Distinguishing Hypertrophic Cardiomyopathy-Associated Mutations from Background Genetic Noise

Jamie D. Kapplinger, BA[1,2,3,*], **Andrew P. Landstrom, MD, PhD**[1,2,3,*,#], **J. Martijn Bos, MD, PhD**[1], **Benjamin A. Salisbury, PhD**[4,¶], **Thomas E. Callis, PhD**[4], and **Michael J. Ackerman, MD, PhD**[1,2,3,5,6]

[1]Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, United States

[2]Mayo Medical School, Mayo Clinic, Rochester, Minnesota, United States

[3]Mayo Clinic Medical Scientist Training Program, Mayo Clinic, Rochester, Minnesota, United States

[4]Transgenomic Inc., New Haven, Connecticut, United States

[5]Department of Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, United States

[6]Department of Pediatrics, Division of Pediatric Cardiology, Mayo Clinic, Rochester, Minnesota, United States

Abstract

Despite the significant progress that has been made in identifying disease-associated mutations, the utility of the Hypertrophic Cardiomyopathy (HCM) genetic test is limited by a lack of understanding of the background genetic variation inherent to these sarcomeric genes in seemingly healthy subjects. This study represents the first comprehensive analysis of genetic variation in 427 ostensibly healthy individuals for the HCM genetic test using the "Gold Standard" Sanger sequencing method validating the background rate identified in the publically available exomes. While mutations are clearly over-represented in disease, a background rate as high as ~5% among healthy individuals prevents diagnostic certainty. To this end, we have identified a number of estimated predictive value-based associations including gene-specific, topology, and conservation methods generating an algorithm aiding in the probabilistic interpretation of an HCM genetic test.

Address for correspondence: Michael J. Ackerman, MD, PhD., Mayo Clinic Windland Smith Rice Sudden Death Genomics Laboratory, Guggenheim 501, Mayo Clinic, Rochester, MN 55905, 507-284-0101 (phone), 507-284-3757 (fax), ackerman.michael@mayo.edu.
*Contributed equally and are co-equal first authors
#Present affiliation: Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital, Houston, Texas, United States
¶Present affiliation: Knome, Inc., Cambridge, Massachusetts, United States

INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is defined by cardiac hypertrophy of generally the left ventricle in the absence of a clinically identifiable etiology. HCM is a significant cause of sudden cardiac death (SCD) in the young, and one of the most common causes of SCD among young athletes [1,2]. Clinically, HCM has a heterogeneous presentation with varying degrees of hypertrophy, fibrosis, myocyte disarray, left ventricular outflow tract obstruction, ventricular septal morphology, associated symptoms, and sudden death susceptibility [3]. This phenotypic heterogeneity is mirrored by variability in the genetic mutations that confer HCM disease susceptibility.

HCM is the most common inherited cardiovascular disease, and hundreds of mutations in at least 27 putative HCM-susceptibility genes have been identified. Clinically-available genetic tests for HCM are available focusing particularly on HCM-susceptibility genes that encode components of the cardiac sarcomere [4,5] including those that encode proteins of the thick myofilament, the intermediate myofilament, and the thin myofilament [6]. Overall, depending on the phenotypes of analyzed cohorts of HCM cases, approximately 30–60% of patients with HCM will host a mutation in one of these genes with the most prevalent HCM genotypes being *MYH7* and *MYBPC3* [7].

Despite the significant progress that has been made in identifying disease-associated mutations, the utility of the HCM genetic test is limited by the background genetic variation inherent to these sarcomeric genes in seemingly healthy subjects [8]. Indeed, identification of a variant of unknown significance (VUS) by the HCM genetic test highlights the dilemma inherent in this clinical test. Recently, the first studies to explore this question have identified a startling number of previously cardiomyopathy-associated genetic variants in population-based exome data [9]. Additionally, new analyses have demonstrated a background rate of rare variants in large cohorts of individuals not enriched for the HCM phenotype, such as the Framingham Heart Study and Jackson Heart Study, NHLBI Exome Sequencing Project (ESP), and the 1000 Genomes Project (1kG) introducing additional complication in the interpretation of a VUS in the HCM genetic test [10–14]. Recently, amino acid level conservation was identified as a potential method to distinguish pathogenic mutations from benign variants [11]. Additionally, gene specific probabilities of pathogenicity have been identified which aids the interpretation of HCM genetic testing. However, these studies have largely relied on publically available whole exome sequenced control cohorts whose results have not been validated for the HCM genes. Additionally, many of these studies have been underpowered to perform detailed analyses as relatively small HCM case cohorts were used for comparisons. To this end, we sought to validate the spectrum and prevalence of background genetic variation that would meet the criteria of a "positive" genetic HCM tes from publically available control data using a set of Sanger-sequenced, unrelated controls as well as identify information to aid in the probabilistic interpretation of a HCM genetic test

result using a large cohort of HCM cases as well as combine the new and current knowledge to provide an algorithm to aid providers in the interpretation an HCM genetic test.

METHODS

Study Cohorts

The Sanger-sequenced control cohort was comprised of 427 unrelated, ostensibly healthy individuals from various racial and ethnic backgrounds who were subjected to HCM genetic testing by Transgenomic Inc. in New Haven, Connecticut, United States (formerly PGxHealth). These control individuals were volunteers recruited and genotyped by Transgenomic as part of the clinical HCM genetic test validation process. Ethnicity, sex, and age at genotyping were recorded for each subject. A normal 12-lead electrocardiogram, Holter-monitoring, and echocardiographic analysis were not a prerequisite for the individuals to be included in this cohort.

Additional estimated frequency of background genetic variation was procured from on-line databases of next-generation sequencing results in large populations of individuals, including the 1000 Genomes Project (1kG) (www.1000genomes.org/) [15] and the National Heart Lung and Blood Institute Exome Sequencing Project (ESP) (esp.gs.washington.edu/drupal/) [16].

The HCM case cohort was comprised of 2178 individuals diagnosed with HCM or referred for HCM genetic testing. This cohort included 1053 unrelated index cases evaluated and diagnosed at the Mayo Clinic HCM Clinic in Rochester, Minnesota, United States, who had consented for HCM genetic testing between April 1997 and April 2007. An additional 1125 cases were referred to Transgenomic Inc. for HCM clinical genetic testing. Ethnicity, sex, and age at genotyping were recorded for these additional cases. The veracity of the clinical diagnosis of HCM in these referred cases could not be confirmed. The demographic and clinical characteristics of the cohorts are summarized in Table 1.

Genetic Analysis

Comprehensive genetic analysis of all subjects was conducted for the coding exons of *MYH7*-encoded myosin heavy chain (MYH7), *MYL2*-encoded regulatory myosin light chain (MYL2), *MYL3*-encoded essential myosin light chain (MYL3), *MYPBC3*-encoded cardiac myosin binding protein C (MYBPC3), *ACTC1*-encoded actin (ACTC), *TNNC1*-encoded cardiac troponin C (TNNC1), *TNNI3*-encoded cardiac troponin I (TNNI3), *TNNT2*-encoded cardiac troponin T (TNNT2), and *TPM1*-encoded alpha-tropomyosin (TPM1).

For cases identified at Mayo Clinic, following receipt of written consent for this Mayo Foundation Institutional Review Board-approved protocol, genetic analysis was performed using PCR, denaturing high performance liquid chromatography (DHPLC) and direct DNA sequencing as previously described [17]

Referral cases sent for commercial FAMILION-HCM genetic testing and the Sanger-sequenced controls were genotyped via Sanger sequencing in both the forward and reverse

direction. For both case cohorts and the Sanger-sequenced controls, the genotyping methods were optimized for the genes analyzed. The 1kG and ESP samples were previously whole exome sequenced and genotypes were procured from online databases. The 1kG and ESP whole exome sequencing was not specifically optimized for the genes analyzed in this study. Additionally, assessment of large gene rearrangements was not able to be performed using the methods in this study.

Genetic Variant Classification

Genetic variants predicted to alter the protein, such as missense, in-frame and frame-shift insertion/deletion, canonical splice-sites (±1–2), and nonsense resulting in a premature truncation, were identified. Variants identified in the case cohort that were **completely** absent in the control cohort, or a variant uniquely identified in only a single control cohort individual across all control cohorts assessed, irrespective of ethnicity, were annotated as rare non-synonymous variants (rNSVs). This strict definition was used in order to polarize the variants for the comparisons between case and control.

The designation of rNSVs herein is not meant to imply pathogenicity or even functional relevance to the respective protein. This designation is intended to reflect that had these variants been discovered during the course of a clinical HCM genetic test, each would be considered a possible pathogenic variant or VUS. All variants were classified using standard HGVS nomenclature. In-frame and frame-shift insertions and deletions, splice junction, and nonsense rNSVs likely resulting in a shortened protein product were designated as "radical" rNSVs. As with the designation of rNSVs, the designation of radical is not meant to imply disease pathogenicity but is used as a descriptor for further analysis.

