

Large-scale genome-wide association analyses identify novel genetic loci and mechanisms in hypertrophic cardiomyopathy

In the format provided by the
authors and unedited

Supplementary Figures and Note

Tadros, et al. Large scale genome-wide association analyses identify novel genetic loci and mechanisms in hypertrophic cardiomyopathy

[Supplementary Figure 1](#): HCM GWAS forest plots

[Supplementary Figure 2](#): Manhattan plots of HCM GWAS for all cases (A) and stratified analyses (B)

[Supplementary Figure 3](#): Manhattan plot for GWAS of global left ventricular longitudinal strain

[Supplementary Figure 4](#): Manhattan plot for GWAS of global left ventricular radial strain

[Supplementary Figure 5](#): Manhattan plot for GWAS of global left ventricular circumferential strain

[Supplementary Figure 6](#): Manhattan plot for GWAS of left ventricular ejection fraction

[Supplementary Figure 7](#): Manhattan plot for GWAS of left ventricular end-systolic volume

[Supplementary Figure 8](#): Manhattan plot for GWAS of left ventricular end-diastolic volume

[Supplementary Figure 9](#): Manhattan plot for GWAS of left ventricular mass

[Supplementary Figure 10](#): Manhattan plot for GWAS of mean left ventricular wall thickness

[Supplementary Figure 11](#): Manhattan plot for GWAS of maximal left ventricular wall thickness

[Supplementary Figure 12](#): Manhattan plot for GWAS of left ventricular concentricity index

[Supplementary Figure 13](#): Comparison of regression coefficients derived from HCM GWAS and MTAG

[Supplementary Figure 14](#): HCM MTAG forest plots

[Supplementary Figure 15](#): Enrichment of 68 cell-state (A) and 9 cell-type (B) gene programs

[Supplementary Figure 16](#): Cohort level quantile-quantile plots for exome-wide burden testing

[Supplementary Figure 17](#): Meta-analysis quantile-quantile plots for exome-wide burden testing

[Supplementary Figure 18](#): Meta-analysis Manhattan plots for exome-wide burden testing

[Supplementary Figure 19](#): Forest plot of rare *SVIL* loss-of-function variants association with HCM

[Supplementary Figure 20](#): SNP effects of contractility on risk of HCM, oHCM and nHCM

