Supplemental Online Content

Biddinger KJ, Jurgens SJ, Maamari D, et al. Rare and common genetic variation underlying the risk of hypertrophic cardiomyopathy in a national biobank. *JAMA Cardiol*. Published online May 18, 2022. doi:10.1001/jamacardio.2022.1061

eMethods

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods

eMethods A. UK Biobank – Data Collection

Analyses were conducted in the UK Biobank, a prospective cohort from the United Kingdom. Nine million individuals were invited to participate, of which 502,629 subjects were enrolled and surveyed at assessment centers across the UK from 2006-2010.[1] Participants ranged in age from 40 to 69 years. Data collection included verbal interviews and touchscreen questionnaires, and all subjects were genotyped using either the Affymetrix UK BiLEVE Axiom array (first 50,000 subjects) or the Affymetrix UK Biobank Axiom® array (remaining 450,000 participants). A subset of 200,000 individuals underwent whole exome sequencing (described below). Subjects also submited blood, urine, and saliva samples for analysis, with blood biochemistry analyses used to determine biomarker levels. Prevalent cardiovascular diseases were recorded at study entry through self-report confirmed in a verbal interview with a trained nurse, or via linked electronic health record data from the National Health Service (NHS). Incident diseases were defined, among those not meeting criteria at baseline, through the application of phenotype definitions to linked, in-patient hospital and death registry data. Participants were censored at whichever came first between disease diagnosis, date of death, or date of last follow-up. The date of last follow-up was February 9, 2016 for participants enrolled in Wales, February 16, 2016 for participants enrolled in England, and October 31, 2015 for participants enrolled in Scotland.

eMethods B. Exclusions

For data quality control, we excluded subjects who had unreliable data or were of 2nd degree relatedness or closer. Unreliable data was defined – per centralized sample quality control performed by UK Biobank – as inferred sex unequal to reported sex, kinship not inferred, putative sex chromosome aneuploidy, withdrawn consent, or excessive heterozygosity or missingness in genetic data.[2] For those of 2nd degree or closer relatedness, one subject was randomly removed from each related pair to prevent individuals with similar lifestyle factors and genetics from skewing the data.

eMethods C. Mass General Brigham Biobank – Data Collection

The Mass General Brigham (MGB) Biobank – a health system-based cohort based in the United States – was used as a secondary cohort to replicate select genetic analyses. We studied 30,716 individuals from the MGB Biobank with genotyping array data. Hypertrophic cardiomyopathy was defined as presence of ICD-10 codes I42.1 or I42.2, and mention of “hypertrophic cardiomyopathy”, “hypertrophic obstructive cardiomyopathy”, “HCM” or “HOCM” in the medical record.

eMethods D. Rare Variant Analysis

Sequencing and quality control

WES was performed on over 200,000 participants from the UK Biobank, and the methods used for sequencing have been described previously.[3,4] The revised version of the IDT xGen Exome Research Panel v1.0 was used to capture exomes with over 20X coverage at 95% of sites. We applied an extensive genotype, variant, and sample level pipeline to produce a high-quality dataset for analysis, which is described in detail elsewhere.[5] Briefly, low-quality genotypes were set to no-calls, after which variants were removed based on call rate (<90%), Hardy-Weinberg equilibrium test (P < 1E-15), presence in low-complexity regions, and minor allele count (<1). Sample-level quality-control consisted of removal of samples that had withdrawn their consent, were duplicates, had a mismatch between WES and genotyping array data, had a mismatch between genetically inferred and self-reported sex, had low call rates or were outliers (outside 8 standard deviations from the mean) for a number of additional metrics. Of the 200,642 individuals with WES who passed the internal quality-control, an additional 305 samples were removed, leaving 200,337 individuals.

Variant annotation

The protein consequences of variants were annotated using dbNSFP[6] (version 4.1a) and the Loss-of-Function Transcript Effect Estimator[7] (LOFTEE) plug-in implemented in the Variant Effect Predictor[8] (VEP; version 95) (https://github.com/konradjk/loftee). VEP was used to ascertain the most severe consequence of a given variant for each gene transcript. LOFTEE was implemented to identify high-confidence loss-of-function variants (LOF), which include frameshift indels, stopgain variants and splice site disrupting variants. We also removed any LOFs flagged by LOFTEE as dubious, such as LOFs affecting poorly conserved exons and splice variants affecting NAGNAG sites or non-canonical splice regions. Missense variants were annotated using 30 in silico prediction tools included in the dbNSFP

database. We collapsed information from these 30 tools into a single value, representing the percentage of tools which predicted a given missense variant was deleterious.[5] A missense variant was considered predicted-deleterious when >90% of tools predicted it to be damaging. We further annotated all variants with adjudications from the clinVar database, as described previously.[5] Briefly, variants with clinVar adjudications from clinical testing laboratories (from 2015 and onward) were included in our annotation. We used the most recent adjudication, removing variants with "Conflicting Interpretation". In this study we focused on variants adjudicated as 'Pathogenic' and 'Likely Pathogenic', termed in the manuscript as 'clinvarPLP'.

To be inclusive, we considered clinvarPLP variants regardless of the phenotypic assertion reported in ClinVar in our primary analyses, given that ClinVar assertions are sometimes broad/vague (e.g. cardiovascular phenotype); we therefore performed additional sensitivity analyses restricting to clinvarPLP variants reported specifically for HCM in the ClinVar database (eTable 4d).

Gene Selection
Potentially pathogenic variants were chosen if they were included on GeneDx, Invitae, or Blueprint gene panels, after removal of mitochondrial genes (eTable 2).[9–11] *GLA* was not included because it is an X-linked gene, and *LAMP2* and *FHL1* variants were not found with the UK Biobank pipeline. A rare variant in any of these genes was termed a HCM-Panel rare variant. Rare variants in these genes were analyzed separately for clinvarPLP, LOF, and predicted-deleterious missense categorizations. Next, a subset of core genes was chosen based on recommendations from the American College of Medical Genetics and Genomics.[12] A high-impact rare variant was defined as (a) a clinvarPLP variant in any of the core genes or (b) an LOF variant in *MYBPC3*, *TNNT2*, or *PLN*. The rationale for these selections is described in eTable 3.[13–15] Given limited evidence for the pathogenicity of *TNNT2* loss-of-function mutations, we conducted sensitivity analyses excluding these LOFs from this classification. For all analyses, the cumulative assessments of risk factors and rare variants were investigated for clinvarPLP variants in the broad HCM panel (HCM-Panel), for the high-priority ACMG variants (HCM-ACMG, described above), and for *MYBPC3* clinvarPLP or LOF variants only, given the well-established high prevalence of *MYBPC3* rare variants among HCM cases.

eMethods E. Polygenic Risk Score

Effect estimates were derived from a previous genome-wide association study of hypertrophic cardiomyopathy.[16] We incorporated 27 independent single nucleotide polymorphisms (SNPs) into a polygenic score by, for each SNP, multiplying the number of risk alleles by the reported beta estimate (log odds ratio) for the association of that SNP with hypertrophic cardiomyopathy. The 27 SNPs were chosen based on their association with HCM at a 5% FDR threshold. 40,283 UK Biobank individuals (healthy individuals from the first 50k release of exome-sequencing of the UK Biobank) were included as controls in this previous GWAS. However, no individuals from the UK Biobank were used as cases, and we empirically found that PRS performance remained strong after removing all samples from the 50k exome subset (eTable 5). Therefore, we chose not to exclude any individuals from our analysis in order to maximize HCM case numbers.

eMethods F. Statistical Analyses

Area under the receiver operator curve and corresponding confidence intervals were estimated using the R package 'pROC'. Survival curves were created as the cumulative incidence and lifetime absolute risks at age 80 years were calculated from cumulative event curves from Cox models adjusted for sex, genotyping array, and PCs 1-5, and using age as the time scale, plotted using the 'survival' package in R.

