BMC Medical Genomics

RESEARCH Open Access

Exploring novel *MYH7* gene variants using in silico analyses in Korean patients with cardiomyopathy

Oc-Hee Kim[1], Jihyun Kim[1], Youngjun Kim[1], Soyoung Lee[2], Beom Hee Lee[3], Bong-Jo Kim[1], Hyun-Young Park[4] and Mi-Hyun Park[1*]

Abstract

Background Pathogenic variants of *MYH7*, which encodes the beta-myosin heavy chain protein, are major causes of dilated and hypertrophic cardiomyopathy.

Methods In this study, we used whole-genome sequencing data to identify *MYH7* variants in 397 patients with various cardiomyopathy subtypes who were participating in the National Project of Bio Big Data pilot study in Korea. We also performed in silico analyses to predict the pathogenicity of the novel variants, comparing them to known pathogenic missense variants.

Results We identified 27 *MYH7* variants in 41 unrelated patients with cardiomyopathy, consisting of 20 previously known pathogenic/likely pathogenic variants, 2 variants of uncertain significance, and 5 novel variants. Notably, the pathogenic variants predominantly clustered within the myosin motor domain of MYH7. We confirmed that the novel identified variants could be pathogenic, as indicated by high prediction scores in the in silico analyses, including SIFT, Mutation Assessor, PROVEAN, PolyPhen-2, CADD, REVEL, MetaLR, MetaRNN, and MetaSVM. Furthermore, we assessed their damaging effects on protein dynamics and stability using DynaMut2 and Missense3D tools.

Conclusions Overall, our study identified the distribution of *MYH7* variants among patients with cardiomyopathy in Korea, offering new insights for improved diagnosis by enriching the data on the pathogenicity of novel variants using in silico tools and evaluating the function and structural stability of the MYH7 protein.

Keywords Cardiomyopathy, MYH7, Whole-genome sequencing, In silico prediction

*Correspondence:
Mi-Hyun Park
mihyun4868@korea.kr
[1]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju 28159, Republic of Korea
[2]Department of Pediatrics, Hallym University Sacred Heart Hospital, Anyang 14068, Republic of Korea
[3]Medical Genetics Center, Asan Medical Center Children's Hospital, University of Ulsan College of Medicines, Seoul 05505, Republic of Korea
[4]National Institute of Health, Cheongju 28159, Republic of Korea

Introduction

Cardiomyopathy, a cardiac muscle disease, is associated with impaired cardiac function and classified into hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), restrictive cardiomyopathy (RCM), and undefined cardiomyopathy. HCM is characterized by unexplained left ventricular hypertrophy and a non-dilated left ventricle with a preserved or increased ejection fraction [1]. DCM is characterized by cardiac dilatation, predominantly of the left ventricle, and a beyond-repair weakening of the heart muscle, resulting in reduced contractile function and inefficient heart pumping [2].

The majority of cardiomyopathies, including HCM and DCM, are genetic disorders inherited in an autosomal-dominant manner [3, 4]. Mutations in genes encoding thick filament components such as myosin heavy chain 7 (*MYH7*) and myosin-binding protein C3 (*MYBPC3*) account for 75% of inherited HCM cases, establishing HCM as a sarcomere disease [3, 5–7]. In contrast, DCM exhibits greater genetic heterogeneity, with mutations occurring in genes encoding cytoskeletal, nucleoskeletal, mitochondrial, and calcium-handling proteins [4, 7, 8]. Over 50 genes have been linked to both HCM and DCM [9].

The *MYH7* gene, located on chromosome 14, encodes β-myosin heavy chain [10]. In cardiac muscle cells, each type II myosin protein consists of two heavy chains (encoded by the *MYH7* gene) and two pairs of regulatory light chains (encoded by several other genes) [10]. The MYH7 protein comprises a head region (also known as the myosin motor domain), which interacts with actin and ATP and is crucial for the protein's motor activity, and a long tail region that interacts with other proteins, including the tail regions of myosin proteins. Most variants are located within the head region, potentially affecting the actin-binding sites [10, 11].

Pathogenic *MYH7* variants rank as the second most common inherited cause of HCM and third most common inherited cause of DCM [12]. These variants, predominantly missense in nature, are dominantly inherited, although instances of *de novo* variants have also been reported [13, 14]. Approximately 2205 missense *MYH7* variants are listed in the ClinVar database, with 372 disease-causing variants and 1833 variants of uncertain significance (VUS) or conflicting interpretations of their pathogenicity despite the set specific pathogenicity criteria for this gene in February 2024 [15].

Here, we present the frequency and distribution of *MYH7* variants in a diverse cohort of unrelated patients with cardiomyopathy from the National Project of Bio Big Data pilot study in Korea. We performed in silico analyses to predict the pathogenicity of novel variants in comparison with known pathogenic/likely pathogenic variants. Furthermore, we assessed the damaging effects of novel *MYH7* variants on protein dynamics and stability in DynaMut2 and Missense3D. The results of this study provide new insights into genetic diagnosis by examining the pathogenicity of novel *MYH7* variants using in silico techniques to evaluate the function and structural stability of the MYH7 protein.

Materials and methods

Study participants and general information

Study participants and clinical information regarding cardiomyopathy were obtained from the National Project of Bio Big Data Pilot Study conducted in Korea between August 2020 and December 2022 [16]. The National Project of Bio Big Data is a national project that aims to implement precision medicine and national health promotion; it has generated whole-genome sequencing (WGS) data, which consisted of singleton, duo, trio, and more than trio proband families, clinical information, and lifestyle data of the Korean population. In the rare disease study of this project, approximately 15 000 patients with rare diseases and their families were recruited for genetic diagnosis as well as for finding new genetic factors related to rare disorders based on the WGS analysis. We used variant calling and filtering data of 397 patients with cardiomyopathy (both familial and nonfamilial cases) in the National Project of Bio Big Data pilot study. Informed consent was obtained from all the participants and parents of patients before conducting the genetic tests. The Institutional Review Board of the Human Research of the Korea National Institute of Health, Korea Disease Control and Prevention Agency approved this study (Approval No. 2022-09-10-P-A, 2022-02-07–2C-A, KDCA-2023-06-06-P-01). This study was performed according to the Declaration of Helsinki.

Identification of *MYH7* variants in patients with cardiomyopathies

The workflow of the study is outlined in Fig. 1. DCM, HCM, undefined cardiomyopathy, ARVC, DCM and conduction defects, left ventricular noncompaction cardiomyopathy, pediatric or syndromic cardiomyopathy, and progressive cardiac conduction disease was detected in 218 (54.9%), 153 (38.5%), 14 (3.5%), 4 (1.0%), 4 (1.0%), 2 (0.5%), 1 (0.3%), and 1 (0.3%) patients, respectively (Table 1). We manually identified all annotated files from the WGS data of 397 patients with cardiomyopathy to identify *MYH7* (NM_000257.4) variants. Variants with allele frequency>0.001 in gnomAD (https://gnomad.broadinstitute.org/) were filtered out. Among them, synonymous, likely benign, benign, and noncoding variants were excluded. Filtered *MYH7* variants were searched in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) and classified as pathogenic, likely pathogenic,