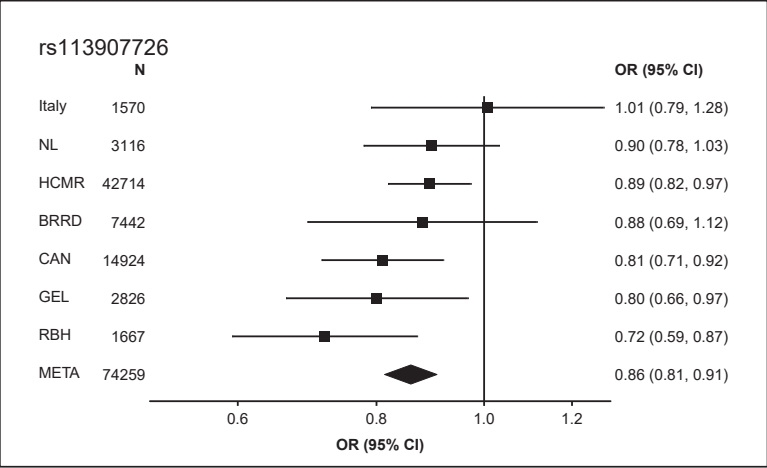
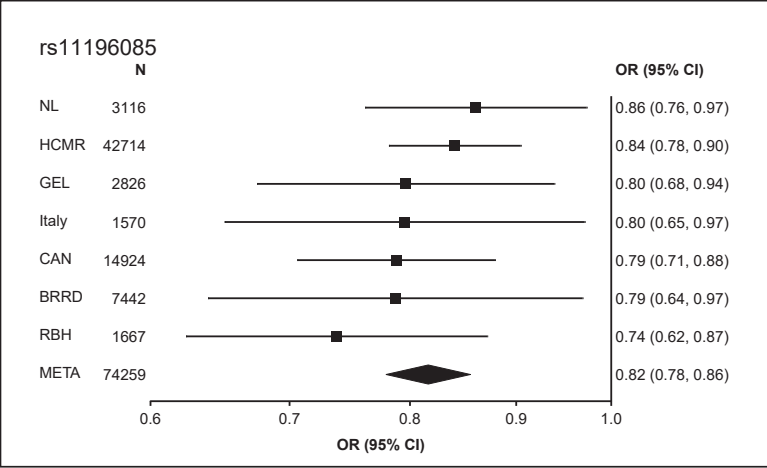
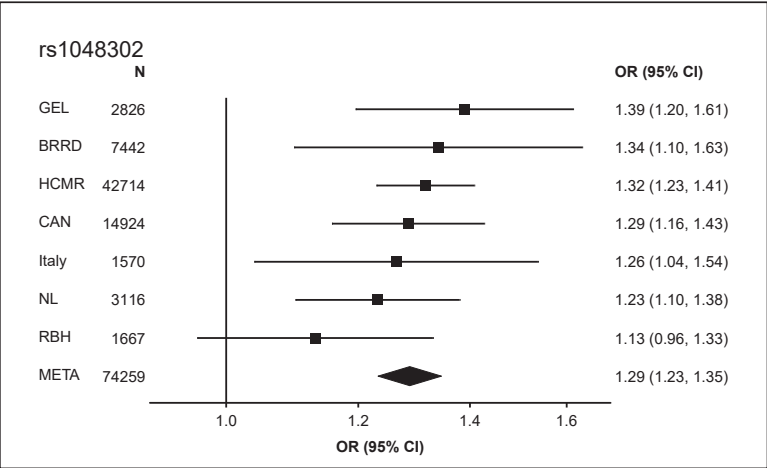
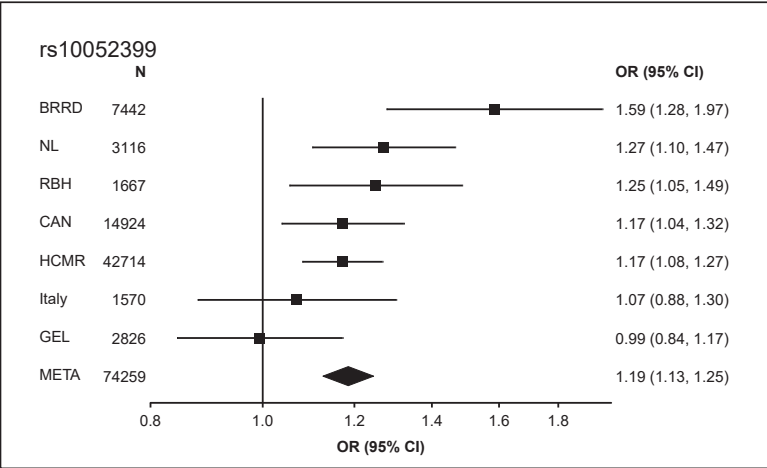
[Supplementary Figure 21](#): Leave-one-out analyses for contractility Mendelian randomization