Primary Sequence Conservation and Linear Topology Analysis

Primary sequence analysis was conducted utilizing primary sequences from the UCSC Genome Browser [18]. To calculate degree of conservation of individual residues across species, primary sequences from 44 species including primates, other placental mammals, and non-mammalian vertebrates were used. The "non-identity score" was determined by calculating the number of primary sequences harboring an amino acid not identical to the human residue at that location (substituted). A non-identity score of 0 indicates that the variant involves an amino acid that is not substituted across the species queried. Scoring range for the non-identity score ranged from 0 (no substitution across species) to 44 (high interspecies variability). Protein domains in linear topology analysis were annotated using published literature [19,20]. Additionally, a moving window analysis was performed using windows examining approximately 5% of the overall protein (60 amino acid and 100 amino acid moving windows for MYH7 and MYBPC3, respectively). The percent of case rNSVs and all control variants (all control variants were assessed for this analysis) were analyzed for each window and windows with a statistical overrepresentation of case rNSVs were identified.

Statistical Analysis

Statistical analysis utilized Fisher's exact test and two-sided t-tests, where appropriate, with a threshold of significance set to $P < 0.05$. Variance was expressed as mean ± standard

deviation for cohort demographics and clinical characteristics. All other statistical analyses expressed variance as mean ± standard error. Sequencing conservation scores were analyzed using Wilcoxon non-parametric 2-sample tests. To determine the likelihood of disease causation, an estimated predictive value (EPV) expressed with 95% confidence interval was determined as previously described [21]. For the purpose of this calculation, only rNSVs were used for both cases and controls while polymorphisms did not contribute to the calculated EPV. The Sanger sequenced and ESP controls were used for the frequencies in order to determine the EPV calculations, unless otherwise stated.

RESULTS

Yield in HCM Cases

Overall, the yield of HCM genetic testing for the Mayo cases was found to be 29.2% (308/1053). The yield among the cases referred for the Transgenomic HCM genetic test was slightly higher at 33.6% (378/1125, $P = 0.03$). Given the similar yields between the two case cohorts, the cases were combined for the remaining analyses

The overall yield in the combined case cohort was found to be 31.5% (686/2178). The majority of case rNSVs were missense with a yield of 20.3% among the cases. rNSVs in MYH7 and MYBPC3 accounted for the vast majority of the rNSVs with yields of 15.6% and 12.0%, respectively. The rNSVs type between the two genes were strikingly different with MYH7 having a missense yield of 11.7%, while MYBPC3 had a radical yield of 10.5%. The remaining 7 genes had a yield of 4.8% with only TNNI3 and TPM1 having an individual yield greater than 1%. These results are summarized in Figure 1.

A total of 392 rNSVs were identified in the cases. Of these, 57 were found to be absent from controls and nominally overrepresented in cases ($p < 0.05$, Table 2). These rNSVs accounted for 49.0% (336/686) of the genotype-positive HCM cases. This overrepresentation makes these rNSVs near-definitive pathogenic mutations. Although accounting for a large portion of genotype-positive cases, these 57 rNSVs only representing 14.5% (57/392) of the rNSVs identified. Due to this disproportion between the number of cases and the number of amino acids altered, these rNSVs heavily skew any case control analysis. Thus in an effort to determine the impact of the next "new" rNSVs throughout these genes, these overrepresented rNSVs were removed from subsequent analyses comparing case and control sample counts. The yield of rNSVs within the HCM cases before and after the removal of the overrepresented rNSVs is listed in Table 3.

rNSVs Frequency in Control Cohorts

In order to validate the background rate of "genetic noise" inherent in the HCM genetic test previously reported from the 1kG and ESP, comprehensive genetic analysis of a Sanger-sequenced control cohort was performed. Among the Sanger-sequenced, ostensibly healthy individuals, 5.2% hosted a rare genetic variant, **unique** among the control cohort. A complete list of all variants found in the Sanger-sequenced control individuals is summarized in Table 4. In comparison, 2.8% of the ESP cohort had a missense rNSVs, a frequency which was significantly lower than the Sanger-sequenced cohort ($P = 0.01$). The

yield of the 1kG cohort was only 0.9%, which was lower than the Sanger-sequenced cohort ($P = 1.6 \times 10^{-6}$) and the ESP cohort ($P = 9.1 \times 10^{-5}$). These results are summarized in Figure 2A.

To determine whether variation in the relative racial/ethnicity composition of the cohorts was the source of the disparity, the yield of each cohort was recalculated based on relative distribution of race/ethnicity. As the ESP cohort is exclusively Caucasian and African-descent, when limiting the Sanger-sequenced cohort to Caucasian and African-descent individuals the yield was 3.3%, which was similar to the ESP yield ($P = $ NS). This held true when the yield was calculated separately for Caucasian and African-descent samples. The 1kG yields remained lower when calculated independently for each ethnic group. While there were differences between the yields across cohorts, there were no differences in yield between ethnicities within each cohort or when the cohorts were combined. These results are summarized in Figure 2B–D and Table 5.

Due to the low background rate in the 1kG samples, this cohort was removed from the remaining analysis moving forward. Therefore, the combined yield of 2.9% from the ESP samples and the Sanger sequenced samples was used for all remaining comparisons. This yield was predominately a result of missense rNSVs; however a background rate of 0.14% was identified for radical rNSVs. The identified control radical rNSVs are listed in Table 6. The control yield for each gene is summarized in Table 3.

Linear Topology and Primary Sequence Conservation

Despite possible clustering in MYH7, no topological region hosted a statistically significant overrepresentation of case-derived variants. Additionally, a moving window analysis did not identify a statistical overrepresentation of case rNSVs when correcting for multiple comparisons. The localization of case rNSVs and control variants for MYH7 and MYBPC3 is illustrated in Figure 3. Due to the low yield in the remaining genes, a topology was not analysed.

Primary sequence analyses revealed overall rNSVs identified among HCM cases localized to amino acids that were more conserved than rNSVs identified in control individuals. Among the 208 distinct missense rNSVs identified in control individuals, 35.6% involved residues unsubstituted across species. In comparison, a significantly larger proportion of HCM case rNSVs, 56.7% (164/289), localized to a universally conserved (i.e. unsubstituted) residue ($P = 3.3 \times 10^{-6}$).

To quantify this further, a non-identity score was calculated, rNSVs identified in control individuals demonstrated weaker conservation resulting in a higher non-identity score (4.1 ± 0.4) compared to cases (2.1 ± 0.3, $P = 3.2 \times 10^{-5}$). While case rNSVs were more conserved than control rNSVs overall, this difference was only driven by missense rNSVs in MYH7 and MYBPC3, as the removal of these rNSVs resulted in only an insignificant trend in the same direction for non-identities score between cases and controls in the remaining genes [Case: 2.1 ± 0.5, Control: 2.8 ± 0.6 (p = ns)]. These results are summarized in Figures 4A and 4B, respectively.

Interpreting a Genetic Test Result Utilizing an Estimated Predictive Value Calculation

Interpretation of genetic testing becomes more complicated in the setting of the identified background rate; therefore estimated predictive values (EPVs) were calculated in an effort to assess a "positive" genetic test result that is reporting a probable disease-causative mutation or a possible one (i.e. a VUS). A radical rNSVs in MYBPC3, resulted in an EPV of 98 (95–99) suggesting high-probability of pathogenicity. The frequency of radical rNSVs in the remaining genes was too low to provide confidence in the pathogenicity of radical rNSVs. A single missense rNSVs yielded a relatively lower estimated predictive value with an EPV of 78 (74–82). Interestingly, there were drastic differences in the EPV between genes when missense rNSVs were assessed by individual genes. By itself, a rare MYH7 missense rNSVs has a moderate EPV of only 87 (74–96). Surprisingly, the identification of a missense rNSVs in MYBPC3 provides a low EPV (69 (58–78)). Among the minor genes, only ACTC1 and TPM1 having EPVs suggesting high probability (EPV>90) of pathogenicity and missense rNSVs in TNNI3 conferred an EPV of 84 (54–95) suggesting intermediate probability of pathogenicity. The remaining genes (MYL2, TNNC1, TNNT2, MYL3) all have lower probabilities of pathogenicity for missense rNSVs (EPVs < 75). Due to the rarity of missense rNSVs in the minor genes, it is difficult to upgrade a VUS in one of these genes without additional evidence.

As conservation helped to distinguish between case and control rNSVs in MYH7 and MYBPC3, we assessed whether conservation across species could upgrade or downgrade the EPVs for missense variants in these genes. Pathogenic rNSVs were identified as those identified at residues unsubstituted across all species examined. The EPV for missense rNSVs in MYH7 was increased from 87 (83–91) to 92 (87–95) when the rNSVs involved an unsubstituted residue and decreased to 79 (68–86) when involving substituted residues. The EPV for a universally conserved MYBPC3 missense rNSVs increased from 69 (58–78) to 85 (74–91) and decreased from 69 to 52 (27–69) when the residue involved was substituted in > 1 of the 44 analyzed species. These results are summarized in Table 7.