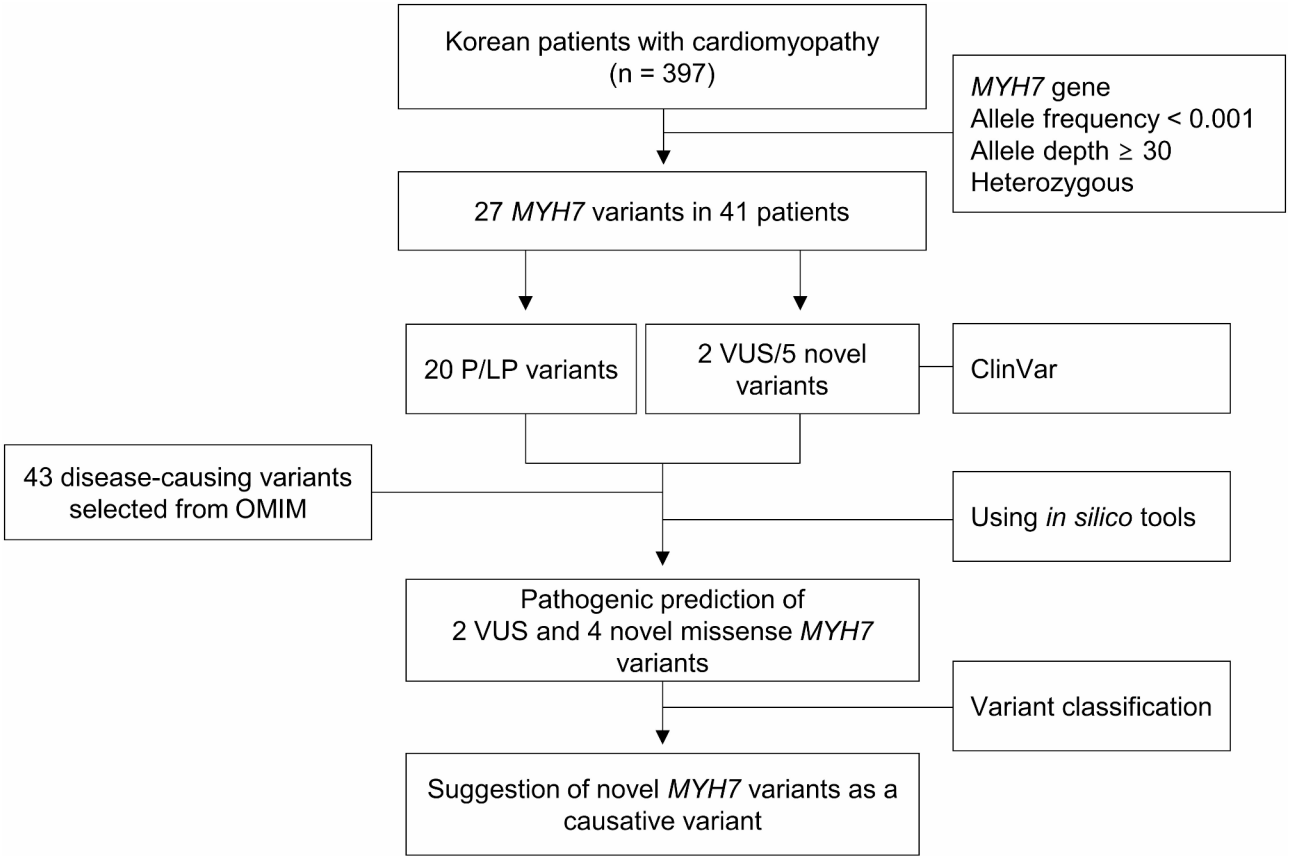

Fig. 1 Workflow of this study. P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance. OMIM, Online Mendelian Inheritance in Man

Table 1 Number of patients with cardiomyopathy carrying *MYH7* variants

Category	Number of patients with MYH7 variants	Total number of patients
Arrhythmogenic right ventricular cardiomyopathy	0	4
Dilated cardiomyopathy	14	218
Dilated cardiomyopathy and conduction defects	0	4
Hypertrophic cardiomyopathy	24	153
Left ventricular noncompaction cardiomyopathy	0	2
Pediatric or syndromic cardiomyopathy	0	1
Progressive cardiac conduction disease	0	1
Undefined	3	14
	41	397

or of uncertain significance, according to the guidelines of the American College of Medical Genetics and Genomics (ACMG). Representative cardiomyopathy-related *MYH7* variants were obtained from the Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org/). Criteria for inclusion include the first mutation to be discovered, high population frequency, distinctive phenotype, historic significance, unusual mechanism of mutation, unusual pathogenetic mechanism, and distinctive inheritance. Most of the variants represent disease-causing mutations, although a few polymorphisms are included. [17].

In silico prediction of *MYH7* missense variants

The pathogenic effects of *MYH7* missense variants were identified using several online tools such as CADD (https://cadd.gs.washington.edu/snv) [18], PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) [19], SIFT [20], MutationAssessor (http://mutationassessor.org/r3/) [21], PROVEAN [22], REVEL [23], MetaLR, MetaRNN [24], and MetaSVM. VarSome [25], a search engine for human genomic variation, was used to classify *MYH7* variants according to the ACMG guidelines [26].

Prediction of changes in protein stability and interaction due to *MYH7* missense variants

Changes in protein stability and interactions of *MYH7* missense variants were predicted using DynaMut2 (http://biosig.unimelb.edu.au/dynamut/) [27] and Missense3D (http://missense3d.bc.ic.ac.uk/missense3d/) [28]. DynaMut2 predicts protein dynamics and flexibility, whereas Missense3D predicts structural changes induced by amino acid substitutions. Experimental (PDB: 5tby)

and predicted (AlphaFold: AF-P12883-F1) structures of the human MYH7 protein were used as inputs.

Statistical analyses

Statistical analyses were performed using GraphPad Prism 8.0.2 for Windows (GraphPad Software; www.graphpad.com). Data are presented as the mean ± standard deviation (SD) of all analyses. Tukey's multiple comparisons test was used for all one-way analyses of variance (ANOVAs) with repeated measures. The level of statistical significance was set at $p < 0.05$. Asterisks * in the figures represent significance levels of $p < 0.05$.

Results

General characteristics of patients with cardiomyopathies carrying *MYH7* variants

To identify *MYH7* variants, we selected 397 patients with cardiomyopathy among the 15 000 participants with diverse rare diseases from the National Project of Bio Big Data pilot study in Korea. Using our filtering strategy, we identified 27 variants within the coding region of *MYH7* in 41 unrelated patients (Fig. 1). Forty-one patients, including 22 men (53.7%) and 19 women (46.3%), had various subtypes of cardiomyopathy, including HCM (59%), DCM (34%), and undefined (7%) (as detailed in Table 1). The mean ages of male and female patients were 47.4 ± 14.6 and 43.9 ± 20.0 years, respectively, with no difference in age between males and females. *MYH7* variants were identified in 33 singletons (80.5%), 4 duos (9.8%), 1 trio (2.4%), and 3 quartets (7.3%) (Supplementary Table 1).