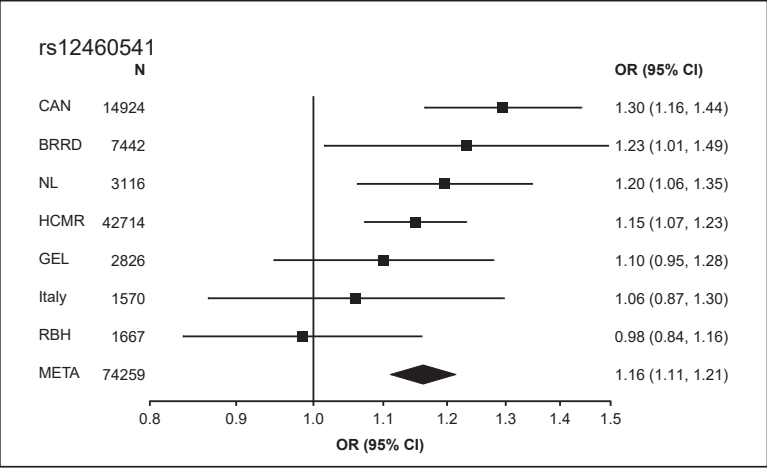
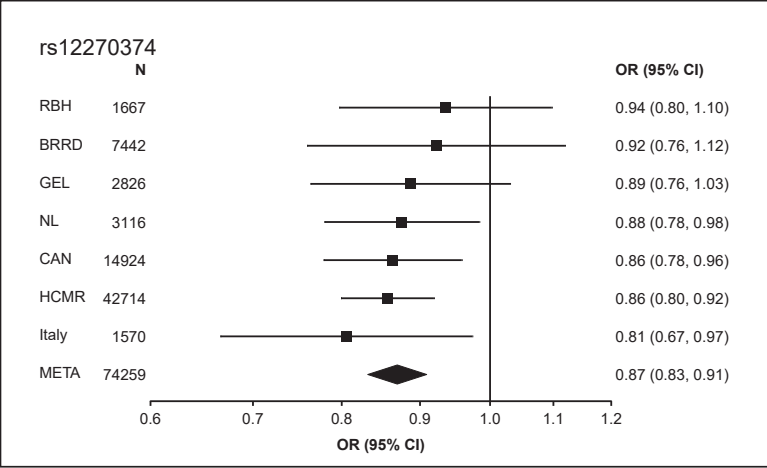
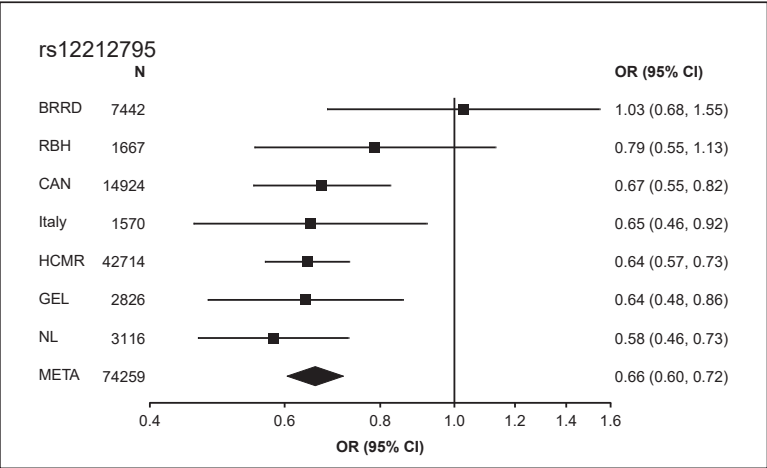
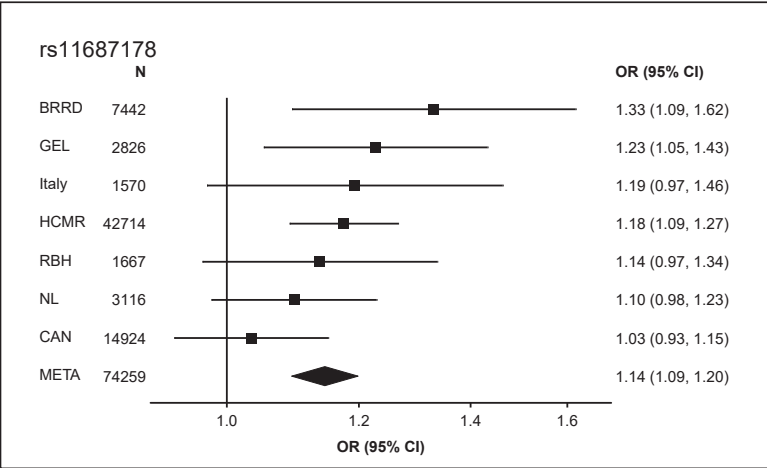
[Supplementary Figure 22](#): SNP effects of blood pressure on risk of HCM, oHCM and nHCM