For analyses of genetic and clinical risk factors, a range of potential clinical risk factors were tested for association with HCM, using Cox proportional-hazards adjusting for age, sex, genotyping array, and PCs 1-5. These variables were physical activity (number of days/week of moderate physical activity 10+ minutes), alcohol use (standard drinks per week), body mass index (BMI, kg/m^2), systolic and diastolic blood pressure (SBP and DBP, mmHg), obesity (BMI > 30 kg/m^2), smoker (having ever smoked), heavy alcohol use (>14 drinks/week for men, >7 drinks/week for women),[17] hypertension (HTN, prevalent disease), atrial fibrillation (AF, prevalent disease), chronic kidney disease (CKD, prevalent disease), type II diabetes (T2D, prevalent disease), and coronary artery disease (CAD, prevalent disease). In order to limit potential reverse causality bias, prevalent cases of HCM were excluded, leaving only incident disease for all analyses of genetic and clinical risk factors. Hazard ratios were calculated from Cox proportional-hazards models, adjusting for age, sex, genotyping array, and PCs 1-5. For these analyses of

genetic and clinical risk factors, AUC values were calculated from logistic regression models with incident HCM as the outcome, and adjusting for age, sex, genotyping array, and PCs 1-5. Variance in disease susceptibility explained was computed as the improvement in R^2 on the liability scale, upon adding either or both predictors to logistic regression models, which included age, sex, genotyping array, and PCs 1-5 as covariates.

Sensitivity analyses were performed with rare variant carrier status defined by the broader HCM-Panel definition or restricting to *MYBPC3* variants only. We also repeated analyses using relevant continuous variables where applicable (systolic blood pressure/diastolic blood pressure, BMI and continuous polygenic score instead of hypertension, obesity and dichotomized polygenic risk, respectively). Finally, primary analyses were repeated after removing participants from the initial tranche of whole exome sequencing data from UK Biobank (N=50,000) as a subset of individuals from this initial tranche were used as controls in the original HCM GWAS.[16]

Net reclassification improvement (NRI) was determined by first constructing a five-year risk score for HCM using as predictors: (1 – clinical risk score) the aforementioned dichotomous clinical risk factors, age, sex, genotyping array, and PCs 1-5, or (2 – genetic and clinical risk score) the previous model in addition to HCM-ACMG carrier status and high polygenic risk (>80th percentile). Risk scores were constructed by (1) transforming linear variables (i.e. age, PCs) into quartiles, (2) multiplying each variable by its beta coefficient from a Cox proportional hazards model with HCM, and taking the sum, and (3) subtracting from 1 the baseline survival raised to the inverse natural log of that sum. NRI was then calculated using the R package 'nricens', taking the median of clinically-predicted 5y risk as the cutoff for reclassification.

eTable 1: Disease definitions in UK Biobank.

Disease	Definition
Atrial Fibrillation	Self-reported history of atrial fibrillation, atrial flutter, or cardioversion during verbal interview with trained nurse; **or** hospitalization with or death due to ICD-10 code for atrial fibrillation or atrial flutter (I48); **or** hospitalization with ICD-9 code for atrial fibrillation or atrial flutter (4273); **or** hospitalization with OPCS-4 code for percutaneous transluminal ablation (K57.1, K 62.1, K62.2, K62.3, K62.4)
Chronic Kidney Disease	Self-reported history of kidney failure ± dialysis, kidney nephropathy, IgA nephropathy, diabetic nephropathy or kidney transplant during verbal interview with trained nurse; **or** hospitalization with or death due to ICD-10 code for hypertensive renal disease, chronic renal failure, end stage renal failure or chronic kidney disease (I12.0, I13.1, I13.2, N18, N18.0-18.5, N18.8, N18.9); **or** hospitalization with ICD-9 code due to chronic renal failure (585, 5859); **or** hospitalization with OPCS-4 coded procedure for kidney transplantation (M01.1-01.5, M01.8, M01.9)
Coronary Artery Disease	Self-reported history of myocardial infarction (MI), coronary artery bypass grafting, coronary artery angioplasty or triple heart bypass during verbal interview with trained nurse; **or** hospitalization for or death due to ICD-10 code for acute or subsequent myocardial infarction (I21, I22, I23, I24.1, I25.2); **or** hospitalization due to ICD-9 code for myocardial infarction (410, 411, 412); **or** hospitalization due to OPCS-4 code for coronary artery bypass grafting (K40, K41, K44, K45, K46), coronary endarterectomy (K47.1), or coronary angioplasty ± stenting (K49, K50.2, K75)
Diabetes Mellitus, Type 2	Self-reported history of type 2 diabetes during verbal interview with trained nurse; **or** hospitalization with or death due to ICD-10 code for non-insulin-dependent diabetes mellitus (E11)
Hypertension	Self-reported history of hypertension, essential hypertension or high blood pressure during verbal interview with trained nurse; **or** hospitalization with or death due to ICD-10 code for essential hypertension, hypertensive heart disease, hypertensive renal disease, or secondary hypertension (I10, I11, I12, I13, I15); **or** hospitalization with ICD-9 code for essential hypertension, hypertensive heart disease, hypertensive renal disease, or secondary hypertension (401, 402, 403, 404, 405)
Hypertrophic Cardiomyopathy	Self-reported history of hypertrophic cardiomyopathy during a verbal interview with a trained nurse; **or** hospitalization or death due to an ICD-10 code for hypertrophic cardiomyopathy (I42.1, I42.2); **or** hospitalization due to an ICD-9 code for hypertrophic cardiomyopathy (425.11, 425.18).

eTable 2: Gene categorizations for hypertrophic cardiomyopathy.

	Description	Included Genes
HCM-Panel$_{RV}$	Inclusive list of candidate genes based on clinical testing panels for hypertrophic cardiomyopathy.[9–11] *FHL1* and *LAMP2* variants were not found in the UK Biobank pipeline, and *GLA* was excluded because it is an X-linked gene.	*ABCC9, ACAD9, ACADVL, ACTA1, ACTC1, ACTN2, AGK, AGL, ALPK3, APOA1, BAG3, BRAF, CACNA1C, CAV3, CBL, COX15, CSRP3, DES, ELAC2, EPG5, FBXL4, FHOD3, FLNC, FXN, GAA, GSK3B, HRAS, JPH2, KLHL24,, MIPEP, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOZ2, NDUFAF2, PLN, PRKAG2, PTPN11, RAF1, RIT1, SLC25A4, SOS1, TCAP, TNNC1, TNNI3, TNNT2, TPM1, TTR, VCL*
HCM-ACMG$_{RV}$	Priority genes from American College of Medical Genetics and Genomics	*MYH7, MYBPC3, TNNT2, TNNC1, TNNI3, TPM1, MYL2, MYL3, ACTC1, ACTN2, CSRP3, PLN, TTR, PRKAG2*

eTable 3: Prioritized rare variants for hypertrophic cardiomyopathy. Genes were selected from the American College of Medical Genetics and Genomics (Hershberger et al.). Rare variants were defined as, for all genes, a variant classified as pathogenic or likely pathogenic in the ClinVar Database. For select genes, in which prior evidence implicated a potential loss-of-function mechanism, predicted loss-of-function mutations were also included.

Gene	Variants included	References
MYH7	ClinVarPLP	
MYBPC3	ClinVarPLP, LOF	Walsh et al., Alfares et al., Morita et al.
TNNT2	ClinVarPLP, LOF	Walsh et al.
TNNC1	ClinVarPLP	
TNNI3	ClinVarPLP	
TPM1	ClinVarPLP	
MYL2	ClinVarPLP	
MYL3	ClinVarPLP	
ACTC1	ClinVarPLP	
ACTN2	ClinVarPLP	
CSRP3	ClinVarPLP	
PLN	ClinVarPLP, LOF	Walsh et al.
TTR	ClinVarPLP	
PRKAG2	ClinVarPLP	
LAMP2	Not included (not found with UKBB pipeline)	
GLA	Not included (X-linked)	

eTable 4: Gene-based associations of rare variants with hypertrophic cardiomyopathy in UK Biobank. Odds ratios, 95% confidence intervals, and probability values are calculated using Firth logistic regression. Bonferroni-corrected significance was defined as >=10 carriers and alpha = 0.05 / 80 testable masks = 6.25E-04. Nominal significance was defined as >=10 carriers and $P < 0.05$.

A) ClinvarPLP (for all diseases) mutations.