General information on the *MYH7* variants identified in this study

In this study, we identified 27 potentially causative *MYH7* variants in 41 patients with cardiomyopathy. We utilized the ClinVar and Clinical Genome Resource (ClinGen) databases, which are public archives of human genetic variants, to confirm the clinical significance of the 27 *MYH7* variants identified in this study for cardiomyopathy. Consequently, the identified *MYH7* variants were classified into 20/27 (74%) known pathogenic or likely pathogenic variants, 2/27 (7%) VUS, and 5/27 (19%) novel variants (Fig. 1 and Supplementary Table 1) based on the ClinVar database. The following five novel variants were not previously reported in ClinVar: p.Glu55Lys, p.Tyr134His, p.Phe488Leu, p.Gln1318Leu, and p.Asn1918_Ala1922del (Supplemental Table 1). The following seven variants were included in the ClinGen Cardiomyopathy Variant Curation Expert Panel: p.Arg369Gln, p.Arg403Gln, p.Arg403Trp, p.Arg663His, p.Arg870Cys, p.Glu1356Lys, and p.Thr1377Met (Supplemental Table 1). Among the 14 patients with DCM, 10 (71%) had pathogenic or likely pathogenic variants, and 4 (29%) had VUS or novel variants (Supplementary Table 1). Moreover, among the 24 patients with HCM, 22 (92%) had pathogenic or likely pathogenic variants and 2 (8%) had novel variants (Supplementary Table 1). One of the three patients with undefined cardiomyopathy had a VUS, and the other had a novel variant. The remaining one patient with left ventricular noncompaction cardiomyopathy and progressive cardiac conduction disease had a likely pathogenic variant (Supplemental Table 1).

Distribution in domain and frequency of *MYH7* variants

The 27 identified *MYH7* variants comprised 26 missense variants and a small-deletion variant of 15 base pairs (see Supplementary Table 1). Known pathogenic or likely pathogenic variants of *MYH7* were predominantly located within the myosin motor domain of the MYH7 protein (15/27, 55.6%) (Fig. 2). Two pathogenic variants were located within the IQ domain (2/27, 7.4%), and seven pathogenic or likely pathogenic variants were located in the coiled-coil domain (9/27, 33.4%) (Fig. 2). Among these variants, p.Arg249Gln emerged as the most prevalent, being present in four unrelated DCM and two unrelated HCM cases in this study, followed by p.Arg663His in four unrelated HCM cases and p.Thr1377Met in three unrelated HCM cases (Fig. 3). The p.Tyr134His variant was identified in two unrelated patients with HCM or undefined cardiomyopathy, with no report of its pathogenicity in ClinVar or other databases (Fig. 3 and Supplementary Table 1). The p.Leu476Phe and p.Glu1356Lys variants were identified in two unrelated patients with HCM or DCM (Fig. 3 and Supplementary Table 1).

In silico prediction of *MYH7* variants

We performed in silico analysis to evaluate the pathogenic effects of the VUS and novel *MYH7* missense variants, including the p.Glu55Lys, p.Tyr134His, p.Ser148Arg, p.Phe488Leu, p.Arg807Cys, and p.Gln1318Leu variants. We compared these missense variants with the known pathogenic *MYH7* variants identified in our study and missense variants selected from the OMIM database. We scored the VUS and novel missense *MYH7* variants using SIFT, Mutation Assessor, and PROVEAN to evaluate their interspecific and evolutionary sequential conservation. The VUS and novel *MYH7* variants had pathogenicity scores < 0.05 in SIFT, with the exception of the p.Gln1318Leu variant (Fig. 4a). The *MYH7* p.Tyr134His and p.Arg807Cys variants were associated with a high risk (scores > 3.5) regarding a deleterious effect on MYH7 function as altered missense variants in Mutation Assessor, whereas the other variants had a low impact (scores ≤ 1.9) (Fig. 4b). In PROVEAN, all variants excluding p.Glu55Lys and p.Ser148Arg indicated pathogenicity with respect to MYH7 function (Fig. 4c). We further

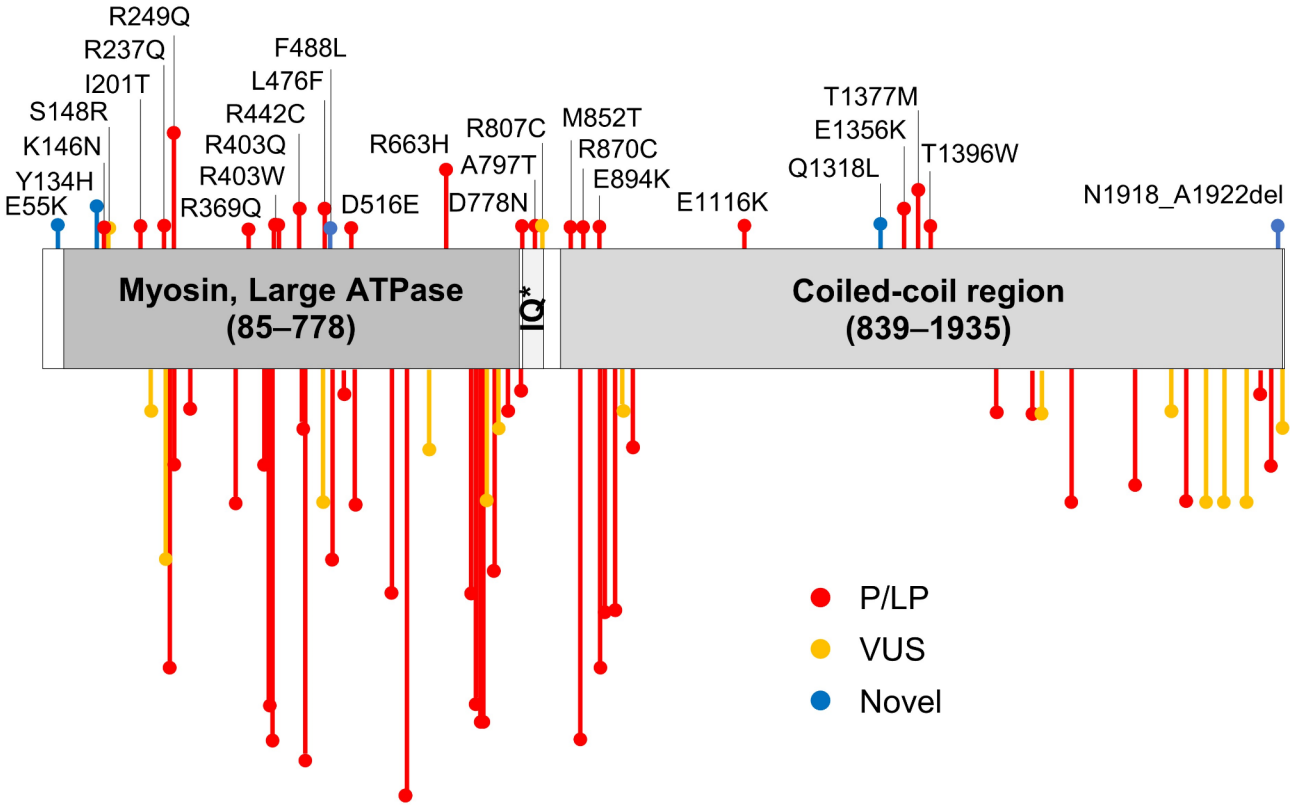

Fig. 2 Distribution of all *MYH7* variants identified in this study. All variants shown are annotated according to the human genome GRCh38, using the MYH7 protein accession number NP_000248.2. Variants identified in this study are shown above the protein domain, and OMIM variants are located below. Pathogenic (P)/likely pathogenic (LP) variants reported as causal variants in ClinVar are shown in red, variants of uncertain significance (VUS) are shown in yellow, and novel variants are shown in blue. MYH7 contains functional domains such as the myosin motor, IQ, and coiled-coil domain

