Common genetic variants, and modifiable risk factors, underpin hypertrophic cardiomyopathy susceptibility and expressivity

Andrew R. Harper[1,2], Anuj Goel[1,2], Christopher Grace[1,2], Kate L. Thomson[1,2,3], Steffen E. Petersen[4], Xiao Xu[5], Adam Waring[2], Elizabeth Ormondroyd[1,2], Christopher M. Kramer[6], Carolyn Y. Ho[7], Stefan Neubauer[1], HCMR Investigators, Rafik Tadros[8], James S. Ware[5*], Connie R. Bezzina[9*], Martin Farrall[1,2*], and Hugh Watkins[1,2,10*]

[1]Radcliffe Department of Medicine, University of Oxford, Division of Cardiovascular Medicine, John Radcliffe Hospital, Oxford, UK.
[2]Wellcome Centre for Human Genetics, Roosevelt Drive, Oxford, UK.
[3]Oxford Medical Genetics Laboratories, Churchill Hospital, Oxford, UK.
[4]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
[5]National Heart & Lung Institute, Imperial College London, London, UK.
[6]University of Virginia Health System, Charlottesville, VA, USA.
[7]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA.
[8]Cardiovascular Genetics Centre, Montreal Heart Institute, Montreal, QC, Canada.
[9]Amsterdam UMC, AMC Heart Center, Amsterdam, The Netherlands.
[10]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK.
*These authors contributed equally to this work.
e-mail: hugh.watkins@rdm.ox.ac.uk

Hypertrophic cardiomyopathy (HCM) is a common, serious, genetic heart disorder. Rare pathogenic variants in sarcomere genes cause HCM, but with unexplained phenotypic heterogeneity. Moreover, most patients do not carry such variants. We report a genome-wide association study of 2,780 cases and 47,486 controls that identified 12 genome-wide significant susceptibility loci for HCM. SNP-heritability indicated a strong polygenic influence, especially for sarcomere-negative HCM (64% of cases; h^2_g = 0.34 ± 0.02). A genetic risk score showed substantial influence on odds of HCM in a validation study, halving odds in the lowest quintile and doubling in the highest quintile, and also influenced phenotypic severity in sarcomere variant carriers. Mendelian randomization identified diastolic blood pressure (DBP) as a key modifiable risk factor for sarcomere-negative HCM, with 1 standard deviation increase in DBP increasing HCM risk four-fold. Common variants and modifiable risk factors have important roles in HCM that we suggest will be clinically actionable.

HCM is common, affecting at least 1 in 500 individuals, and presents substantial unmet medical need[1]. It is a leading cause of sudden death, embolic stroke and heart failure in early and mid-adult life. Sarcomeric HCM, caused by mutations in myofilament genes, is inherited as an autosomal dominant disorder. However, as commonly seen in adult onset heterozygous disorders, HCM is characterized by reduced penetrance and variable expressivity, providing challenges for diagnosis and prognosis[2,3]. In the more common sarcomere-negative setting, cases are often isolated, but clustering in nuclear families is still frequent, requiring clinical surveillance in families[4,5]. To investigate the contribution of common genetic variants to HCM risk, we performed two independent multi-ancestry case-control genome-wide

association studies (GWAS) of unrelated HCM patients recruited to the Hypertrophic Cardiomyopathy Registry (HCMR, 2,541 unselected cases vs. 40,283 UK Biobank controls) and the BioResource for Rare Disease (BRRD, 239 sarcomere-negative cases vs. 7,203 controls) (Fig. 1, Supplementary Table 1 and Supplementary Note). SNP-heritability (h^2_g) estimates calculated using GREML-LDMS indicated that a significant proportion of HCM risk was attributable to the additive effects of common (minor allele frequency (MAF) > 0.01) SNPs (HCMR $h^2_g = 0.35 \pm 0.01$; BRRD $h^2_g = 0.68 \pm 0.16$).

We performed fixed-effects inverse-variance meta-analysis of the HCMR and BRRD GWAS datasets for 8,590,397 single nucleotide polymorphisms (SNPs) across a total of 2,780 HCM cases and 47,486 age and sex-matched controls. All-comer analysis (i.e. inclusive of sarcomere-positive and sarcomere-negative HCM cases) identified 13 independent genome-wide significant variants in 12 loci ($P < 5 \times 10^{-8}$) using a stepwise model selection procedure with Genome-wide Complex Trait Analysis (GCTA) and confirmed with conditional analysis (Table 1, Supplementary Table 2 and Online Methods). We identified an additional 16 independent variants at a 5% FDR significance threshold ($P < 1.82 \times 10^{-6}$) (Supplementary Table 3). We replicated 11 of the 13 genome-wide significant variants and 4 of the 16 FDR variants in a smaller, independent HCM meta-analysis (n = 1,643 cases and 6,628 controls; Table 1, Supplementary Table 3 and Online Methods). Additionally, we obtained similar discovery findings with a European-only analysis (Supplementary Table 2 and Supplementary Note).

The *FHOD3* locus was found to harbor two independent genome-wide significant variants, rs4799426 and rs118060942, in linkage equilibrium ($r^2 = 0.01$). Sentinel SNPs in the HCM susceptibility loci conferred relatively large susceptibility effect sizes (median OR = 1.25, range 1.18–2.16) across a range of effect allele frequencies (range 0.012–0.83) (Supplementary Table 4). Tissue enrichment tests, performed in FUMA using gene-level data

(Supplementary Table 5) and tissue expression data from GTEx (v8.0) showed enrichment in left ventricular myocardium (beta = 0.04 ± 0.01; $P = 7.46 \times 10^{-6}$), skeletal muscle (beta = 0.03 ± 0.01; $P = 1.13 \times 10^{-5}$), and atrial appendage (beta = 0.04 ± 0.01; $P = 1.18 \times 10^{-5}$) (Supplementary Table 6 and Supplementary Note)[6,7]. Functional GWAS supported these findings and revealed cell types where sentinel SNPs were most enriched (Supplementary Tables 7 and 8).

We dichotomized HCM cases in HCMR into sarcomere-positive (34.3%) and sarcomere-negative (64.3%) groups using a published framework (Supplementary Tables 9 and 10)[8]. The GREML heritability estimate for sarcomere-negative HCM exceeded that of sarcomere-positive HCM ($h^2_g = 0.34 \pm 0.02$ vs. 0.16 ± 0.04) (Supplementary Table 11). This supports the hypothesis that where there is familial aggregation that is not explained by co-segregation with a rare variant, as in sarcomere-negative HCM, a greater role for common variants may be expected. This applies in particular to the BRRD samples, which were enriched for positive family history despite negative gene-panel testing, where heritability was indeed greatest. A meta-analysis of sarcomere-negative HCM (1,874 HCM cases vs. 27,344 controls) identified 10 independent genome-wide significant variants in 9 loci and a further 15 independent variants in 13 loci below a 5% FDR threshold ($P < 1.56 \times 10^{-6}$) (Supplementary Table 12). Three loci (*FHOD3*, *TBX3* and *PLN*) harbored a secondary independent variant following conditional association analysis (Supplementary Table 2). Sarcomere-positive HCM GWAS analysis (871 HCMR cases vs. 20,142 UKBB controls) yielded 7 independent genome-wide significant variants and a further 11 independent variants below a 5% FDR threshold ($P < 1.50 \times 10^{-6}$) from 12 loci. This includes 7 variants in the peri-centromeric region of chromosome 11 neighboring *MYBPC3*, a prominent cause of monogenic HCM (Supplementary Tables 2 and 13). Haplotype analysis of individual-level sequence data demonstrated long-range linkage disequilibrium and potential spurious

associations between frequently observed rare pathogenic variants in *MYBPC3* (NM_000256.3), specifically p.R502W and p.Trp792ValfsTer41, and common imputed variants in the chr11:44,976,681-57,917,265 genomic interval[9–11]. Modelling the impact of both rare and common variants with multiple logistic regression confirmed that HCM risk could be entirely attributed to the rare variants (Supplementary Note). Common variants in chr11:44,000,000-58,000,000 were masked from subsequent analyses, leaving 2 independent variants of genome-wide significance and 9 below a 5% FDR threshold. Excluding chr11:44,976,681-57,917,265 had a trivial effect on the heritability estimate.

Bivariate GREML analysis revealed a strong positive genetic correlation between sarcomere-positive and sarcomere-negative HCM ($r_g = 1.00 \pm 0.12$). Pairwise GWAS comparison revealed overlapping signals between sarcomere-negative and sarcomere-positive loci for 59% of regions ($n = 22/37$) (Supplementary Table 14). Most of the sarcomere-negative GWAS loci were not reproduced at the genome-wide significance level in the sarcomere-positive GWAS, which could potentially be explained by a relative lack of power (Supplementary Table 15). Four SNPs (rs2312403, rs35469308, rs12299450 and rs2758215) showed association only in the sarcomere-positive GWAS, and may represent modifier loci.

All loci were novel, apart from *FHOD3*, which has been previously reported in a HCM GWAS[12]. Previous candidate gene studies have reported rare variant associations in different forms of cardiomyopathy for *BAG3* and *FHOD3*, and common variant associations with dilated cardiomyopathy (DCM) have been reported for *BAG3* and *HSPB7* loci[13–17]. At these loci shared by HCM and DCM, the direction of effect is opposite, with the HCM risk allele being previously shown to decrease risk of DCM. The involvement of *BAG3*, *HSPB7* and *FHOD3* points to the importance of homeostatic pathways for sarcomeric structural integrity during mechanical stress (Supplementary Note). While some of the other loci also encode known cardiomyopathy genes (*PLN*, *TTN*), the major HCM and DCM myofilament

loci are not represented, consistent with the cardiomyopathy-causing changes in these genes altering protein structure rather than expression level. In the remaining loci, some early clues implicate specific genes and mechanisms: a deleterious missense variant implicates *ADPRHL1*, important for Z-disc and actin dynamics, and a *cis*-eQTL implicates *SLC6A6*, which encodes a taurine transporter known to be responsible for cardiomyopathy in dogs (Supplementary Note).