Gene	Carriers (%)	Cases among Carriers (%)	Cases among Noncarriers (%)	OR [95% CI]	*P*-value
MYBPC3	205 (0.11%)	13 (6.34%)	191 (0.1%)	71.99 [38.66-123.64]	1.90E-43
MYH7	130 (0.07%)	7 (5.38%)	197 (0.11%)	60.67 [26.07-121.09]	1.20E-11
GAA	1021 (0.55%)	2 (0.2%)	202 (0.11%)	2.45 [0.51-6.95]	2.20E-01
AGL	438 (0.24%)	1 (0.23%)	203 (0.11%)	3.16 [0.36-11.4]	0.238
DES	23 (0.01%)	1 (4.35%)	203 (0.11%)	63.53 [6.99-253.62]	2.00E-03
PTPN11	25 (0.01%)	1 (4%)	203 (0.11%)	50.16 [5.4-205.86]	3.00E-03
TNNI3	31 (0.02%)	1 (3.23%)	203 (0.11%)	45.13 [5-175.63]	0.004
TTR	27 (0.01%)	1 (3.7%)	203 (0.11%)	71.89 [7.92-285.44]	0.002
ACAD9	154 (0.08%)	0 (0%)	204 (0.11%)	3.06 [0.02-21.13]	0.506
ACADVL	359 (0.19%)	0 (0%)	204 (0.11%)	1.4 [0.01-9.6]	0.821
ACTA1	15 (0.01%)	0 (0%)	204 (0.11%)	33.97 [0.26-264.66]	0.112
AGK	54 (0.03%)	0 (0%)	204 (0.11%)	9.82 [0.08-69.2]	0.24
ALPK3	10 (0.01%)	0 (0%)	204 (0.11%)	61.19 [0.47-489.59]	0.078
ELAC2	42 (0.02%)	0 (0%)	204 (0.11%)	11.19 [0.09-80.08]	0.221
EPG5	20 (0.01%)	0 (0%)	204 (0.11%)	14.7 [0.11-122.32]	0.19
FBXL4	237 (0.13%)	0 (0%)	204 (0.11%)	2.08 [0.02-14.23]	0.646
FLNC	13 (0.01%)	0 (0%)	204 (0.11%)	20.12 [0.15-167.61]	0.156
FXN	23 (0.01%)	0 (0%)	204 (0.11%)	20.66 [0.16-151.41]	0.151
NDUFAF2	26 (0.01%)	0 (0%)	204 (0.11%)	20.04 [0.16-145.41]	0.154

B) Predicted loss-of-function mutations.

Gene	Carriers (%)	Cases among Carriers (%)	Cases among Noncarriers (%)	OR [95% CI]	*P*-value
MYBPC3	77 (0.04%)	8 (10.39%)	196 (0.11%)	89.39 [37.64-187.2]	2.78E-14

ALPK3	213 (0.12%)	3 (1.41%)	201 (0.11%)	13.33 [4.37-28.38]	2.89E-04
PLN	17 (0.01%)	1 (5.88%)	203 (0.11%)	105.16 [11.36-440.71]	1.00E-03
ACTN2	22 (0.01%)	1 (4.55%)	203 (0.11%)	59.85 [6.58-238.81]	0.002
MYH6	162 (0.09%)	1 (0.62%)	203 (0.11%)	7.89 [0.96-26.14]	5.30E-02
RIT1	43 (0.02%)	0 (0%)	204 (0.11%)	68.67 [0.52-620.49]	0.074
TNNC1	11 (0.01%)	0 (0%)	204 (0.11%)	12.97 [0.58-59.74]	0.077
TCAP	11 (0.01%)	0 (0%)	204 (0.11%)	61.93 [0.47-524.55]	0.078
PRKAG2	26 (0.01%)	0 (0%)	204 (0.11%)	17.54 [0.41-77.04]	0.09
ACTC1	17 (0.01%)	0 (0%)	204 (0.11%)	39.89 [0.31-301.38]	0.101
KLHL24	28 (0.02%)	0 (0%)	204 (0.11%)	16.36 [0.29-71.94]	0.109
APOA1	12 (0.01%)	0 (0%)	204 (0.11%)	33.58 [0.26-269.06]	0.113
SLC25A4	16 (0.01%)	0 (0%)	204 (0.11%)	31.53 [0.25-239.19]	0.117
CAV3	14 (0.01%)	0 (0%)	204 (0.11%)	28.46 [0.22-218.54]	0.124
CSRP3	57 (0.03%)	0 (0%)	204 (0.11%)	23.19 [0.18-169.7]	0.14
ACTA1	38 (0.02%)	0 (0%)	204 (0.11%)	11.7 [0.2-44.32]	0.142
CBL	27 (0.01%)	0 (0%)	204 (0.11%)	20.21 [0.16-147.64]	0.153
JPH2	26 (0.01%)	0 (0%)	204 (0.11%)	15.82 [0.12-115.72]	0.178
SOS1	45 (0.02%)	0 (0%)	204 (0.11%)	7.62 [0.13-29.66]	0.187
DES	39 (0.02%)	0 (0%)	204 (0.11%)	8.17 [0.12-36.45]	0.198
FLNC	45 (0.02%)	0 (0%)	204 (0.11%)	10.42 [0.1-52.09]	0.208
MYL2	65 (0.04%)	0 (0%)	204 (0.11%)	6.82 [0.1-26.92]	0.219
TNNI3	39 (0.02%)	0 (0%)	204 (0.11%)	11.03 [0.09-78.99]	0.223
MYOZ2	55 (0.03%)	0 (0%)	204 (0.11%)	8.7 [0.08-44.81]	0.24
CACNA1C	39 (0.02%)	0 (0%)	204 (0.11%)	9.73 [0.08-71.63]	0.243
HRAS	70 (0.04%)	0 (0%)	204 (0.11%)	6.44 [0.07-27.58]	0.258
VCL	57 (0.03%)	0 (0%)	204 (0.11%)	7.6 [0.06-42.9]	0.273
ACAD9	71 (0.04%)	0 (0%)	204 (0.11%)	7.26 [0.06-42]	0.283
FHOD3	69 (0.04%)	0 (0%)	204 (0.11%)	6.81 [0.05-47.58]	0.303

TNNT2	92 (0.05%)	0 (0%)	204 (0.11%)	5.28 [0.05-29.16]	0.345
NDUFAF2	192 (0.1%)	0 (0%)	204 (0.11%)	5.23 [0.04-31.11]	0.352
AGK	124 (0.07%)	0 (0%)	204 (0.11%)	4.58 [0.04-28.21]	0.386
ELAC2	140 (0.08%)	0 (0%)	204 (0.11%)	3.54 [0.03-17.28]	0.432
MYH7	140 (0.08%)	0 (0%)	204 (0.11%)	3.74 [0.03-21.84]	0.436
MIPEP	176 (0.1%)	0 (0%)	204 (0.11%)	3.15 [0.03-16.76]	0.479
EPG5	153 (0.08%)	0 (0%)	204 (0.11%)	3.27 [0.03-21]	0.482
ACADVL	174 (0.09%)	0 (0%)	204 (0.11%)	2.98 [0.03-14.27]	0.483
ABCC9	144 (0.08%)	0 (0%)	204 (0.11%)	3.14 [0.03-21.64]	0.498
COX15	218 (0.12%)	0 (0%)	204 (0.11%)	2.44 [0.02-11.89]	0.555
FBXL4	239 (0.13%)	0 (0%)	204 (0.11%)	2.26 [0.02-15.53]	0.612
GAA	304 (0.16%)	0 (0%)	204 (0.11%)	1.74 [0.02-9.66]	0.711
AGL	481 (0.26%)	0 (0%)	204 (0.11%)	0.99 [0.01-6.16]	0.991

C) Predicted deleterious missense mutations.

Gene	Carriers (%)	Cases among Carriers (%)	Cases among Noncarriers (%)	OR [95% CI]	*P*-value
FLNC	75 (0.04%)	1 (1.33%)	203 (0.11%)	12.46 [1.39-47.4]	3.00E-02
MYH7	54 (0.03%)	1 (1.85%)	203 (0.11%)	31.93 [3.59-119.6]	6.00E-03
PTPN11	16 (0.01%)	1 (6.25%)	203 (0.11%)	147.07 [15.97-609.42]	5.89E-04
ABCC9	105 (0.06%)	0 (0%)	204 (0.11%)	4.25 [0.03-29.42]	0.41
ACAD9	36 (0.02%)	0 (0%)	204 (0.11%)	15.82 [0.13-113.17]	1.78E-01
ACADVL	82 (0.04%)	0 (0%)	204 (0.11%)	6.91 [0.06-48.13]	3.00E-01
ACTA1	11 (0.01%)	0 (0%)	204 (0.11%)	41.02 [0.32-331.93]	0.1
ACTN2	19 (0.01%)	0 (0%)	204 (0.11%)	17.83 [0.14-140.35]	0.167
AGL	16 (0.01%)	0 (0%)	204 (0.11%)	35.71 [0.28-272.2]	0.108
CACNA1C	43 (0.02%)	0 (0%)	204 (0.11%)	11.01 [0.09-78.75]	0.224
CAV3	48 (0.03%)	0 (0%)	204 (0.11%)	10.16 [0.08-71.78]	0.235

CBL	14 (0.01%)	0 (0%)	204 (0.11%)	34.12 [0.27-262.55]	0.111
COX15	55 (0.03%)	0 (0%)	204 (0.11%)	7 [0.06-49.84]	0.298
CSRP3	59 (0.03%)	0 (0%)	204 (0.11%)	7 [0.06-49.29]	0.298
DES	11 (0.01%)	0 (0%)	204 (0.11%)	32.39 [0.25-255.69]	0.115
GAA	66 (0.04%)	0 (0%)	204 (0.11%)	6.83 [0.05-47.81]	0.303
MYH6	30 (0.02%)	0 (0%)	204 (0.11%)	16.13 [0.13-116.8]	0.176
SOS1	38 (0.02%)	0 (0%)	204 (0.11%)	8.95 [0.07-64.44]	0.255
TPM1	11 (0.01%)	0 (0%)	204 (0.11%)	28.61 [0.22-234.23]	0.125

d) ClinvarPLP (HCM assertions only) mutations.