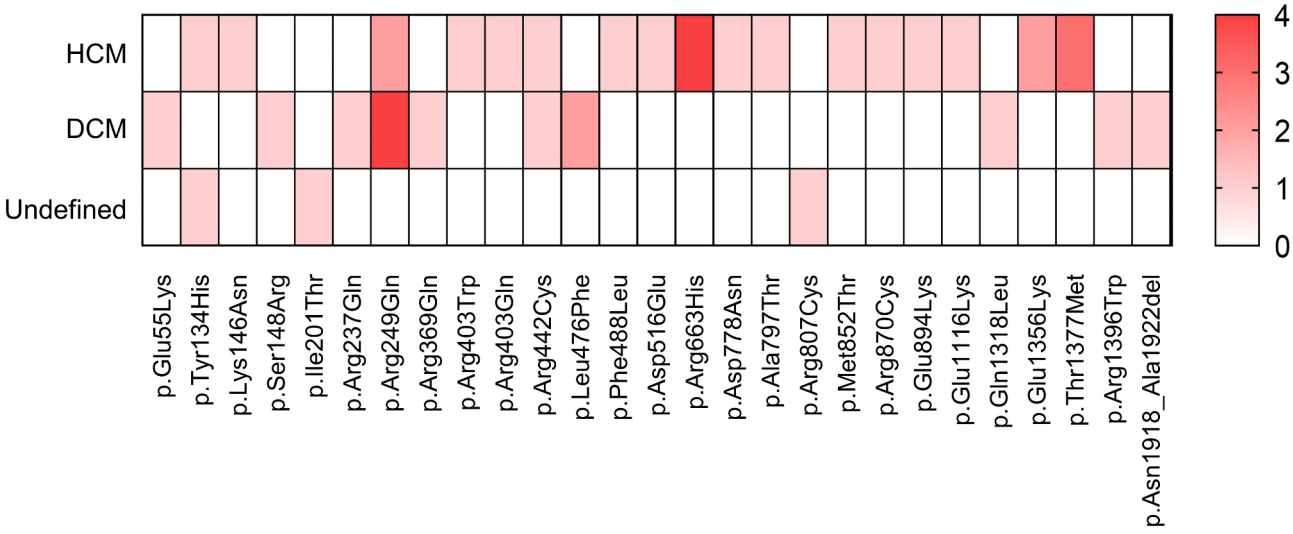

Fig. 3 Heat map demonstrating the frequency of *MYH7* variants identified in this study. The frequency of *MYH7* variants ranges from white (0) to red (4), with diverse subtypes of cardiomyopathy indicated on the *y*-axis. All *MYH7* variants are clustered across the *x*-axis. DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy

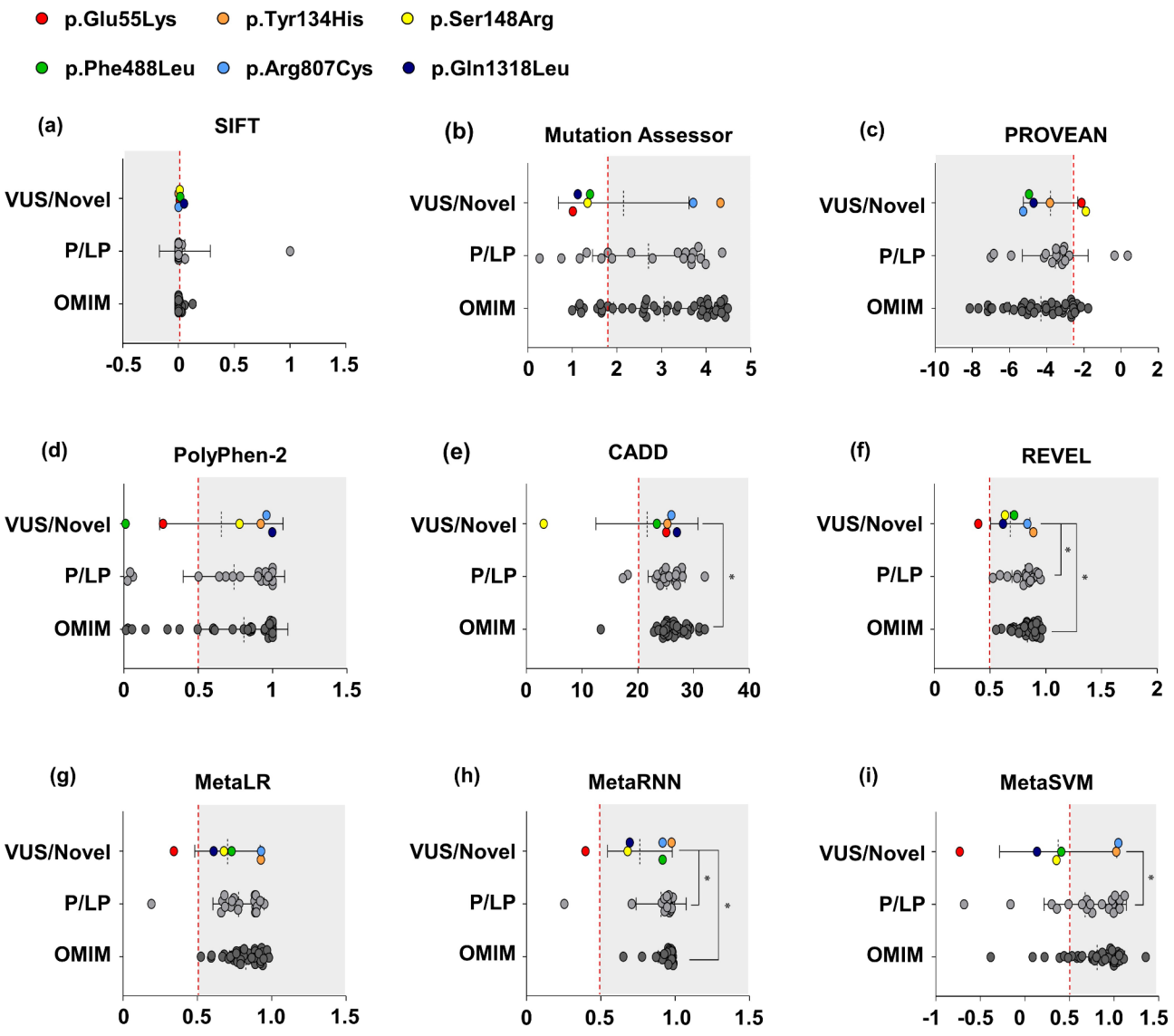

Fig. 4 In silico predictive scores of *MYH7* variants. Comparison of predicted pathogenicity scores for pathogenic (P)/likely pathogenic (LP) and VUS (variants of uncertain significance)/novel missense *MYH7* variants in this study, with missense variants identified in OMIM. The scatter dot plots show the values for various predictive scores: (a) SIFT, (b) Mutation assessor, (c) PROVEAN, (d) PolyPhen-2, (e) CADD (GRCh38-v1.6), (f) REVEL, (g) MetaLR, (h) MetaRNN, and (i) MetaSVM scores. The pathogenicity score threshold based on a previous study [41] is presented in the gray area of each graph. Data are presented as the mean ± standard deviation. Statistical significance in e, f, h, and i was calculated using ordinary one-way ANOVA in GraphPad Prism 8.0.2. * = $p < 0.05$

