrophic cardiomyopathy: Current status and clinical application. Can. J. Cardiol. (2024).

4. D. García-Giustiniani, M. Arad, M. Ortíz-Genga, R. Barriales-Villa, X. Fernández, I. Rodríguez-García, A. Mazzanti, E. Veira, E. Maneiro, P. Rebolo, et al., Phenotype and prognostic correlations of the converter region mutations affecting the beta myosin heavy chain. Heart (2015).

5. J.R. Homburger, E.M. Green, C. Caleshu, M.S. Sunitha, R.E. Taylor, K.M. Ruppel, R.P.R. Metpally, S.D. Colan, M. Michels, S.M. Day, et al., Multidimensional structure-function relationships in human β-cardiac myosin from population-scale genetic variation. Proc. Natl. Acad. Sci. U.S.A. 113, 6701–6706 (2016).

6. A.J. Marian, R. Roberts, The molecular genetic basis for hypertrophic cardiomyopathy. J. Mol. Cell. Cardiol. 33, 655–670 (2001).

7. J.D. Kapplinger, A.P. Landstrom, J.M. Bos, B.A. Salisbury, T.E. Callis, M.J. Ackerman, Distinguishing hypertrophic cardiomyopathy-associated mutations from background genetic noise. J. Cardiovasc. Transl. Res. 7, 347–361 (2014).

8. A.S. Helms, A.D. Thompson, A.A. Glazier, N. Hafeez, S. Kabani, J. Rodriguez, J.M. Yob, H. Woolcock, F. Mazzarotto, N.K. Lakdawala, et al., Spatial and functional distribution of MYBPC3 pathogenic variants and clinical outcomes in

patients with hypertrophic cardiomyopathy. Circ. Genomic Precis. Med. 13, 396–405 (2020).

9. R. Walsh, F. Mazzarotto, N. Whiffin, R. Buchan, W. Midwinter, A. Wilk, N. Li, L. Felkin, N. Ingold, R. Govind, 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. 11, 1–18 (2019).

10. A. Waring, A. Harper, S. Salatino, C. Kramer, S. Neubauer, K. Thomson, H. Watkins, M. Farrall, Data-driven modelling of mutational hotspots and in silico predictors in hypertrophic cardiomyopathy. J. Med. Genet. 58, 556–564 (2021).

11. L.M. Kurzlechner, E.G. Jones, A.M. Berkman, H.J. Tadros, J.A. Rosenfeld, Y. Yang, H. Tunuguntla, H.D. Allen, J.J. Kim, A.P. Landstrom, Signal-to-noise analysis can inform the likelihood that incidentally identified variants in sarcomeric genes are associated with pediatric cardiomyopathy. J. Pers. Med. 12, 733 (2022).

12. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritzel, L.H. Wong, M. Zielinski, T. Sargeant, et al., Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381, eadg7492 (2023).

13. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

14. J.R. Moore, L. Leinwand, D.M. Warshaw, Understanding cardiomyopathy phenotypes based on the functional impact of mutations in the myosin motor. Circ. Res. 111, 375–385 (2012).

15. M. Colegrave, M. Peckham, Structural implications of β-cardiac myosin heavy chain mutations in human disease. Anat. Rec. 297, 1670–1680 (2014).

16. M. Kawana, J.A. Spudich, K.M. Ruppel, Hypertrophic cardiomyopathy: Mutations to mechanisms to therapies. Front. Physiol. (2022).

17. J.A. Spudich, The myosin mesa and a possible unifying hypothesis for the molecular basis of human hypertrophic cardiomyopathy. Biochem. Soc. Trans. (2015).

18. D.N. Kim, T.M. Jacobs, B. Kuhlman, Boosting protein stability with the computational design of β-sheet surfaces. Protein Sci. 25, 702–710 (2016).

19. N. Louros, J. Schymkowitz, F. Rousseau, Mechanisms and pathology of protein misfolding and aggregation. Nat. Rev. Mol. Cell Biol. 24, 912–933 (2023).