Gene	Carriers (%)	Cases among Carriers (%)	Cases among Noncarriers (%)	OR [95% CI]	*P*-value
MYBPC3	199 (0.11%)	12 (6.03%)	192 (0.1%)	2.59 [2.16-3.02]	2.28E-12
MYH7	128 (0.07%)	7 (5.47%)	197 (0.11%)	12.06 [1.04-38.96]	4.80E-02
PTPN11	2 (0%)	1 (50%)	203 (0.11%)	318.93 [2.17-6070.76]	3.20E-02
TNNI3	31 (0.02%)	1 (3.23%)	203 (0.11%)	30.91 [4.14-230.73]	8.22E-04
ABCC9	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ACAD9	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ACADVL	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ACTA1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ACTC1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ACTN2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
AGK	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
AGL	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ALPK3	1 (0%)	0 (0%)	204 (0.11%)	6.61 [0.71-16.41]	0.076
APOA1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
BAG3	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
BRAF	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
CACNA1C	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
CAV3	0 (0%)	0 (0%)	204 (0.11%)	NA	NA

CBL	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
COX15	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
CSRP3	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
DES	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
ELAC2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
EPG5	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
FBXL4	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
FHOD3	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
FLNC	13 (0.01%)	0 (0%)	204 (0.11%)	2.71 [0.54-5.35]	0.153
FXN	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
GAA	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
GSK3B	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
HRAS	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
JPH2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
KLHL24	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
LAMP2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
MIPEP	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
MYH6	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
MYL2	9 (0%)	0 (0%)	204 (0.11%)	5.41 [0.96-12.33]	0.054
MYL3	1 (0%)	0 (0%)	204 (0.11%)	NA	NA
MYOZ2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
NDUFAF2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
PLN	1 (0%)	0 (0%)	204 (0.11%)	NA	NA
PRKAG2	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
RAF1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
RIT1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
SLC25A4	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
SOS1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
TCAP	3 (0%)	0 (0%)	204 (0.11%)	28.89 [1.63-129.07]	0.034
TNNC1	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
TNNT2	5 (0%)	0 (0%)	204 (0.11%)	NA	NA
TPM1	4 (0%)	0 (0%)	204 (0.11%)	NA	NA

TTR	0 (0%)	0 (0%)	204 (0.11%)	NA	NA
VCL	0 (0%)	0 (0%)	204 (0.11%)	NA	NA

eTable 5: Associations of the polygenic score with hypertrophic cardiomyopathy in carriers and noncarriers of an HCM-ACMG rare variant in the UK Biobank.

A) UK Biobank (N = 184,511)

Population	N. Cases	Odds Ratio [95% CI]	Std. Error	*P*-value
UK Biobank	204	1.556 [1.361-1.778]	0.068	8.22E-11
Noncarriers	178	1.585 [1.375-1.828]	0.073	2.39E-10
Carriers	25	1.345 [0.911-1.987]	0.199	0.136

B) UK Biobank, excluding first 50,000 exomes (N = 138,304)

Population	N. Cases	Odds Ratio [95% CI]	Std. Error	*P*-value
UK Biobank	153	1.641 [1.407-1.914]	0.079	2.72E-10
Noncarriers	131	1.683 [1.425-1.988]	0.085	8.85E-10
Carriers	22	1.318 [0.870-1.998]	2.120	0.192

eTable 6: Survey of clinical risk factors for HCM in the UK Biobank. Hazard ratios are calculated using Cox proportional hazards regression adjusting for age, sex, genotyping array, and PCs 1-5. Associations of HCM-ACMG status and the PRS are provided for comparison.

Risk Factor	HR [95% CI]	SE	*P*-value
HCM-ACMG	49.896 [29.082-85.605]	0.275	9.64E-46
PRS	1.795 [1.521-2.117]	0.084	4.04E-12
Phys. Act.	0.97 [0.892-1.054]	0.043	4.70E-01
Alcohol	0.99 [0.968-1.012]	0.012	3.70E-01
BMI	1.034 [0.999-1.07]	0.018	6.00E-02
SBP	1 [0.99-1.009]	0.005	9.70E-01
DBP	0.987 [0.97-1.004]	0.009	1.30E-01
Obesity	1.561 [1.082-2.252]	0.187	2.00E-02
Smoker	1.064 [0.747-1.514]	0.18	7.30E-01
Heavy Alcohol Use	0.814 [0.524-1.264]	0.225	3.60E-01
HTN	2.541 [1.731-3.731]	0.196	1.93E-06
AF	3.838 [1.843-7.994]	0.374	3.26E-04
CKD	3.8 [0.935-15.45]	0.716	6.00E-02
T2D	1.218 [0.492-3.016]	0.462	6.70E-01
CAD	2.961 [1.696-5.17]	0.284	1.34E-04

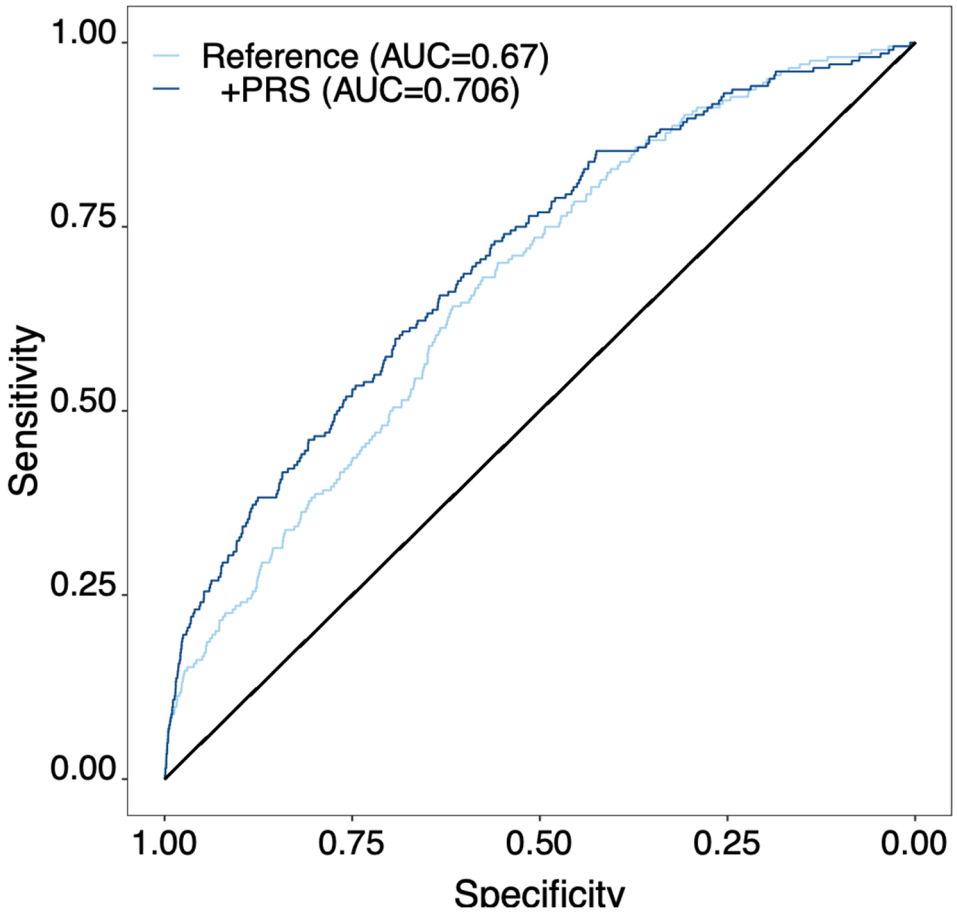
eTable 7: Net reclassification improvement of clinical and genetic factors over clinical factors. A clinical 5-year predicted risk score for HCM were constructed with age, sex, genotyping array, the first 5 principal components of genetic ancestry, obesity, prevalent hypertension, prevalent atrial fibrillation, and prevalent coronary artery disease. A second 5-Year risk score was then constructed with HCM-ACMG rare variant carrier status and high polygenic risk score, defined as >80th percentile, added to the previous model. Net reclassification for HCM was then calculated to compare the benefit of adding genetic factors to a risk prediction model.

a) Reclassification of 5-year predicted risk of HCM. The threshold for 5-year predicted risk was calculated as the median of clinically-predicted 5-year risk.

