Supplemental Material

Index

Table S6. Percentage of cases carriers of IEVs in sarcomeric and non-sarcomeric genes

Table S7. IEVs candidates that were non-validated in PCA analysis (non-pass IEVs)

Table S8. Small effect variants (risk-factors); enriched with ORs < 2

Table S9. Enrichment analysis in *TRIM63* null variants

Table S10. Homozygous/heterozygous IEVs cases described in the literature and identified in this study

Table S11. List of patients with IEV and monogenic variants in the same gene, stratified by variant (analysis to exclude linkage disequilibrium – LD)

Table S12. Epidemiological and Functional Additional evidence on selected IEVs

Table S13. Different IEV selection and key methodological differences with Meisner at al. study

Statistical Analysis and Data Availability

Statistical Analysis

Normal distribution variables were expressed as mean ± standard deviation (SD), while non-normal distribution variables were represented by the median and interquartile range (IQR; 25th–75th percentiles); categorical data were represented by frequencies (percentage of total).

Group comparisons were performed using the appropriate tests: Student's T-test for normally distributed continuous variables and chi-square test for categorical variables. To compare means across multiple groups of normally distributed data, ANOVA was applied. To account for multiple comparisons, the Benjamini-Hochberg procedure was applied for continuous and time-to-event data to control the false discovery rate (FDR).

For non-normally distributed variables, non-parametric tests were employed: Mann-Whitney U test for comparisons between two independent groups, and Kruskal-Wallis test for more than two groups whit Bonferroni correction used for categorical comparisons.

For age-related penetrance and survival analysis, Kaplan-Meier curves were generated to assess time-to-diagnosis and time-to-event outcomes; log-rank test was used to compare survival between groups. Pairwise comparisons were also conducted using post-hoc pairwise statistics.

To address potential biases from diverging survival curves (particularly evident around age 40) arising from decreasing numbers at risk or survival effects, we performed supplementary Cox proportional hazards regression analyses. Both univariate models and age-of-diagnosis-adjusted models were implemented to evaluate the independent impact of covariates on MACE-free survival. The proportional hazards (PH) assumption was tested using Schoenfeld residuals

All statistical tests were two-tailed, and a p-value of less than 0.05 was considered statistically significant. Statistical analysis was performed using R Studio version 4.3.2.

NGS Sequencing Methods

All the probands were sequenced in our center by NGS customized sequencing libraries. The number of genes varied through time, ranging from 242 genes in the first library in 2014, to 368 genes in the library that is currently used, being updated regularly to include genes with new evidence of association with inherited cardiac diseases (Table S1).

The libraries were enriched using a specificity hybridization probe kit SureSelect XT HS Low input (Agilent) and the obtained fragments were sequenced in paired-end mode (2x150bp) on the NovaSeq X Plus Sequencing System (Illumina). Sequencing data analysis was performed using a proprietary bioinformatics pipeline comprising sample demultiplexing, alignment refinement and adjustment, variant calling, normalization, sequence quality control, generation of coverage statistics by region of interest and quantification of the region of interest and quantification of copy number variation: HIC-Mutations, Version 11.8.8326.16811.

The read depth (number of times that a base was sequenced by independent reads) of every nucleotide of genes related to the referring phenotype was >30× (mean 250× to 400×). Bioinformatics analysis was performed by means of a custom pipeline including software for variant calling, genotyping, and annotation.

This library also includes whole gene sequencing (both coding and non-coding regions in their entirety), in the case of *MYBPC3, FLNC, DSP, LMNA, BAG3, PK*P2 and *SCN5A* and intronic regions +/- 200bp from the exon-intron transition of several other genes (*NKX2-5, TGFB3, ALPK3, DSC2, DSG2, EMD, FHL1, GLA, KCNH2, LAMP2, MYBPC3, PLN, TBX20, KCNQ1, PRDM16, DES, MYH7, TRIM63, DMD, RBM20, TTN, FBN1, TGFB2, TGFBR1, TGFBR2*).

An in-house software was used for detecting copy-number variation (CNVs), by means of a coverage normalization method, which compares the behavior of a sample case with the average behavior of the rest of the samples in the same run. Depending on the detection of split-reads, dosing plausibility, among other factors, a confidence score is obtained. These deviations, when they occur in genes relevant to the phenotype with a high score, are always confirmed by an alternative orthogonal technique (MLPA, Sanger, dPCR).

Tailored ACMG Criteria for Genetic Variants in Hypertrophic Cardiomyopathy

Pathogenicity Criteria

PVS1: Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease.

We applied this criterion to the following genes (inheritance pattern in brackets), all of which are part of our comprehensive HCM gene panel (primary and genocopy genes):

MYBPC3 (AD), ALPK3 (AD, AR), TRIM63 (AR), FHL1 (XL), GLA (XL), LAMP2 (XL), NF1 (AR), LZTR1 (AR), AARS2 (AR), ACAD9 (AR), ACADVL (AR), AGK (AR), AGL (AR), AGPAT2 (AR), ATPAF2 (AR), COA5 (AR), COA6 (AR), COQ2 (AR), COX15 (AR), COX6B1 (AR), DLDH (AR), DYSF (AR), ELAC2 (AR), FAH (AR), FNIP1 (AR), FOXRED1 (AR), FXN (AR), GAA (AR), GFM1 (AR), GLB1 (AR), GNPTAB (AR), GUSB (AR), GYG1 (AR), GYS1 (AR), KLHL24 (AR), LIAS (AR), MLYCD (AR), MRPL3 (AR), MRPL44 (AR), MRPS22 (AR), MTO1 (AR), NAA10 (XL), NDUFB11 (XL), PMM2 (AR), QRSL1 (AR), SCO2 (AR), SLC22A5 (AR), SLC25A3 (AR), SLC25A4 (AR), SPRED1 (AD), SURF1 (AR), TAZ (XL), TMEM70 (AR), TSFM (AR), VARS2 (AR).

In the genes not included in this list, this criterion was considered as not applicable.

We used this criterion as VERY STRONG, in those situations the variant was highly likely to be associated to NMD and the exon is incorporated into the reference transcript and also, in those splice-site variant when exon skipping is the expected impact preserving the reading frame (symmetric exon) in the case it would remove a region stablished as critical to protein function.

This criterion was modulated in the following situations:

PVS1_STRONG.

For nonsense, frameshift and/or splice-site variant when NMD is not the predicted mechanism and would remove >10% of the protein.

PVS1_MODERATE.

For nonsense, frameshift and/or splice-site variant when NMD is not the predicted mechanism and it would remove <10% of the protein, and it would remove a region without a critical established function.

For start-loss variants, when no alternative functional gene transcripts with different alternative start codon, and one or more pathogenic variant(s) have been reported 5’ of the next downstream putative

in-frame start codon (Methionine), we used PVS1_moderate. If no pathogenic variant(s) occurs upstream of the new Methionine, we used PVS1_Supporting.

PS1: Different nucleotide change (same amino acid) to a previously established pathogenic variant.

PS1 was also used in the context of exonic and intronic variants based on similarity of predicted RNA effects for a variant under assessment in comparison with a known pathogenic variant, according to the last recommendations in this regard. For variants located outside splice donor/acceptor ±1,2 dinucleotide positions, in the case there is another variant affecting the same nucleotide classified as pathogenic we apply PS1_strong (PVS1_moderate whether the other variant is LP); and in the case of there is another variant within same splice donor/acceptor motif (including at ±1,2 positions) classified as pathogenic we apply PS1_moderate (PVS1_supporting whether the other variant is LP).

PS3:

Functional studies of mammalian knock-in models supportive of a damaging effect on the gene or gene product.

According to the latest ClinGen agreement, for missense variants functionally characterized, PS3 should be modulated as moderate.

> MODERATE: "Mammalian variant-specific knock-in animal models that produce a phenotype consistent with the clinical phenotype in humans (e.g., structural and/or functional cardiac abnormalities, premature death, arrhythmia)".
>
> SUPPORTING: In vitro assays (e.g., biochemical assays of myofilament function, motility assays, human iPSC-CM)
>
> While some in vitro assays may provide evidence that a variant in a cardiomyopathy gene has an effect on protein and/or myofilament function, at present, there are no validated "gold-standard" assays that are considered to reliably predict the clinical phenotype.

In the case of spliceogenic variants, when a splicing assay data–assay demonstrating that a variant leads to aberrant splicing profile existed, PS3 was replaced by PVS1 with the corresponding strength as previously discussed.

PS2: De novo (paternity confirmed) in a patient with disease and no family history.

PM2: Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes or Exome Aggregation Consortium.

A threshold of ≤0.00004 in the subpopulation with the highest frequency in gnomAD was used for activating this rule. Used by default as SUPPORTING.

PS4: The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls.

The count of unrelated probands was only applied when the variant was very rare in the control population, and/or had a low phenotypic dispersion in our database.

Strong: ≥ 10 probands with consistent phenotypes.

- ≥ 5 probands with highly specific phenotypes such as RASopathies suspicion.

Moderate: ≥ 6 probands with consistent phenotypes.

- 3-4 probands with highly specific phenotypes such as RASopathies suspicion.

Supporting: ≥ 2 probands with consistent phenotypes.

- 1-2 probands with highly specific phenotypes such as RASopathies suspicion.

PP1: Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease.

STRONG: ≥7 segregations (LOD score of 2.1).

MODERATE: ≥5 segregations (LOD score of 1.5).

SUPPORTING: ≥3 segregations (LOD score of 0.9).

PM1: Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation.

Used by default as MODERATE. In the genes not included in this list, this criterion was considered as not applicable.

- *MYH7*, (amino acids 181–937)[20]
- *MYBPC3* (amino acids 485-502 & 1248-1266) (using the reference isoform ENST00000545968 & transcript version NM_000256.3).
- *TNNT2* (amino acids 79-179) (using the reference isoform ENST00000367318 & transcript version NM_001001430.2)[20]
- *TNNI3* (amino acids 141-209) (using the reference isoform ENST00000344887 & transcript version NM_000363.5)[20]
- *BRAF* (residues 459-474, P-loop; & 594-627, CR3 activation segment).
- *MAP2K1* (residues 43-61, 124-134)

- *MAP2K2* (residues 47-65, 128-138)
- *PTPN11* (residues 4, 7-9, 58-63, 69-77, 247, 251, 255, 256, 258, 261, 265, 278-281,284, interacting residues between N-SH2 and PTPN domains).
- *RAF1* (residues 251-266, CR2 domain).
- *SOS1* (residues 420-500).
- *GAA* (residues D282, W376, D404, L405, I441, W481, W516, D518, M519, R600, W613, D616, W618, F649, L650, H674, active site architecture and substrate binding of GAA).