20. J. Abramson, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 1–3 (2024).

21. M. Mirdita, et al., Colabfold: making protein folding accessible to all. Nat. Methods 19, 679–682 (2022).

22. A. Grinzato, et al., Cryo-EM structure of the folded-back state of human β-cardiac myosin. Biophys. J. 122, 258a–259a (2023).

23. M.H. Doran, et al., Myosin loop-4 is critical for optimal tropomyosin repositioning on actin during muscle activation and relaxation. J. Gen. Physiol. 155, e202213274 (2022).

24. D. Dutta, V. Nguyen, K.S. Campbell, R. Padrón, R. Craig, Cryo-EM structure of the human cardiac myosin filament. Nature 623, 853–862 (2023).

25. M.H. Doran, et al., Conformational changes linked to ADP release from human cardiac myosin bound to actin-tropomyosin. J. Gen. Physiol. 155, e202213267 (2023).

26. C.E. Friedman, et al., Multiplexed functional assessments of MYH7 variants in human cardiomyocytes. Circ. Genomic Precis. Med. 17, e004377 (2024).

27. Z. Cao, Y. Jia, B. Zhu, BNP and NT-proBNP as diagnostic biomarkers for cardiac dysfunction in both clinical and forensic medicine. Int. J. Mol. Sci. 20, 1820 (2019).

28. M. Varadi, et al., AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2022).

29. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. 35, 1026–1028 (2017).

30. C. Risi, et al., N-terminal domains of cardiac myosin binding protein C cooperatively activate the thin filament. Structure 26, 1604–1611 (2018).

31. C.M. Risi, et al., Interaction of the C2 Ig-like domain of cardiac myosin binding protein-C with F-actin. J. Mol. Biol. 433, 167178 (2021).

32. C.M. Risi, et al., Cryo-electron microscopy reveals cardiac myosin binding protein-C M-domain interactions with the thin filament. J. Mol. Biol. 434, 167879 (2022).

33. L. Alamo, et al., Effects of myosin variants on interacting-heads motif explain distinct hypertrophic and dilated cardiomyopathy phenotypes. Elife 6, e24634 (2017).

34. Y. Yamada, K. Namba, T. Fujii, Cardiac muscle thin filament structures reveal calcium regulatory mechanism. Nat. Commun. 11, 153 (2020).

35. E. Pavadai, W. Lehman, M.J. Rynkiewicz, Protein-protein docking reveals dynamic interactions of tropomyosin on actin filaments. Biophys. J. 119, 75–86 (2020).

36. A. Henrie, et al., ClinVar miner: demonstrating utility of a web-based tool for viewing and filtering ClinVar data. Hum. Mutat. 39, 1051–1060 (2018).

37. K Nieznanski, H Nieznanska, K Skowronek, AA Kasprzak, D Stepkowski, Ca2+ binding to myosin regulatory light chain affects the conformation of the n-terminus of essential light chain and its binding to actin. *Arch. biochemistry biophysics* 417, 153–158 (2003).

38. D Szczesna-Cordary, PP de Tombe, Myosin light chain phosphorylation, novel targets to repair a broken heart? (2016).

39. J Lossie, et al., Mutations of ventricular essential myosin light chain disturb myosin binding and sarcomeric sorting. *Cardiovasc. Research* 93, 390–396 (2012).

40. CC Yuan, et al., Hypercontractile mutant of ventricular myosin essential light chain leads to disruption of sarcomeric structure and function and results in restrictive cardiomyopathy in mice. *Cardiovasc. Research* 113, 1124–1136 (2017).

41. D Dumka, et al., E22k mutation of rlc that causes familial hypertrophic cardiomyopathy in heterozygous mouse myocardium: effect on cross-bridge kinetics. *Am. J. Physiol. Circ. Physiol*. 291, H2098–H2106 (2006).