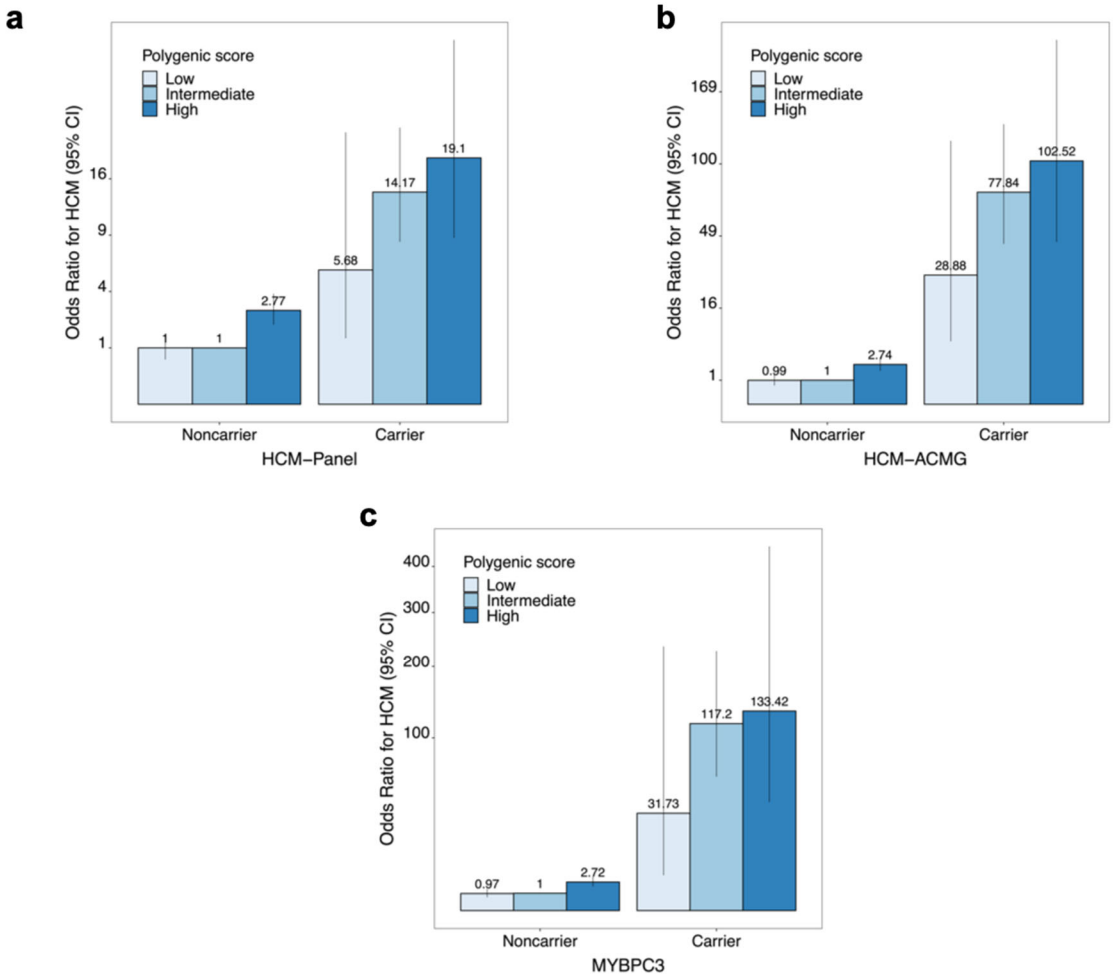
		Clinical + Genetic Model	
	Clinical Model	**< 0.008%**	**>=0.008%**
HCM-CASES	**<0.008%**	1	2
	>=0.008%	0	18
HCM-Controls	**<0.008%**	73 994	10 445
	>=0.008%	33 689	50 764

b) Net reclassification improvement for HCM. The threshold for 5-year predicted risk was calculated as the median of clinically-predicted 5-year risk.

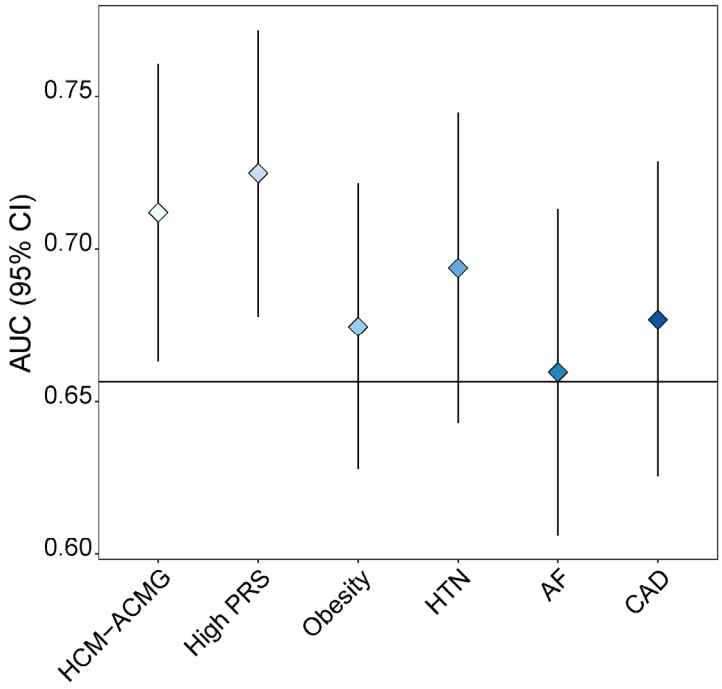
	NRI	**Lower**	**Upper**
Overall	0.233	0.136	0.401
Cases	0.095	0.000	0.263
Non-cases	0.138	0.135	0.140