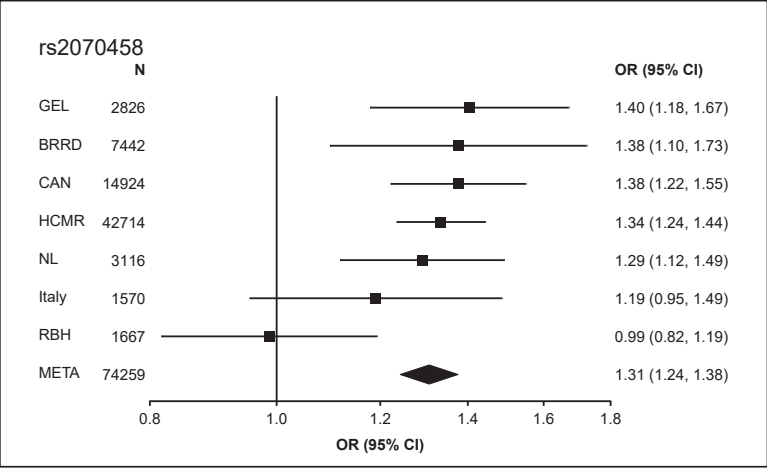
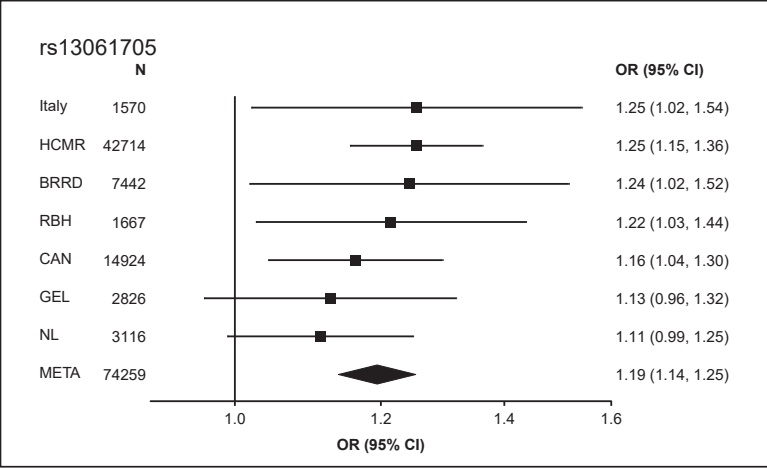
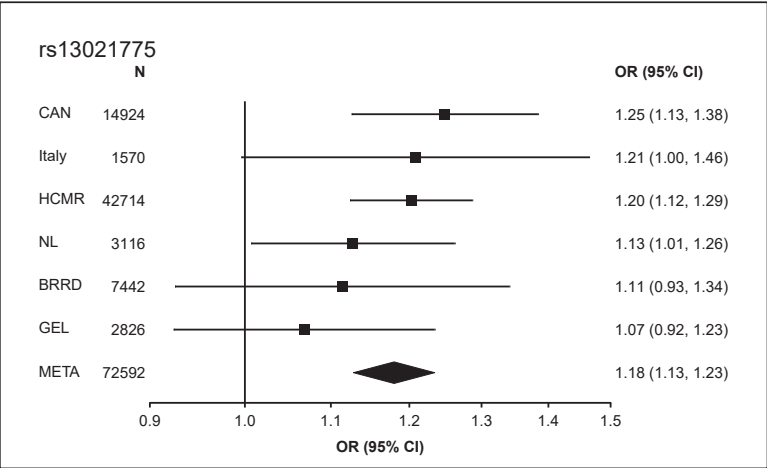
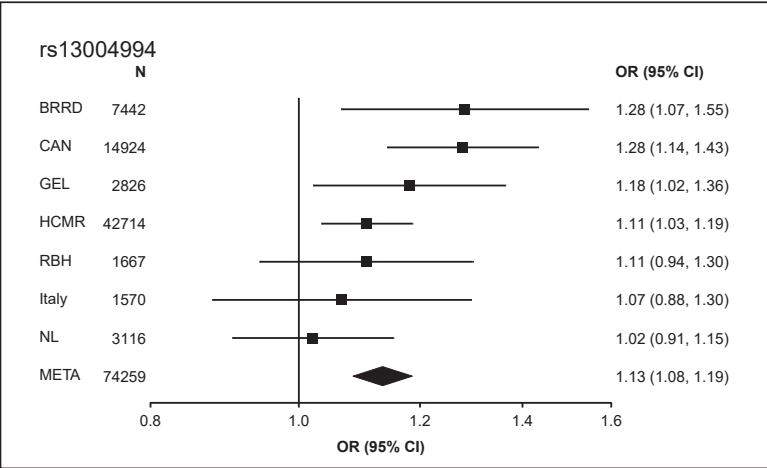
[Supplementary Note](#)