After excluding rs28768976 for extreme pleiotropy (Supplementary Note) and rs78310129 due to long-range linkage disequilibrium with pathogenic *MYBPC3* variants (Supplementary Note), 27 SNPs demonstrating independent associations with HCM at a 5% FDR threshold in the all-comer HCM meta-analysis were aggregated into a scaled (i.e. per-standard deviation effects) weighted genetic risk score (GRS) (Table 2 and Supplementary Tables 16 and 17). The GRS predicts odds of HCM in a validation meta-analysis of three independent HCM cohorts comprising 1,769 cases and 39,828 controls (OR = 1.73 per SD (95%CI 1.63-1.83)) (Fig. 2). Using the largest replication cohort, we conducted a sensitivity analysis and confirmed a 5% FDR threshold as representative of alternate SNP significance thresholds (Supplementary Table 16).

Stratification of the HCMR cases by their average genetic ancestry, as determined by principal components analysis, demonstrates similar effect sizes across all ancestry groups (Supplementary Table 18). Using the central 60% of the population as the reference group, there is a protective effect for individuals in the lowest quintile (OR = 0.53 (95%CI 0.45-0.63)) and a greater than a two-fold increased odds of HCM for individuals in the highest quintile (OR = 2.30 (95%CI 2.02–2.62)). In alignment with h^2_g estimates, the GRS demonstrated larger effects in the sarcomere-negative subgroup (Fig. 2 and Supplementary Note). Nevertheless, in young individuals carrying a pathogenic sarcomere mutation, who

might typically have a ~50% chance of developing overt cardiomyopathy in adulthood, a halving or doubling of average risk of developing the cardiomyopathy is likely to be clinically meaningful.

To determine whether the common susceptibility variants also influence disease severity in sarcomere-positive HCM, i.e. through a modifier effect, we assessed the impact of the GRS on LV hypertrophy in groups of cases with similar mutational mechanisms. A 1 SD unit increase in GRS conferred a 0.71 ± 0.35 mm increase in maximum left ventricular wall thickness (P = 0.048) in carriers of *MYBPC3* truncating variants (n = 232) and a 0.73 ± 0.36 mm increase (P = 0.037) in carriers of *MYH7* missense variants (n = 186) (Fig. 3). Allelic heterogeneity currently limits single variant expressivity estimates; the most frequently observed pathogenic variant in HCM (*MYBPC3*, p.R502W) is associated with a larger GRS effect size (1.61 ± 0.80 mm increase per 1 SD unit increase in GRS) but is currently modestly powered (n = 48).

Observational studies show that hypertension, obesity and type 2 diabetes are more prevalent in individuals with HCM, but these could be secondary to reduced exercise[8,18,19]. We performed two-sample Mendelian randomization (2SMR) to leverage large-scale GWAS for these heritable traits[20–23]. We inferred causal relationships with HCM for hypertension and obesity, but not diabetes (Fig. 4 and Supplementary Table 19). Most notably, diastolic blood pressure appears to be a substantial risk factor for the development of sarcomere-negative HCM (Fig. 4 and Supplementary Table 19). A 1 SD unit increase in diastolic blood pressure (11.3 mmHg) confers a four-fold increased risk of HCM (OR = 3.93 (95%CI 2.86–5.41); $P = 3.74 \times 10^{-16}$), more than double the risk typically observed for other diseases associated with diastolic blood pressure (Fig. 4 and Supplementary Table 20)[24–28]. The strong association with hypertension raises the possibility that sarcomere-negative HCM may

represent, in part, an exaggerated response to hypertension in genetically susceptible individuals. The association specifically with diastolic blood pressure likely reflects that this is the dominant form of hypertension in young and mid-adult life[29,30].

The individual loci identified in this study hold great potential for driving new insights into cardiomyopathy pathogenesis. Many of the association signals have already been replicated; others will need further study to guard against false positive findings. Collectively, our findings highlight the important influence of common variants on the risk of developing HCM. The polygenic contribution is weaker in individuals with pathogenic sarcomeric variants, but a common variant GRS may still be particularly useful here because the high prior risk means that the modest (e.g. four-fold) changes in individual-specific penetrance, which will apply to 40% of individuals, will have a large absolute effect on outcome. Additionally, it appears that common variants explain part of the variable expressivity of pathogenic sarcomeric variants. The clinical utility of a GRS now needs study in adequately powered longitudinal surveys of HCM disease progression, especially in sarcomere-positive individuals who were limited in number in the current study. In individuals lacking cardinal pathogenic mutations in sarcomeric genes, we suggest that extremes of the polygenic risk distribution (e.g. the top 1% of the population), combined with causal risk factors, drive individual susceptibility. Managing sarcomere-negative HCM patients and their relatives may be greatly facilitated by awareness of the strong influence of polygenic risk and of diastolic blood pressure as a major modifiable risk factor.

HCMR Investigators

Paul Kolm[11], Raymond Kwong[12], Sarahfaye F. Dolman[11], Patrice Desvigne-Nickens[13], John P. Dimarco[14], Nancy Geller[13], Dong-Yun Kim[13], Cheng Zhang[11], William Weintraub[11], Theodore Abraham[15], Lisa Anderson[16], Evan Appelbaum[17], Camillo Autore[18], Colin Berry[19],

Elena Biagini[20], William Bradlow[21], Chiara Bucciarelli-Ducci[22], Amedeo Chiribiri[23], Lubna Choudhury[24], Andrew Crean[25], Dana Dawson[26], Milind Y. Desai[27], Eleanor Elstein[28], Andrew Flett[29], Matthias Friedrich[30], Stephen Heitner[31], Adam Helms[32], Daniel L. Jacoby[33], Han Kim[34], Bette Kim[35], Eric Larose[36], Masliza Mahmod[37], Heiko Mahrholdt[38], Martin Maron[39], Gerry McCann[40], Michelle Michels[41], Saidi Mohiddin[42], Sherif Nagueh[43], David Newby[44], Iacopo Olivotto[45], Anjali Owens[46], F. Pierre-Mongeon[47], Sanjay Prasad[48], Ornella Rimoldi[49], Michael Salerno[50], Jeanette Schulz-Menger[51], Mark Sherrid[52], Peter Swoboda[53], Albert van Rossum[54], Jonathan Weinsaft[55], James White[56], and Eric Williamson[57]

[11]MedStar Heart and Vascular Institute, Washington, DC, USA.
[12]Cardiovascular Division, Department of Medicine and Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA.
[13]National Heart, Lung, and Blood Institute, Bethesda, MD, USA.
[14]Cardiovascular Division, University of Virginia Health System, Charlottesville, VA, USA.
[15]Hypertrophic Cardiomyopathy Center of Excellence, Johns Hopkins University, Baltimore, MD, USA.
[16]St Georges University Hospitals NHS Trust, London, UK.
[17]Departments of Medicine, Cardiovascular Division and Radiology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA.
[18]Division of Cardiology, Department of Clinical and Molecular Medicine, St. Andrea Hospital, Sapienza University, Rome, Italy.
[19]British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK.
[20]Cardio-Thoraco-Vascular Department, University Hospital of Bologna, Policlinico S. Orsola-Malpighi, Bologna, Italy.
[21]Department of Cardiology, New Queen Elizabeth Hospital Birmingham, Birmingham, UK.
[22]Bristol Heart Institute, Bristol National Institute of Health Research (NIHR) Biomedical Research Centre, University Hospitals Bristol NHS Trust and University of Bristol, Bristol, UK.
[23]Cardiovascular Division, King's College London British Heart Foundation Centre of Excellence, The Rayne Institute, St. Thomas' Hospital Campus, London, UK.
[24]Division of Cardiology, Department of Medicine, Bluhm Cardiovascular Institute, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.
[25]Division of Cardiology, Peter Munk Cardiac Center, University Health Network, University of Toronto, Toronto, ON, Canada.
[26]Aberdeen Cardiovascular and Diabetes Centre, University of Aberdeen, Aberdeen, UK.
[27]Department of Cardiovascular Medicine, Center for Radiation Heart Disease, Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH, USA.
[28]Division of Cardiology, Department of Medicine, Royal Victoria Hospital, McGill University Health Centre, Montreal, QC, Canada.
[29]Department of Cardiology, University Hospital Southampton NHS Foundation Trust, Southampton, UK.
[30]Department of Medicine, Heidelberg University, Heidelberg, Germany.
[31]Oregon Health and Sciences University (OHSU), Division of Cardiovascular Medicine, Knight Cardiovascular Institute, Portland, OR, USA.
[32]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA.
[33]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA.
[34]Duke Cardiovascular Magnetic Resonance Center and Division of Cardiology, Duke University Medical Center, Durham, NC, USA.

[35]Mount Sinai West, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[36]Quebec Heart and Lung Institute, Laval University, QC, Canada.
[37]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.
[38]Department of Cardiology, Robert-Bosch-Krankenhaus, Stuttgart, Germany.
[39]Hypertrophic Cardiomyopathy Center and Research Institute, Tufts Medical Center, Boston, MA, USA.
[40]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK.
[41]Erasmus University, Rotterdam, The Netherlands.
[42]Barts Heart Center, The Cardiovascular Magnetic Resonance Imaging Unit, St Bartholomew's Hospital, London, UK.
[43]Methodist DeBakey Heart and Vascular Center, Houston, TX, USA.
[44]Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK.
[45]Cardiomyopathy Unit and Genetic Unit, Careggi University Hospital, Florence, Italy.
[46]Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
[47]Montréal Heart Institute, Montréal, QC, Canada.
[48]National Heart and Lung Institute, Imperial College London, and Royal Brompton Hospital, London, UK.
[49]Vita Salute University and San Raffaele Hospital, Milan, Italy.
[50]Department of Medicine, University of Virginia, Charlottesville, VA, USA.
[51]Charité, Medical Faculty of the Humboldt University, Experimental and Clinical Research Center, Berlin, and Helios Clinics, Cardiology, Berlin, Germany.
[52]Hypertrophic Cardiomyopathy Program, Leon Charney Division of Cardiology, Department of Medicine, New York University School of Medicine, New York, NY, USA.
[53]Department of Cardiovascular Imaging Science, Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK.
[54]Department of Cardiology, Amsterdam UMC, Amsterdam, The Netherlands.
[55]Departments of Medicine and Radiology, Weill Cornell Medical College, New York, NY, USA.
[56]Calgary Foothills Medical Centre, University of Calgary, Calgary, AB, Canada.
[57]Department of Radiology, Mayo Clinic, Rochester, MN, USA.