scored the *MYH7* variants using PolyPhen-2 to evaluate protein structural alterations. We found that all variants excluding p.Glu55Lys and p.Phe488Leu had a "probably" (scores ≥ 0.957) or "possibly" ($0.453 \leq x \leq 0.956$) damaging effect on MYH function (Fig. 4d). Finally, we scored the *MYH7* variants using supervised machine learning analysis with the CADD, REVEL, MetaLR, MetaRNN, and MetaSVM tools (Fig. 4e–i). The novel *MYH7* variants, excluding p.Glu55Lys, were predicted to have damaging effects on protein function in REVEL, MetaLR, and MetaRNN (Fig. 4f-h). The *MYH7* p.Ser148Arg variant was predicted to have a benign effect in CADD, whereas other variants exhibited pathogenic scores in CADD (Fig. 4e). The *MYH7* p.Tyr134His and p.Arg807Cys variants had pathogenic scores in MetaSVM (Fig. 4i). The pathogenicity scores of the VUS and novel variants were similar to those of the known pathogenic variants in SIFT, Mutation Assessor, PROVEAN, PolyPhen-2, and MetaLR (Fig. 4a–d **and g**). However, the VUS and novel variants had significantly lower scores compared with those of known pathogenic variants in CADD ($p=0.0190$ vs. OMIM), REVEL ($p=0.0058$ vs. OMIM; $p=0.0334$ vs. P/LP), MetaRNN ($p=0.0019$ vs. OMIM; $p=0.0298$ vs. P/LP), and MetaSVM ($p=0.0338$ vs. P/LP)(Fig. 4, Supplementary Tables 2 and 3). In particular, the p.Tyr134His and p.Arg807Cys variants showed high pathogenicity in

the in silico analyses, suggesting that they may be possible causative variants of cardiomyopathy (Fig. 4 and Table 2).

Next, we utilized the DynaMut2 and Missense3D web servers to predict the effect of these *MYH7* variants on protein structure and visualize the residues of *MYH7* missense variants. The p.Tyr134His (−1.23 kcal/mol), p.Ser148Arg (−0.36 kcal/mol), p.Phe488Leu (−2.08 kcal/mol), and p.Arg807Cys (−0.49 kcal/mol) variants had relatively lower ΔΔG values, implying reduced protein stability, although the p.Glu55Lys (0.57 kcal/mol) and p.Gln1318Leu (0.4 kcal/mol) variants were predicted to not affect protein stability in DynaMut2 (Fig. 5; Table 2, and Supplementary Table 4). In addition, the p.Tyr134His (156.816 Å^3) and p.Ser148Arg (111.672 Å^3) variants led to alterations in cavity volume, whereas the other variants had no effect on MYH7 protein structure as predicted in Missense3D (Supplementary Table 4).

Classification of novel *MYH7* variants

Based on population data, location, in silico pathogenic prediction, and phenotype associations, we manually classified the VUS and novel *MYH7* variants using the ACMG/Association Molecular Pathology (AMP) framework (Supplementary Table 5). In patient 1, the p.Glu55Lys variant was predicted as a VUS with levels of evidence PM2 and BP4. In patients 2 and 3, the p.Tyr134His variant was predicted as likely pathogenic with levels of evidence PP3, PM1, PM2, and PS4. In patient 4, the p.Ser148Arg variant was predicted as pathogenic with levels of evidence PS1, PM1, PM5, and PM2. In patient 20, the p.Phe488Leu variant was predicted as a VUS with levels of evidence PM1, PP3, and PM2. In patient 28, the p.Arg807Cys variant was predicted as a VUS with levels of evidence PM5, PP3, and PM2. In patient 34, the p.Gln1318Leu variant was predicted as a VUS with levels of evidence PM2. In patient 42, the p.Asn1918_Ala1922del variant was predicted as a VUS with levels of evidence PM4 and PM2. We also used VarSome [25], an automated engine, for classifying the VUS and novel *MYH7* variants according to ACMG guidelines [26]. The novel p.Tyr134His, p.Ser148Arg, p.Arg807Cys, and p.Asn1918_Ala1922del variants were predicted as likely pathogenic, whereas the p.Glu55Lys, p.Phe488Leu, and p.Gln1318Leu variants were predicted as VUS (Supplementary Table 5).

Discussion

In this study, we identified 27 *MYH7* variants in 41 unrelated cases in a cohort of 397 patients with cardiomyopathy from the National Project of Bio Big Data pilot study in Korea. These *MYH7* variants included 20 known pathogenic/likely pathogenic, 2 VUS and 5 novel variants of cardiomyopathy. We performed in silico analysis

Table 2 In silico scores and pathogenic prediction of novel *MYH7* variants

Patient ID	Protein change	CADD (score/prediction)	PolyPhen-2 (score/prediction)	SIFT (score/prediction)	MetaRNN (score/prediction)	DynaMut (score/prediction)
1	p.Glu55Lys	25.1/damaging	0.264/benign	0.011/damaging	0.3973/benign	0.57/stabilizing
2, 3	p.Tyr134His	25.3/damaging	0.920/probably damaging	0/damaging	0.9752/damaging	−1.23/ destabilizing
4	p.Ser148Arg	3.1/benign	0.778/possibly damaging	0.012/damaging	0.6805/damaging	−0.36/ destabilizing
20	p.Phe488Leu	23.4/damaging	0.012/benign	0.016/damaging	0.9151/damaging	−2.08/ destabilizing
28	p.Arg807Cys	26.0/damaging	0.959/probably damaging	0/damaging	0.9158/damaging	−0.49/ destabilizing
34	p.Gln1318Leu	27.0/damaging	0.998/probably damaging	0.051/benign	0.6946/damaging	0.4/stabilizing

The pathogenicity score threshold is based on a previous study [41]

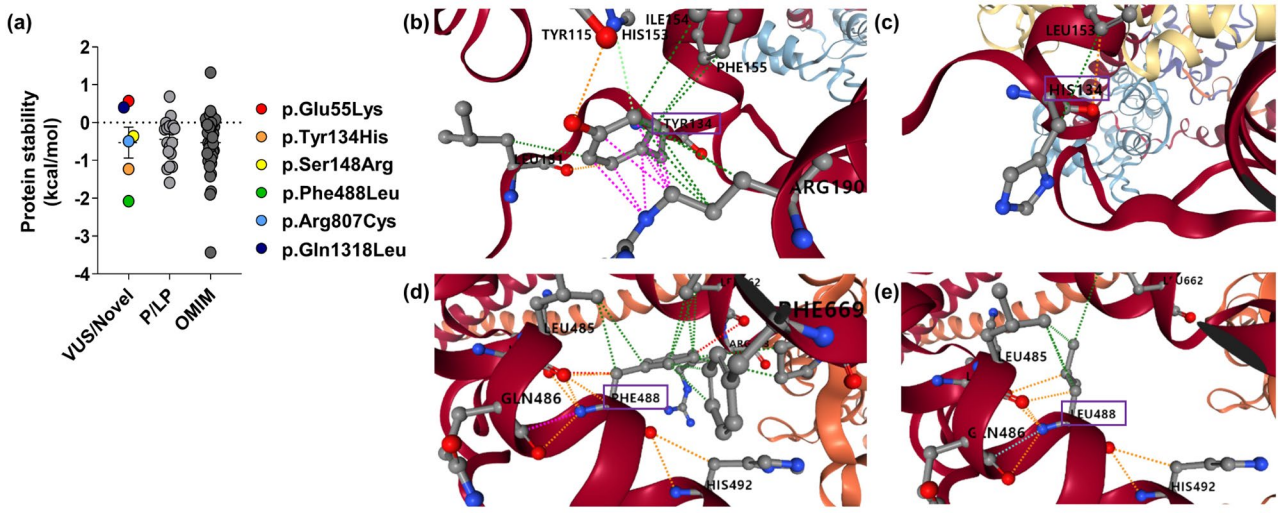

Fig. 5 DynaMut2 prediction of interatomic interactions of the native *MYH7* vs. the variants. **(a)** Prediction scores of changes in protein stability using the DynaMut2 server. Interatomic interaction of *MYH7* variants visualized using DynaMut2 in (**b**, **c**) p.Tyr134His and (**d**, **e**) p.Phe488Leu. Wild-type and mutant residues are indicated in a purple rectangle and represented as sticks along with the surrounding residues involved in the interaction. Dot points represent the interaction with the following color codes: pink, clash; light blue, VDW; red, hydrogen bond; yellow, ionic; light green, aromatic; green, hydrophobic; blue, carbonyl contacts

of the VUS and novel *MYH7* variants to understand their potential pathogenic contributions to cardiomyopathy.