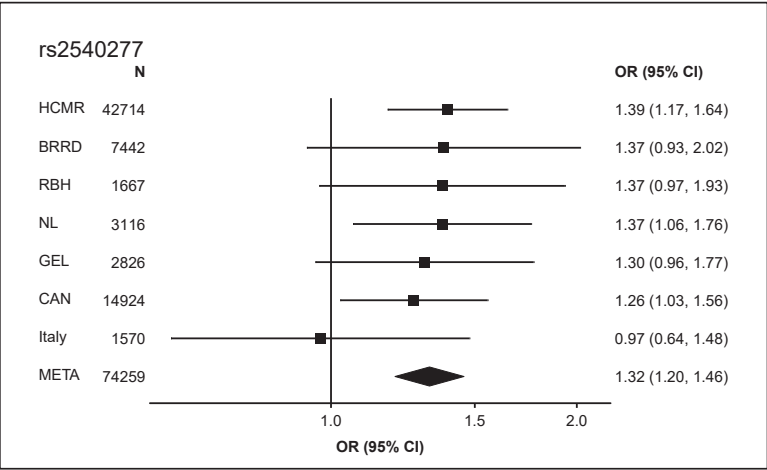
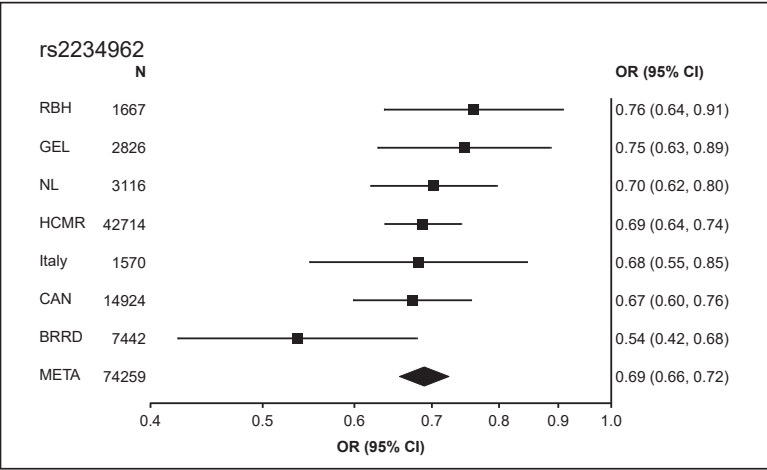
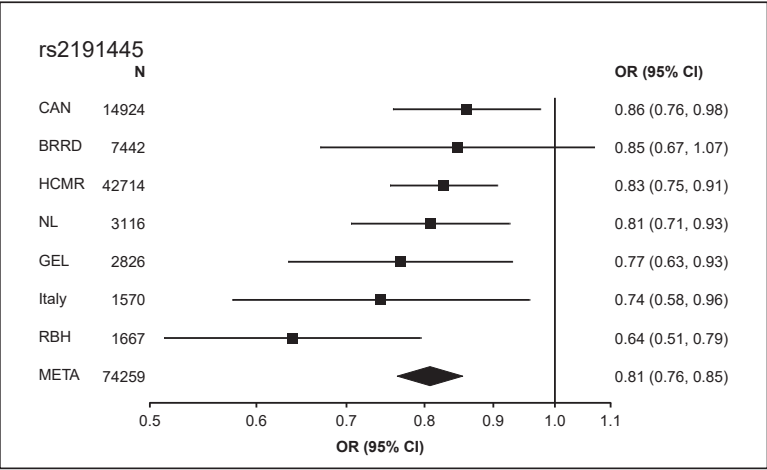
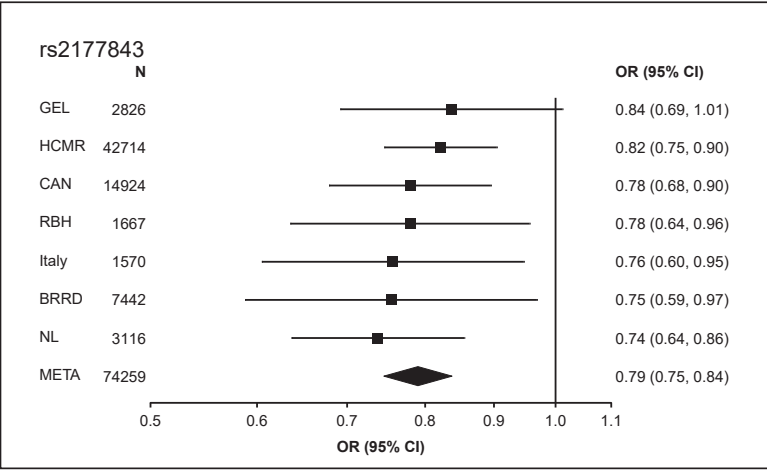
[Supplementary References](#)