DISCUSSION

Genetic Testing Noise in Sudden Cardiac Death-Predisposing Diseases

Only within the past decade have the first systematic studies exploring background genetic variation in diseases that predispose individuals to SCD become available. Ostensibly healthy individuals subjected to genetic analysis of the three canonical genes of long QT syndrome (LQTS), *KCNQ1, KCNH2,* and *SCN5A*, had a background rate of ~5% [21]. Analysis of healthy individuals subjected to genotyping of canonical ARVC desmosomal genes (*PKP2*, *DSP*, *DSG2*, *DSC2*, *JUP*, and *TMEM43*) identified a background rate of ~16%, three times greater than the background genetic variation in the 3 LQTS-susceptibility genes [22].

Recently, studies have identifed an alarming background rate of rare variants within the nine HCM-susceptibility genes examined in this study, which are estimated to account for at least 90–95% of currently genetically identifiable HCM. Similar to the previous studies, we identify a conservative background rate of ~3% that may be as high as ~5%, in these nine

sarcomeric genes. Using a large cohort of Sanger-sequenced controls as a 'gold standard' of genetic analysis, we were able to validate the background rate among the ESP samples.

Differences in Reported Genetic Variation in Healthy Individuals

The background rate found among Sanger-sequenced control individuals was 5.2%, which is significantly higher than found in on-line databases and previously published manuscripts on this topic [23]. This was largely due to differences in ethnicity; however the low yield among the 1kG may be due to another possible, more concerning, explanation. As the publically available control cohorts are all based on next-generation sequencing technologies, there may be issues of varying degrees of coverage and issues of inefficient exon capture. A recent publication highlighted the degree of variation in coverage across a number of genes associated with HCM when using next-generation sequencing technologies [24]. This would suggest caution when using the publically available next-generation sequencing based control data to determine background rates or assigning disease pathogenicity, especially when using the 1kG population.

Interpretation of the HCM Genetic Test

The HCM genetic test is recommended for patients with clinically suspected HCM primarily for the purpose of mutation-specific cascade genetic testing of the index case's relatives [6,25]. In this way, a clear biomarker of disease can risk stratify family members without current echocardiographic or clinical manifestation of disease and guide future clinical screening of genotype positive family members [25,26]. In addition, there is emerging evidence that a positive genetic test, regardless of the type of rNSVs identified, is an independent prognostic indicator of outcome in HCM [7,27]. Based on this, the need to interpret the pathogenic significance of a positive genetic test result is of critical importance.

Recent studies have begun to address this dilemma identifying a number of distinguishing features of pathogenic mutations. The overrepresentation of rNSVs in cases over controls has been identified as characteristic of disease-causing genes for the interpretation of whole exome data, however there is strikingly different levels of intolerance among different disease genes [28]. This highlights the necessity of disease specific examination. Recent studies have begun to examine the HCM genes specifically. A recent study found that genes associated with HCM had a high intolerance to genetic variation and found that amino acid conservation may help distinguish pathogenic variants from the benign genetic variants identified in up to 3% of the population [11]. A second study identified gene-specific probabilities of pathogenicity and highlighted four genes (*MYH7*, *MYBPC3*, *TNNT2*, and *TNNI3*) with the highest pathogenic probability. While these studies have begun to address the issue of the background genetic noise, most were underpowered to perform many assessments.

In our study, we harness the power of over 2000 genetically tested HCM cases. We identify 3 of the same genes as Lopes *et al.*, with only TNNT2 being identified with a low probability of pathogenicity. This may be due to cohort specific differences, additionally, our removal of the near definitive, overrepresented variants may lower the probability of TNNT2. Lopes *et al.* identified 7 cases with TNNT2 mutations with 3 of these samples

hosting the same *TNNT2* p.Ala104Val mutation. Removal of this clearly pathogenic variant may reduce the probability of the next new variant to a similarly low probability identified in our study.

Additionally, we have identified a large number of near-definitive pathogenic rNSVs, as well as several characteristics of rNSVs identified in HCM cases compared with controls that may aid the cardiologist or genetic counselor in this interpretation. A recent report identified that radical rNSVs in the HCM genes are likely pathogenic [11]. While we identified radical rNSVs in 11.2% of all HCM cases versus 0.14% of control subjects which strongly supports that frame-shift insertions and deletions, splice junction, and nonsense rNSVs are likely pathogenic, over 90% of these radical rNSVs fall in MYBPC3 suggesting that radical rNSVs within this gene are likely pathogenic, while caution must be used when assessing radical rNSVs in the remaining genes. However, since all the radical control rNSVs were identified in the ESP cohort, our interpretation may be conservative as a recent paper identified that a much larger percentage of radical variants versus missense variants identified by Next-Generation sequencing were in-fact false-positives, as they were not confirmed by Sanger sequencing [29]. Conversely, while missense rNSVs were over-represented 7.5-fold in HCM cases, the 2.8% background rate of mutation-like, but presumably benign, variants is too high to ignore. This creates a challenge in interpreting a genetic test that has identified a missense VUS, especially in a number of the minor genes where the identified EPV was less than 75.

Previous studies have indicated that both LQTS- and ARVC-associated mutations preferentially localize to specific structure-function domains (regions). Here, however, when correcting for multiple comparisons, we do not observe an overrepresentation in any of the previously reported regions within either the *MYH7*-encoded beta myosin heavy chain or the *MYBPC3*-encoded myosin binding protein C.

In agreement with studies on other diseases, HCM case rNSVs preferentially involved amino acids that were conserved across species. This would suggest the use of conservation as a potential mechanism for distinguishing case rNSVs from benign control variants. However, the statistical difference between case and control rNSVs was only driven by variants in *MYH7* and *MYBPC3* as reflected in a lower non-identity score than control rNSVs for both genes, but not for the remaining minor genes. This has implications for many in-silico prediction tools such as SIFT and PolyPhen, as these tools largely rely on conservation. Our data suggests limiting the use of conservation-based tools to only the two major HCM genes *MYH7* and *MYBPC* or at least extreme caution when applying the various in-silico prediction tools for variants involving a minor gene.

Based on our findings and incorporating current literature on background noise in the HCM genetic test, we have created an algorithm to aid in the interpretation of a so-called positive HCM genetic test (Figure 5). rNSVs with evidence (e.g. segregates with disease, functional data) of pathogenicity for that specific rNSVs should be considered probably pathogenic mutations. Additionally, we provide evidence that the absence of a rNSVs in a large set of controls as well as overrepresentation in cases provides sufficient evidence to elevate a rNSVs to a probably pathogenic mutation. Here in, we have identified 57 rare rNSVs listed

in Table 2 that meet these criteria and should be considered actionable for predictive mutation-specific genetic testing among family members. However, pathogenic uncertainty persists for the next rare rNSVs identified. It is reasonable to assume that radical rNSVs in MYBPC3 are near-definite HCM-susceptibility mutation, while greater caution must be applied to the remaining genes. When/if a missense rNSVs is identified in MYH7 or MYBPC3, it should be interpreted in the context of the conservation of the residue. Currently, a rare rNSVs involving one of the minor genes, where we have found conservation does not distinguish case mutations from control variants, should be interpreted cautiously taking into context the EPV as well as the pre-test probability. Without additional evidence for causality, such a genetic test result should not be used for predictive testing of the index case's relatives. Importantly, the results of the HCM genetic test, or any other genetic test, should never completely usurp clinical judgment.

Given the transition of clinical genetic testing to whole exome sequencing platforms, the American College of Medical Genetics recently provided a list of genes for which they recommend returning incidental findings of which all 9 of the HCM genes analyzed in this study are included. Given our identification that the up to 5% of the general population may host a rare variant in these 9 genes, the ability to assess the pathogenicity of these rare variants is key. Assessing an incidental variant in the context we provide may provide physicians with a statistical framework to distinguish pathogenic from benign variants. With this transition and the increased utilization of clinical genetic testing, the identification of overrepresentation in cases for a particular rare variant can carry great weight in the interpretation of the variants pathogenicity. To that end, a publically available database providing HCM case counts for a particular mutation, such as ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), would be of immense value to clinicians. This would allow physicians to interpret genetic results with constantly evolving information.