PM4: Protein length changes due to in-frame deletions/insertions of any size in a nonrepeat region or stop-loss variants.

PM5: Missense change at an amino acid residue where a different missense change was previously established as pathogenic. This criterion was upgraded to STRONG whether there two or more disease-causing variants in the same residue. PM1 and PM5 were not concomitantly used, and if PM5 applies to any strength, we applied PM5 because it is a more variant-specific criterion.

PM6: Confirmed de novo without confirmation of paternity.

PP3: Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.). Used by default as SUPPORTING.

- For non-synonymous variants, we use REVEL with a thresholds of ≥0.70 for PP3.[37]
- For evaluating splicing, we have used agreement with the meta-scores with the most specificity/sensitivity with a value of SpliceAI ≥ 0.2[38]

PP2: Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease.

Used by default as SUPPORTING. Only applicable to *TPM1*, AD RASopathies genes according to the specific guidelines (*BRAF, CBL, HRAS, KRAS, NRAS, MAP2K1, MAP2K2, PTPN11, RAF1, RIT1, SHOC2, SOS1, SOS2*),

Benignity Criteria

BA1: Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes or Exome Aggregation Consortium.

Allele frequency is ≥ 0.1% based on the filtering allele frequency in ExAC. STAND-ALONE criteria.

BS1: Allele frequency is greater than expected for disorder.
This criteria is considered gene-specific.
Allele frequency is ≥ 0.01% based on the filtering allele frequency in gnomAD.
For *MYBPC3*, according to last ClinGen specifications, BS1 was activated when the allele frequency is ≥0.02%.
For recessive disorders, highest minor allele frequency >0.005 (>0.5%) in any continental population in gnomAD with >2000 alleles.

BS3: Functional studies of mammalian knock-in models supportive of no damaging effect on protein function or splicing.

BS4: No segregation in affected members of a family.

BP2: Observed as comp het (in trans) or double het in genes with overlapping function (e.g., sarcomere genes) without increased disease severity or observed in cis with a pathogenic variant in any inheritance pattern.

BP4: Multiple lines of computational evidence suggest no impact on gene or gene product.

- REVEL score <0.4 for missense variants.
- No predicted impact on splicing by SpliceAI (score <0.2)

BP5: Variant found in a case with an alternate molecular basis for disease.

- Not applicable for primary HCM genes.

BP6: Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation.
This benign criterion was not applied to any variant in any gene.

BP7: A silent variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site -AND- the nucleotide is not highly conserved.

Penetrance estimation

Penetrance, defined as the probability of disease manifestation given the presence of given variant, is quantified on a scale from 0 to 1 or as a percentage. The estimation process utilizes case-population data within a Binomial framework, applying Bayes' theorem:

$$P(D|A) = \frac{P(D) \times P(A|D)}{P(A)}$$

Where:

Penetrance: P(D|A), probability of disease (D) given the presence of a IEV (A).

Population Prevalence: P(D), baseline risk of disease in the general population, adjusted to the most widely accepted estimated prevalence for HCM, which is 0.002 (1 in 500 individuals).

Case Allele Frequency: P(A|D), Allele frequency of the IEV in HCM NFE cohort.

Population Allele Frequency: P(A), Allele frequency of the IEV in the NFE population in gnomAD.

This approach allows for the calculation of penetrance by integrating population-level disease prevalence with allele frequencies observed in case and control cohorts. The 95% confidence interval is likely computed using standard statistical methods to provide a range of plausible penetrance values, accounting for sampling variability in the observed allele.[20,39]

PCA Analysis

To conduct ancestry analysis of the HCM and internal control samples, single nucleotide polymorphisms (SNPs) included in the Health in Code cardiac gene panel were identified (using only variants in genes present in all iterations of the panel). SNPs located on non-autosomal chromosomes, those with a minor allele frequency (MAF) in all samples < 0.01 and those with a large difference in MAF (>0.1) compared to gnomAD-NFE samples were excluded, the latter based on the assumption most case and control samples in this study are likely to be of European ancestry. A total of 379 SNPs were included for the ancestry analysis.

Individual level genotyping data for these 379 SNPs was extracted from 2,504 samples in the 1000 Genomes Project (1KG). Multi-dimensional scaling (MDS) analysis using PLINK software was performed and principal component analysis (PCA) plots were derived to assess ability of this selected list of panel-derived SNPs to distinguish between the major ancestral groups in 1KG. As shown in Figure S1A, European samples can still be clearly differentiated from East Asian and African samples but some overlap is observed with South Asian samples (admixed American samples displayed typically less discrete clustering including substantial overlap with European samples).

The HCM case and non-cardiomyopathy control samples were then added to the 1KG PCA analysis. A subset of samples were distributed along a different principal component to the ancestry differences, likely due to technical artefacts or missing data, and were therefore removed as their ancestry could not be accurately assessed. For the remaining samples, the clustering of the case and control samples with the 1KG data is shown in Figure S1B. As expected, most of the samples broadly overlap with the 1KG European samples, with a minority of samples clustering with 1KG African, East Asian and South Asian cohorts. European ancestry samples were selected as indicated by the red circle in Figure S1B to enable ancestry informed case-control analysis and comparison with data from gnomAD non-Finnish Europeans. A total of 10990 HCM cases and 4031 internal controls were selected from this analysis.

Estimation of the contribution of IEV to HCM (PAF)

To further explore the contribution of intermediate-effect variants to the overall burden of HCM, we analyzed the relative risk and **population attributable fraction (PAF)** associated with these variants through an adjusted etiological fraction analysis to the global proportion of these variants in our NFE case cohort weighting the estimated penetrance of each variant, (Levin`s formula).

The relative risk (RR) was calculated by comparing the estimated penetrance of each variant (risk under exposure) to the global prevalence of the disease in the general population (1/500) (risk under non-exposure). This estimation follows previously established methodologies.[39-42] Although it is not exclusive to a single reference, it is based on concepts widely used in population genetics and genetic epidemiology studies, such as the calculation of the Population Attributable Fraction (PAF) or the genetic contribution of variants to disease risk.

$$\text{PAF } IEVs = \frac{\sum_{i=1}^{n} fi * (RRi - 1)}{\sum_{i=1}^{n} fi * (RRi - 1) + 1}$$

Where:

fi = Frequency in HCM cases carrying the variant i in the case cohort.

RRi = Relative Risk (RR) associated with variant i, which can be estimated as $RRi = Pi$ / Pbackground, where:

- Pi = Penetrance of variant i in individuals carrying the variant (as proportion 0-1).
- Pbackground = Baseline risk of the disease in the general population. (0.002).

The addition of 1 in the denominator ensures that the attributable fraction is correctly adjusted and does not exceed a value of 1, accurately reflecting the relative impact of risk factors. This adjustment is particularly useful to prevent bias when there is a high prevalence in the unexposed group.

In some versions of the attributable fraction formula, adding 1 to the denominator transforms the term into a relative ratio that adjusts for total risk. This adjustment is especially useful when evaluating models with a high prevalence of the disease, as the risk in non-exposed individuals may not be zero. By adding 1 to the denominator, we ensure that the calculation is adjusted, making comparisons between exposed and non-exposed groups meaningful within the context of the overall disease events.

Supplemental Figures and Figure Legends

Figure S1: Principal Component Analysis (PCA). A. PCA of 1000 Genomes samples using 379 SNPs present on the Health in Code cardiac gene panel. **B.** PCA of 1000 Genomes samples and HCM cases and internal controls using 379 SNPs present on the Health in Code cardiac gene panel. Samples circled in red were selected as likely European ancestry samples for ancestry specific case-control comparison.

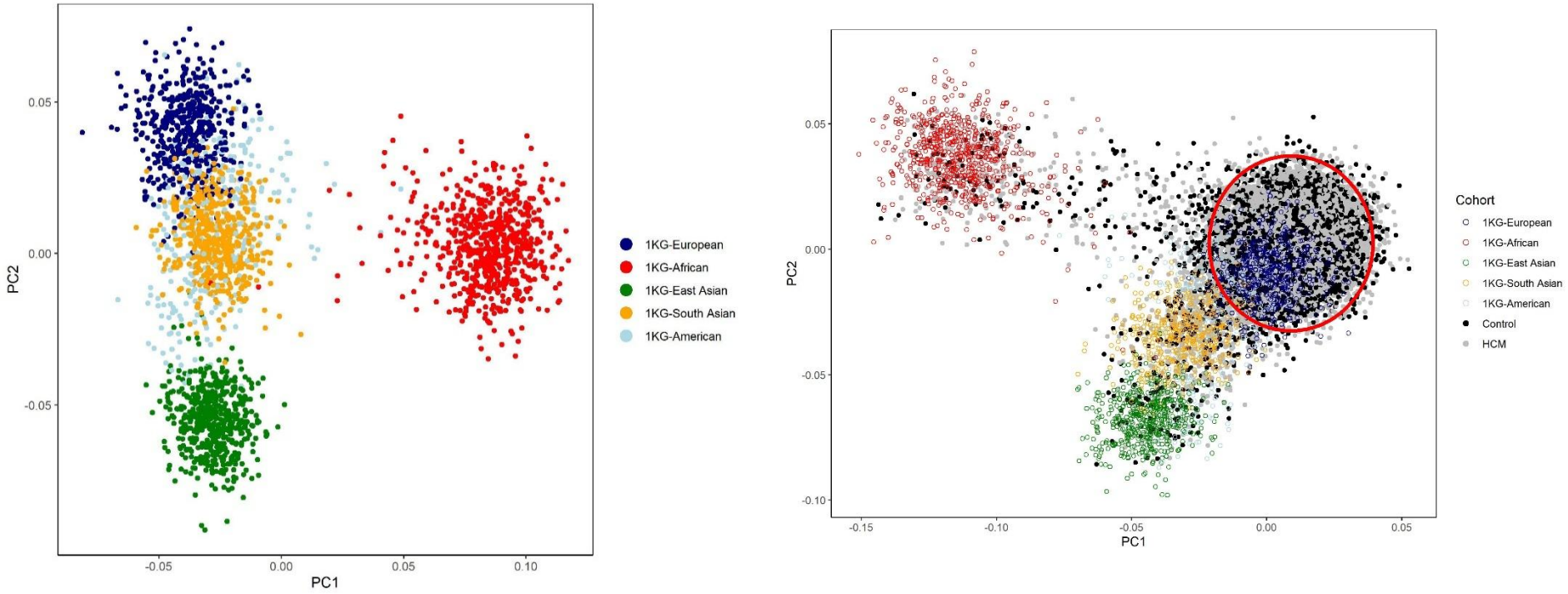


Figure S2: Comparisson of internal vs external ORs after PCA analysis

Comparisson of internal vs external ORs after PCA analysis. HCM versus control odds ratios (ORs) for candidate intermediate effect variants (IEVs) showing European cases versus internal controls (y-axis) and European cases versus gnomAD-NFE (x-axis). The OR threshold utilised in this study (OR=2) is shown by the red dashed line. The final selected IEVs are shown in green. The variants annotated are those with substantially higher ORs using gnomAD-NFE, i.e. the population frequency of these variants in gnomAD-NFE (strongly enriched for Northwest Europeans) is lower than the internal controls (enriched for Southern Europeans).