Of the 27 variants identified, 26 were missense variants. Cardiomyopathy is well-established to be predominantly caused by missense variants in the *MYH7* gene [29]. Notably, 15 variants in this study clustered within the myosin motor domain spanning positions 85 to 778, which is critical for MYH7 function. Variants in the myosin motor domain are likely to negatively affect the interaction with actin, which is essential for converting ATP into mechanical energy. Supporting this finding, previous studies showed that the p.Arg369His variant led to decreased maximal myosin binding to F-actin without affecting ATPase activity in vitro. This variant also reduced locomotion in adult *Drosophila*, indicating that the p.Arg369His variant is critical for both actin interaction and muscle function [30]. In addition, a zebrafish knockout model of the myosin motor domain of *MYH7* generated using the CRISPR/Cas9 genome editing system exhibited abnormal heart development [11]. A study on induced pluripotent stem cell-derived cardiomyocytes from a patient with the p.Arg663His variant reported disease characteristics and abnormal calcium handling as the mechanisms underlying HCM development at the cellular level [31]. This implied that variants in this region are more likely to disrupt protein function than those located elsewhere.

We identified the same *MYH7* variants in several unrelated patients in this study. These patients were recruited nationwide from Korea as part of the National Project of Bio Big Data pilot study, and it was confirmed that they were unrelated individuals with cardiomyopathy. The high frequency of these variants in our study is consistent with reports in the literature [6, 32–38] and databases such as ClinVar, ClinGen, and OMIM DB, where these variants have been identified as causative for cardiomyopathy. It is plausible that these variants are also frequently observed in other cohorts. For instance, the p.Arg249Gln variant was found in six unrelated individuals in our study. Richard et al. reported this variant in three unrelated cases of HCM from their French cohort [6], and Woo et al. detected the variant in three individuals with HCM from a Canadian cohort [32]. Notably, the p.Arg249Gln variant appears to be particularly prevalent in our Korean population. To determine whether the p.Arg249Gln variant is especially common within the Korean population, further population studies are necessary. Additionally, the p.Arg663His variant was identified in four individuals in our study and has been identified in a substantial number of HCM patients across ethnically diverse populations [33–38].

The *MYH7* p.Asn1918_Ala1922del variant identified in this study is an in-frame deletion. Interestingly, *MYH7* variants causing the distal phenotypes are located within the tail region of the slow myosin molecule [39]. This location is in the region of the reported domains for interaction with myomesin and titin [39]. Variants in the ultimate C-terminus of slow myosin are known to cause body myopathy, whereas variants in the proximal neck, head part, and coiled-coil domain may cause cardiomyopathy [6, 40]. Most variants for the distal phenotype are targeted to lysine residues; however, further mechanistic explanations for the molecular pathogenesis are lacking. Therefore, monitoring and updating the clinical

phenotypes of patients with C-terminus variants, such as the distal myopathy of the patient with the p.Asn1918_Ala1922del variant identified in this study, is required.

We discovered 7 *MYH7* variants that were reported as VUS or not previously reported in the ClinVar database. We compared the VUS and novel missense *MYH7* variants identified in this study with those catalogued in OMIM to assess variant pathogenicity using in silico tools. Representative variants of OMIM are mostly disease-causing variants [17]. Available in silico tools can help researchers establish the pathogenicity of novel variants identified in disease-causing genes [41–43]. Many approaches have been developed to enable these predictions, with their number growing rapidly in recent years [18–24, 27, 28]. To date, such studies have mainly focused on predicting the effect of missense variants on the structure, function, and evolutionary conservation of proteins [41, 44]. This strategy could provide a genetic diagnosis in addition to variant classification. The missense variants identified in our study, including p.Tyr134His, p.Ser148Arg, p.Phe488Leu, and p.Arg807Cys, were predicted to have pathogenic effects on normal MYH7 function and structure. In CADD, REVEL, MetaRNN, and MetaSVM analyses, the in silico scores of the novel variants were significantly lower than those of known pathogenic variants, further suggesting a benign effect of the p.Glu55Lys variant on protein function. As *MYH7* is a key contributor to cardiomyopathy, our results from the in silico analyses of novel variants may be sufficient to suggest that they are causative variants. Nevertheless, functional in vitro or in vivo experiments of these novel variants are required to validate the damaging effects on the *MYH7* gene or protein in further studies.

Genetic variants are mainly reported in ClinGen [15, 45], which is a Food and Drug Administration-recognized human genetic variant database containing expert-curated assertions regarding variant pathogenicity and supporting evidence summaries. Several *MYH7* variants, including p.Arg369Gln, p.Arg403Gln, p.Arg403Trp, p.Arg663His, p.Arg870Cys, p.Glu1356Lys, and p.Thr1377Met, were analysed in our study and are registered in the Inherited Cardiomyopathy Expert Panel (CMP-EP) in ClinGen. The CMP-EP established adjusted ACMG/AMP variant classification rules that can be used for all *MYH7*-associated cardiomyopathies [15].

We manually classified two *MYH7* variants as potential pathogenic or likely pathogenic according to the ACMG/AMP framework for *MYH7* [15]. One novel variant, p.Tyr134His, is located within the myosin motor domain (PM1_moderate) and was first identified in two patients with HCM in this study (PM2_moderate and PS4_supporting). The p.Tyr134His variant was predicted to have pathogenic effects on protein function and structural stability using all *in silico* tools (PP3_supporting). The p.Ser148Arg variant was also located in the myosin motor domain (PM1_moderate) and is absent in the normal population (PM2_moderate). The equivalent variant c.442 A>C (p.Ser148Arg) was classified as likely pathogenic by ClinVar (PS1_strong), and the alternative variant c.441 C>A (p.Ser148Ile) was classified as pathogenic by UniProt (PM5_moderate). Other variants were classified as VUS. Sharing information on patients with cardiomyopathy harboring the same variant in other studies or adding details of the clinical phenotypes and family histories of patients can potentially upgrade the rule-based classification of VUS to likely pathogenic.