Acknowledgements

This work was supported by funding from the British Heart Foundation (BHF), the Medical Research Council (MRC), the National Heart, Lung, and Blood Institute (NIH grant U01HL117006-01A1), the Wellcome Trust (201543/B/16/Z), Wellcome Trust core awards (090532/Z/09/Z, 203141/Z/16/Z) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. A.R.H. has received support from the Medical Research Council Doctoral Training Partnership. A.G. has received support from the BHF, European

Commission [LSHM-CT- 2007-037273, HEALTH-F2-2013-601456] and TriPartite Immunometabolism Consortium [TrIC]- NovoNordisk Foundation [NNF15CC0018486]. S.E.P. acknowledges support from the NIHR Barts Biomedical Research Centre. A.W. has received support from the Wellcome Trust. S.N., M.F. and H.W. are members of the Oxford BHF Centre for Research Excellence (RE/13/1/30181). We are grateful for access to the high-performance Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health or the Department of Health and Social Care.

We thank the NIHR BioResource volunteers for their participation, and gratefully acknowledge NIHR BioResource centres, NHS Trusts and staff for their contribution. We thank the National Institute for Health Research and NHS Blood and Transplant. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care) and funded by the National Institute for Health Research and NHS England with research infrastructure funding from the Wellcome Trust, Cancer Research UK and the Medical Research Council. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

We acknowledge the contribution of the Oxford Medical Genetics Laboratories.

Author Contributions

A.R.H., M.F., and H.W. conceived and designed the study. A.R.H., A.G., C.G., K.L.T., S.E.P., A.W., E.O., C.M.K., S.N., and C.Y.H. acquired, analyzed, and interpreted the data.

X.X. and R.T. provided assistance with replication. A.R.H., M.F., and H.W. wrote the manuscript. J.S.W., C.R.B., and R.T. critically revised the manuscript for important intellectual content.

Competing Interests

As of April 2020, A.R.H. is an employee of AstraZeneca.

References

1. Yotti, R., Seidman, C. E. & Seidman, J. G. Advances in the genetic basis and pathogenesis of sarcomere cardiomyopathies. *Annu. Rev. Genomics Hum. Genet.* **20**, 129–153 (2019).
2. Harper, A. R., Parikh, V. N., Goldfeder, R. L., Caleshu, C. & Ashley, E. A. Delivering clinical grade sequencing and genetic test interpretation for cardiovascular medicine. *Circ. Cardiovasc. Genet.* **10**, e001221 (2017).
3. Watkins, H., Ashrafian, H. & Redwood, C. Inherited cardiomyopathies. *N. Engl. J. Med.* **364**, 1643–1656 (2011).
4. Thomson, K. L. *et al.* Analysis of 51 proposed hypertrophic cardiomyopathy genes from genome sequencing data in sarcomere negative cases has negligible diagnostic yield. *Genet. Med.* **21**, 1576–1584 (2019).
5. Ingles, J. *et al.* Nonfamilial hypertrophic cardiomyopathy: prevalence, natural history and clinical implication. *Circ. Cardiovasc. Genet.* **10**, e001620 (2017).
6. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
7. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv* (2019) doi:10.1101/787903.
8. Neubauer, S. *et al.* Distinct subgroups in hypertrophic cardiomyopathy in the NHLBI HCM Registry. *J. Am. Coll. Cardiol.* **74**, 2333–2345 (2019).
9. Wray, N. R., Purcell, S. M. & Visscher, P. M. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol.* **9**, e1000579 (2011).
10. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
11. Orozco, G., Barrett, J. C. & Zeggini, E. Synthetic associations in the context of genome-wide association scan signals. *Hum. Mol. Genet.* **19**, R137–R144 (2010).
12. Wooten, E. C. *et al.* Formin homology 2 domain containing 3 variants associated with hypertrophic cardiomyopathy. *Circ. Cardiovasc. Genet.* **6**, 10–18 (2013).
13. Ochoa, J. P. *et al.* Formin Homology 2 Domain Containing 3 (FHOD3) is a genetic basis for hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **72**, 2457–2467 (2018).
14. Esslinger, U. *et al.* Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy. *PLoS One* **12**, e0172995 (2017).
15. Selcen, D. *et al.* Mutation in BAG3 causes severe dominant childhood muscular

dystrophy. *Ann. Neurol.* **65**, 83–89 (2009).
16. Odgerel, Z. *et al.* Inheritance patterns and phenotypic features of myofibrillar myopathy associated with a BAG3 mutation. *Neuromuscul. Disord.* **20**, 438–442 (2010).
17. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur. Heart J.* **32**, 1065–1076 (2011).
18. Fumagalli, C. *et al.* Association of obesity with adverse long-term outcomes in hypertrophic cardiomyopathy. *JAMA Cardiol.* **5**, 65–72 (2019).
19. Ho, C. Y. *et al.* Genotype and lifetime burden of disease in hypertrophic cardiomyopathy. *Circulation* **138**, 1387–1398 (2018).
20. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
21. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
22. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
23. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
24. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
25. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
26. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
27. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
28. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
29. Franklin, S. S. *et al.* Predictors of new-onset diastolic and systolic hypertension: the Framingham Heart Study. *Circulation* **111**, 1121–1127 (2005).
30. Franklin, S. S., Jacobs, M. J., Wong, N. D., L'Italien, G. J. & Lapuerta, P. Predominance of isolated systolic hypertension among middle-aged and elderly US hypertensives: analysis based on National Health and Nutrition Examination Survey (NHANES) III. *Hypertension* **37**, 869–874 (2001).

Figure legends

Figure 1 | Study design for the HCM genome-wide association analysis. Two independent HCM genome-wide association studies (GWAS) were performed before fixed-effects inverse-variance meta-analysis was conducted. Genetic risk scores were generated and stratified by sarcomere variant status. Findings were validated using three independent cohorts (GeL, Genomics England; RBH, Royal Brompton Hospital; the Netherlands cohort (Amsterdam, Rotterdam and Groningen)). Two-sample Mendelian randomization was performed, stratified by sarcomere variant status, to provide insight into heritable risk factors for HCM. SNP-heritability (h^2_g) estimates were compared between component GWAS studies using GREML-LDMS and stratified by sarcomere variant status. Standard errors for h^2_g estimates are presented in square brackets.

Figure 2 | Validation of an HCM genetic risk score. A genetic risk score was generated from 27 SNPs with <5% FDR and weighted by the beta estimate from the multi-ancestry meta-analysis joint model GCTA results. The genetic risk score was evaluated in all-comers, sarcomere-positive and sarcomere-negative HCM cases, in three validation cohorts. **a**, A quintile-based analysis demonstrates the protective effects of the genetic risk score in the lowest 20% of the population, compared with the middle 60%. Similarly, the upper 20% demonstrate increased susceptibility towards risk of developing HCM, when compared with the middle 60%. Odds ratios (*x*-axis) reported with error bars denoting 95% confidence intervals. **b**, To facilitate comparison between other genetic risk scores, a per standard deviation estimate is reported. Odds ratios (*x*-axis) reported with error bars denoting 95% confidence intervals. Validation cohorts: GeL (Genomics England HCM cases (n = 435) vs.

controls (*n* = 36,500)); RBH (Royal Brompton Hospital HCM cases (*n* = 359) vs. controls (*n* = 1,211)); and the Netherlands HCM cases (*n* = 975) vs. controls (*n* = 2,117)).

Figure 3 | Relationship between standardized genetic risk score (GRS) and maximum left ventricular (LV) wall thickness. Linear regression was performed to assess the most frequently observed HCM variant classes: truncating variants in *MYBPC3, MYH7* missense variants, and most frequently observed pathogenic variant (*MYPBC3* p.R502W). **a**, Carriers of pathogenic or likely pathogenic *MYH7* missense variants (*n* = 186, beta = 0.73 ± 0.35, *P* = 0.036). **b**, Carriers of *MYBPC3* truncating variants (*n* = 232, beta = 0.71 ± 0.35, *P* = 0.048). **c**, Carriers of the most frequently observed pathogenic variant in HCM, *MYBPC3* p.R502W (*n* = 48, beta = 1.61 ± 0.80, *P* = 0.051) evaluated in HCMR cases (*n* = 36) and participants from the UK Biobank (*n* = 12). Linear regression line is denoted in blue, alongside 95% confidence interval in gray. *P* values are uncorrected for multiple testing.

Figure 4 | Two-sample inverse variance weighted Mendelian randomization identifies modifiable risk factors for HCM. a, Effect of presumed risk phenotypes, based on prior observational evidence, on sarcomere-positive (*n* = 871) and sarcomere-negative (*n* = 1,635) HCM. Odds ratios represented per standard deviation for SBP, DBP, BMI and WHRadjBMI. Error bars represent 95% confidence intervals. As T2D is a binary phenotype, risk is represented as the per log odds unit of T2D. **b**, Relative impact of DBP on sarcomere-positive and sarcomere-negative HCM susceptibility, in relation to other established hypertension-associated phenotypes. Odds ratio measured per standard deviation of DBP (11.3 mmHg). Error bars represent 95% confidence intervals. Ischemic stroke reflects all TOAST subtypes. Summary of cases/controls included: heart failure (47,309 cases and 930,014 controls), atrial fibrillation (65,446 cases and 522,744 controls), cardioembolic stroke (9,006 cases and 454,450 controls), ischemic stroke (60,341 cases and 454,450 controls), T2D (74,124 cases and 824,006 controls), chronic kidney disease (64,164 cases and 625,219 controls), and coronary artery disease (122,733 cases and 424,528 controls). Abbreviations: BMI, body mass index; DBP, diastolic blood pressure; HCM, hypertrophic cardiomyopathy; SBP, systolic blood pressure; T2D, type 2 diabetes mellitus; WHRadjBMI, waist-hip ratio adjusted for BMI.