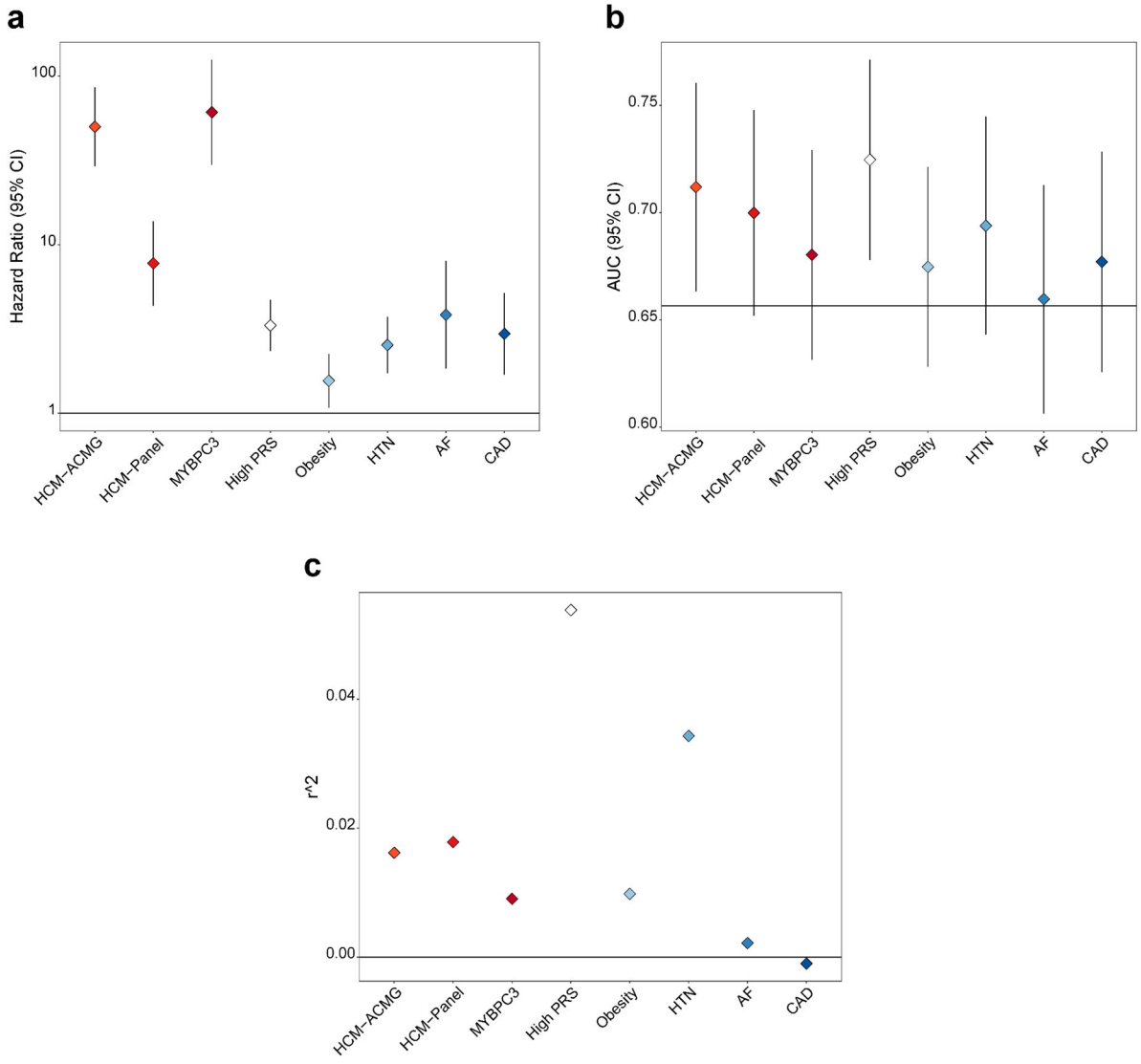
eFigure 1: Receiver-operator curve for HCM PRS. The reference model consisted of age, sex, and the top 5 components of genetic ancestry. The polygenic score was then added to this model to measure the discriminative benefit of common genetic variants.



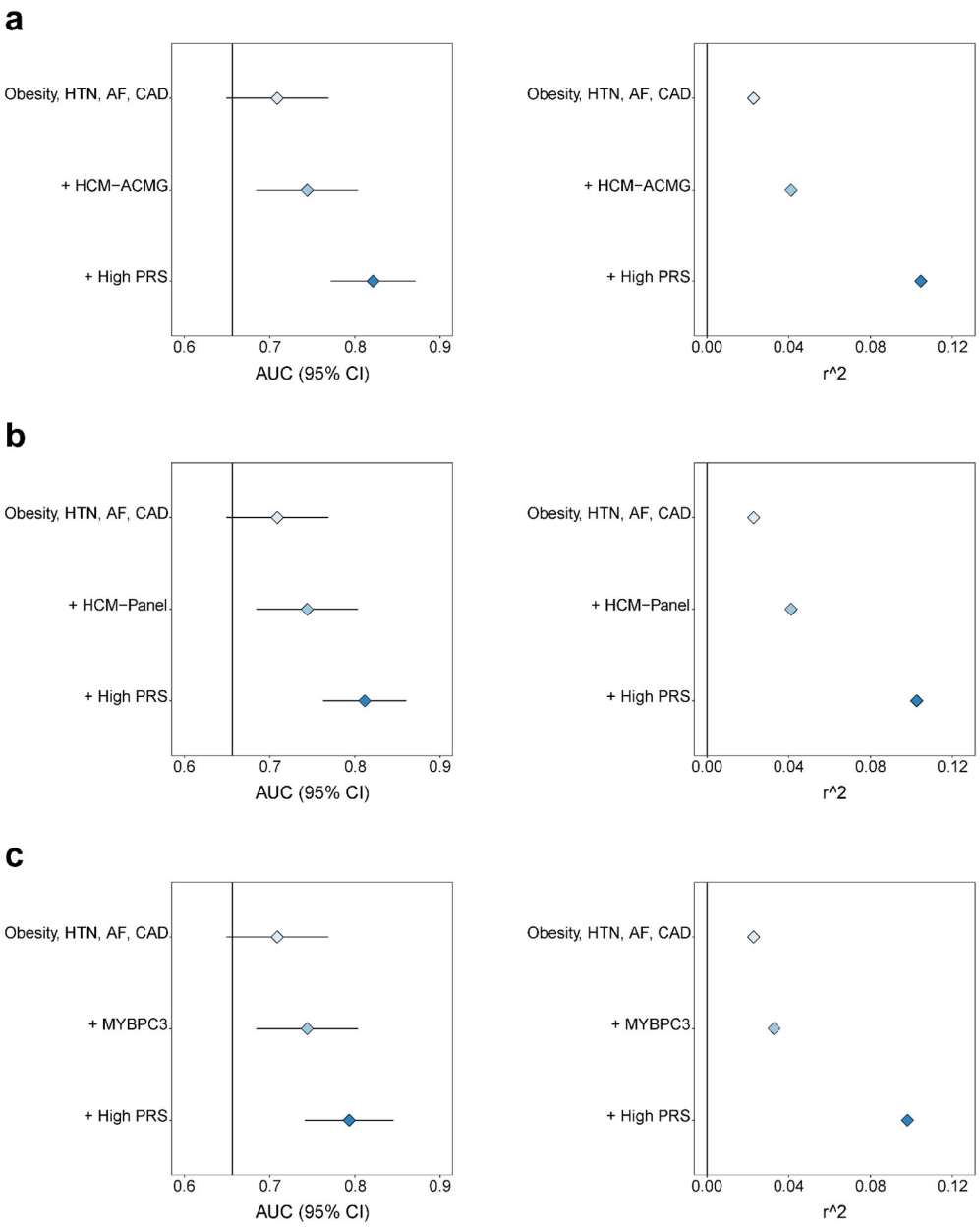
eFigure 2: Combined odds of polygenic risk score and carrier status in the UK Biobank. Low polygenic score was defined as the lowest quintile, intermediate as quintiles 2-4, and high as the top quintile. The reference group is defined as noncarriers with an intermediate polygenic risk score. Carrier status is presented as presence of (a) a HCM-Panel rare variant, (b) a HCM-ACMG rare variant, or (c) a clinvarPLP or LOF variant in *MYBPC3*. Population-based HCM-ACMG variant carriers are at significantly higher odds of HCM than noncarriers. Individuals with high PRS are at significantly higher odds of HCM compared to lower PRS. Within rare variant carriers, there is a trend towards increased risk with higher PRS, although error bars are large at current sample sizes.