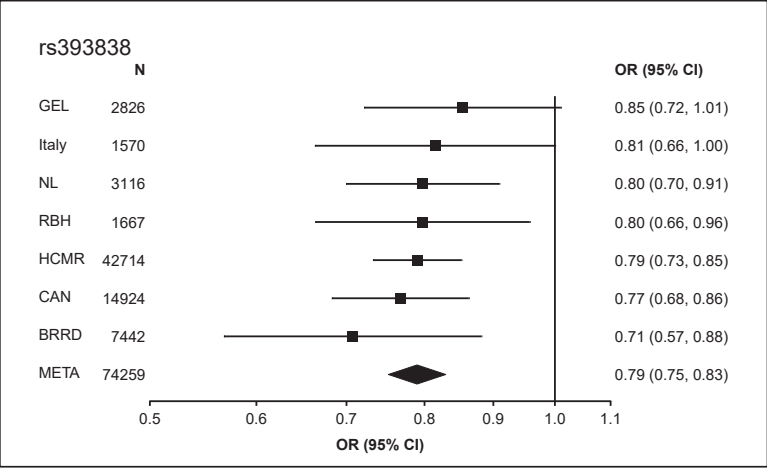
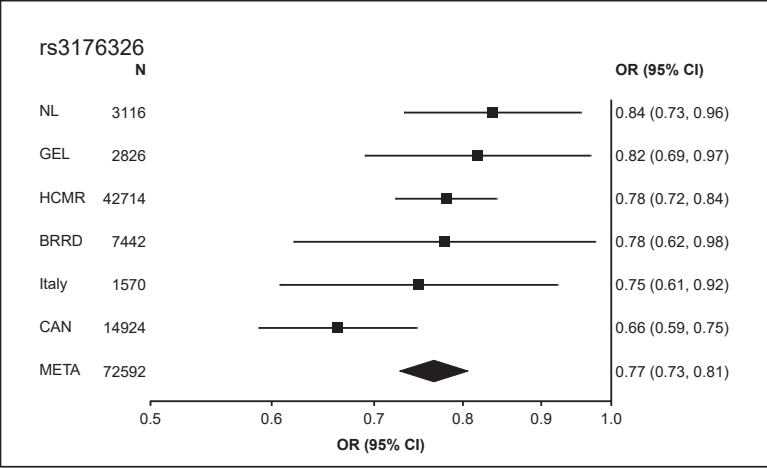
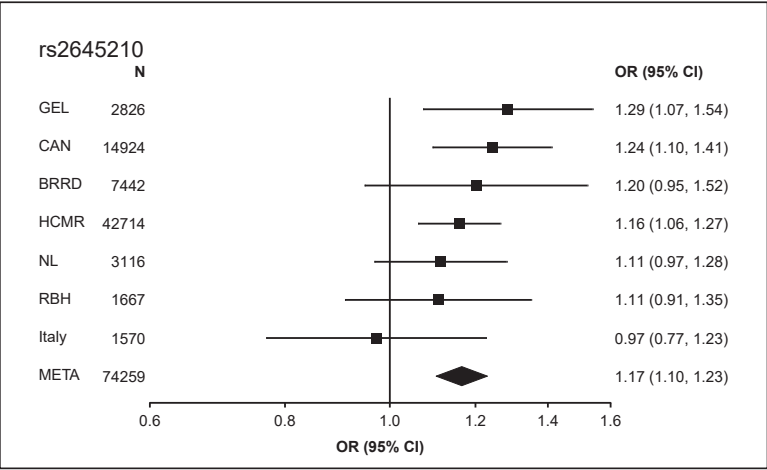
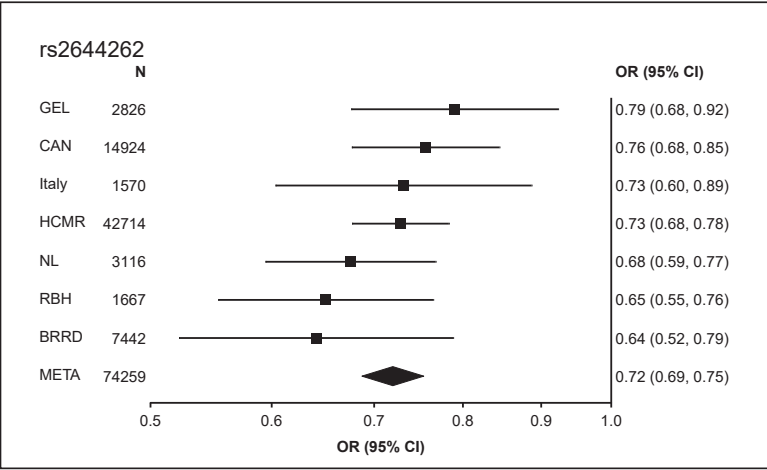
↓ **Supplementary Figure 1:** Forest plots of 34 GWAS lead variants showing the point estimate of the odds ratio (OR) and 95% confidence interval (CI) for each of the 7 case-control GWAS datasets (shown as filled squares and error bars, respectively), as well as OR and 95% CI estimated from the fixed effects model inverse variance weighted meta-analysis (shown as a diamond). BRRD, Bioresource Rare Disease; CAN, Canada; GEL, Genomics England; HCMR, Hypertrophic Cardiomyopathy Registry; META, meta-analysis; N, total sample size (cases + controls); NL, Netherlands; RBH, Royal Brompton Hospital.