ONLINE METHODS

GWAS in multi-ancestry HCMR cases vs. UKBB controls. As described in Neubauer et al.[8], 2,755 incident HCM cases were recruited from 44 sites across 6 countries in North America and Europe. Cases were 18 to 65 years of age with evidence of unexplained left ventricular hypertrophy (wall thickness > 15 mm)[31]. All participants provided written informed consent (South Central - Oxford A Research Ethics Committee approval: 14/SC/0190; clinicaltrials.gov identifier: NCT01915615). Genotyping was performed using the Axiom™ Precision Medicine Research Array (Affymetrix/ThermoFisher). Following quality control, 2,541 individuals, not closely related (i.e. > 3 degrees of relatedness), were available for analysis. Gene panel sequence data, generated using a custom-designed TruSeq kit (Illumina), were available on 2,636 HCMR cases, as previously reported[8,32]. Variant classification was performed for the 8 core sarcomere genes (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1* and *TPM1*) using the American College of Medical Genetics and Genomics (ACMG) guidelines[33]. Cases were systematically dichotomized into sarcomere-positive (n = 871) or sarcomere-negative (n = 1,635) groups using a published, evidence-based framework (Supplementary Tables 9 and 10)[8,34]. Details of the rare variants used to partition cases are reported (Supplementary Tables 21 and 22). Access to the UK Biobank (UKBB) genotypes was provided through application 11223 (UK REC approval: 11/NW/0382). Genotyping was performed using the UK Biobank Axiom® array (Affymetrix). Individuals who underwent genotyping using the UK BiLEVE array, or had asked to be withdrawn from the UK Biobank, as of 16 October 2018, were excluded. Individuals with an ICD10 code indicating HCM (I420 or I421), or other phenotypes that may confound HCM analyses (Supplementary Table 23) in Hospital Episode Statistics (HES) data or self-reported questionnaire fields, were excluded (n = 15,901).

Individuals in the UKBB exome sequencing subset (n = 49,959) that harbor variants of uncertain significance (VUS), likely pathogenic or pathogenic variants in the core sarcomere genes were excluded. Closely related individuals, within 3 degrees of relatedness, and gender mismatches were excluded. Of the remaining 270,260 individuals, 40,283 were randomly selected for subsequent analysis, sampled using a 20:1 allocation against HCMR cases (n = 2,541), with approximate age (per decade) and genotype-assigned sex matching.

The HCMR (Precision Medicine Research Array (Affymetrix)) and UKBB (UK Biobank Axiom® array (Affymetrix)) cohorts were genotyped on partially overlapping arrays. 174,974 single nucleotide polymorphisms (SNPs) (MAF > 0.01, genotype missing rate 1%, Hardy-Weinberg equilibrium with mid-P correction of 1×10^{-9}) present in both the HCMR and UKBB cohorts were extracted for subsequent analysis. The UKBB and HCMR SNPs were aligned to the HRC reference panel, using HRC-1000G-check-bim.pl from https://www.well.ox.ac.uk/~wrayner/tools/, before being merged.

Principal components analysis was then performed using FlashPCA2 on a subset of SNPs in approximate linkage equilibrium ($r^2 < 0.05$), determined using the –indep-pairwise function in PLINK (version 1.90b3). Ancestry was inferred by projecting principal components, derived from the 1000 Genomes Project (Phase 3), onto HCMR/UKBB genotypes. A multinomial logistic regression model, performed using the *nnet* CRAN package in R (https://CRAN.R-project.org/package=nnet), classified ancestral groups as specified by the International Genome Sample Resource (http://www.internationalgenome.org/category/population/) (Supplementary Table 1). The Michigan Imputation Server[35] (https://imputationserver.sph.umich.edu/) performed haplotype phasing with Eagle[36], and imputation against the Haplotype Reference Consortium (HRC.r1.1.2016 reference panel)[37], generating genotypes for 38,954,302 imputed variants.

Imputed variants with an INFO score > 0.3 and MAF > 0.01 were retained for subsequent analysis.

An all-comer analysis (2,541 HCM cases vs. 40,283 controls) and separate sarcomere-positive (871 vs. 20,142) and sarcomere-negative HCM analyses (1,635 vs. 20,141) were performed. The UKBB controls were randomly allocated to either the sarcomere-positive or sarcomere-negative GWAS.

Analyses were performed with logistic regression to fit an additive case-control association model, using the SNPTEST v2.5.4-beta3 newml function, adjusting for the first ten ancestry-informative principal components. As HCM is a disease of a relatively low prevalence (~1 in 500), statistical power was maximized by not adjusting for age or sex[38]. There was no evidence of extreme population stratification in genomic control analyses (all-comers: original λ = 1.191, sarcomere-positive: λ = 1.089, sarcomere-negative: pre-λ_{GC} = 1.142). A genomic control adjustment was performed when λ exceeded 1.1.

Multi-ancestry BioResource for Rare Disease case-control GWAS. Details regarding the BioResource for Rare Disease (BRRD) cohort, a pilot study of the Genomics England 100,000 Genomes Project (GeL), have been described elsewhere[39]. All participants provided written informed consent (East of England - Cambridge South REC approval: 13/EE/0325). Briefly, 13,037 individuals from 20 rare disease areas underwent genome sequencing, including 243 individuals diagnosed with sarcomere-negative HCM[4]. Individuals clinically diagnosed with HCM, with diagnostic criteria as for HCMR, were recruited via inherited cardiac condition (ICC) clinics within the UK (Oxford University Hospitals NHS Foundation Trust, Royal Brompton & Harefield NHS Foundation Trust, Guy's and St Thomas' NHS Foundation Trust and the Newcastle upon Tyne Hospitals NHS Foundation Trust). Individuals recruited were aged 18 to 70 years old, or >70 years when there was a positive

family history, with an absence of likely pathogenic or pathogenic variants across 13 well-established HCM genes (sarcomeric genes: *MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1*, *TPM1*; other non-sarcomeric, but robustly associated HCM genes: *CSRP3*, *PLN*; and phenocopy genes: *PRKAG2*, *GLA*, *FHL1*)[4].

Reference controls were recruited from the other BRRD rare disease participants, or their family members. Individuals recruited via the GeL pilot study for the purposes of investigating an ICC were excluded. Overall, 239 cases and 7,203 controls were available for analysis, and high-quality variants were extracted from the respective genome sequencing variant call format (VCF) files. High quality variants were defined as those that had: PASS filter status; MAF > 1%, a depth of at least 10 informative reads per site (DP > 10); a genotype quality score of at least 20 (GQ > 20); and a genotype missingness of no more than 10% (CR > 0.9). Multiallelic sites were split. Ancestrally informative principal components were derived using FastPCA2 and 1000 Genomes Phase 3 data (Supplementary Table 1). Association analysis was performed using SAIGE (v0.29.4.2) with the first three principal components included as covariates[40]. SAIGE step 1 was performed using 123,903 genotypes following a linkage disequilibrium pruning procedure in PLINK (v1.9), with a 500-kb window, a step size of 50 markers, and a pairwise r^2 threshold of 0.2[41,42]. SAIGE step 2 analysis was performed using genotypes with a minor allele count > 5 and MAF > 0.01. Summary genetic association statistics for 9,341,129 autosomal variants were then computed using a mixed logistic regression model; a genomic control analysis showed little evidence of over-dispersion (λ = 1.049). The BRRD GWAS was included in the all-comer and the sarcomere-negative meta-analyses.

HCM sarcomere carrier stratification. Up to two-thirds of variants of uncertain significance (VUS) in confirmed sarcomere genes are considered causal of HCM[34]. In order to contrast the common-variant genetic architecture of HCM patients carrying pathogenic variants in sarcomeric genes with non-carriers, individuals were assigned sarcomere-positive status if they harbored a variant classified as either VUS-indeterminate, VUS-favors pathogenic, likely pathogenic or pathogenic in *ACTC1*, *MYH7*, *MYL2*, *MYL3*, *TNNT2*, *TNNI3* and *TPM1*, or a VUS-favors pathogenic, likely pathogenic or pathogenic in *MYBPC3* (Supplementary Tables 9 and 10)[8].

Heritability estimates. SNP-heritability (h^2_g) was estimated using GREML (Genomic relatedness matrix Restricted Maximum Likelihood)[43] for SNPs demonstrating an INFO score > 0.3 and MAF > 0.01. LD scores were assigned to SNPs from 200-kb blocks across the genome, before SNPs were stratified into quartiles based on SNP LD scores to generate genomic relatedness matrices (GRM). The GRMs were subjected to REML analysis of case-control status with the first ten ancestry-informative principal components as covariates. h^2_g estimates were approximated on a liability scale; this represents binary traits (i.e. cases vs. controls) on a continuous scale, and above a liability threshold an individual will be affected. Representation of a binary trait on this classic multifactorial liability scale is dependent on both the population prevalence of disease (0.2% based on population-based epidemiological estimates) and the sample prevalence of disease. The prevalence of sarcomere-negative and positive HCM was set as 0.0012 and 0.0008[34].

Quality control. EasyQC (v9.2)[44] was used for genotype quality control. The HRC reference panel was used for mapping and allele frequencies. Variants were removed if they were monomorphic, demonstrated a minor allele count < 6, were absent from the HRC reference

panel or duplicated, or when the observed allele frequency deviated by > 0.2 from the HRC allele frequency.