LIMITATIONS

There are several limitations inherent to this study. While we classify any variant identified in the controls as benign, there may be control variants that are pathogenic. However, a recent study examining the impact of rare genetic variation in HCM genes among well phenotyped samples from the Framingham and Jackson Heart Studies identified a very small subset of these general population samples carrying these rare variants had evidence of HCM and many of the previously reported pathogenic mutations identified in these samples much smaller estimated effect size than previously reported from HCM families [14]. Given this recent data and the overall prevalence of 1 in 500 for HCM, the vast majority of the rNSVs s identified in the controls can be assumed to have little effect individually on the HCM phenotype. Further, if some of the control rNSVs are in fact pathogenic, the EPVs would be underestimated, making our estimates a conservative assumption. The lack of phenotypic information on the referral cohort presents another limitation. As the yield for the HCM genetic test can be impacted by the phenotypic strength of the cases assessed, the absence of clinical phenotype for the referral cases is another limitation of this study. However, when examining the cohorts separately we see that there is a similar yield between the Mayo cases and the referral cases (29.2% and 33.6%, respectively). While the yields are lower than the expected published yield of ~50%, this is expected given the stringent

requirement of absence of identified variants among the controls examined. Additionally, the previously published yield is likely slightly inflated with benign variants that have now been identified in controls given that most previous studies required absence in small control cohorts. Additionally, these samples were all subject to genetic testing based off of a physician's supposition of an HCM diagnosis. Given the number of cohorts examined in this study, there is heterogeneity in the sequencing technology, depth of sequencing, accuracy, depth, etc., we have made every attempt to control for these issue while maintaining a cohort size powered to do the comparisons. Further, there are differences in the racial/ethnic composition of our control and case cohorts. Specifically, our HCM case cohort is predominantly Caucasian due to the geographic location of subject recruitment. Given there were no statistical differences between ethnicities we decided to use the overall yield among all ESP and Sanger controls. This may inflate the relative control rNSVs yield due to the greater diversity of that cohort relative to the cases, but would again result in a more conservative estimate. Despite these limitations, several statistical observations appear robust. Future validation of these observations in independent studies utilizing phenotypically robust cohorts of healthy individuals as well as subjects with HCM should help to mathematically upgrade or downgrade a VUS.

CONCLUSIONS

This study provides a validates the ESP derived background rate of genetic variation in ostensibly healthy individuals for the HCM genetic test using cohort analyzed via the "Gold Standard" Sanger sequencing. While mutations are clearly over-represented in disease, a background rate of up to ~5% among healthy individuals prevents pathogenic certainty. We have identified a number of EPV-based associations that can guide a clinician in the interpretation of an HCM genetic test.

Acknowledgments

ABBREVIATIONS

1kG	1000 Genomes Sequencing Project
ACTC	actin
ARVC	arrhythmogenic right ventricular cardiomyopathy
DHPLC	denaturing high performance liquid chromatography heteroduplex analysis
EPV	estimated predictive value
ESP	National Heart Lung and Blood Institute Exome Sequencing Project
HCM	hypertrophic cardiomyopathy
LQTS	long QT syndrome

MYH7	beta myosin heavy chain
MYL2	regulatory myosin light chain
MYL3	essential myosin light chain
MYPBC3	cardiac myosin binding protein C
PCR	polymerase chain reaction
SCD	sudden cardiac death
TNNC1	cardiac troponin C
TNNI3	cardiac troponin I
TNNT2	cardiac troponin T
TPM1	alpha-tropomyosin
rNSV	rare non-synonymous variant
VUS	variant of undetermined/uncertain significance

References

1. Maron BJ, Epstein SE, Roberts WC. Causes of sudden death in competitive athletes. JACC. 1986; 7 (1):204–214. [PubMed: 3510233]

2. Maron BJ, Doerer JJ, Haas TS, Tierney DM, Mueller FO. Sudden deaths in young competitive athletes: analysis of 1866 deaths in the United States, 1980–2006. Circulation. 2009; 119 (8):1085–1092.10.1161/CIRCULATIONAHA.108.804617 [PubMed: 19221222]

3. Maron BJ. Hypertrophic cardiomyopathy: a systematic review. JAMA. 2002; 287 (10):1308–1320. [PubMed: 11886323]

4. Bos JM, Ommen SR, Ackerman MJ. Genetics of hypertrophic cardiomyopathy: one, two, or more diseases? Current Opinion in Cardiology. 2007; 22 (3):193–199.10.1097/HCO.0b013e3280e1cc7f [PubMed: 17413275]

5. Landstrom AP, Ackerman MJ. Mutation type is not clinically useful in predicting prognosis in hypertrophic cardiomyopathy. Circulation. 2010; 122 (23):2441–2449. discussion 2450. 10.1161/CIRCULATIONAHA.110.954446 [PubMed: 21135372]

6. Ackerman MJ, Priori SG, Willems S, Berul C, Brugada R, Calkins H, Camm AJ, Ellinor PT, Gollob M, Hamilton R, Hershberger RE, Judge DP, Le Marec H, McKenna WJ, Schulze-Bahr E, Semsarian C, Towbin JA, Watkins H, Wilde A, Wolpert C, Zipes DP. HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies: this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). Europace. 2011; 13 (8):1077–1109.10.1093/europace/eur245 [PubMed: 21810866]

7. Van Driest SL, Ommen SR, Tajik AJ, Gersh BJ, Ackerman MJ. Yield of genetic testing in hypertrophic cardiomyopathy. Mayo Clinic Proceedings. 2005; 80 (6):739–744.10.1016/S0025-6196(11)61527-9 [PubMed: 15945527]

8. Landstrom AP, Ackerman MJ. GWAS or Gee Whiz, PSAS or Pshaw: elucidating the biologic and clinical significance of genetic variation in cardiovascular disease. Heart Rhythm. 2009; 6 (12): 1751–1753.10.1016/j.hrthm.2009.09.016 [PubMed: 19959124]

9. Andreasen C, Nielsen JB, Refsgaard L, Holst AG, Christensen AH, Andreasen L, Sajadieh A, Haunso S, Svendsen JH, Olesen MS. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur J Hum Genet. 2013 ejhg2012283 [pii]. 10.1038/ejhg.2012.283

10. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. Circ Cardiovasc Genet. 2012; 5(4):391–399. CIRCGENETICS.112.962928 [pii]. 10.1161/CIRCGENETICS.112.962928 [PubMed: 22763267]

11. Pan S, Caleshu CA, Dunn KE, Foti MJ, Moran MK, Soyinka O, Ashley EA. Cardiac structural and sarcomere genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. Circ Cardiovasc Genet. 2012; 5(6):602–610. CIRCGENETICS.112.963421 [pii]. 10.1161/CIRCGENETICS.112.963421 [PubMed: 23074333]

12. Lopes LR, Zekavati A, Syrris P, Hubank M, Giambartolomei C, Dalageorgou C, Jenkins S, McKenna W, Plagnol V, Elliott PM. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. J Med Genet. 2013; 50(4):228–239. jmedgenet-2012-101270 [pii]. 10.1136/jmedgenet-2012-101270 [PubMed: 23396983]

13. Norton N, Robertson PD, Rieder MJ, Zuchner S, Rampersaud E, Martin E, Li D, Nickerson DA, Hershberger RE. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circ Cardiovasc Genet. 2012; 5(2):167–174. CIRCGENETICS.111.961805 [pii]. 10.1161/CIRCGENETICS.111.961805 [PubMed: 22337857]

14. Bick AG, Flannick J, Ito K, Cheng S, Vasan RS, Parfenov MG, Herman DS, DePalma SR, Gupta N, Gabriel SB, Funke BH, Rehm HL, Benjamin EJ, Aragam J, Taylor HA Jr, Fox ER, Newton-Cheh C, Kathiresan S, O'Donnell CJ, Wilson JG, Altshuler DM, Hirschhorn JN, Seidman JG, Seidman C. Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. Am J Hum Genet. 2012; 91 (3):513–519.10.1016/j.ajhg.2012.07.017 [PubMed: 22958901]

15. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491 (7422):56–65.10.1038/nature11632 [PubMed: 23128226]

16. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, Akey JM. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013; 493 (7431):216–220.10.1038/nature11690 [PubMed: 23201682]

17. Van Driest SL, Will ML, Atkins DL, Ackerman MJ. A novel TPM1 mutation in a family with hypertrophic cardiomyopathy and sudden cardiac death in childhood. Am J Cardiol. 2002; 90 (10):1123–1127. [PubMed: 12423715]

18. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002; 12 (6):996–1006. Article published online before print in May 2002. 10.1101/gr.229102 [PubMed: 12045153]

19. Kulikovskaya I, McClellan GB, Levine R, Winegrad S. Multiple forms of cardiac myosin-binding protein C exist and can regulate thick filament stability. J Gen Physiol. 2007; 129 (5):419–428.10.1085/jgp.200609714 [PubMed: 17470661]

20. Gautel M, Zuffardi O, Freiburg A, Labeit S. Phosphorylation switches specific for the cardiac isoform of myosin binding protein-C: a modulator of cardiac contraction? Embo J. 1995; 14 (9):1952–1960. [PubMed: 7744002]

21. Kapa S, Tester DJ, Salisbury BA, Harris-Kerr C, Pungliya MS, Alders M, Wilde AAM, Ackerman MJ. Genetic Testing for Long-QT Syndrome: Distinguishing Pathogenic Mutations From Benign Variants. Circulation. 2009; 120 (18):1752–1760.10.1161/circulationaha.109.863076 [PubMed: 19841300]