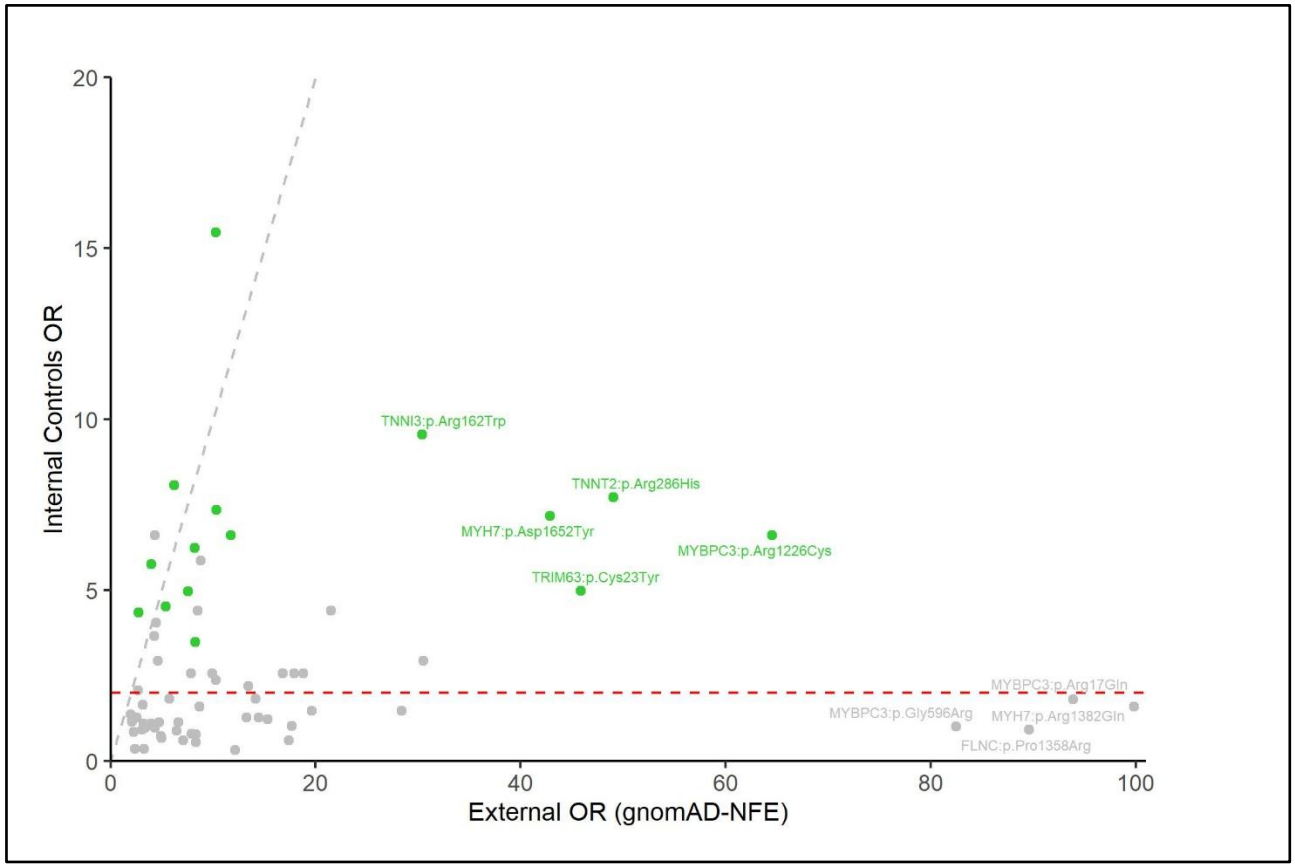


Figure S3. Comparative phenotype analysis of carriers of the most prevalent intermediate-effect variant (FHOD3, p.Arg637Gln) versus other IEVs, in isolation. Clinical characteristics of individuals carrying the most prevalent intermediate-effect variant (IEV) in the cohort (FHOD3 p.Arg637Gln; Group 1) versus carriers of any other IEV (Group 0). Only carriers of IEVs variants in isolation were selected (those with other P/LP disease-causing variants were exluded). Top row: Kaplan–Meier curves for age at diagnosis and MACE-free survival. Bottom left: Violin plots showing left ventricular maximum wall thickness (LVMWT) distribution across groups. Bottom center and right: Bar plots showing the proportion of individuals with LVMWT >25 mm and >30 mm, respectively.

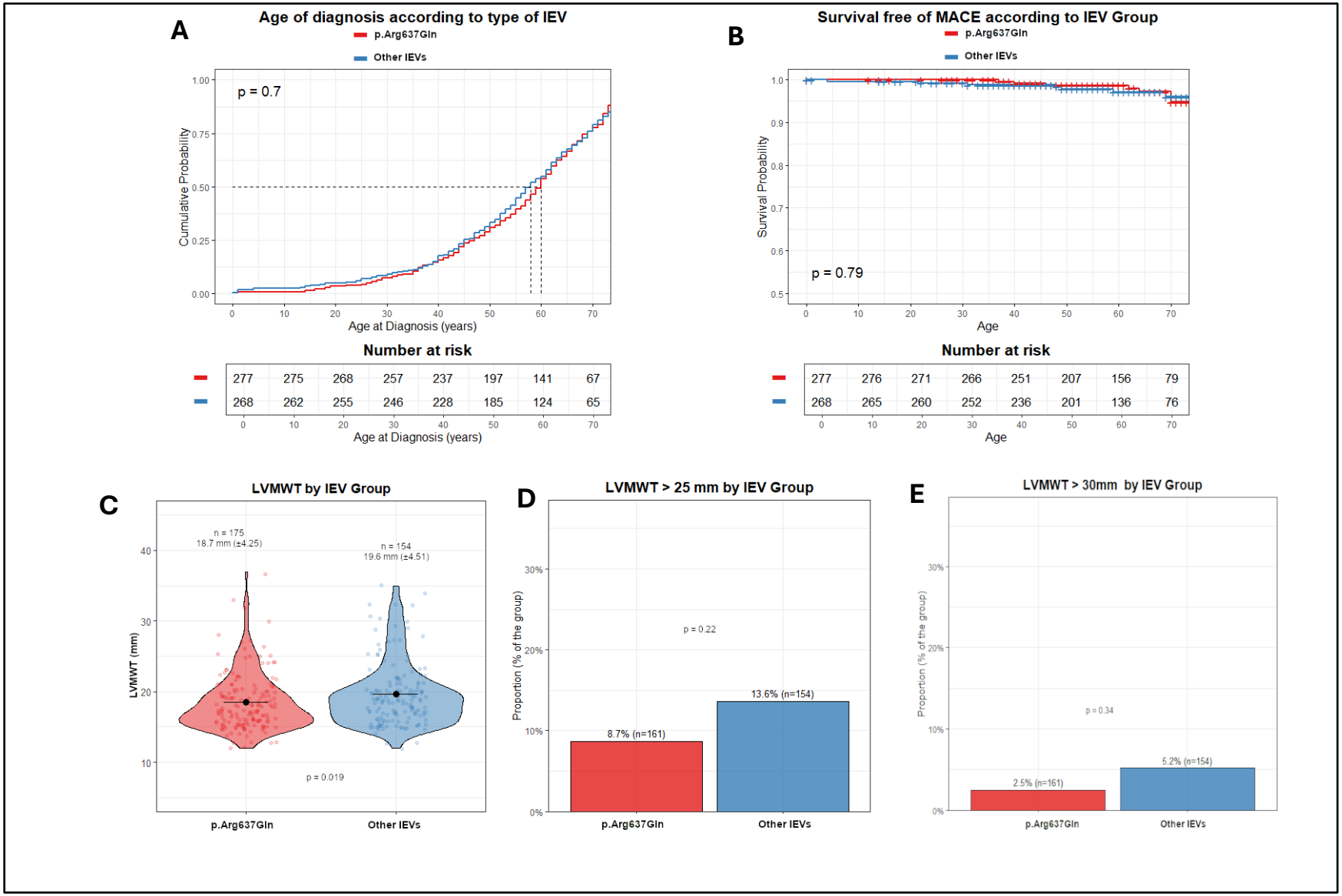
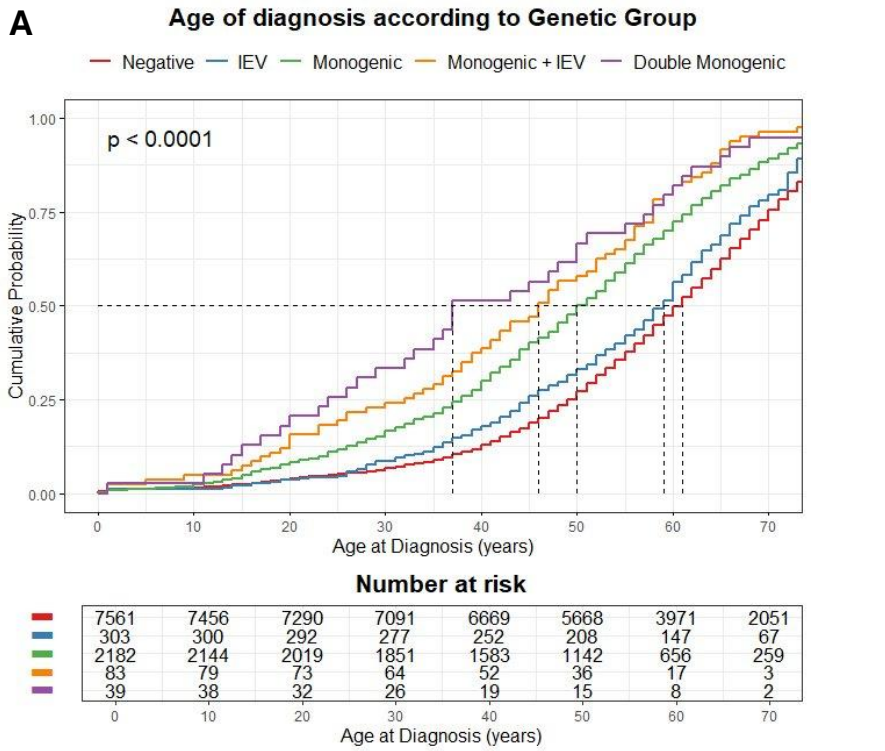
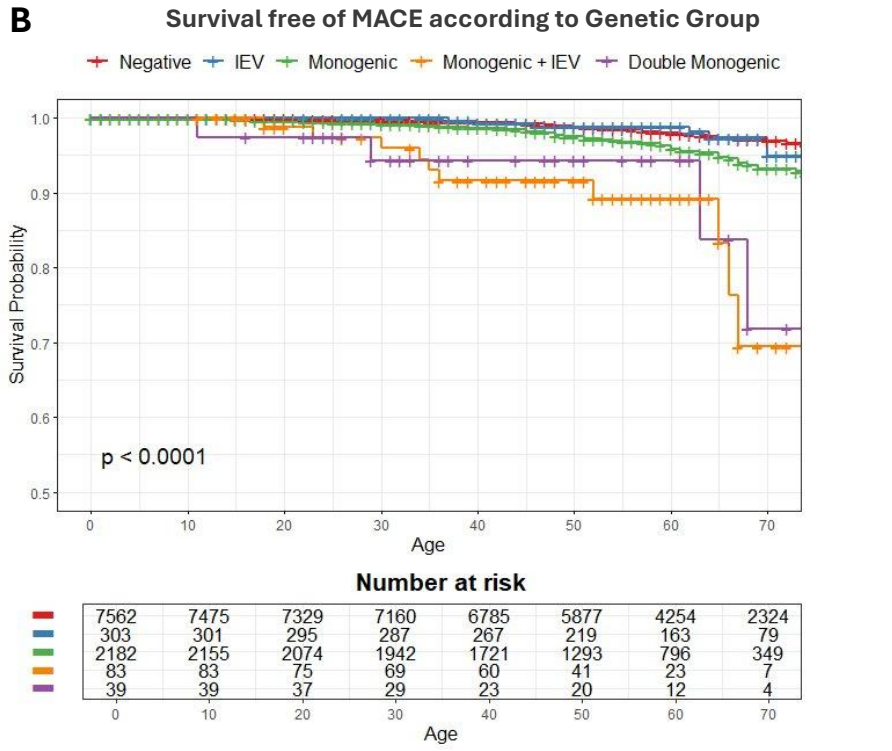