42. Kawana, Masataka, et al. Biophysical properties of human β-cardiac myosin with converter mutations that cause hypertrophic cardiomyopathy. Science advances 3.2 (2017)

43. Adhikari, Arjun S., et al. "β-Cardiac myosin hypertrophic cardiomyopathy mutations release sequestered heads and increase enzymatic activity." Nature communications 10.1 (2019)

44. Sommese, Ruth F., et al. "Molecular consequences of the R453C hypertrophic cardiomyopathy mutation on human β-cardiac myosin motor function." Proceedings of the National Academy of Sciences 110.31 (2013)

45. Yadav, Sunil, et al. "Hereditary heart disease: pathophysiology, clinical presentation, and animal models of HCM, RCM, and DCM associated with

mutations in cardiac myosin light chains." Pflügers Archiv-European Journal of Physiology 471 (2019)

46. Farman, Gerrie P., et al. "Impact of familial hypertrophic cardiomyopathy-linked mutations in the NH2 terminus of the RLC on β-myosin cross-bridge mechanics." Journal of Applied Physiology 117.12 (2014)

47. Kelly, Melissa 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." Genetics in Medicine 20.3 (2018)

48. Olivotto, Iacopo, et al. "Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, phase 3 trial." The Lancet 396.10253 (2020): 759-769

49. Maron, Martin S., et al. "Aficamten for symptomatic obstructive hypertrophic cardiomyopathy." New England Journal of Medicine 390.20 (2024): 1849-1861

50. Richards, Sue, 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." Genetics in medicine 17.5 (2015): 405-423.

51. Chen, Liang, et al. "Structure of mavacamten-free human cardiac thick filaments within the sarcomere by cryoelectron tomography." Proceedings of the National Academy of Sciences 121.9 (2024)

52. Risi, Cristina M., et al. "High-resolution cryo-EM structure of the junction region of the native cardiac thin filament in relaxed state." PNAS nexus 2.1 (2023)

53. Thompson, Andrea D., Adam S. Helms, Anamika Kannan, Jaime Yob, Neal K. Lakdawala, Samuel G. Wittekind, Alexandre C. Pereira, et al. "Computational Prediction of Protein Subdomain Stability in MYBPC3 Enables Clinical Risk

Stratification in Hypertrophic Cardiomyopathy and Enhances Variant Interpretation.” Genetics in Medicine 23, no. 7 (July 2021): 1281–87

54. Houdusse, Anne, and H. Lee Sweeney. “How Myosin Generates Force on Actin Filaments.” Trends in Biochemical Sciences 41, no. 12 (December 2016): 989–97

55. Grabarek, Zenon. "Structural basis for diversity of the EF-hand calcium-binding proteins." Journal of molecular biology 359.3 (2006): 509-525.

56. Robert-Paganin, Julien, et al. "Force generation by myosin motors: a structural perspective." Chemical reviews 120.1 (2019): 5-35.

57. Robert-Paganin, Julien, Daniel Auguin, and Anne Houdusse. “Hypertrophic Cardiomyopathy Disease Results from Disparate Impairments of Cardiac Myosin Function and Auto-Inhibition.” Nature Communications 9, no. 1 (December 2018): 4019

58. Olson, Timothy M., Margaret L. Karst, Frank G. Whitby, and David J. Driscoll. “Myosin Light Chain Mutation Causes Autosomal Recessive Cardiomyopathy With Mid-Cavitary Hypertrophy and Restrictive Physiology.” Circulation 105, no. 20 (May 21, 2002): 2337–40

59. Caleshu, Colleen, Rahul Sakhuja, Robert L. Nussbaum, Nelson B. Schiller, Philip C. Ursell, Celeste Eng, Teresa De Marco, Dana McGlothlin, Esteban González Burchard, and J. Eduardo Rame. “Furthering the Link between the Sarcomere and Primary Cardiomyopathies: Restrictive Cardiomyopathy Associated with Multiple Mutations in Genes Previously Associated with Hypertrophic or Dilated Cardiomyopathy.” American Journal of Medical Genetics Part A 155, no. 9 (2011): 2229–35

60. Rehm, Heidi L., et al. "ClinGen—the clinical genome resource." New England Journal of Medicine 372.23 (2015): 2235-2242

61. Tran, T.N., Fong, C., Pichotta, K. *et al.* AI cancer driver mutation predictions are valid in real-world data. *Nat Commun* 16, 8509 (2025). https://doi.org/10.1038/s41467-025-63461-8

62. Zhang, X., Walsh, R., Whiffin, N. *et al.* Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *Genet Med*23, 69–79 (2021). https://doi.org/10.1038/s41436-020-00972-3