22. Kapplinger JD, Landstrom AP, Salisbury BA, Callis TE, Pollevick GD, Tester DJ, Cox MGPJ, Bhuiyan Z, Bikker H, Wiesfeld ACP, Hauer RNW, van Tintelen JP, Jongbloed JDH, Calkins H, Judge DP, Wilde AAM, Ackerman MJ. Distinguishing Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia-Associated Mutations From Background Genetic Noise. JACC. 2011; 57 (23):2317–2327.10.1016/j.jacc.2010.12.036 [PubMed: 21636032]

23. Van Driest SL, Vasile VC, Ommen SR, Will ML, Tajik AJ, Gersh BJ, Ackerman MJ. Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. JACC. 2004; 44 (9):1903–1910.10.1016/j.jacc.2004.07.045 [PubMed: 15519027]

24. Ng D, Johnston JJ, Teer JK, Singh LN, Peller LC, Wynter JS, Lewis KL, Cooper DN, Stenson PD, Mullikin JC, Biesecker LG. Interpreting Secondary Cardiac Disease Variants in an Exome Cohort. Circ Cardiovasc Genet. 201310.1161/CIRCGENETICS.113.000039

25. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2011; 124 (24):2761–2796.10.1161/CIR.0b013e318223e230 [PubMed: 22068435]

26. Bos JM, Towbin JA, Ackerman MJ. Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. JACC. 2009; 54 (3):201–211.10.1016/j.jacc.2009.02.075 [PubMed: 19589432]

27. Olivotto I, Girolami F, Ackerman MJ, Nistri S, Bos JM, Zachara E, Ommen SR, Theis JL, Vaubel RA, Re F, Armentano C, Poggesi C, Torricelli F, Cecchi F. Myofilament protein gene mutation screening and outcome of patients with hypertrophic cardiomyopathy. Mayo Clinic Proceedings. 2008; 83 (6):630–638.10.4065/83.6.630 [PubMed: 18533079]

28. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genetics. 2013; 9 (8):e1003709.10.1371/journal.pgen.1003709 [PubMed: 23990802]

29. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IH, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG, Wang J, Li Y, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012; 335 (6070):823–828.10.1126/science.1215040 [PubMed: 22344438]
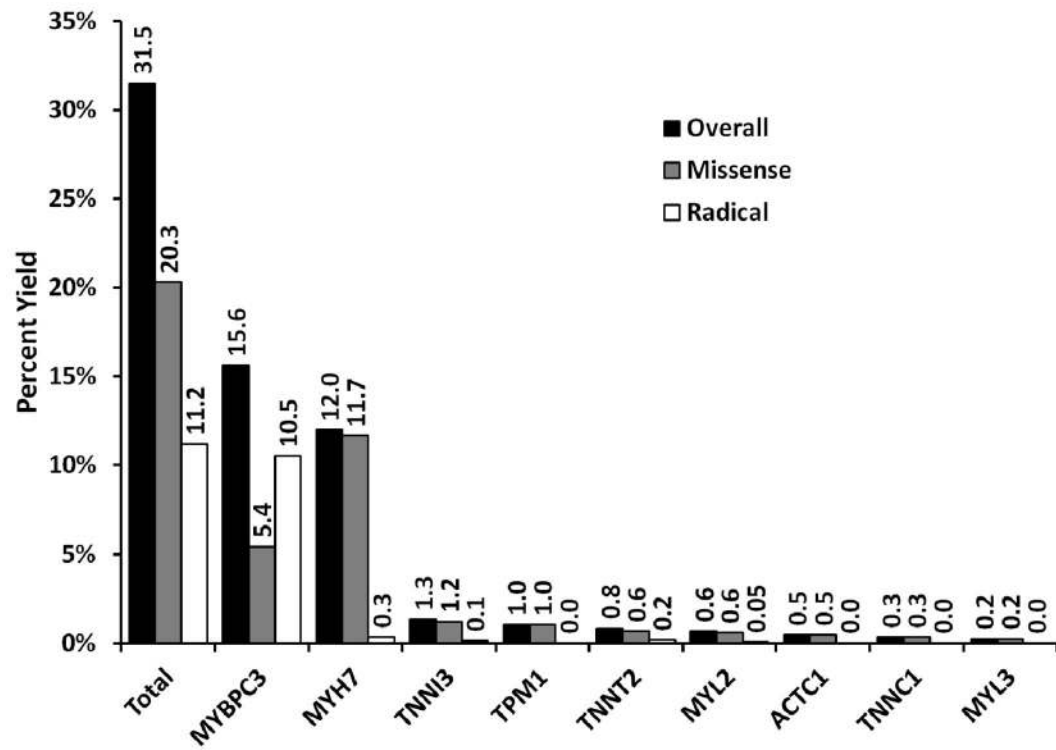
Fig 1. Yield of the HCM Genetic Test in Cases

Bar graph representing the percent yield of the HCM genetic test in HCM cases by gene. The overall yield (black) reflects all rNSVs identified. Missense rNSVs (gray) include rNSVs altering a single amino acid. Radical rNSVs (white) include all in-frame and frame-shift insertions and deletions, splice junction, and nonsense rNSVs. The numbers above the bars represent the percentage yield.
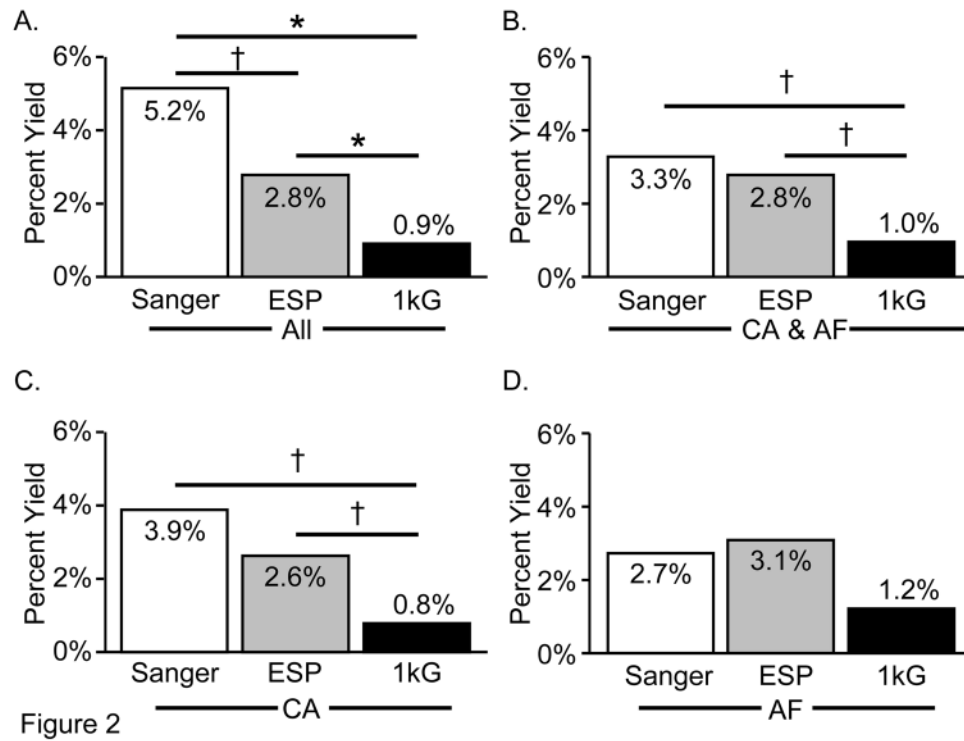
Fig 2. Frequency of Rare Genetic Variants Present in the Nine Genes Associated with Sarcomeric/Myofilament-HCM across Cohorts of Healthy Individuals

Bar graph representing the percent yield of the HCM genetic test among the Sanger-sequenced control cohort (white), the ESP (light gray), and the 1kG (dark gray) cohorts. **A.** Percent yield among all control samples. **B.** Percent yield among Caucasian and African-descent samples. **C.** Percent yield among Caucasian samples. **D.** Percent yield among samples of African-descent. * $P < 10^{-4}$ compared with respective control cohort percentage; †, $10^{-4} < P < 0.05$.