Figure S4A. Clinical impact and modifier effect of intermediate-effect variants (IEVs) after FHOD3:p.Arg637Gln exclusion: sensitivity analysis. Age-related penetrance and MACE. Sensitivity analysis evaluating the isolated clinical impact and modifier effect of intermediate-effect variants (IEVs) excluding FHOD3 p.Arg637Gln. Groups of comparison were: Negative, IEV (excluding FHOD3 p.Arg637Gln), Monogenic, Monogenic + IEV (excluding FHOD3 p.Arg637Gln), and Double Monogenic. Kaplan–Meier curves for age at diagnosis and MACE-free survival.





Pairwise comparisons (long-rank test; adj. BH)

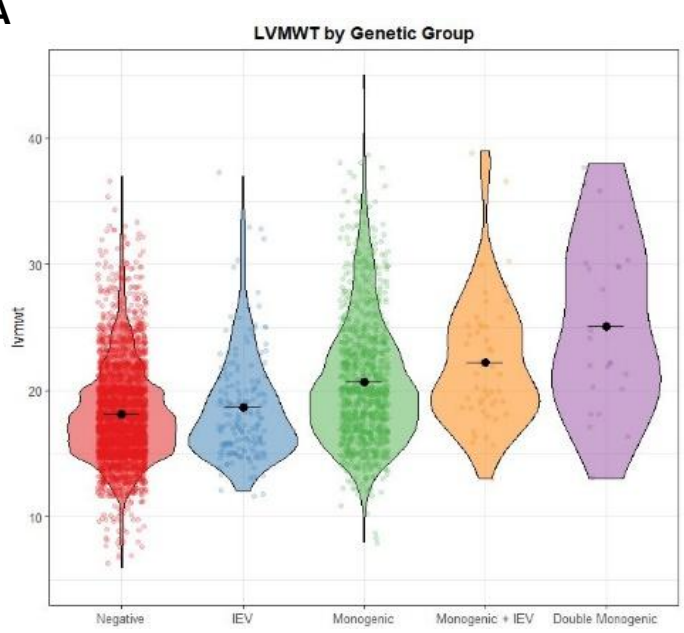
	Negative	IEV	Monogenic	Monogenic + IEV	Double Monogenic
Negative	-				
IEV	0.0072	-			
Monogenic	<2e-16	1.7e-08	-		
Monogenic + IEV	4.0e-13	5.6e-07	0.0653	-	
Double Monogenic	2.8e-09	1.4e-05	0.0328	0.5901	-

Pairwise comparisons (long-rank test; adj. BH)

	Negative	IEV	Monogenic	Monogenic + IEV	Double Monogenic
Negative	-				
IEV	0.4080	-			
Monogenic	3.3e-07	0.2684	-		
Monogenic + IEV	1.5e-13	1.4e-05	1.5e-05	-	
Double Monogenic	1.5e-05	0.0019	0.0169	0.7108	-

Figure S4B. Clinical impact and modifier effect of intermediate-effect variants (IEVs) after FHOD3 p.Arg637Gln exclusion. Sensitivity analysis evaluating the isolated clinical impact and modifier effect of intermediate-effect variants (IEVs) excluding FHOD3 p.Arg637Gln. Groups of comparison were: Negative, IEV (excluding FHOD3 p.Arg637Gln), Monogenic, Monogenic + IEV (excluding FHOD3 p.Arg637Gln), and Double Monogenic. **Panel A**: Violin plots displaying left ventricular maximum wall thickness (LVMWT) distribution across genetic groups. **Panels B and C:** Bar plots showing the proportion o f individ uals with LVMWT >25 mm and >30 mm.

A

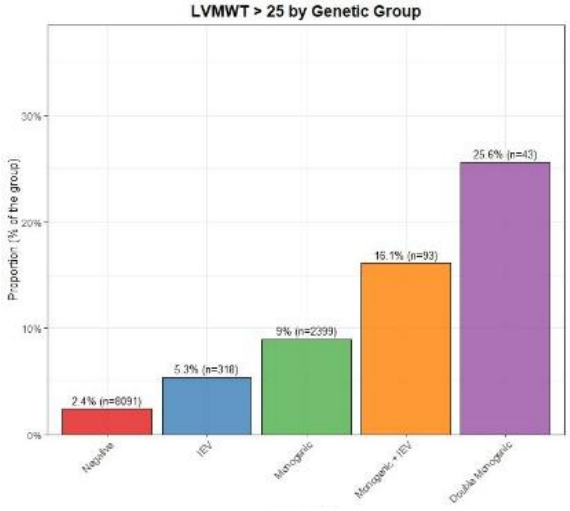


Group	n	Mean	Median	SD	Range
Negative	3339	18.1	18	3.74	6–37
IEV	175	18.7	18	4.25	12–37
Monogenic	1144	20.7	21	4.98	13–37
Monogenic + IEV	47	22.2	21	5.13	13–39
Double Monogenic	24	25	23	6.91	15–40

Pairwise comparisons

Group1	Group2	p value (adj. Bonferroni)
Negative	IEV	0.0072
Negative	Monogenic	<2E-16
Negative	Monogenic + IEV	4.00E-13
Negative	Double Monogenic	2.80E-09
IEV	Monogenic	1.70E-08
IEV	Monogenic + IEV	5.60E-07
IEV	Double Monogenic	1.40E-05
Monogenic	Monogenic + IEV	0.0653
Monogenic	Double Monogenic	0.0328
Monogenic + IEV	Double Monogenic	0.5901

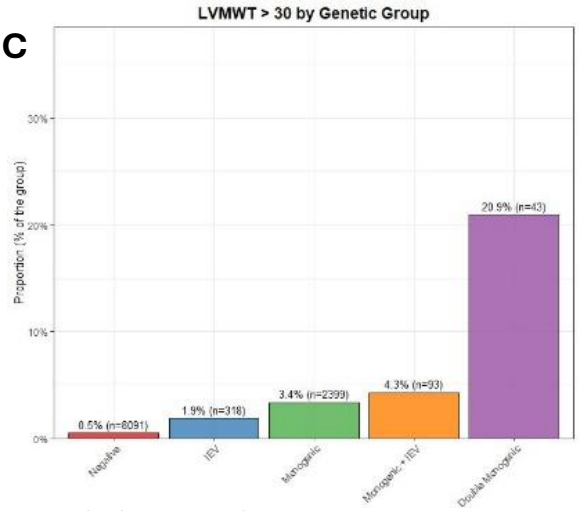
B



Pairwise comparisons

Group1	Group2	p value (adj. Bonferroni)
IEV	Negative	0.02122
Monogenic	Negative	<2e-16
Monogenic	IEV	0.99325
Monogenic + IEV	Negative	<2e-16
Monogenic + IEV	IEV	0.89216
Monogenic + IEV	Monogenic	0.30746
Double Monogenic	Negative	<2e-16
Double Monogenic	IEV	0.16014
Double Monogenic	Monogenic	0.005316
Double Monogenic	Monogenic + IEV	1