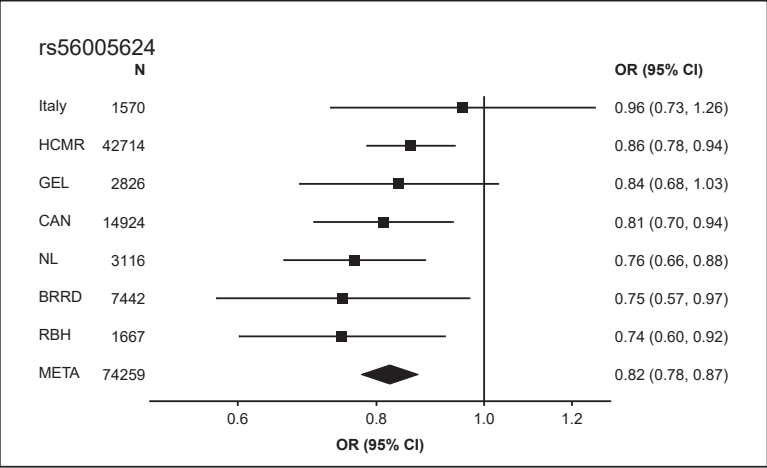
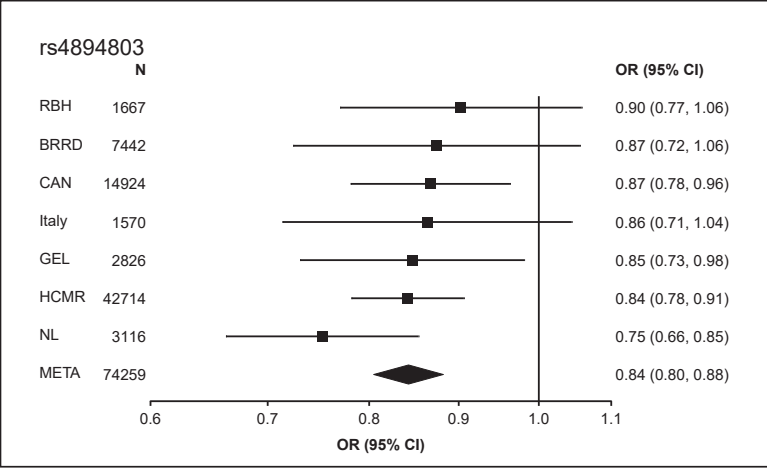
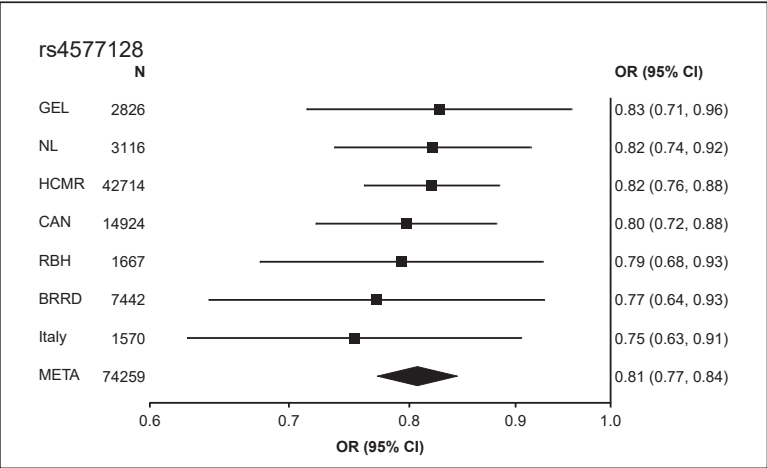