A) MYH7

B) MYBPC3

Fig 3. MYH7 and MYBPC3 Linear Topology
Linear topology of MYH7 (A) and MYBPC3 (B) with the location of all control missense variants (lines below topology) and HCM case rNSVs (lines above topology) identified.
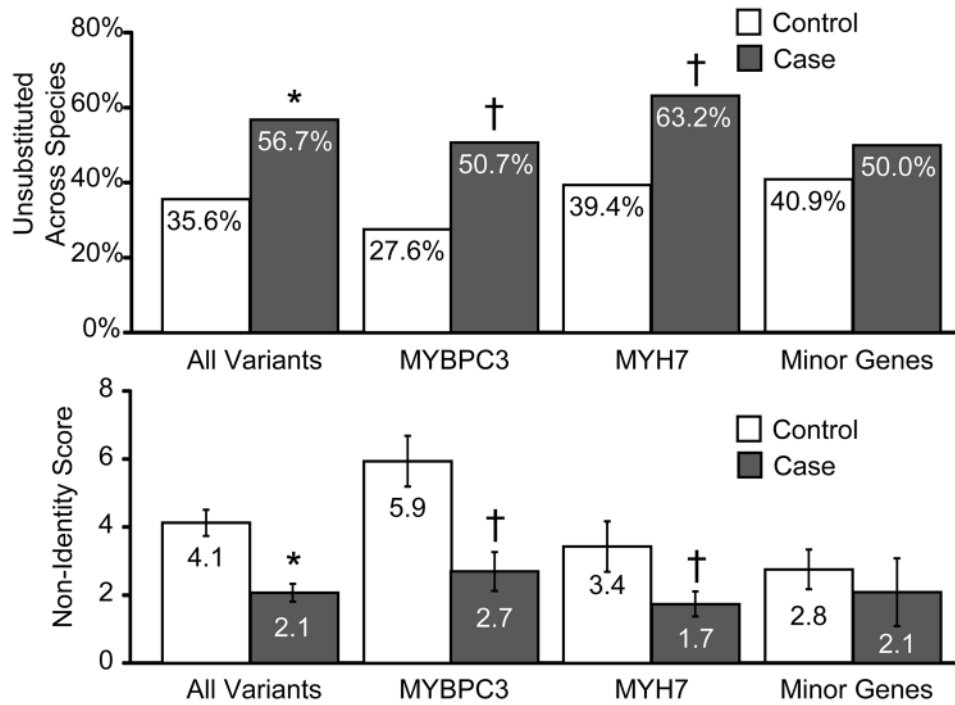
Fig 4. Amino Acid Conservation

A. Bar graph of the percentage of rNSVs identified in the control (white) and HCM case (gray) cohorts, respectively, that involve residues completely conserved across species. The numbers within the bars represent the percentage of rNSVs at unsubstituted residues. **B**. Bar graph of the non-identity score of rNSVs identified in the control (white) and HCM case (gray) cohorts, respectively. The numbers within the bars represent the non-identity score. Error bars indicate standard error. *, $P < 10^{-4}$ compared with respective control cohort percentage; †, $10^{-4} < P < 0.05$.
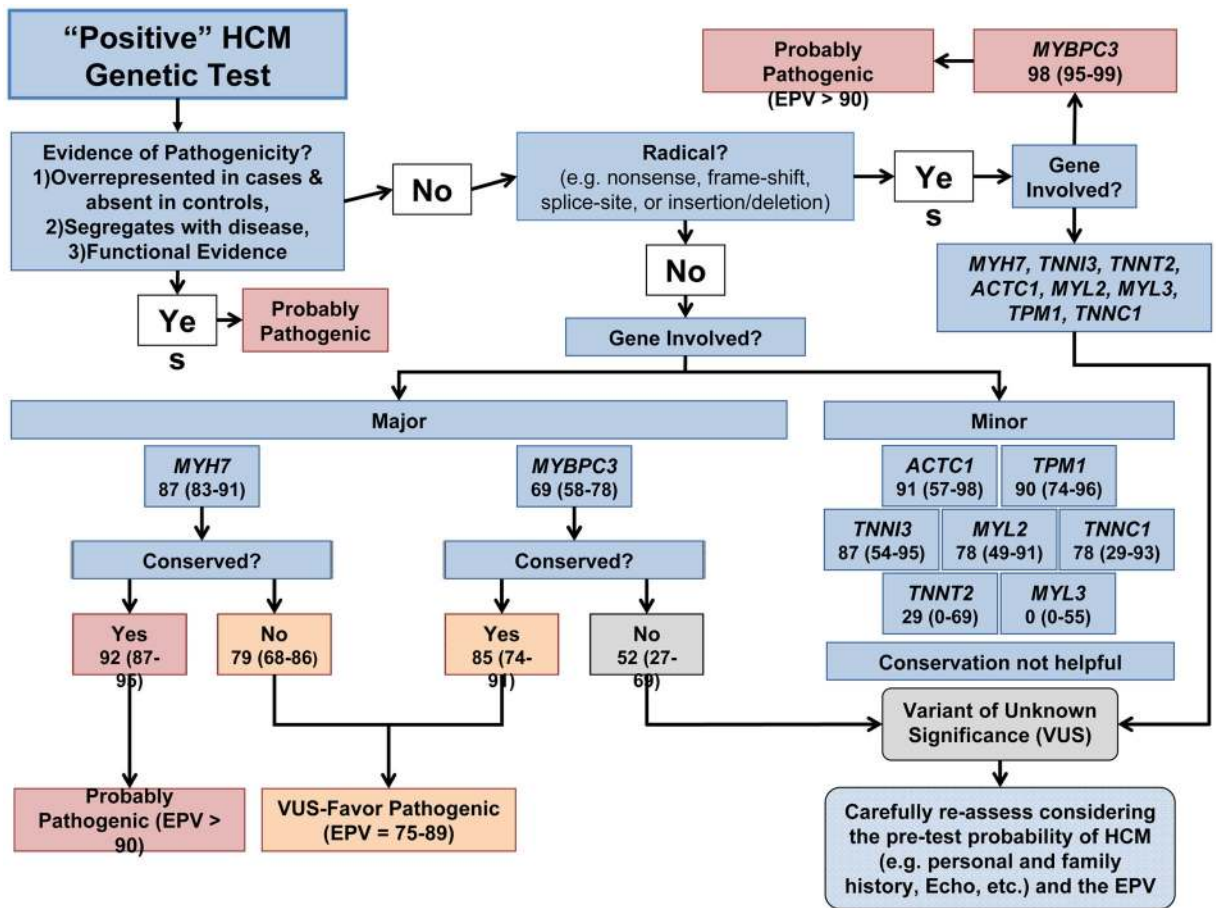
Fig 5. Interpretation of HCM Genetic Test Schematic

An algorithm to guide the interpretation of a "positive" HCM genetic test. A positive genetic test result involving an rNSVs with additional evidence of pathogenicity (i.e. 1) Overrepresented in cases & absent in controls, 2) Segregates with disease, 3) Functional Evidence) is a high probability disease mutation. Additionally, a radical mutation in *MYBPC3* is also a high probability disease mutation. If a single missense rNSVs is identified, there is less statistical rigor to classify this rNSVs as a definite HCM-causative mutation. Here, the missense rNSVs' probability of pathogenicity is influenced by the gene in which it is located. Additionally, conservation across species can enhance the interpretation for *MYBPC3* and *MYH7* missense rNSVs.

Cohort Demographics

Table 1

	Controls				HCM Cases		
	Sanger	ESP	1kG*	Mayo	Referral	All Cases
N	427	6503	1093	1053	1125	2178
Age (years)	41.1 ± 15.4	NA	NA	44.4 ± 18.6	37.3 ± 23.0	41.3 ± 20.9
Female (%)	50.1%	NA	52.1%	40.4%	37.0%	38.9%
Caucasian (%)	24.1%	66%	34.8%	96.5%	70.4%	88.4%
African (%)	25.8%	34%	22.5%	1.4%	17.0%	6.3%
Asian (%)	29.0%	NA	26.2%	1.0%	1.1%	1.1%
Hispanic (%)	9.8%	NA	16.6%	0.5%	10.1%	3.5%
Other (%)	11.2%	NA	NA	0.5%	1.3%	0.8%
Septal Thickness (mm)	NA	NA	NA	20.9 ± 5.9	NA	NA
Resting LVOTO (mmHg)	NA	NA	NA	43.8 ± 43.7	NA	NA

*
in the case of the 1kG cohort, races/ethnicities designate country of origin. NA, not available/applicable; LVOTO, left ventricular outflow tract obstruction.