C

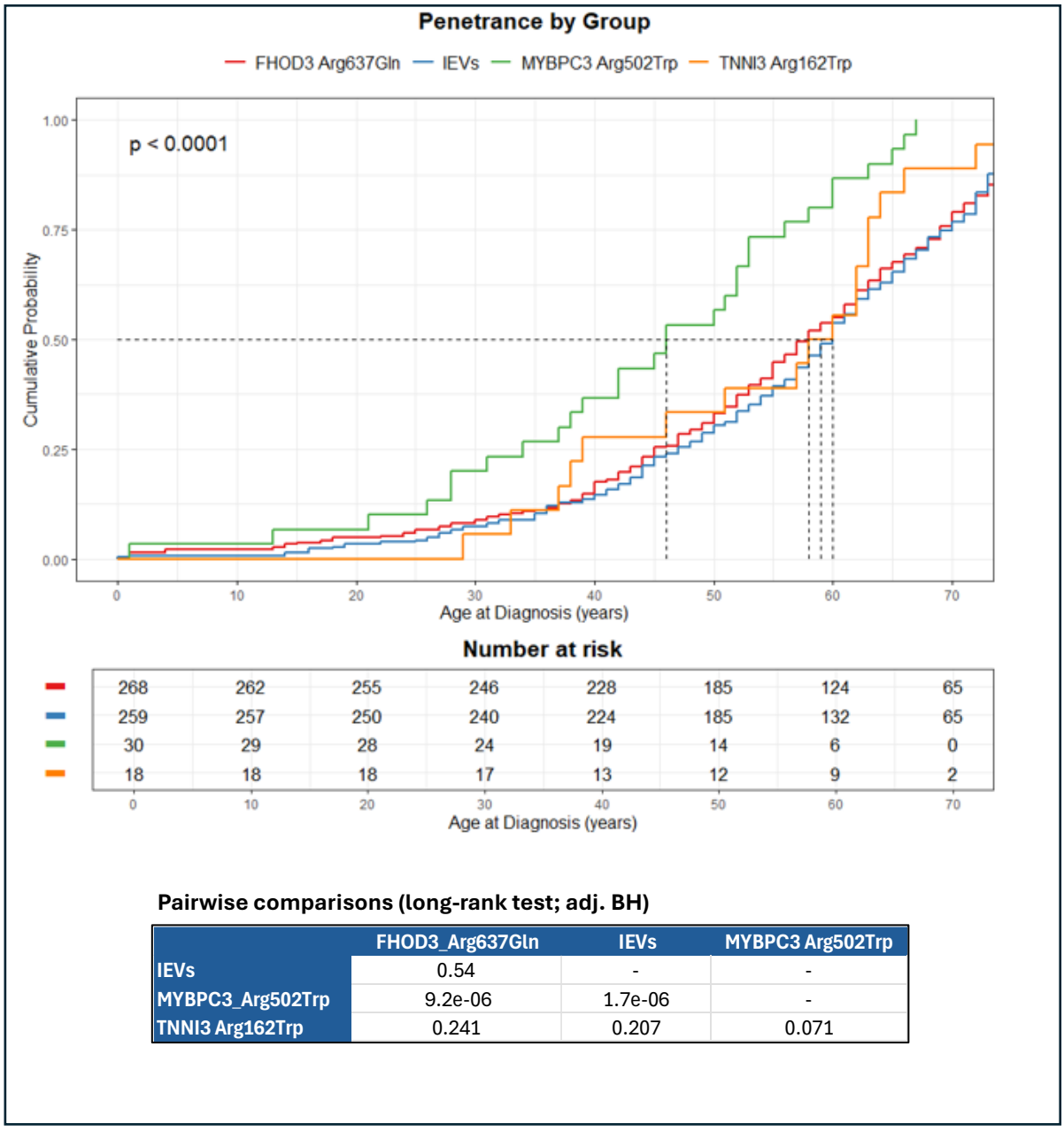


Pairwise comparisons

Group1	Group2	p value (adj. Bonferroni)
IEV	Negative	0.02198
Monogenic	Negative	<2e-16
Monogenic	IEV	1
Monogenic + IEV	Negative	0.000164
Monogenic + IEV	IEV	1
Monogenic + IEV	Monogenic	0.05901
Double Monogenic	Negative	<2e-16
Double Monogenic	IEV	0.06713
Double Monogenic	Monogenic	1.6e-07
Double Monogenic	Monogenic + IEV	0.05901

Figure S5: Age-Dependent penetrance analysis comparing *FHOD3* p.Arg637Gln, *TNNI3* p. Arg162Gln, and *MYBPC3* p.Arg502Trp with the rest of IEVs

Kaplan-Meier curves depict the cumulative probability of HCM diagnosis as a function of age for individuals carrying the FHOD3 Arg637Gln, IEVS, MYBPC3 Arg502Trp, and TNNI3 Arg162Trp variants. The table below each curve provides the number of individuals at risk at different ages, and the p-values for difference across variants, as indicated in the figure. Compared to the FHOD3 Arg637Gln, IEVS, and TNNI3 Arg162Trp variants, the MYBPC3 Arg502Trp variant demonstrates a notably higher age-dependent penetrance, while no statistically significant differences in penetrance were observed among the remaining variants.



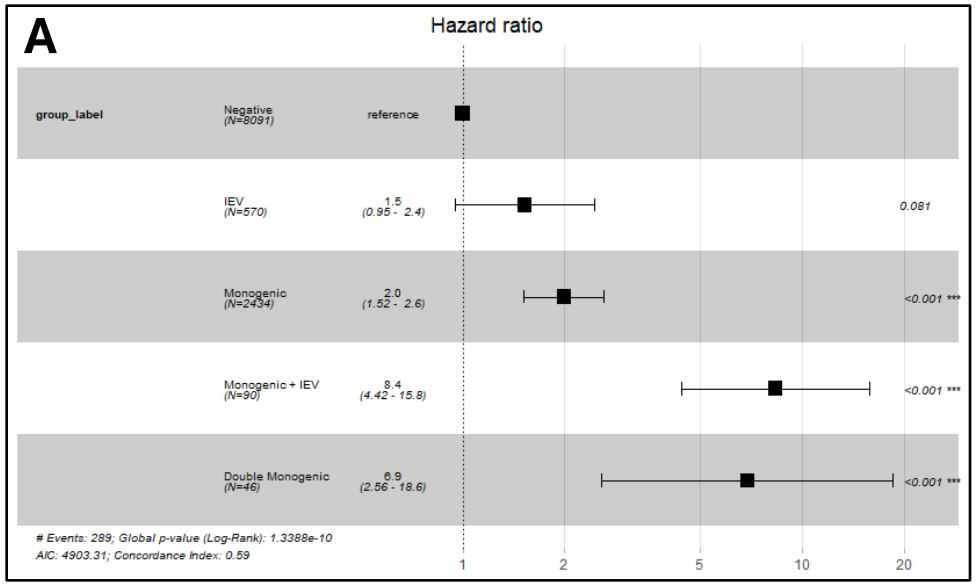
Pairwise comparisons (long-rank test; adj. BH)

	FHOD3_Arg637Gln	IEVs	MYBPC3 Arg502Trp
IEVs	0.54	-	-
MYBPC3_Arg502Trp	9.2e-06	1.7e-06	-
TNNI3 Arg162Trp	0.241	0.207	0.071

Figure S6. Forest plots of Cox regression Analyses for MACE by genetic group

Panel A. Univariate Cox model for MACE according to the Genetic Group. Each square marks the hazard ratio (HR) for the specified genetic group relative to the Negative reference, and horizontal lines show the 95 % confidence intervals. Monogenic + IEV carriers exhibit a significantly elevated HR, also observed in the Monogenic and the Double Monogenic Groups, with a non-significant tendency in the IEV Group (p=0.081).

Panel B. Multivariable Cox model adjusted for age at diagnosis. This plot uses the same layout but additionally controls for the effect of age at diagnosis. The Monogenic + IEV group remains the only category with a significant increase in MACE risk after adjustment. Proportional hazards assumption testing revealed a violation for age at diagnosis in this model, but not for the genetic group variable. This limitation was addressed in an additional age-stratified Cox model (<50, 50–65, >65 years; Supplementary Table S7), which confirmed that the monogenic + IEV group remained with significantly increased MACE risk (HR 3.91; 95% CI 2.05–7.43; p<0.001).



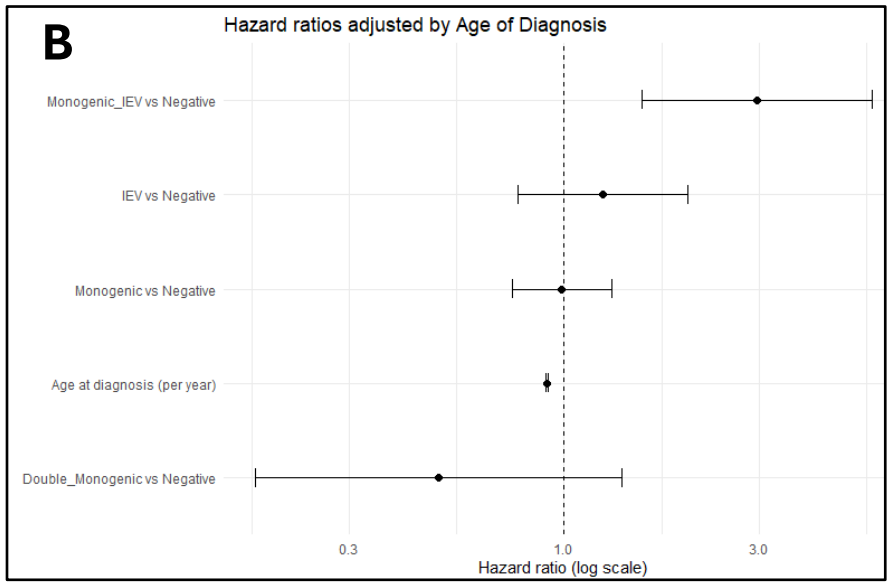
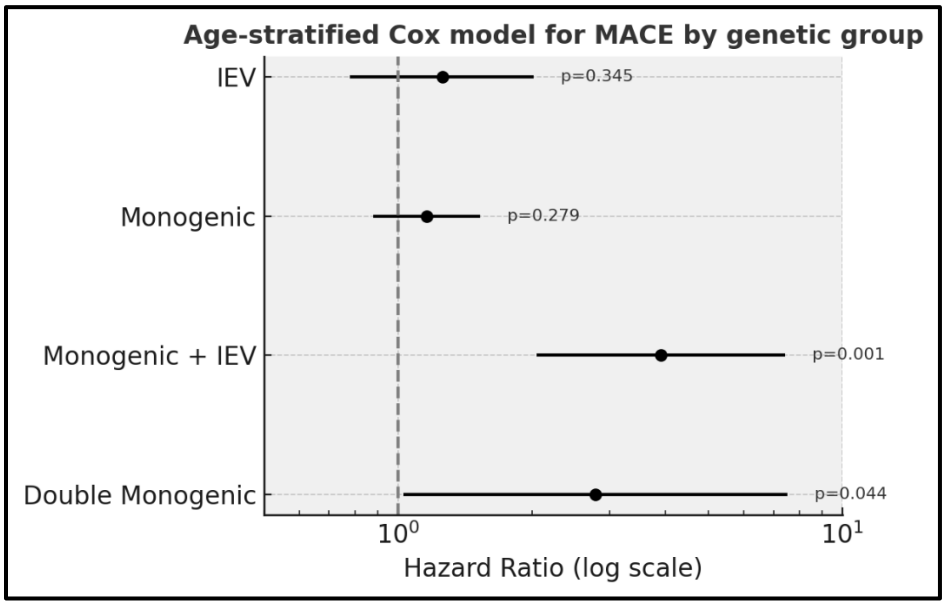


Figure S7. Age-stratified (<50, 50–65, >65 years) Cox regression model for MACE by genetic group.