In the current study, we identified three patients with variants in genes other than *MYH7*. Patient 1, who has DCM, carries both the *MYH7* p.Glu55Lys and *LMNA* p.Phe113Leu variants. Whereas *LMNA* is well-known as a causative gene for DCM [46, 47], the p.Phe113Leu variant is reported as a VUS for Charcot-Marie-Tooth disease type 2 in the ClinVar database. Patient 20, also with DCM, has both the *MYH7* p.Leu476Phe and *MYBPC3* p.Arg502Gln variants. *MYBPC3* is recognized as a causative gene for both HCM and DCM [48, 49] and the *MYBPC3* p.Arg502Gln variant is frequently reported as pathogenic or likely pathogenic for cardiomyopathy in the ClinVar database. Patient 40, with DCM, has the *MYH7* p.Arg1396Trp and *TNNT2* p.Arg151Gln variants. The *TNNT2* is reported as a causative gene for cardiomyopathy [50], and the *TNNT2* p.Arg151Gln variant is also reported as pathogenic or likely pathogenic for cardiomyopathy in the ClinVar database. Given the reported possibility of oligogenic inheritance in cardiomyopathy [51, 52], it is necessary to investigate whether the phenotypes of patients with both *MYH7* and other causative gene variants for cardiomyopathy, such as *LMNA*, *MYBPC3*, and *TNNT2*, are more severe compared to those with only *MYH7* variants. Further studies will be needed to update additional clinical data.

Structural variants (SVs) such as copy number variants, insertions, and inversions have greater effects on human genome functions than single nucleotide variants (SNVs) or small insertions/deletions. Long-read sequencing (LRS) produces tens to thousands of kilobase reads and detects the breakpoints of complex SVs. Using LRS, Sonoda et al. [53] confirmed a large deletion extending from *MYH6* to *MYH7*, including the breakpoints, in a family with the inherited atrial septal defect. Another recent study used LRS on five sarcomeric protein genes (*MYH7*, *MYBPC3*, *TPM1*, *TNNT2*, and *TNNI3*) in four patients with HCM [54]. Thus, LRS is a useful approach to detect SVs in undiagnosed patients with suspected inherited diseases, including cardiomyopathy, that do not harbor any reported causative SNVs.

This study had several limitations. First, there are limitations in the clinical information collected from the large-scale National Project of Bio Big Data pilot study to analyse the genotype–phenotype relationships of the novel *MYH7* variants identified in our study. These limitations include a lack of human phenotype ontology terms, information on age at first diagnosis, family history, and phenotypic information on patients with oligogenic inheritance or modifier variants. A new project to update the insufficient clinical information in this study will be needed in the future. Second, we could not perform trio-based genetic analysis owing to limitations in acquiring patient family information. The average age of patients with *MYH7* variants was 46 years; thus, it was difficult for the parents of patients to participate in this project. Last, we assessed SNVs and small deletions only and did not evaluate SVs, large rearrangements, or exon-range alterations even when using the WGS data. Hence, additional pathogenic SVs of *MYH7* that were not detected in this study might be present, the detection of which will require further analyses.

Conclusions

In this study, we identified 27 *MYH7* variants in 41 unrelated patients with diverse cardiomyopathies from the National Project of Bio Big Data pilot study in Korea. These variants comprised 20 known pathogenic/likely pathogenic, 2 VUS, and 5 novel variants. Notably, known pathogenic *MYH7* variants were predominantly situated within the myosin motor domain of the MYH7 protein. In silico analyses revealed that these novel *MYH7* variants had pathogenic effects on protein function and structural stability. In summary, this study is the first to identify the distribution of *MYH7* variants in patients with cardiomyopathy in Korea and could provide new insights into genetic diagnosis through the assessment of novel variants' pathogenicity using in silico tools.

Abbreviations

HCM	Hypertrophic Cardiomyopathy
DCM	Dilated Cardiomyopathy
ARVC	Arrhythmogenic right ventricular cardiomyopathy
RCM	Restrictive cardiomyopathy
MYH7	Myosin heavy chain 7
MYBPC3	Myosin-binding protein C3
VUS	Variants of uncertain significance
WGS	Whole-genome sequencing
ACMG	American College of Medical Genetics and Genomics
AMP	Association for Molecular Pathology
OMIM	Online Mendelian Inheritance in Man
ClinGen	Clinical Genome Resource
CMP-EP	Cardiomyopathy Expert Panel
SVs	Structural variants
SNVs	Single nucleotide variants
LRS	Long-read sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-024-02000-8.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

OHK and MHP wrote the main manuscript text. OHK, JHK, YJK and SYL prepared genetic data collection and analysis. BHL prepared clinical data collection. BJK, HYP, and MHP provided concept of the study. All authors reviewed the manuscript.

Funding

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Health and Welfare, Ministry of Science and ICT, Ministry of Trade Industry and Energy, Korea Disease Control and Prevention Agency (No. 2020M3E5D7086780, The National Project of Bio Big Data), and supported by an intramural research program of the Korea National Institute of Health (2022-NI-060-01).

Data availability

All variants and in silico datasets in the current study are included in supplementary tables. Databases used in this study were gnomAD Browser (https://gnomad.broadinstitute.org/), ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar), ClinGen database (https://clinicalgenome.org/), OMIM database (https://www.omim.org/), DynaMut2 (http://biosig.unimelb.edu.au/dynamut/), and Missense3D (http://missense3d.bc.ic.ac.uk/missense3d/).

Declarations

Ethics approval and consent to participate

This study was approved by The Institutional Review Board of the Human Research of the Korea National Institute of Health, Korea Disease Control and Prevention Agency (Approval No. 2022-09-10-P-A, 2022-02-07–2 C-A, KDCA-2023-06-06-P-01). Informed consent was obtained from all the participants and parents of patients before conducting the genetic tests.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 February 2024 / Accepted: 29 August 2024
Published online: 05 September 2024

References

1. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. Lancet. 2013;381(9862):242–55.
2. de Frutos F, Ochoa JP, Navarro-Penalver M, Baas A, Bjerre JV, Zorio E, et al. Natural history of MYH7-Related dilated cardiomyopathy. J Am Coll Cardiol. 2022;80(15):1447–61.
3. Velicki L, Jakovljevic DG, Preveden A, Golubovic M, Bjelobrk M, Ilic A, et al. Genetic determinants of clinical phenotype in hypertrophic cardiomyopathy. BMC Cardiovasc Disord. 2020;20(1):516.
4. McNally EM, Golbus JR, Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. J Clin Invest. 2013;123(1):19–26.
5. Olivotto I, Girolami F, Ackerman MJ, Nistri S, Bos JM, Zachara E, et al. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. Mayo Clin Proc. 2008;83(6):630–8.
6. Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, et al. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of