Table 2

Statistically Near Definite HCM-Causative Missense Mutations

Gene	Mutation	Case Count	Mutation Type
ACTC1	NM_005159.4:c.301G>A:p.Glu101Lys	3	Missense
MYBPC3	NM_000256.3:c.26-2A>G	6	Splicing
MYBPC3	NM_000256.3:c.177-187delAGAGGGCACAC:p.Glu60Alafs*49	8	Frameshift
MYBPC3	NM_000256.3:c.236dupA:p.Tyr79Ter	3	Frameshift
MYBPC3	NM_000256.3:c.481C>A:p.Pro161Thr	3	Missense
MYBPC3	NM_000256.3:c.655G>C:p.Val219Leu	7	Missense
MYBPC3	NM_000256.3:c.772G>A:p.Glu258Lys	9	Missense
MYBPC3	NM_000256.3:c.821+1G>A	9	Splicing
MYBPC3	NM_000256.3:c.927-2A>G	3	Splicing
MYBPC3	NM_000256.3:c.1235-1236delTT:p.Phe412Ter	3	Frameshift
MYBPC3	NM_000256.3:c.1484G>A:p.Arg495Gln	13	Missense
MYBPC3	NM_000256.3:c.1505G>A:p.Arg502Gln	6	Missense
MYBPC3	NM_000256.3:c.1700-1701delAG:p.Glu567Glyfs*4	3	Frameshift
MYBPC3	NM_000256.3:c.1928-2A>G	5	Splicing
MYBPC3	NM_000256.3:c.2459G>A:p.Arg820Gln	3	Missense
MYBPC3	NM_000256.3:c.2490dupT:p.His831Serfs*2	3	Frameshift
MYBPC3	NM_000256.3:c.2541C>G:p.Tyr847Ter	5	Nonsense
MYBPC3	NM_000256.3:c.2556-2557CG>TGT:p.Gly853Valfs*31	3	Frameshift
MYBPC3	NM_000256.3:c.2670G>A:p.Trp890Ter	6	Nonsense
MYBPC3	NM_000256.3:c.2827C>T:p.Arg943Ter	6	Nonsense
MYBPC3	NM_000256.3:c.2864-2865delCT:p.Pro955Argfs*95	15	Frameshift
MYBPC3	NM_000256.3:c.2905+1G>A	3	Splicing
MYBPC3	NM_000256.3:c.3083C>T:p.Thr1028Ile	3	Missense
MYBPC3	NM_000256.3:c.3226-3227insT:p.Asp1076Valfs*6	8	Frameshift
MYBPC3	NM_000256.3:c.3288delG:p.Glu1096Aspfs*93	3	Frameshift
MYBPC3	NM_000256.3:c.3330+2T>G	29	Splicing
MYBPC3	NM_000256.3:c.3697C>T:p.Gln1233Ter	13	Nonsense
MYBPC3	NM_000256.3:c.3742-3759dupGGGGGCATCTATGTCTGC:p.Gly1248_Cys1253dup	13	In/Del
MYBPC3	NM_000256.3:c.3771C>A:p.Asn1257Lys	3	Missense
MYH7	NM_000257.2:c.427C>T:p.Arg143Trp	6	Missense
MYH7	NM_000257.2:c.1208G>A:p.Arg403Gln	8	Missense
MYH7	NM_000257.2:c.1357C>T:p.Arg453Cys	8	Missense
MYH7	NM_000257.2:c.1491G>T:p.Glu497Asp	3	Missense
MYH7	NM_000257.2:c.1750G>A:p.Gly584Ser	3	Missense
MYH7	NM_000257.2:c.1816G>A:p.Val606Met	5	Missense
MYH7	NM_000257.2:c.2155C>T:p.Arg719Trp	4	Missense
MYH7	NM_000257.2:c.2156G>A:p.Arg719Gln	10	Missense

Gene	Mutation	Case Count	Mutation Type
MYH7	NM_000257.2:c.2167C>T:p.Arg723Cys	3	Missense
MYH7	NM_000257.2:c.2221G>T:p.Gly741Trp	6	Missense
MYH7	NM_000257.2:c.2221G>C:p.Gly741Arg	4	Missense
MYH7	NM_000257.2:c.2539-2541delAAG:p.Lys847del	3	In/Del
MYH7	NM_000257.2:c.2681A>G:p.Glu894Gly	4	Missense
MYH7	NM_000257.2:c.2717A>G:p.Asp906Gly	5	Missense
MYH7	NM_000257.2:c.2722C>G:p.Leu908Val	16	Missense
MYH7	NM_000257.2:c.2770G>A:p.Glu924Lys	5	Missense
MYH7	NM_000257.2:c.3133C>T:p.Arg1045Cys	3	Missense
MYH7	NM_000257.2:c.3169G>A:p.Gly1057Ser	4	Missense
MYH7	NM_000257.2:c.4130C>T:p.Thr1377Met	8	Missense
MYH7	NM_000257.2:c.5135G>A:p.Arg1712Gln	7	Missense
MYH7	NM_000257.2:c.5302G>A:p.Glu1768Lys	3	Missense
TNNI3	NM_000363.4:c.407G>A:p.Arg136Gln	4	Missense
TNNI3	NM_000363.4:c.422G>A:p.Arg141Gln	4	Missense
TNNI3	NM_000363.4:c.485G>A:p.Arg162Gln	5	Missense
TNNI3	NM_000363.4:c.497C>T:p.Ser166Phe	3	Missense
TNNT2	NM_001001430.1:c.236T>A:p.Ile79Asn	3	Missense
TNNT2	NM_001001430.1:c.274C>T:p.Arg92Trp	3	Missense
TPM1	NM_001018005.1:c.574G>A:p.Glu192Lys	4	Missense

rNSVs absent from controls and overrepresented in cases vs. controls

Table 3

Yield by Gene in Cases and Controls

	Control (N = 6930)			Case (N = 2178)				Case with overrepresented removed (N = 2178)		
	All	Missense	Radical	All	Missense	Radical		All	Missense	Radical
All Genes	203 (2.93%)	193 (2.78%)	10 (0.14%)	686 (31.50%)	442 (20.29%)	244 (11.20%)		364 (16.71%)	277 (12.72%)	94 (4.32%)
MYBPC3	77 (1.11%)	71 (1.02%)	6 (0.09%)	340 (15.61%)	118 (5.42%)	229 (10.51%)		151 (6.93%)	73 (3.35%)	82 (3.76%)
MYH7	59 (0.85%)	58 (0.84%)	1 (0.01%)	261 (11.98%)	254 (11.66%)	7 (0.32%)		146 (6.70%)	142 (6.52%)	4 (0.18%)
TNNI3	5 (0.07%)	5 (0.07%)	0 (0.00%)	29 (1.33%)	26 (1.19%)	3 (0.14%)		13 (0.60%)	10 (0.46%)	3 (0.14%)
TPM1	6 (0.09%)	6 (0.09%)	0 (0.00%)	22 (1.01%)	22 (1.01%)	0 (0.00%)		18 (0.83%)	18 (0.83%)	0 (0.00%)
TNNT2	19 (0.27%)	18 (0.26%)	1 (0.01%)	18 (0.83%)	14 (0.64%)	4 (0.18%)		12 (0.55%)	8 (0.37%)	4 (0.18%)
MYL2	10 (0.14%)	9 (0.13%)	1 (0.01%)	14 (0.64%)	13 (0.60%)	1 (0.05%)		14 (0.64%)	13 (0.60%)	1 (0.05%)
ACTC1	2 (0.03%)	2 (0.03%)	0 (0.00%)	10 (0.46%)	10 (0.46%)	0 (0.00%)		7 (0.32%)	7 (0.32%)	0 (0.00%)
TNNC1	6 (0.09%)	5 (0.07%)	1 (0.01%)	7 (0.32%)	7 (0.32%)	0 (0.00%)		7 (0.32%)	7 (0.32%)	0 (0.00%)
MYL3	19 (0.27%)	19 (0.27%)	0 (0.00%)	5 (0.23%)	5 (0.23%)	0 (0.00%)		5 (0.23%)	5 (0.23%)	0 (0.00%)

Yield of rNSVs in cases and controls. The number of cases or controls hosting a rNSV is given with the percentage in parenthesis. Additionally, the case yield is given when the overrepresented rNSV's are removed. Bold numbers indicate yields that are statistically higher in cases than in controls ($p < 0.05$).

Table 4

Frequency and Type of Missense Variant in Sanger Sequence Controls

Gene	Variant	Sanger Count	1kb Count	ESP Count	Total Control Count	Status
MYBPC3	NM_000256.3:c.122G>A:p.Arg41His	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.283A>C:p.Ile95Leu	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.364G>A:p.Ala122Thr	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2197C>T:p.Arg733Cys	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2761C>G:p.Gln921Glu	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2771C>T:p.Thr924Ile	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2834G>A:p.Arg945Gln	1	0	0	1	rNSVs
MYH7	NM_000257.2:c.11C>T:p.Ser4Leu	1	0	0	1	rNSVs
MYH7	NM_000257.2:c.3382G>A:p.Ala1128Thr	1	0	0	1	rNSVs
MYH7	NM_000257.2:c.1066A>C:p.Ile356Leu	1	0	0	1	rNSVs
MYH7	NM_000257.2:c.3622G>A:p.Asp1208Asn	1	0	0	1	rNSVs
MYH7	NM_000257.2:c.5326A>G:p.Ser1776Gly	1	0	0	1	rNSVs
MYL2	NM_000432.3:c.173G>A:p.Arg58Gln	1	0	0	1	rNSVs
MYL3	NM_000258.2:c.92G>A:p.Arg31His	1	0	0	1	rNSVs
MYL3	NM_000258.2:c.404A>C:p.Glu135Ala	1	0	0	1	rNSVs
MYL3	NM_000258.2:c.457C>G:p.Leu153Val	1	0	0	1	rNSVs
TNNT2	NM_001001430.1:c.70C>A:p.Gln24Lys	1	0	0	1	rNSVs
TNNT2	NM_001001430.1:c.206C>T:p.Ser69Leu	1	0	0	1	rNSVs
TNNT2	NM_001001430.1:c.361G>A:p.Val121Ile	1	0	0	1	rNSVs
TPM1	NM_001018005.1:c.511A>G:p.Ile171Val	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2308G>A:p.Asp770Asn	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.2602G>A:p.Gly868Ser	1	0	0	1	rNSVs
MYL2	NM_000432.3:c.4G>C:p.Ala2Pro	1	0	0	1	rNSVs
MYBPC3	NM_000256.3:c.74G>A:p.Ser25Asn	1	0	5	6	Polymorphism
MYBPC3	NM_000256.3:c.787G>A:p.Gly263Arg	1	0	1	2	Polymorphism
MYBPC3	NM_000256.3:c.841C>T:p.Arg281Trp	1	0	1	2	Polymorphism
MYBPC3	NM_000256.3:c.1468G>A:p.Gly490Arg	1	0	4	5	Polymorphism