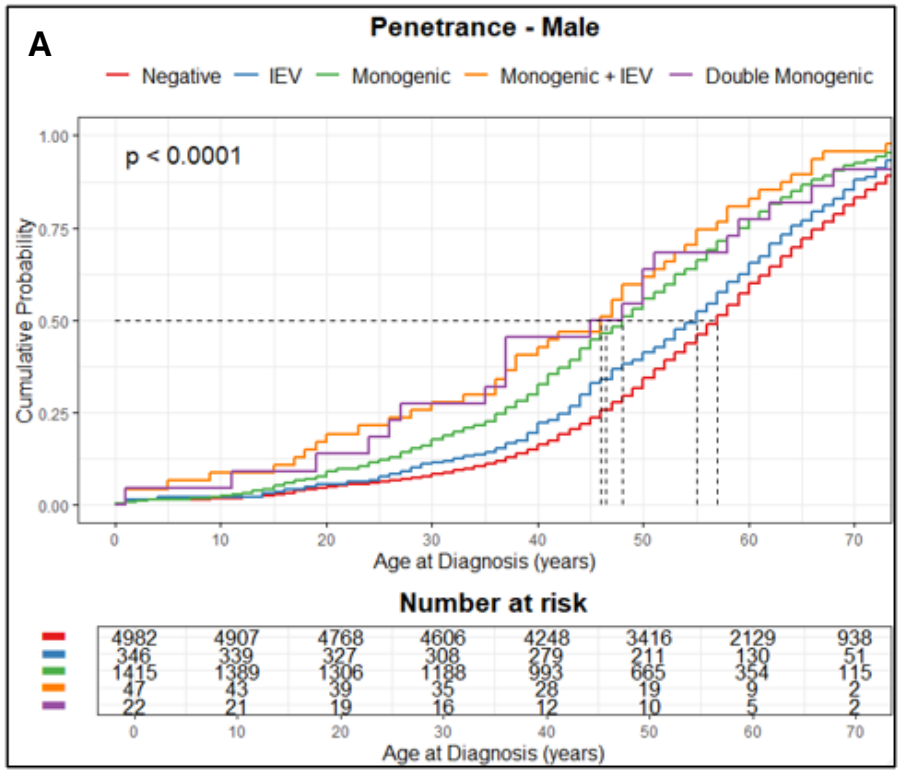
Cox model stratified by age at diagnosis (<50, 50–65, >65 years) to address violation of the PH assumption observed for age at diagnosis in the multivariable model (Supplementary Figure S6, Panel B). Hazard ratios (HR) and 95% confidence intervals (CI) are shown on a log scale. The monogenic + IEV group remains with significantly increased MACE risk (HR 3.91; 95% CI 2.05–7.43; p<0.001). Reference category = Negative.

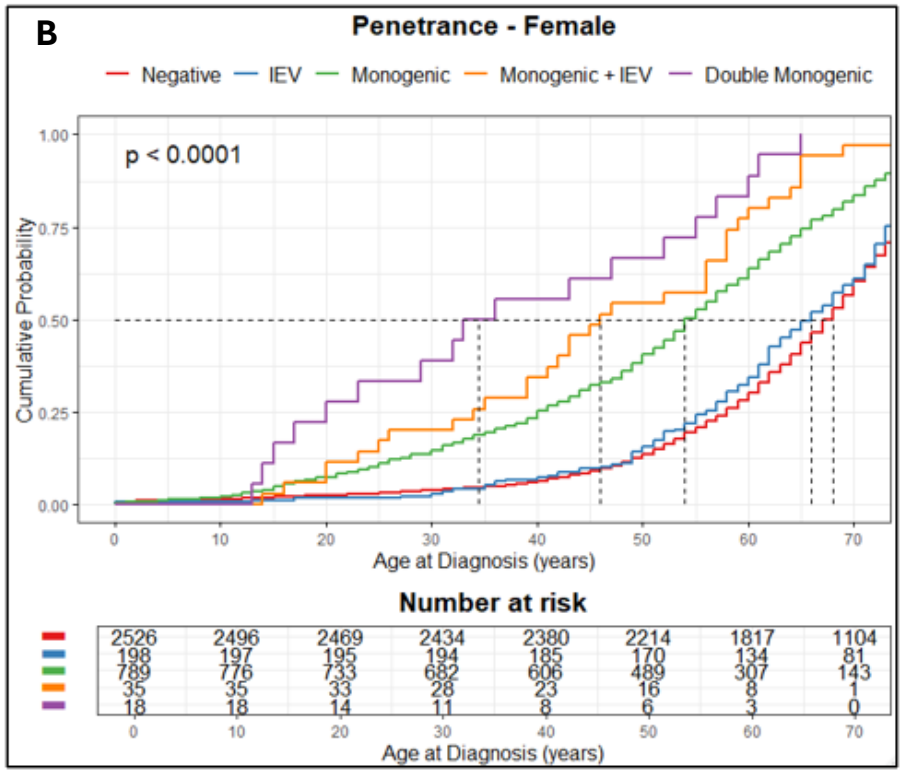


Genetic group	HR	95% CI	p-value
IEV	1.26	0.78 – 2.02	0.345
Monogenic	1.16	0.88 – 1.53	0.279
Monogenic + IEV	3.91	2.05 – 7.43	<0.001
Double Monogenic	2.78	1.03 – 7.53	0.044
Global PH test	—	—	0.76

Figure S8A: Age-related penetrance stratified by Sex

Kaplan-Meier curves showing the cumulative probability of HCM diagnosis by age in males (left) and females (right) for different genetic groups: Negative, IEV, Monogenic, Monogenic + IEV, and Double Monogenic. The tables below show the number of individuals at risk by age and the p-values for pairwise comparisons between groups. The global differences between the genetic groups remain statistically significant (Log-Rank p < 0.0001) in both males and females.





Pairwise comparisons (long-rank test; adj. BH)

	Negative	IEV	Monogenic	Monogenic + IEV
IEV	0.00246	-	-	-
Monogenic	<2e-16	1.5e-05	-	-
Monogenic + IEV	0.0019	0.005485	0.4622	-
Double Monogenic	0.04376	0.31458	0.93196	0.66247

Pairwise comparisons (long-rank test; adj. BH)

	Negative	IEV	Monogenic	Monogenic + IEV
IEV	0.10174	-	-	-
Monogenic	<2e-16	1.2e-11	-	-
Monogenic + IEV	<1.2e-15	4.8e-10	0.02651	-
Double Monogenic	<2e-16	9.9e-15	0.00051	0.10174

Figure S8B: Age-related penetrance stratified by Sex

Kaplan-Meier curves depict the cumulative probability of HCM diagnosis by age, stratified by gender (females in red, males in blue), for different genetic groups: Negative, IEV, Monogenic, Monogenic + IEV, and Double Monogenic. P-values indicate the statistical significance of the difference between males and females within each group. Overall, penetrance is higher in males compared to females across most genetic groups; however, this trend is attenuated or reversed in the Monogenic + IEV and Double Monogenic groups, suggesting that the impact of these specific variant combinations may diminish the effect of sex on HCM penetrance.