Genomic inflation was assessed across all cohorts through calculation of the genomic control, λ, and by evaluating the overall *P*-value distribution. Genomic control correction was performed when $\lambda > 1.1$, by adjusting the standard error ($SE_{gc} = SE \times \sqrt{\lambda}$) and re-calculating adjusted χ^2 statistics ($\chi^2_{gc} = (\frac{\beta}{SE_{gc}})^2$) and associated *P*-values (under 1 degree of freedom). The overall *P*-value distribution, generated from each component study and meta-analysis, were plotted and assessed. Local false discovery rates (FDR) were computed by the qvalue R package (https://github.com/StoreyLab/qvalue)[45]. The FDR provides a frequentist equivalent to the empirical Bayesian posterior probability that the null hypothesis is true, based on the distribution of generated p-values. For genome-wide significance, an *a priori* alpha threshold of 5×10^{-8} was set, and a FDR threshold of 5% was calculated for each study[46].

Meta-analysis. All-comer (2,780 HCM cases vs. 47,486 controls) and sarcomere-negative (1,874 cases vs. 27,344 controls) fixed-effects inverse-variance meta-analysis analyses, incorporating the HCMR vs. UKBB and BRRD vs. BRRD component GWAS, were conducted using GWAMA[47]. Effect sizes, standard errors, effect allele frequency estimates and heterogeneity statistics, specifically Cochran's statistic (*Q*), were reported alongside q-values and FDR values.

Replication. Replication of HCM loci was performed in a smaller, independent dataset composed of three HCM case-control studies from the Netherlands (975 cases, 2,117 controls), Royal Brompton Hospital (359 cases, 1,211 controls) and Canada (313 cases, 3,300 controls). No cases recruited to HCMR or BRRD were present in these replication cohorts.

Meta-analysis of these three GWAS studies was performed using METAL (λ = 1.074). Detailed methods regarding these replication cohorts are available in Tadros et al.[48].

Conditional association analysis. To identify genetic variants that confer independent risk effects, conditional association analysis was performed using a stepwise model selection procedure (--cojo-slct) using GCTA[49,50]. Summary statistics were extracted from meta-analyses results, or component GWAS studies, and linkage disequilibrium metrics were based on 62,018 unrelated, European individuals randomly selected from the UKBB. For each analysis, the *P*-value threshold was determined by a 5% FDR level.

Assessment of pleiotropy. All independently associated variants, confirmed in a conditional analysis, were evaluated for association with other diseases or traits by cross-referencing publicly available databases via OpenTargets[51] (https://genetics.opentargets.org/), a web resource that synthesises data from both the NHGRI-EBI GWAS catalogue (https://www.ebi.ac.uk/gwas/) and previously published UK Biobank summary statistics (Supplementary Table 24)[40]. Associations between sentinel SNPs and their tagging SNPs (r^2 > 0.8) and gene expression, proteins, metabolites or epigenetics were evaluated using PhenoScanner v2.0 (Supplementary Table 25)[52].

Genetic correlation. To measure the genetic correlation between sarcomere-positive and sarcomere-negative HCM, bivariate GREML analysis was performed. The prevalence of sarcomere-negative and sarcomere-positive HCM were set at 0.0012 and 0.0008. The first ten ancestry-informative principal components were included as covariates. To assess the shared genetic effects between sarcomere-positive and sarcomere-negative HCM at individual loci, pairwise GWAS (GWAS-PW) calculated the probability a variant located within locus

contributes to either one, both, or neither traits, or whether two separate signals within the same region contribute to each trait independently. GWAS-PW was performed using genomic regions 500 kb upstream and downstream of independent loci from the all-comer, sarcomere-positive and sarcomere-negative GWAS[53].

Long-range linkage disequilibrium and spurious association. Independent, genome-wide significant, variants on chromosome 11 (chr11:44,976,681-57,917,265) were identified due to their close proximity to *MYBPC3*, an HCM gene known to contain pathogenic founder variants, raising the possibility of long-range linkage disequilibrium and spurious association. Haplotypes were constructed using genotyped and imputed common variants (MAF > 0.2), in combination with rare *MYBPC3* variants (p.R502W, p.Trp792ValfsTer41 and c.1224-52G>A) derived from gene panel data (HCM cases) or exome data (UKBB controls), in PLINK (version 1.90b3). HCM cases were limited to those individuals in whom a pathogenic or likely pathogenic variant in a core sarcomere gene had previously been identified (n = 851). Controls were drawn from the UK Biobank (n = 19,851). Haplotype structure was evaluated using a maximum likelihood method in Haploview (version 4.2)[54]. Given the presence of numerous zero-value genotype counts, multiple logistic regression association analysis was performed using the R *logistf* function. This method was used to model the independent effects of a *rare pathogenic variant and a common* variant on HCM risk, while allowing for linkage disequilibrium between the two variants.

Genetic risk score. After removing SNPs demonstrating extreme pleiotropy or extreme ancestral bias (see Supplementary Note), a genetic risk score (GRS) that combines independent (i.e. in linkage equilibrium) SNPs identified through the conditional analysis of the multi-ancestry meta-analysis was first tested in a component GWAS (HCMR vs. UKBB)

then validated in three independent studies: Genomics England's 100,000 Genomes Project (GeL) (REC: 14/EE/1112), the Royal Brompton Hospital's HCM case-control series, and a Netherlands HCM case-control series (Supplementary Table 16). Individuals recruited to both a discovery cohort and a validation cohort were identified and excluded from the validation cohort (51 individuals from the Royal Brompton Hospital series and 24 individuals from the Netherlands series). The cumulative genetic effect of the SNPs was calculated for each individual using the allelic scoring function in PLINK. The relative weight assigned to each SNP was the beta estimate from the multi-ancestry meta-analysis joint model COJO results. Raw GRSs were plotted and evaluated, before standardizing the GRS distribution to a mean of 0 and variance of 1. A logistic regression model was fitted with affection status as the outcome variable and standardized GRS score as an explanatory variable, with covariates including the first ten principal components, age and gender. Cases from each validation dataset were dichotomized based on the presence of a rare causal variant in an established sarcomere gene into sarcomere-positive and sarcomere-negative cases (see Supplementary Note).

Expressivity analysis. The two predominant classes of pathogenic HCM variants are *MYBPC3* truncating variants and *MYH7* missense variants, the mechanisms of which have been previously demonstrated[55,56]. The most frequently detected pathogenic variant, present in ~2% of HCM cases, is *MYBPC3* p.R502W. Variants, classified using ACMG guidelines as pathogenic or likely pathogenic in *MYBPC3* and *MYH7* were identified[33,57]. Ensembl's Variant Effect Predictor was used to define variant consequences. Truncating variants included frameshift, stop gained and splice acceptor/donor variants.

For individuals in the HCMR cohort, maximum wall thickness measurements were derived from cardiac magnetic resonance (CMR) imaging (either 1.5-T or 3.0-T) performed

using a standardized protocol, multichannel phased-array chest coils and electrocardiographic gating, as previously reported[8].

Variant carriers for *MYBPC3* p.R502W were identified in both the HCMR and UKBB cohorts. In the UKBB cohort, heterozygous *MYBPC3* p.R502W carriers, who had also undertaken UKBB based CMR imaging, were identified using array-based genotypes (*n* = 12). Exome sequence data was available for six of these individuals; in all cases the presence of *MYBPC3* p.R502W was confirmed, supporting prior analysis indicating that UKBB-based genotyping for *MYBPC3* p.R502W was satisfactory[58,59]. Demographic and phenotypic details were reviewed for *MYBPC3* p.R502W carriers, including ICD10 classifications and self-reported co-morbidities for HCM codes (I421 and I422). *MYBPC3* p.R502W carriers were age and sex matched 1:1 with a non-variant carrier from a sample of unrelated UKBB participants who had undertaken exome sequencing and CMR imaging and demonstrated no potentially disease-causing variant (i.e. no HCM associated variant of uncertain significance, likely pathogenic variant or pathogenic variant) or a previous diagnosis, reported via HES data, that might confound CMR analyses (Supplementary Table 23). Long and short axis cine-tagged CMR data, generated by the UKBB as previously reported, were reviewed by an investigator blinded to variant carrier status, to minimise bias, when reporting maximum left ventricular wall thickness[60]. A linear regression model was used to approximate the effect of a standardised GRS (mean of 0 and variance of 1) against maximum left ventricular wall thickness in mm.

Mendelian randomization. Mendelian randomization (MR) uses the random meiotic segregation of alleles to assess whether an association between a risk factor and an outcome is consistent with a causal effect. Two-sample Mendelian randomization (2SMR) leverages data from large-scale genome-wide association studies to infer causal relationships between

two heritable traits. Observational data suggest several modifiable risk factors that may influence the phenotypic variability observed in HCM[18,19,61,62]. Genome-wide significant loci for blood pressure (systolic blood pressure, diastolic blood pressure)[21], body mass index/waist-hip ratio[22] and type 2 diabetes[20] were identified and summary statistics were collated as instrumental variables. Effect alleles were harmonized between instrumental variables and the HCM summary statistics. Analyses were performed using MR-base[63]. The corresponding SNPs were extracted from the sarcomere-positive and sarcomere-negative summary statistics and MR estimates generated using fixed and random-effects inverse-variance weighted (IVW) MR. Sensitivity analysis to test for horizontal pleiotropy was performed using MR-Egger, and for robustness by unweighted and weighted median regression. Betas and standard errors were compared, for all risk factors, between sarcomere-positive and sarcomere-negative HCM. Additional 2SMR was performed to further evaluate the relative effect of diastolic blood pressure on HCM, relative to other well-established risk factor-disease relationships (Supplementary Table 20).

Functional mapping and annotation. Functional annotation of GWAS summary statistics was undertaken using FUMA (v1.3.5e) (https://fuma.ctglab.nl/snp2gene)[64] to link genotypes, eQTLs and chromatin interactions. Tissue enrichment was performed using MAGMA (Multi-marker Analysis of GenoMic Annotation) (v1.07), a gene-level analysis tool provided by FUMA, with tissue expression data from GTEx (version 8.0) (https://www.gtexportal.org/home)[7]. eQTLs were evaluated in heart tissue (atrial appendage and left ventricle). Chromatin interaction data was evaluated in left and right ventricular tissue, derived from previously reported Hi-C data (GSE87112)[65].