mutations, and implications for a molecular diagnosis strategy. Circulation. 2003;107(17):2227–32.
7. Tanjore R, Rangaraju A, Vadapalli S, Remersu S, Narsimhan C, Nallari P. Genetic variations of beta-MYH7 in hypertrophic cardiomyopathy and dilated cardiomyopathy. Indian J Hum Genet. 2010;16(2):67–71.
8. McNally EM, Mestroni L. Dilated cardiomyopathy: genetic determinants and mechanisms. Circ Res. 2017;121(7):731–48.
9. Yamada T, Nomura S. Recent findings related to Cardiomyopathy and Genetics. Int J Mol Sci. 2021;22(22).
10. Walsh R, Rutland C, Thomas R, Loughna S. Cardiomyopathy: a systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic manifestations. Cardiology. 2010;115(1):49–60.
11. Hesaraki M, Bora U, Pahlavan S, Salehi N, Mousavi SA, Barekat M, et al. A novel missense variant in actin binding domain of MYH7 is Associated with Left Ventricular Noncompaction. Front Cardiovasc Med. 2022;9:839862.
12. Kim MJ, Cha S, Baek JS, Yu JJ, Seo GH, Kang M, et al. Genetic heterogeneity of cardiomyopathy and its correlation with patient care. BMC Med Genomics. 2023;16(1):270.
13. Li X, Tang J, Li J, Lin S, Wang T, Zhou K, et al. Genetic clues on Implantable Cardioverter-Defibrillator Placement in Young-Age Hypertrophic Cardiomyopathy: a Case Report of Novel MYH7 Mutation and Literature Review. Front Cardiovasc Med. 2021;8:810291.
14. Vadgama N, Ameen M, Sundaram L, Gaddam S, Genomics England Research C, Gifford C, et al. De novo and inherited variants in coding and regulatory regions in genetic cardiomyopathies. Hum Genomics. 2022;16(1):55.
15. Kelly MA, Caleshu C, Morales A, Buchan J, Wolf Z, Harrison SM, 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):351–9.
16. KISTI. The National Project of Bio Big. Data research portal 2021 [https://www.cirn.re.kr
17. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online mendelian inheritance in man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43(Database issue):D789–98.
18. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886–94.
19. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chap. 7:Unit7 20.
20. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812–4.
21. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011;39(17):e118.
22. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.
23. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an Ensemble Method for Predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877–85.
24. Li C, Zhi D, Wang K, Liu X, MetaRNN. Differentiating Rare Pathogenic and Rare Benign Missense SNVs and InDels Using Deep Learning. 2021.
25. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35(11):1978–80.
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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):405–24.
27. Rodrigues CHM, Pires DEV, Ascher DB. DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci. 2021;30(1):60–9.
28. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide Reliable insights into whether missense variants are Disease Associated? J Mol Biol. 2019;431(11):2197–212.
29. Mattivi CL, Bos JM, Bagnall RD, Nowak N, Giudicessi JR, Ommen SR, et al. Clinical utility of a phenotype-enhanced MYH7-Specific variant classification Framework in Hypertrophic Cardiomyopathy Genetic Testing. Circ Genom Precis Med. 2020;13(5):453–9.
30. Trujillo AS, Hsu KH, Viswanathan MC, Cammarato A, Bernstein SI. The R369 myosin residue within Loop 4 is critical for actin binding and muscle function in Drosophila. Int J Mol Sci. 2022;23(5).
31. Lan F, Lee AS, Liang P, Sanchez-Freire V, Nguyen PK, Wang L, et al. Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. Cell Stem Cell. 2013;12(1):101–13.
32. Sole AWHRJCLM-SZC-CLTGPMZEDWMJ. Mutations of the beta myosin heavy chain gene in hypertrophic cardiomyopathy: critical functional sites determine prognosis. Heart. 2003;89(10):7.
33. Van Driest SL, Jaeger MA, Ommen SR, Will ML, Gersh BJ, Tajik AJ, et al. Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. J Am Coll Cardiol. 2004;44(3):602–10.
34. Song L, Zou Y, Wang J, Wang Z, Zhen Y, Lou K, et al. Mutations profile in Chinese patients with hypertrophic cardiomyopathy. Clin Chim Acta. 2005;351(1–2):209–16.
35. Bales ND, Johnson NM, Judge DP, Murphy AM. Comprehensive Versus targeted genetic testing in children with hypertrophic cardiomyopathy. Pediatr Cardiol. 2016;37(5):845–51.
36. Bashyam MD, Purushotham G, Chaudhary AK, Rao KM, Acharya V, Mohammad TA, et al. A low prevalence of MYH7/MYBPC3 mutations among familial hypertrophic cardiomyopathy patients in India. Mol Cell Biochem. 2012;360(1–2):373–82.
37. Brito D, Miltenberger-Miltenyi G, Vale Pereira S, Silva D, Diogo AN, Madeira H. Sarcomeric hypertrophic cardiomyopathy: genetic profile in a Portuguese population. Rev Port Cardiol. 2012;31(9):577–87.
38. Chiou KR, Chu CT, Charng MJ. Detection of mutations in symptomatic patients with hypertrophic cardiomyopathy in Taiwan. J Cardiol. 2015;65(3):250–6.
39. Meredith C, Herrmann R, Parry C, Liyanage K, Dye DE, Durling HJ, et al. Mutations in the slow skeletal muscle fiber myosin heavy chain gene (MYH7) cause laing early-onset distal myopathy (MPD1). Am J Hum Genet. 2004;75(4):703–8.
40. Blair E, Redwood C, de Jesus Oliveira M, Moolman-Smook JC, Brink P, Corfield VA, et al. Mutations of the light meromyosin domain of the beta-myosin heavy chain rod in hypertrophic cardiomyopathy. Circ Res. 2002;90(3):263–9.
41. Garcia FAO, de Andrade ES, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. Front Genet. 2022;13:1010327.
42. Katsonis P, Wilhelm K, Williams A, Lichtarge O. Genome interpretation using in silico predictors of variant impact. Hum Genet. 2022;141(10):1549–77.
43. Accetturo M, Bartolomeo N, Stella A. In-silico analysis of NF1 missense variants in ClinVar: translating variant predictions into variant interpretation and classification. Int J Mol Sci. 2020;21(3).
44. Licata L, Via A, Turina P, Babbi G, Benevenuta S, Carta C, et al. Resources and tools for rare disease variant interpretation. Front Mol Biosci. 2023;10:1169109.
45. Preston CG, Wright MW, Madhavrao R, Harrison SM, Goldstein JL, Luo X, et al. ClinGen variant curation interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. Genome Med. 2022;14(1):6.
46. Wang Y, Dobreva G. Epigenetics in LMNA-Related cardiomyopathy. Cells. 2023;12(5).
47. Hasselberg NE, Haland TF, Saberniak J, Brekke PH, Berge KE, Leren TP, et al. Lamin A/C cardiomyopathy: young onset, high penetrance, and frequent need for heart transplantation. Eur Heart J. 2018;39(10):853–60.
48. Ehlermann P, Weichenhan D, Zehelein J, Steen H, Pribe R, Zeller R, et al. Adverse events in families with hypertrophic or dilated cardiomyopathy and mutations in the MYBPC3 gene. BMC Med Genet. 2008;9:95.
49. Tudurachi BS, Zavoi A, Leonte A, Tapoi L, Ureche C, Birgoan SG et al. An update on MYBPC3 gene mutation in hypertrophic cardiomyopathy. Int J Mol Sci. 2023;24(13).
50. Hershberger RE, Pinto JR, Parks SB, Kushner JD, Li D, Ludwigsen S, et al. Clinical and functional characterization of TNNT2 mutations identified in patients with dilated cardiomyopathy. Circ Cardiovasc Genet. 2009;2(4):306–13.
51. Li L, Bainbridge MN, Tan Y, Willerson JT, Marian AJ. A potential oligogenic etiology of hypertrophic cardiomyopathy: a Classic single-gene disorder. Circ Res. 2017;120(7):1084–90.
52. Gifford CA, Ranade SS, Samarakoon R, Salunga HT, Soysa TYd, Huang Y et al. Oligogenic inheritance of a human heart disease involving a genetic modifier. Science. 2019;364.
53. Sonoda K, Ishihara H, Sakazaki H, Suzuki T, Horie M, Ohno S. Long-read sequence confirmed a large deletion including MYH6 and MYH7 in an infant of atrial septal defect and Atrial Arrhythmias. Circ Genom Precis Med. 2021;14(4):e003223.

54. Salakhov RR, Golubenko MV, Valiakhmetov NR, Pavlyukova EN, Zarubin AA, Babushkina NP et al. Application of Long-Read Nanopore Sequencing to the search for mutations in hypertrophic cardiomyopathy. Int J Mol Sci. 2022;23(24).

Publisher's note