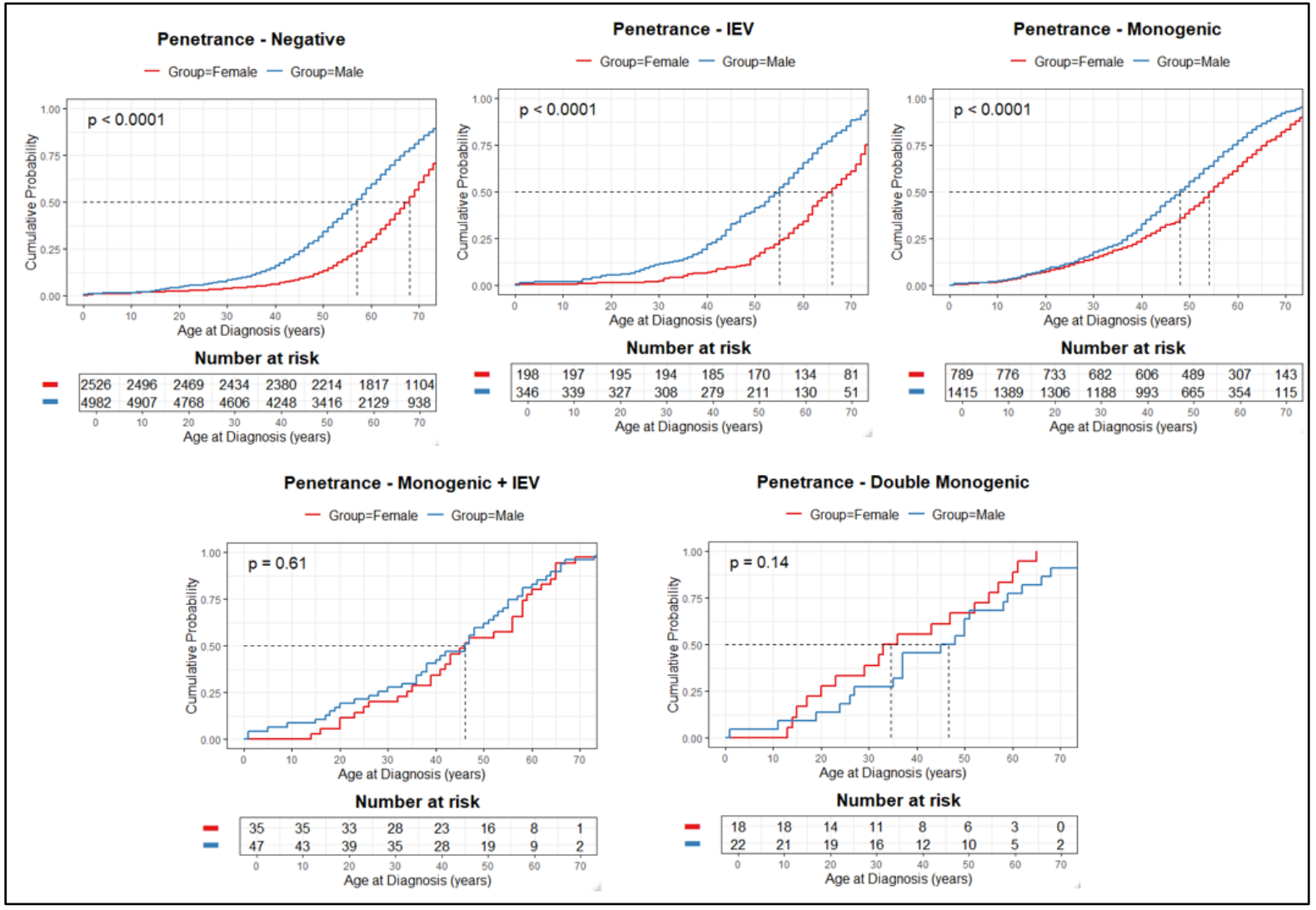
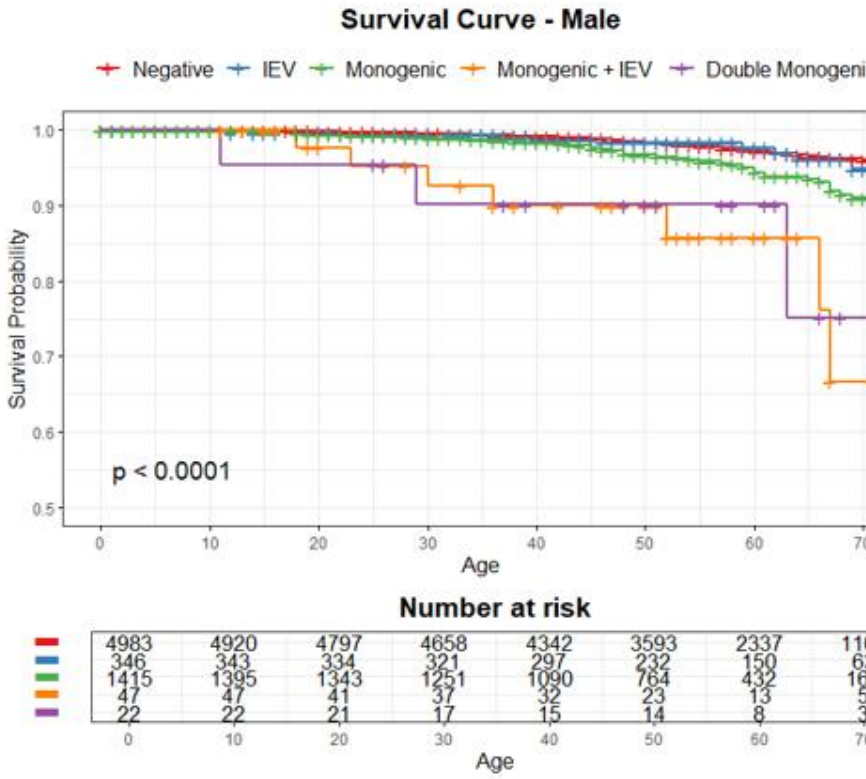
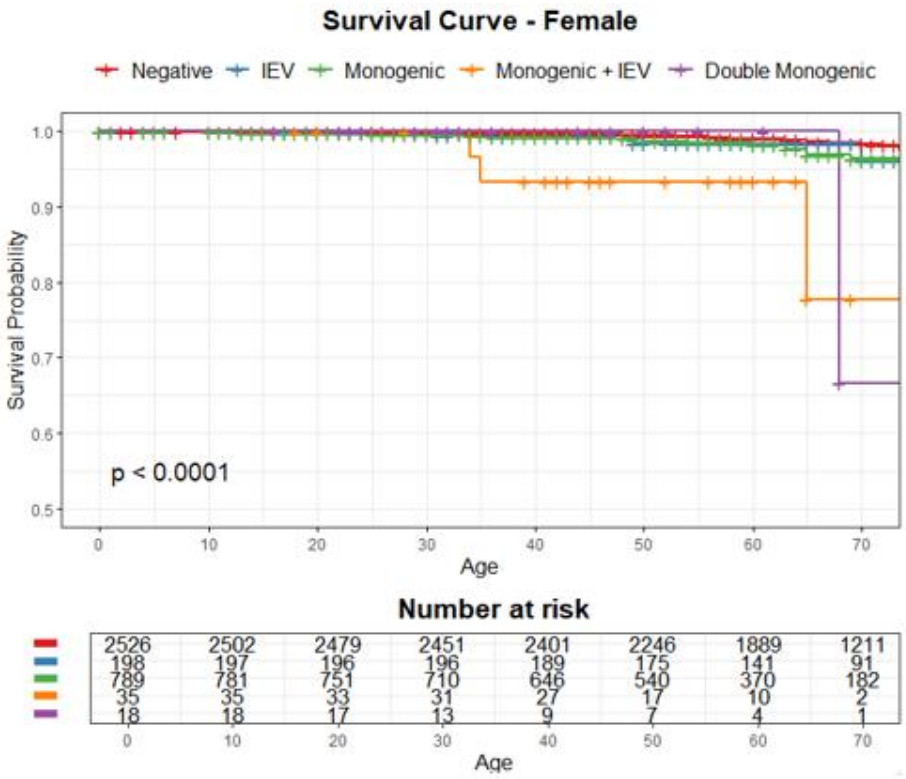


Figure S9: Survival analysis for MACE stratified by Sex

Kaplan-Meier curves illustrating MACE-free survival probability by age, stratified by gender (males on the left, females on the right) across different genetic groups: Negative, IEV, Monogenic, Monogenic + IEV, and Double Monogenic. Log-Rank p-values ($p < 0.0001$) displayed on each plot indicate statistically significant overall differences between genetic groups. Tables below the curves detail the number of individuals at risk at given ages, along with p-values for selected pairwise comparisons between specific genetic groups.





Pairwise comparisons (long-rank test; adj. BH)

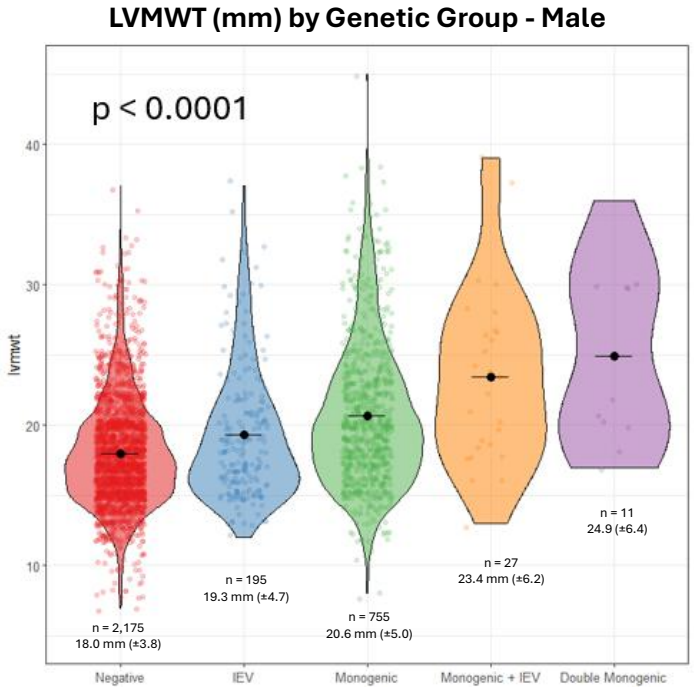
	Negative	IEV	Monogenic	Monogenic+IEV
IEV	0.75508	-	-	-
Monogenic	6.0e-06	0.09186	-	-
Monogenic+IEV	9.8e-09	3.4e-05	0.00083	-
Double_Monogenic	0.00069	0.00298	0.06814	0.82661

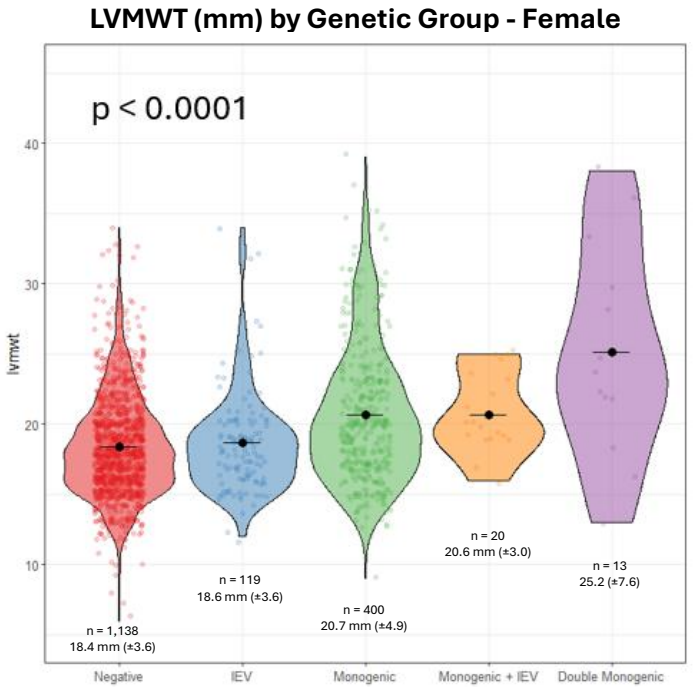
Pairwise comparisons (long-rank test; adj. BH)

	Negative	IEV	Monogenic	Monogenic+IEV
IEV	0.0204	-	-	-
Monogenic	0.0544	0.4284	-	-
Monogenic+IEV	7.6e-07	0.0221	0.0061	-
Double_Monogenic	0.0221	0.2697	0.1443	0.7126

Figure S10: LVMWT Analysis stratified by Sex

Violin plots display the distribution of Left Ventricular Maximmum Wall Thickness (LVMWT) values for **Males (left)** and **Females (right)** across different genetic groups: Negative, IEV, Monogenic, Monogenic + IEV, and Double Monogenic. Horizontal lines within each violin indicate the mean LVMWT. P-values displayed above each plot indicate the overall significance ($p < 0.0001$) of differences in LVMWT between genetic groups (Kruskal-Wallis test). The tables below each violin plot provide p-values (adjusted by Bonferroni method) for select pairwise comparisons.