Functional GWAS. Functional GWAS (fGWAS) is software that assesses enrichment of

functional sites (such as histone marks and methylation data) within GWAS summary statistics[66]. fGWAS then uses these enrichment parameters to fine map and re-weight GWAS loci.

Using multi-ancestry HCM meta-analysis summary statistics, fGWAS was performed using chromatin marks (enhancers, flanking/active TSS, active TSS, genetic enhancers, repressed polycomb, bivalent enhancers, transcription at gene 5' and 3', flanking/bivalent TSS/enhancers and bivalent/poised TSS) from the ChromHMM dataset of the Roadmap Epigenomic project, for cardiac tissues (left ventricle (E095), fetal heart (E083), right ventricle (E105), right atrium (E104))[67,68]. Enrichment estimates were generated for each tissue type, together with a list of loci below a 5% FDR significance threshold.

Data availability

We confirm that all relevant data are included in the paper and/or its supplementary information files. The datasets generated during this study are available from the corresponding author upon reasonable request. The following institutional domain (www.well.ox.ac.uk/hcm) will provide summary level statistics.

Code availability

Publicly available software tools were used to analyze these data. This includes: SAIGE (https://github.com/weizhouUMICH/SAIGE); SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html); GCTA (https://cnsgenomics.com/software/gcta/); PLINK (https://www.cog-genomics.org/plink/1.9/data); BGENIX (https://bitbucket.org/gavinband/bgen/wiki/bgenix); QCTOOL (https://www.well.ox.ac.uk/~gav/qctool_v2/); GWAMA (https://genomics.ut.ee/en/tools/gwama); and MR-base (http://www.mrbase.org/).

Methods-only references

31. Gersh, B. J. *et al.* 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **124**, e783-831 (2011).
32. Harper, A. R. *et al.* A re-evaluation of the South Asian *MYBPC3*$^{\Delta 25bp}$ intronic deletion in hypertrophic cardiomyopathy. *Circ. Genomic Precis. Med.* **13**, e002783 (2020).
33. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants : a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
34. Walsh, R. *et al.* Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet. Med.* **19**, 192–203 (2017).
35. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
36. Loh, P. R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
37. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
38. Pirinen, M., Donnelly, P. & Spencer, C. C. A. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat. Genet.* **44**, 848–851 (2012).
39. Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96–102 (2020).
40. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
41. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
42. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
43. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
44. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
45. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
46. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
47. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
48. Tadros, R. *et al.* Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat. Genet.* **[CE: add reference and doi for NG-LE54093R1 Tadros]**
49. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
50. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

51. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
52. Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
53. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
54. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
55. Toepfer, C. N. *et al.* Hypertrophic cardiomyopathy mutations in *MYBPC3* dysregulate myosin. *Sci. Transl. Med.* **11**, eaat1199 (2019).
56. Toepfer, C. N. *et al.* Myosin sequestration regulates sarcomere function, cardiomyocyte energetics, and metabolism, informing the pathogenesis of hypertrophic cardiomyopathy. *Circulation* **141**, 828–842 (2020).
57. Kelly, M. A. *et al.* Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet. Med.* **20**, 351–359 (2018).
58. Wright, C. F. *et al.* Assessing the pathogenicity, penetrance, and expressivity of putative disease-causing variants in a population setting. *Am. J. Hum. Genet.* **104**, 275–286 (2019).
59. Van Hout, C. V. *et al.* Exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).
60. Petersen, S. E. *et al.* UK Biobank's cardiovascular magnetic resonance protocol. *J. Cardiovasc. Magn. Reson.* **18**, 8 (2016).
61. Karam, R., Lever, H. M. & Healy, B. P. Hypertensive hypertrophic cardiomyopathy or hypertrophic cardiomyopathy with hypertension? A study of 78 patients. *J. Am. Coll. Cardiol.* **13**, 580–584 (1989).
62. Wasserstrum, Y. *et al.* The impact of diabetes mellitus on the clinical phenotype of hypertrophic cardiomyopathy. *Eur. Heart J.* **40**, 1671–1677 (2019).
63. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
64. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
65. Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
66. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
67. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **3**, 215–216 (2012).
68. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

Tables

Chr	SNP	Position (GRCh37)	NEA/ EA	Freq EA	Discovery			Replication			Locus name
					OR	95% CI	*P*	OR	95% CI	*P*	
Genome-wide significant ($P < 5 \times 10^{-8}$)											
1	rs1048302	16,340,879	G/T	0.33	1.32	1.24-1.40	2.54×10^{-17}	1.26	1.16-1.37	1.06×10^{-7}*	*HSPB7*
3	rs13061705	14,291,129	T/C	0.69	1.25	1.16-1.34	9.18×10^{-9}	1.13	1.04-1.24	4.49×10^{-3}	*SLC6A6*
6	rs3176326	36,647,289	G/A	0.21	1.28	1.19-1.38	2.22×10^{-11}	1.27	1.15-1.40	1.87×10^{-6}*	*CDKN1A*
6	rs12212795	118,654,308	G/C	0.05	1.48	1.31-1.67	2.51×10^{-10}	1.72	1.45-2.05	8.19×10^{-10}*	*PLN*
10	rs72840788	121,415,685	G/A	0.21	1.52	1.42-1.64	5.06×10^{-29}	1.42	1.29-1.56	4.90×10^{-13}*	*BAG3*
12	rs7301677	115,381,147	T/C	0.73	1.24	1.15-1.33	1.26×10^{-8}	1.19	1.08-1.31	2.74×10^{-4}*	*TBX3*
13	rs41306688	114,078,558	A/C	0.03	1.82	1.53-2.17	1.08×10^{-11}	1.38	1.10-1.73	5.03×10^{-3}	*ADPRHL1*
15	rs8033459	85,253,258	C/T	0.47	1.21	1.14-1.29	3.41×10^{-9}	1.18	1.09-1.28	5.49×10^{-5}*	*ALPK3*
17	rs28768976	43,688,317	A/G	0.23	1.29	1.20-1.39	4.12×10^{-12}	1.29	1.17-1.42	2.11×10^{-7}*	*SPPL2C*
17	rs7210446	64,307,014	G/A	0.58	1.25	1.16-1.34	6.82×10^{-10}	1.25	1.15-1.35	8.93×10^{-8}*	*PRKCA*
18	rs4799426	34,280,891	A/G	0.35	1.38	1.29-1.47	4.00×10^{-23}	1.44	1.32-1.57	1.13×10^{-16}*	*FHOD3*
18	rs118060942	34,280,732	C/T	0.01	1.79	1.45-2.20	2.35×10^{-8}	2.70	1.80-4.04	1.49×10^{-6}*	*FHOD3*
22	rs2070458	24,159,307	T/A	0.22	1.34	1.25-1.44	7.12×10^{-15}	1.25	1.12-1.38	2.81×10^{-5}*	*MMP11*

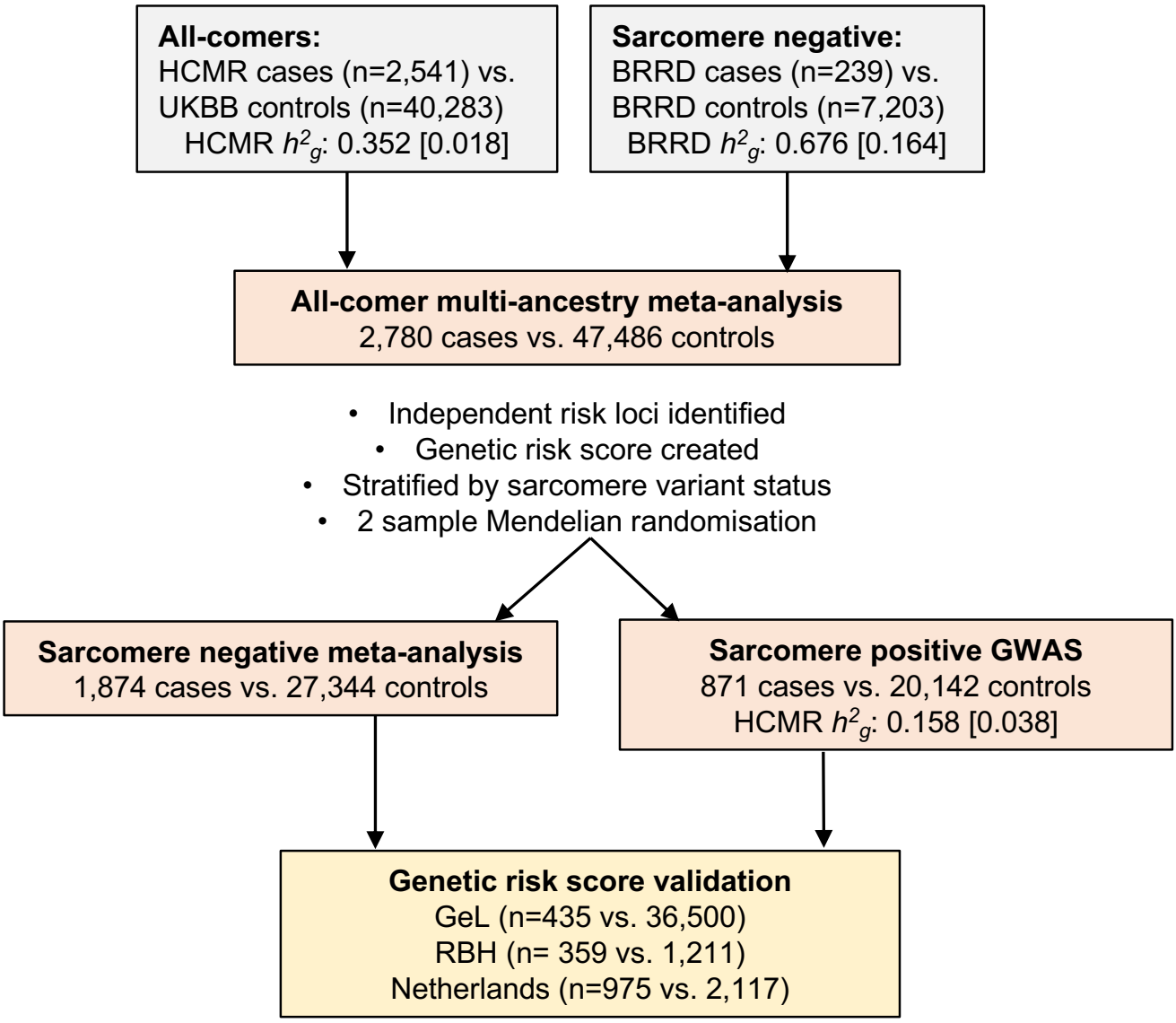
Table 1 | Variants independently associated with hypertrophic cardiomyopathy beneath the genome-wide significant threshold. Fixed-effects inverse-variance meta-analysis was performed for two multi-ancestry genome-wide association studies (Hypertrophic Cardiomyopathy Registry (HCMR) cases (*n* = 2,541) vs. UK Biobank (*n* = 40,283)) and BRRD (BioResource for Rare Disease hypertrophic cardiomyopathy cases (*n* = 239) vs. controls (*n* = 7,203)) using 8,590,397 SNPs. Variants conferring independent risk effects were identified through a stepwise model selection procedure. Other than *FHOD3*, all loci are novel. Most loci do not overlap with the myofilament genes known to carry rare variants causal for hypertrophic cardiomyopathy, but instead highlight important homeostatic pathways. 10 of the 12 genome-wide significant variants replicate. Replication was performed in non-overlapping cases from the Netherlands, Canada and Royal Brompton discovery series (*n* = 1,643 cases, 6,628 controls) with significant loci (alpha threshold = 0.05/29 (total number of independent variants beneath 5% FDR threshold, i.e. $P < 1.72 \times 10^{-3}$) highlighted (*). No significant heterogeneity was detected between the discovery and replication studies ($P > 0.05$) (Supplementary Table 3). Abbreviations: CI, confidence interval; EA, effect allele; FDR, false discovery rate; Freq EA, effect allele frequency; GWAS, genome-wide association study; NEA, non-effect allele; OR, odds ratio; SNP, single nucleotide polymorphism. *P* values are uncorrected for multiple testing.