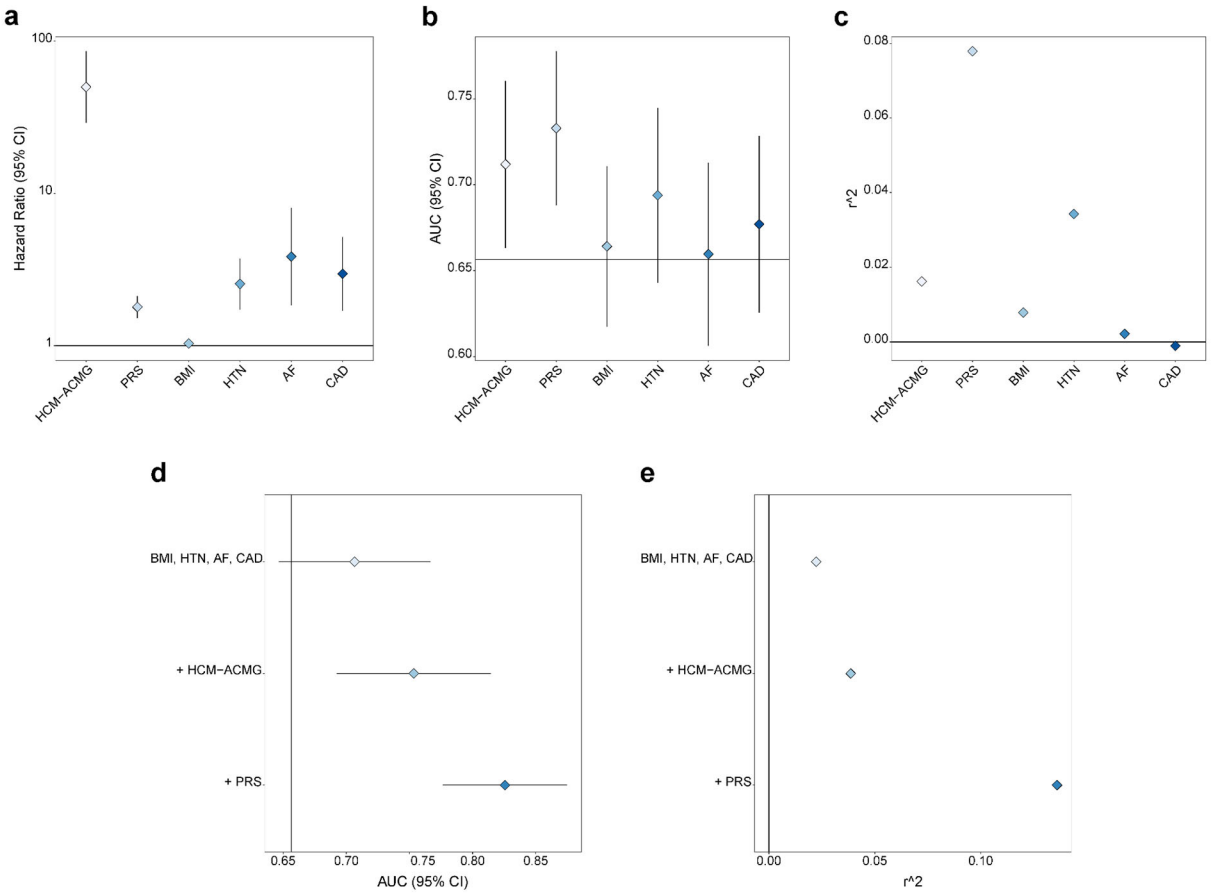
eFigure 3: Discriminative benefit of individual risk factors for HCM in the UK Biobank. Area under the receiver-operator characteristic curve, calculated by logistic regression adjusting for age, sex, genotyping array, and PCs 1-5. The reference for the area under the receiver-operator curve is a model consisting of age, sex, and PCs 1-5; additional predictors were added individually to the model, and those AUC values are plotted. Abbreviations: HCM, hypertrophic cardiomyopathy; ACMG, American College of Medical Genetics; PRS, polygenic risk score; HTN, hypertension; AF, atrial fibrillation; CAD, coronary artery disease.



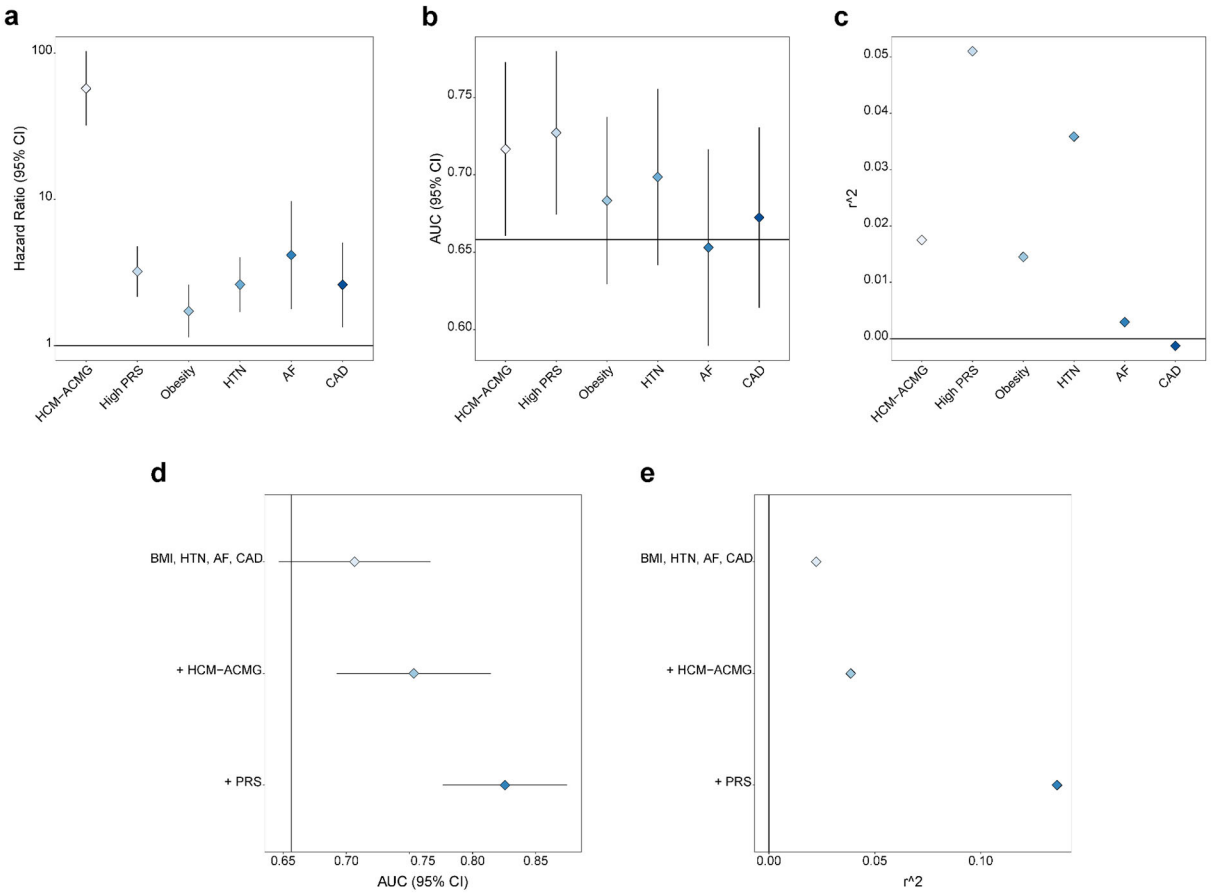
eFigure 4: Comparison of genetic and nongenetic risk factors for hypertrophic cardiomyopathy in the UK Biobank. (a) Hazard ratios, calculated using Cox proportional-hazards models adjusting for age, sex, genotyping array, and PCs 1-5. (b) Area under the receiver-operator curve, calculated by logistic regression adjusting for age, sex, genotyping array, and PCs 1-5. The reference for the area under the receiver-operator curve is a model consisting of age, sex, and PCs 1-5; additional predictors were added individually to the model, and those AUC values are plotted. (c) Disease variance explained for genetic and nongenetic factors. High polygenic risk score is defined as an individual with a score value above the 80th percentile. Error bars denote 95% confidence intervals. Different definitions of rare variant carrier status (HCM-ACMG, HCM-Panel, MYBPC3) are shown in red, PRS is shown in white, and clinical risk factors are shown in blue. Abbreviations: PRS, polygenic risk score; HTN, hypertension; AF, atrial fibrillation; CAD, coronary artery disease.