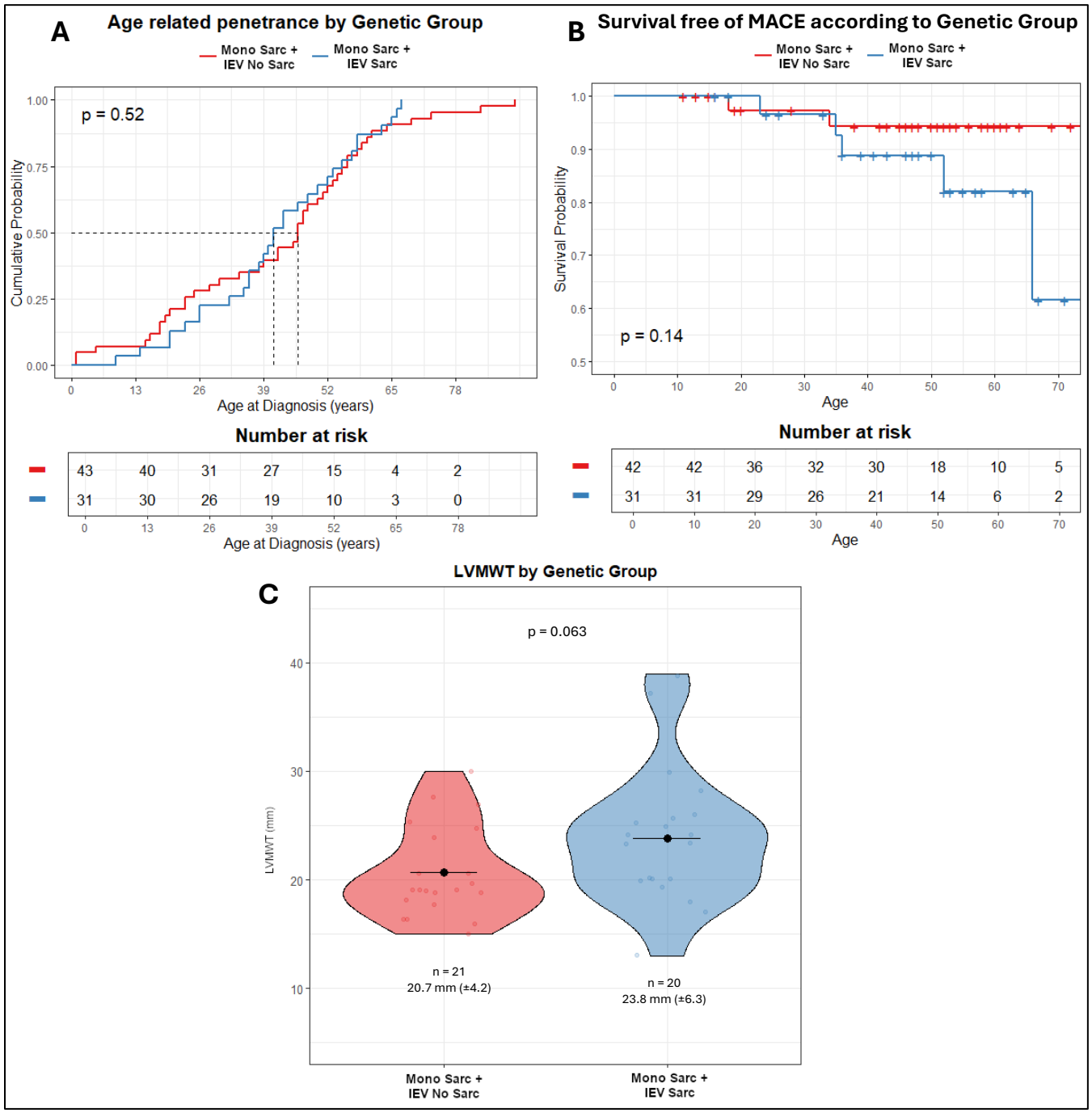




Comparison	p value (.adj Bonferroni)
Double Monogenic vs. IEV	2.34149e-02
Double Monogenic vs. Monogenic	4.36469e-02
IEV vs .Monogenic	4.73371e-05
Double Monogenic vs. Monogenic+IEV	1.00000e+00
IEV vs. Monogenic+IEV	2.88557e-02
Monogenic vs. Monogenic+IEV	2.98424e-01
Double Monogenic vs. Negative	9.58552e-04
IEV vs. Negative	1.35304e-02
Monogenic vs. Negative	1.00259e-55
Monogenic+IEV vs. Negative	4.03364e-06

Comparison	p value (.adj Bonferroni)
Double Monogenic vs. IEV	6.82299e-03
Double Monogenic vs. Monogenic	5.97288e-01
IEV vs .Monogenic	2.59648e-04
Double Monogenic vs. Monogenic+IEV	1.00000e+00
IEV vs. Monogenic+IEV	7.03450e-02
Monogenic vs. Monogenic+IEV	1.00000e+00
Double Monogenic vs. Negative	2.15557e-03
IEV vs. Negative	1.00000e+00
Monogenic vs. Negative	1.94677e-16
Monogenic+IEV vs. Negative	2.47739e-02

Figure S11. Comparison of phenotypic impact of additional sarcomeric versus non-sarcomeric intermediate-effect variants in carriers of single sarcomeric variants. This figure evaluates whether the combination of a sarcomeric or a non-sarcomeric intermediate-effect variant (IEV) a single sarcomeric variant (MonoSarc) exerts a different modifying effect on the clinical phenotype**. Panel A:** Kaplan–Meier analysis of age at diagnosis according to IEV type (sarcomeric vs. non-sarcomeric). **Panel B:** Kaplan–Meier analysis of survival free from major adverse cardiac events (MACE) by group. **Panel C:** Violin plot showing the distribution of left ventricular maximum wall thickness (LVMWT) by group.



STROBE Checklist for Observational Studies (Cohort)

Item	Recommendation	Where in manuscript
1. Title and abstract	Indicate the study's design with a commonly used term in the title or the abstract; Provide an informative and balanced summary of what was done and what was found.	Title/Abstract; Introduction (objectives).
2. Background/rationale	Explain the scientific background and rationale for the investigation being reported.	Introduction.
3. Objectives	State specific objectives, including any prespecified hypotheses.	End of Introduction.
4. Study design	Present key elements of study design early in the paper.	Methods – 'Study population…', adherence to STROBE; dual focus: case-control enrichment and retrospective phenotype analysis.
5. Setting	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	'Between May 2014 and June 2024… sequenced in Health in Code S.L.'
6. Participants	Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. For matched studies, give matching criteria and number of exposed and unexposed.	Methods – 'Study population…', n=14,113 HCM, n=8,144 controls; IRB and informed consent.

7. Variables	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Methods – 'Phenotypic Analysis…' (diagnosis age, LVMWT, MACE); genetic group definitions (Negative, IEV, Monogenic, Monogenic+IEV, Double Monogenic).
8. Data sources/measurement	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	'Variant Genotyping and Classification', 'NGS Sequencing Methods' (Supplement); ACMG-adapted criteria.
9. Bias	Describe any efforts to address potential sources of bias.	PCA for ancestry; internal controls and gnomAD-NFE; exclusion of genocopies for enrichment/phenotype analyses.
10. Study size	Explain how the study size was arrived at.	Reported actual cohort sizes; no formal a priori sample size calculation.
11. Quantitative variables	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.	Methods – 'Statistical Analysis' (treatment of continuous variables; LVMWT; age; multiple testing adjustments).
12. Statistical methods	Describe all statistical methods, including those used to control for confounding. Describe any methods used to examine subgroups and interactions. Explain how missing data were addressed. If applicable, describe analytical methods taking account of sampling	'Statistical Analysis' (t-test/ANOVA/Mann-Whitney/Kruskal-Wallis, BH and Bonferroni; Kaplan-Meier, log-rank; Cox univariable and adjusted; PH assumption checks; stratified model by age in Supplement). Sensitivity excluding FHOD3 p.Arg637Gln.

	strategy. Describe any sensitivity analyses.	
13. Participants	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.	Results – numbers by genetic groups; Figures/Tables of distribution. No flowchart included.
14. Descriptive data	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders.	Table 2 (clinical characteristics by group); Results text.
15. Outcome data	Report numbers of outcome events or summary measures over time.	Results – primary outcomes (diagnosis age, LVMWT, MACE); Kaplan-Meier curves; Cox models.
16. Main results	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included.	Results – Cox adjusted by age; PH verification (Supplement).
17. Other analyses	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses.	Sex-stratified analyses (Fig S8–S10), IEV type (sarcomeric vs non-sarcomeric; Fig S11), zygosity (Fig 6), variant-specific comparisons (Fig S5).
18. Key results	Summarise key results with reference to study objectives.	Discussion – first paragraphs.
19. Limitations	Discuss limitations of the study, taking into account	Limitations section.

	sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	
20. Interpretation	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	Discussion – interpretation in context of SHaRe, genetic architecture framework.
21. Generalisability	Discuss the generalisability (external validity) of the study results.	Discussion – ancestry (NFE) and population differences.
22. Funding	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	'Sources of Funding' (BMS; EIC support); 'Disclosures'.
23. Ethics and data availability	State IRB approval, informed consent, and whether data/materials are available.	Methods ('IRB approval… informed consent'; 'Data Availability').