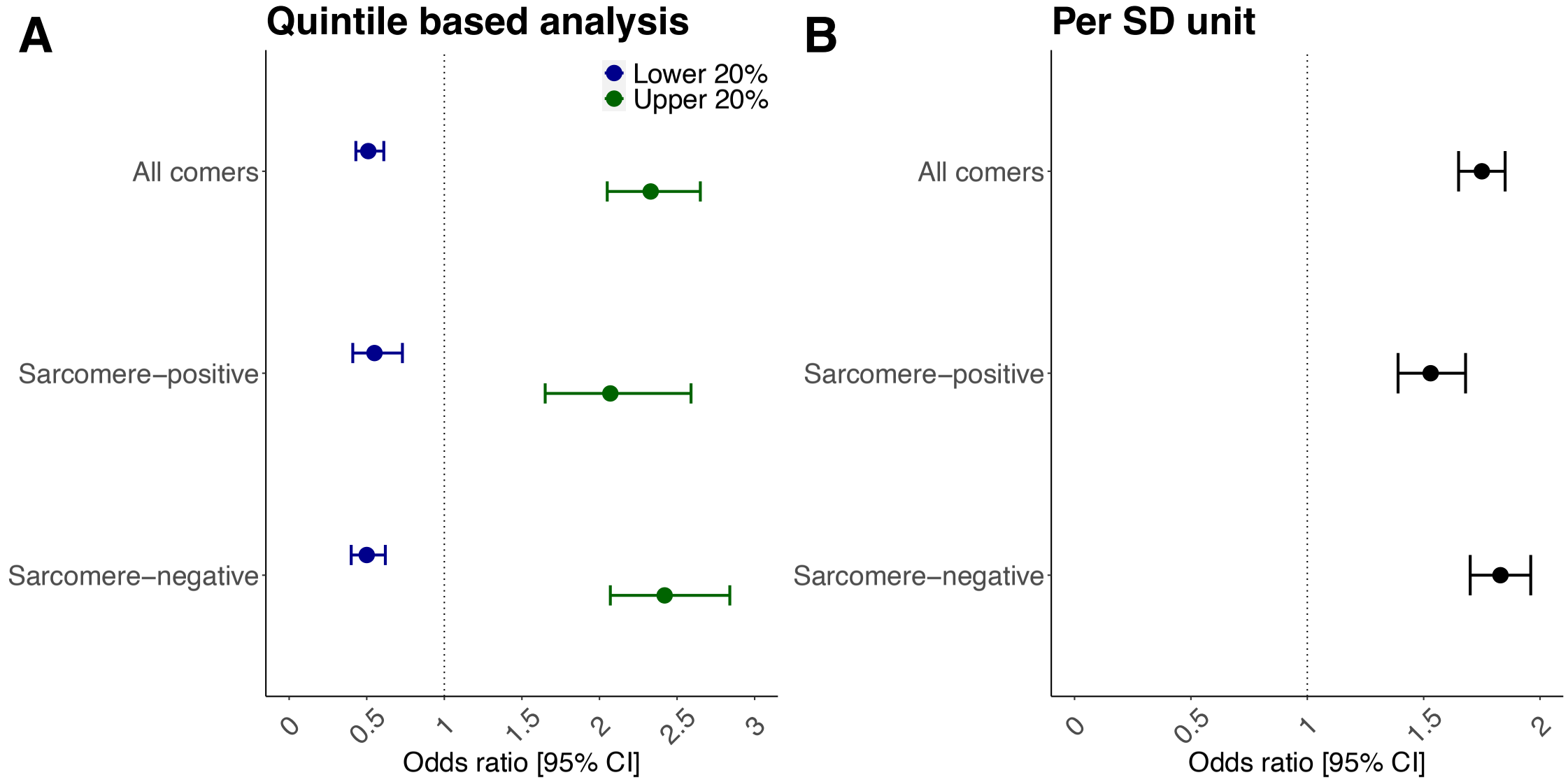
Chr	SNP	NEA/ EA	Beta	Discovery *P*	Replication *P*	Locus name
10	rs72840788	A/G	-0.421	5.06×10^{-29}	4.90×10^{-13}*	*BAG3*
18	rs4799426	G/A	-0.321	4.00×10^{-23}	1.13×10^{-16}*	*FHOD3*
1	rs1048302	G/T	0.277	2.54×10^{-17}	1.06×10^{-7}*	*HSPB7*
22	rs2070458	T/A	0.293	7.12×10^{-15}	2.81×10^{-5}*	*MMP11*
13	rs41306688	C/A	-0.601	1.08×10^{-11}	5.03×10^{-3}	*ADPRHL1*
6	rs3176326	A/G	-0.247	2.22×10^{-11}	1.87×10^{-6}*	*CDKN1A*
6	rs12212795	C/G	-0.393	2.51×10^{-10}	8.19×10^{-10}*	*PLN*
17	rs7210446	A/G	-0.22	6.82×10^{-10}	8.93×10^{-8}*	*PRKCA*
15	rs8033459	T/C	-0.19	3.41×10^{-9}	5.49×10^{-5}*	*ALPK3*
3	rs13061705	T/C	0.224	9.18×10^{-9}	4.49×10^{-3}	*SLC6A6*
12	rs7301677	T/C	0.214	1.26×10^{-8}	2.74×10^{-4}*	*TBX3*
18	rs118060942	T/C	-0.799	2.35×10^{-8}	1.49×10^{-6}*	*FHOD3*
6	rs9320939	A/G	-0.174	5.78×10^{-8}	4.48×10^{-2}	*TRDN*
5	rs10052399	C/T	0.206	6.21×10^{-8}	4.42×10^{-6}*	*PROB1*
10	rs11196085	C/T	-0.19	7.30×10^{-8}	1.99×10^{-6}*	*TCF7L2*
2	rs2003585	C/T	0.174	8.60×10^{-8}	1.12×10^{-2}	*STRN*
9	rs734638	G/C	-0.186	1.09×10^{-7}	3.39×10^{-1}	*RAPGEF1, POMT1*
5	rs66761011	G/A	-0.349	1.45×10^{-7}	5.18×10^{-1}	*AK098570*
15	rs1814880	C/T	0.179	1.59×10^{-7}	4.84×10^{-3}	*CHRNB4*
12	rs1480036	C/T	-0.211	3.50×10^{-7}	2.60×10^{-1}	*SSPN*
3	rs4894803	G/A	-0.179	3.51×10^{-7}	9.91×10^{-6}*	*FNDC3B*
2	rs7556984	A/G	0.186	5.21×10^{-7}	4.94×10^{-2}	*E2F6, ROCK2*
2	rs62177303	T/C	0.175	7.00×10^{-7}	7.79×10^{-3}	*TTN*
19	rs117710064	T/C	-0.222	8.55×10^{-7}	9.36×10^{-1}	*AZU1*
21	rs2832230	T/G	0.251	9.57×10^{-7}	2.49×10^{-3}	*MAP3K7CL*
8	rs7003871	T/C	0.169	1.14×10^{-6}	4.21×10^{-2}	*MTSS1*
11	rs1390519	G/A	-0.203	1.16×10^{-6}	2.44×10^{-2}	*CYP2R1*