Gene	Variant	Sanger Count	1kb Count	ESP Count	Total Control Count	Status
MYBPC3	NM_000256.3:c.2980C>T:p.Leu994Phe	1	0	1	2	Polymorphism
MYH7	NM_000257.2:c.3981C>A:p.Asn1327Lys	1	0	1	2	Polymorphism
MYH7	NM_000257.2:c.4909G>A:p.Ala1637Thr	1	0	6	7	Polymorphism
MYBPC3	NM_000256.3:c.1544A>G:p.Asn515Ser	1	1	3	5	Polymorphism
MYBPC3	NM_000256.3:c.478C>T:p.Arg160Trp	1	3	2	6	Polymorphism
MYBPC3	NM_000256.3:c.472G>A:p.Val158Met	30	95	762	887	Polymorphism
MYBPC3	NM_000256.3:c.530G>A:p.Arg177His	3	0	43	46	Polymorphism
MYBPC3	NM_000256.3:c.565G>A:p.Val189Ile	2	0	30	32	Polymorphism
MYBPC3	NM_000256.3:c.706A>G:p.Ser236Gly	52	165	1222	1439	Polymorphism
MYBPC3	NM_000256.3:c.833G>A:p.Gly278Glu	3	10	63	76	Polymorphism
MYBPC3	NM_000256.3:c.977G>A:p.Arg326Gln	4	4	32	40	Polymorphism
MYBPC3	NM_000256.3:c.1144C>T:p.Arg382Trp	12	31	125	168	Polymorphism
MYBPC3	NM_000256.3:c.1519G>A:p.Gly507Arg	5	5	17	27	Polymorphism
MYBPC3	NM_000256.3:c.1564G>A:p.Ala522Thr	3	4	22	29	Polymorphism
MYBPC3	NM_000256.3:c.2498C>T:p.Ala833Val	3	14	98	115	Polymorphism
MYBPC3	NM_000256.3:c.2686G>A:p.Val896Met	1	6	46	53	Polymorphism
MYBPC3	NM_000256.3:c.3137C>T:p.Thr1046Met	2	0	1	3	Polymorphism
MYBPC3	NM_000256.3:c.3392T>C:p.Ile1131Thr	2	0	7	9	Polymorphism
MYBPC3	NM_000256.3:c.2992C>G:p.Gln998Glu	3	19	0	22	Polymorphism
MYH7	NM_000257.2:c.4472C>G:p.Ser1491Cys	6	12	112	130	Polymorphism
TNNI3	NM_000363.4:c.235C>T:p.Arg79Cys	2	0	0	2	Polymorphism
TNNI3	NM_000363.4:c.244C>T:p.Pro82Ser	8	5	71	84	Polymorphism
TNNT2	NM_001001430.1:c.758A>G:p.Lys253Arg	58	113	722	893	Polymorphism

Bold font, variants considered rNSVs and used for background rate calculations in the Sanger-sequenced cohort

Table 5

Yield of Control Cohorts by Race and Ethnicity

Race	Sanger	ESP	1kG
Overall	5.2%	2.8%	0.9%
Caucasian	3.9%	2.6%	0.8%
Non-Caucasian	5.8%	3.1%	1.0%
Caucasian + African	3.3%	2.8%	1.0%
African	2.7%	3.1%	1.2%
Asian	5.7%	NA	1.1%
Hispanic	9.5%	NA	0.6%
Other	8.3%	NA	NA

NA, not available/applicable

Table 6

Radical Variants identified in Controls

Genomic Position	Genes	Variant	Type	Count
chr11:47359280	**MYBPC3**	**NM_000256.3:c.2373dupG:p.Trp792Valfs*41**	**Frameshift**	**1**
chr11:47362772	**MYBPC3**	**NM_000256.3:c.1812_1813dupCG:p.Asp605Alafs*59**	**Frameshift**	**1**
chr14:23884257	MYH7	NM_000257.2:c.5504_5505dupAG:p.Val1837Argfs*2	Frameshift	2
chr12:111356985	**MYL2**	**NM_000432.3:c.11_15delAGAAA:p.Lys4Serfs*25**	**Frameshift**	**1**
chr3:52485523	**TNNC1**	**NM_003280.2:c.337delG:p.Asp113Thrfs*5**	**Frameshift**	**1**
chr19:55668006	TNNI3	NM_000363.4:c.114dupA:p.Ser39Ilefs*2	Frameshift	4
chr1:201333463	TNNT2	NM_001001430.1:c.451delC:p.Arg151Glyfs*41	Frameshift	2
chr1:201334750	**TNNT2**	**NM_001001430.1:c.281dupG:p.Met95Hisfs*7**	**Frameshift**	**1**
chr15:63336227	TPM1	NM_001018005.1:c.117delG:p.Glu40Lysfs*46	Frameshift	6
chr11:47369031	**MYBPC3**	**NM_000256.3:c.852-1G/T**	**Splice**	**1**
chr11:47373058	**MYBPC3**	**NM_000256.3:c.26-2A/G**	**Splice**	**1**
chr11:47360070	**MYBPC3**	**NM_000256.3:c.2308+1G/A**	**Splice**	**1**
chr11:47353795	**MYBPC3**	**NM_000256.3:c.3642G/A:p.Trp1214Ter**	**Nonsense**	**1**
chr14:23889431	**MYH7**	**NM_000257.2:c.3349G/T:p.Glu1117Ter**	**Nonsense**	**1**

Bold font, rNSVs used to calculate the background rate of radical rNSVs in the ESP cohort.

Table 7

Summary of Calculated EPVs

Category	All			Overrepresented removed		
	EPV (CI)	Case	Control	EPV (CI)	Case	Control
Overall Test	91 (89–92)	686/2178	203/6930	NA	NA	NA
Radical only (All Genes)	99 (98–99)	244/2178	10/6930	NA	NA	NA
MYBPC3 radical	99 (98–100)	229/2178	6/6930	98 (95–99)	82/2178	6/6930
TNNI3 radical	100 (0–100)	3/2178	0/6930	100 (0–100)	3/2178	0/6930
MYH7 radical	96 (64–99)	7/2178	1/6930	92 (30–99)	4/2178	1/6930
MYL2 radical	69 (0–98)	1/2178	1/6930	69 (0–98)	1/2178	1/6930
MYL3 radical	NA	0/2178	0/6930	NA	0/2178	0/6930
TNNC1 radical	0 (0–100)	0/2178	1/6930	0 (0–100)	0/2178	1/6930
TNNT2 radical	92 (30–99)	4/2178	1/6930	92 (30–99)	4/2178	1/6930
TPM1 radical	NA	0/2178	0/6930	NA	0/2178	0/6930
ACTC1 radical	NA	0/2178	0/6930	NA	0/2178	0/6930
Missense only (All Genes)	86 (84–88)	442/2178	193/6930	78 (74–82)	277/2178	193/6930
MYBPC3 missense	81 (75–86)	118/2178	71/6930	69 (58–78)	73/2178	71/6930
- MYBPC3 unsubstituted	91 (85–95)	65/2178	18/6930	85 (74–91)	38/2178	18/6930
- MYBPC3 substituted	70 (57–80)	56/2178	53/6930	52 (27–69)	35/2178	53/6930
MYH7 missense	93 (90–95)	254/2178	58/6930	87 (83–91)	142/2178	58/6930
- MYH7 unsubstituted	96 (93–97)	164/2178	23/6930	92 (87–95)	90/2178	23/6930
- MYH7 substituted	88 (83–92)	93/2178	35/6930	79 (68–86)	52/2178	35/6930
TNNI3 missense	94 (84–98)	26/2178	5/6930	84 (54–95)	10/2178	5/6930
MYL2 missense	78 (49–91)	13/2178	9/6930	78 (49–91)	13/2178	9/6930
MYL3 missense	0 (0–55)	5/2178	19/6930	0 (0–55)	5/2178	19/6930
TNNC1 missense	78 (29–93)	7/2178	5/6930	78 (29–93)	7/2178	5/6930
TNNT2 missense	60 (19–80)	14/2178	18/6930	29 (0–69)	8/2178	18/6930
TPM1 missense	91 (79–97)	22/2178	6/6930	90 (74–96)	18/2178	6/6930
ACTC missense	94 (71–99)	11/2178	2/6930	91 (57–98)	7/2178	2/6930

EPVs are expressed as estimated predictive percentage value with 95% confidence interval.