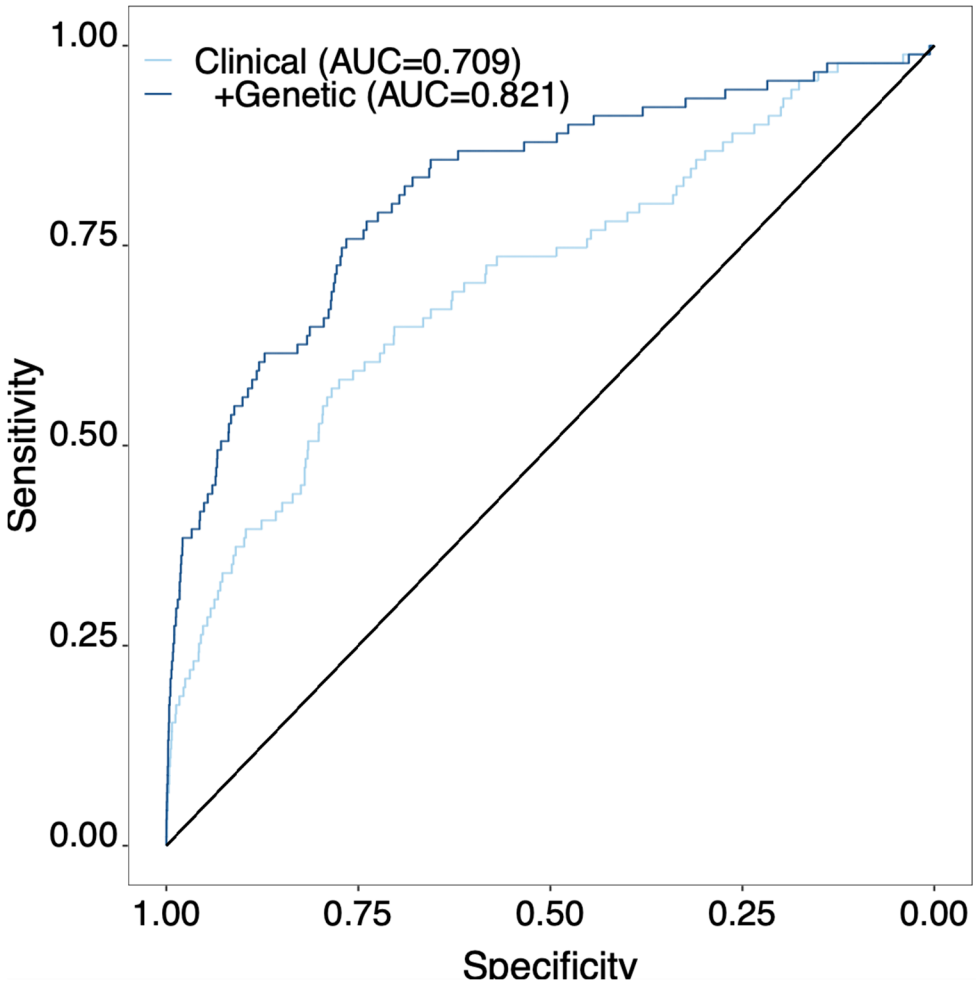
eFigure 5: Cumulative predictive capabilities of genetic and nongenetic risk factors for hypertrophic cardiomyopathy. Left panels show cumulative area-under-the-receiver-operator curve (AUC) for different prediction models. The vertical line shows the AUC for the reference model, calculated from logistic regression with age, sex, and PCs 1-5 as predictors. Clinical risk factors, rare variant status, and polygenic risk were then cumulatively added as predictors on top of this base model. Right panels show cumulative addition in variance explained (pseudo r^2 values) for different prediction models. Vertical bars represent r^2 of 0. In each row, the addition in r^2 values are shown for cumulative addition of clinical risk factors, rare variant status and polygenic risk, as compared to the baseline logistic regression model. (**a**) shows results for rare variant carrier status defined as HCM-ACMG variants, (**b**) shows results for HCM-Panel, and (**c**) shows results for MYBPC3 rare variants only. High polygenic risk is defined as an individual with a polygenic score above the 80th percentile. Error bars denote 95% CI. Abbreviations: PRS, polygenic risk score; HTN, hypertension; AF, atrial fibrillation; CAD, coronary artery disease.



eFigure 6: Comparisons of genetic and clinical risk factors, using continuous values for risk factors where applicable. Analyses of genetic and nongenetic risk factors were rerun, this time using continuous risk factors instead of dichotomized risk factors where applicable (if continuous risk factor exists and also shows association with HCM). Panels a-c show comparisons of individual genetic and nongenetic risk factors for hypertrophic cardiomyopathy, looking at (**a**) Hazard ratios, calculated using Cox proportional-hazards models adjusting for age, sex, genotyping array, and PCs 1-5; (**b**) Area under the receiver-operator curve (AUC) values, calculated from logistic regression adjusting for age, sex, genotyping array, and PCs 1-5. The reference for the area under the receiver-operator curve is a model consisting of age, sex, and PCs 1-5; additional predictors were added individually to the model, and those AUC values are plotted; (**c**) Disease variance explained for genetic and nongenetic factors. Panels d-e show cumulative predictive capabilities of models adding different nongenetic and genetic risk factors to the baseline model, with (**d**) showing cumulative AUC values; (**e**) showing cumulative pseudo r^2 values over baseline model. Error bars denote 95% confidence intervals. Abbreviations: PRS, polygenic risk score; BMI, body-mass-index; HTN, hypertension; AF, atrial fibrillation; CAD, coronary artery disease.



eFigure 7: Comparisons of genetic and clinical risk factors, excluding the first UK Biobank 50,000 exome-sequencing tranche. Analyses of genetic and nongenetic risk factors were rerun, this time removing any samples who were included in the initial 50,000 tranche of UK Biobank exome sequencing data. Panels a-c show comparisons of individual genetic and nongenetic risk factors for hypertrophic cardiomyopathy, looking at (**a**) Hazard ratios, calculated using Cox proportional-hazards models adjusting for age, sex, genotyping array, and PCs 1-5; (**b**) Area under the receiver-operator curve (AUC) values, calculated from logistic regression adjusting for age, sex, genotyping array, and PCs 1-5. The reference for the area under the receiver-operator curve is a model consisting of age, sex, and PCs 1-5; additional predictors were added individually to the model, and those AUC values are plotted; (**c**) Disease variance explained for genetic and nongenetic factors. Panels d-e show cumulative predictive capabilities of models adding different nongenetic and genetic risk factors to the baseline model, with (**d**) showing cumulative AUC values; (**e**) showing cumulative pseudo r^2 values over baseline model. Error bars denote 95% confidence intervals. Abbreviations: PRS, polygenic risk score; HTN, hypertension; AF, atrial fibrillation; CAD, coronary artery disease.



eFigure 8: Receiver-operator curve for the clinical vs. genetic and clinical factors for HCM. Area under the receiver-operator curve (AUC) values, calculated from logistic regression adjusting for age, sex, genotyping array, and PCs 1-5. The reference for the area under the receiver-operator curve is a model consisting of obesity, prevalent hypertension, prevalent atrial fibrillation, and prevalent coronary artery disease, age, sex, and PCs 1-5; HCM-ACMG carrier status and high polygenic score (>80th percentile) were then added to the model; Error bars denote 95% confidence intervals. HCM is defined as incident HCM.

eReferences

1. About UK Biobank | UK Biobank n.d. https://www.ukbiobank.ac.uk/about-biobank-uk/ (accessed June 9, 2020).
2. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–209. doi:10.1038/s41586-018-0579-z.
3. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature 2020;586:749–756. doi:10.1038/s41586-020-2853-0.
4. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat Genet 2021;53:942–948. doi:10.1038/s41588-021-00885-0.
5. Jurgens SJ, Choi SH, Morrill VN, et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet*. Published online February 17, 2022:1-11. doi:10.1038/s41588-021-01011-w
6. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat 2016;37:235–241. doi:10.1002/humu.22932.
7. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–443. doi:10.1038/s41586-020-2308-7.
8. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biology 2016;17:122. doi:10.1186/s13059-016-0974-4.
9. GeneDx. GeneDx n.d. https://www.genedx.com/test-catalog/available-tests/hcm-panel/ (accessed April 7, 2021).
10. Genetic testing for hypertrophic cardiomyopathy. Blueprint Genetics n.d. https://blueprintgenetics.com/tests/panels/cardiology/hypertrophic-cardiomyopathy-hcm-panel/ (accessed April 7, 2021).
11. Test | Invitae Hypertrophic Cardiomyopathy Panel n.d. https://www.invitae.com/en/physician/tests/02261/ (accessed April 7, 2021).
12. Hershberger RE, Givertz MM, Ho CY, Judge DP, Kantor PF, McBride KL, et al. Genetic evaluation of cardiomyopathy: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine 2018;20:899–909. doi:10.1038/s41436-018-0039-z.
13. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genetics in Medicine 2017;19:192–203. doi:10.1038/gim.2016.90.
14. Alfares AA, Kelly MA, McDermott G, Funke BH, Lebo MS, Baxter SB, et al. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. Genetics in Medicine 2015;17:880–888. doi:10.1038/gim.2014.205.
15. Morita H, Rehm HL, Menesses A, McDonough B, Roberts AE, Kucherlapati R, et al. Shared Genetic Causes of Cardiac Hypertrophy in Children and Adults. Http://DxDoiOrg/101056/NEJMoa075463 2009. doi:10.1056/NEJMoa075463.
16. Harper AR, Goel A, Grace C, Thomson KL, Petersen SE, Xu X, et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. Nature Genetics 2021;53:135–142. doi:10.1038/s41588-020-00764-0.
17. Drinking Levels Defined. National Institute on Alcohol Abuse and Alcoholism (NIAAA) 2011. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking (accessed June 9, 2020).