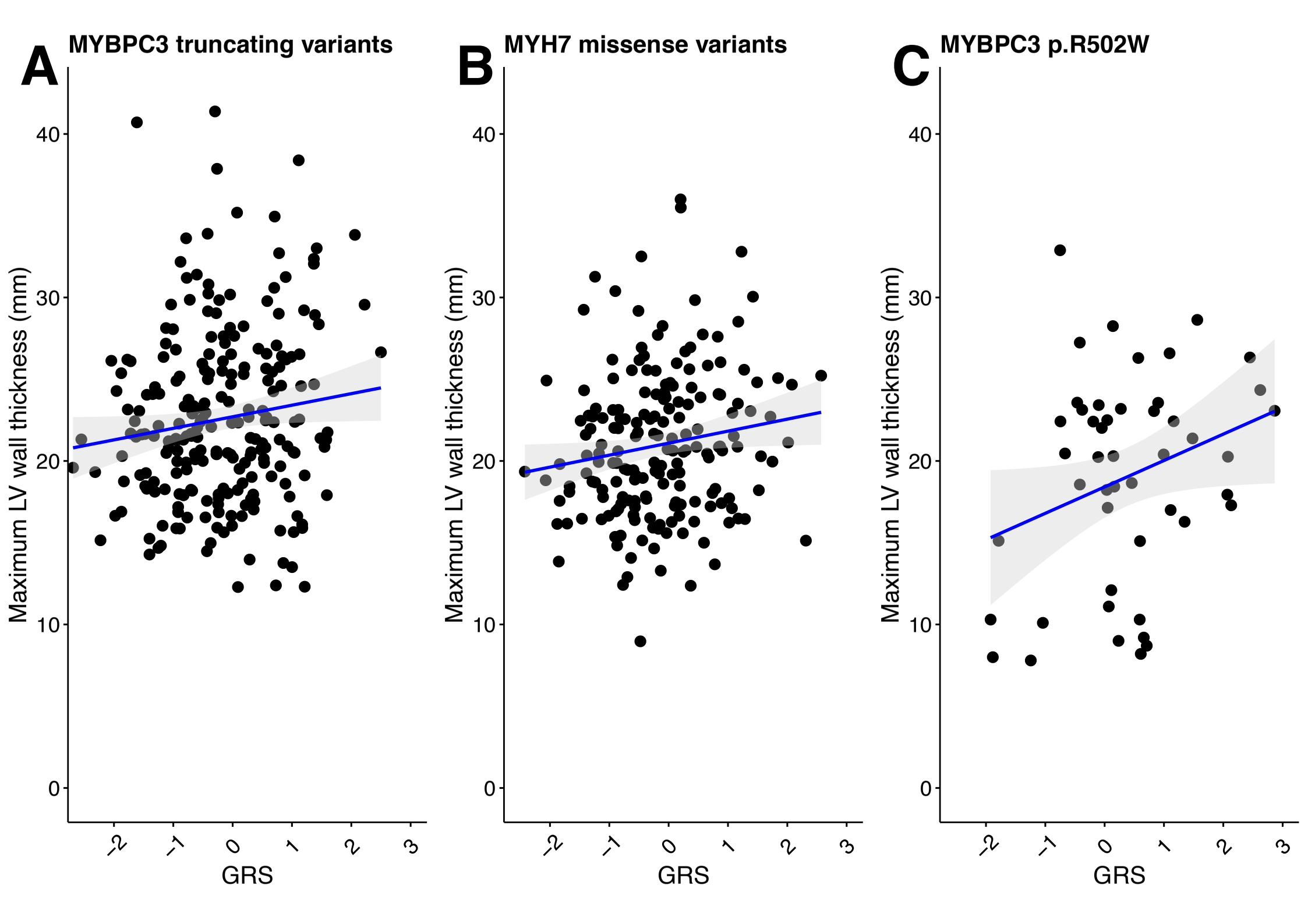
Table 2 | Variants independently associated with hypertrophic cardiomyopathy beneath the 5% FDR threshold used in a genetic risk score instrument. Beta estimates derived from a fixed-effects inverse-variance meta-analysis that incorporated two multi-ancestry genome-wide association studies (Hypertrophic Cardiomyopathy Registry (HCMR) cases (*n* = 2,541) vs. UK Biobank (*n* = 40,283)) and BRRD (BioResource for Rare Disease hypertrophic cardiomyopathy cases (*n* = 239) vs. controls (*n* = 7,203)) using 8,590,397 SNPs. Variants conferring independent

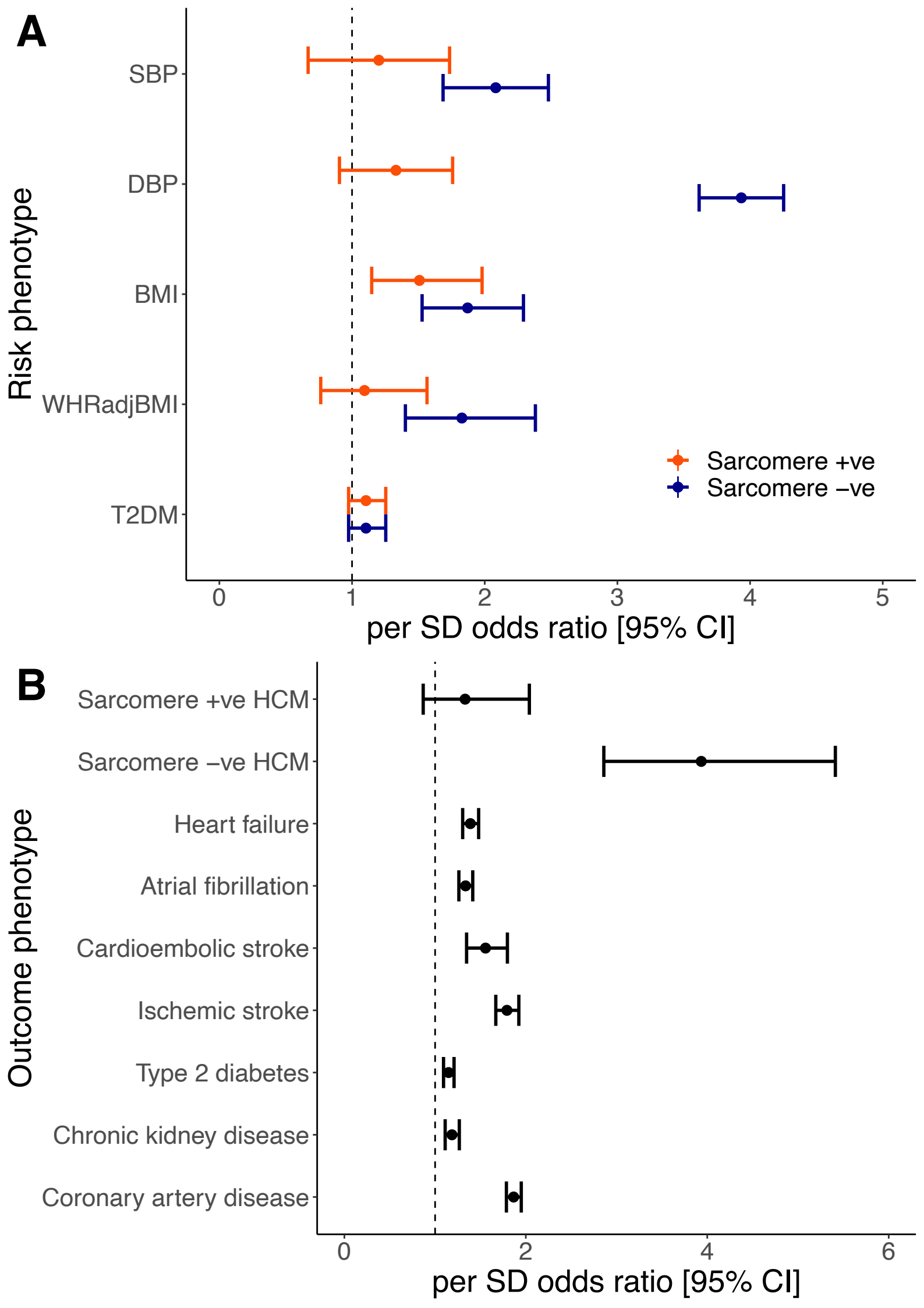
risk effects beneath a 5% FDR threshold ($P < 1.82 \times 10^{-6}$) were identified through a stepwise model selection procedure. Abbreviations: Chr, chromosome; EA, effect allele; FDR, false discovery rate; NEA, non-effect allele; SNP, single nucleotide polymorphism. *denotes independent replication at Bonferroni-corrected significance level ($P < 1.72 \times 10^{-3}$), as described in Table 1. *P* values are uncorrected for multiple testing.

866


All-comers:
HCMR cases (n=2,541) vs.
UKBB controls (n=40,283)
HCMR h^2_g: 0.352 [0.018]
Sarcomere negative:
BRRD cases (n=239) vs.
BRRD controls (n=7,203)
BRRD h^2_g: 0.676 [0.164]
All-comer multi-ancestry meta-analysis
2,780 cases vs. 47,486 controls
• Independent risk loci identified
• Genetic risk score created
• Stratified by sarcomere variant status
• 2 sample Mendelian randomisation
Sarcomere negative meta-analysis
1,874 cases vs. 27,344 controls
Sarcomere positive GWAS
871 cases vs. 20,142 controls
HCMR h^2_g: 0.158 [0.038]
Genetic risk score validation
GeL (n=435 vs. 36,500)
RBH (n= 359 vs. 1,211)
Netherlands (n=975 vs. 2,117)


A
Quintile based analysis
Lower 20%
Upper 20%
All comers
Sarcomere−positive
Sarcomere−negative
0
0.5
1
1.5
2
2.5
3
Odds ratio [95% CI]
B
Per SD unit
All comers
Sarcomere−positive
Sarcomere−negative
0
0.5
1
1.5
2
Odds ratio [95% CI]


A
MYBPC3 truncating variants
B
MYH7 missense variants
C
MYBPC3 p.R502W
Maximum LV wall thickness (mm)
0
10
20
30
40
GRS
-2
-1
0
1
2
3


A
SBP
DBP
BMI
WHRadjBMI
T2DM
Risk phenotype
Sarcomere +ve
Sarcomere –ve
0
1
2
3
4
5
per SD odds ratio [95% CI]
B
Sarcomere +ve HCM
Sarcomere –ve HCM
Heart failure
Atrial fibrillation
Cardioembolic stroke
Ischemic stroke
Type 2 diabetes
Chronic kidney disease
Coronary artery disease
Outcome phenotype
0
2
4
6
per SD odds ratio [95% CI]