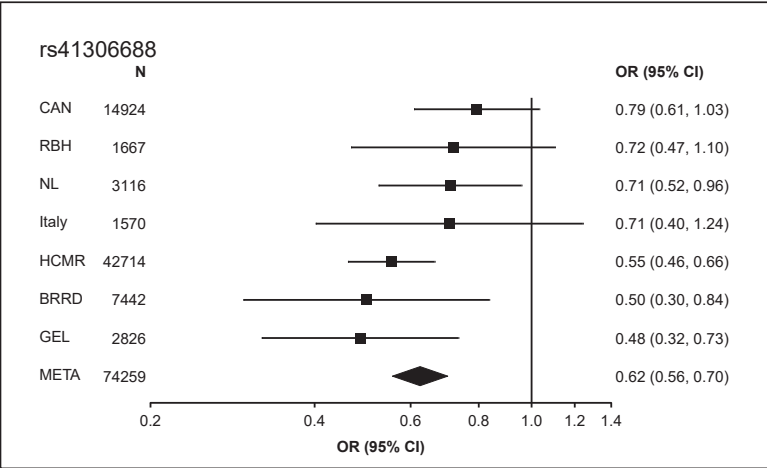


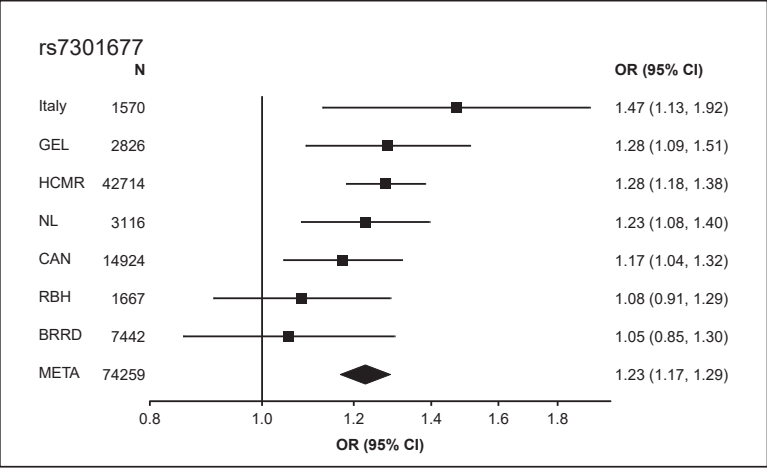
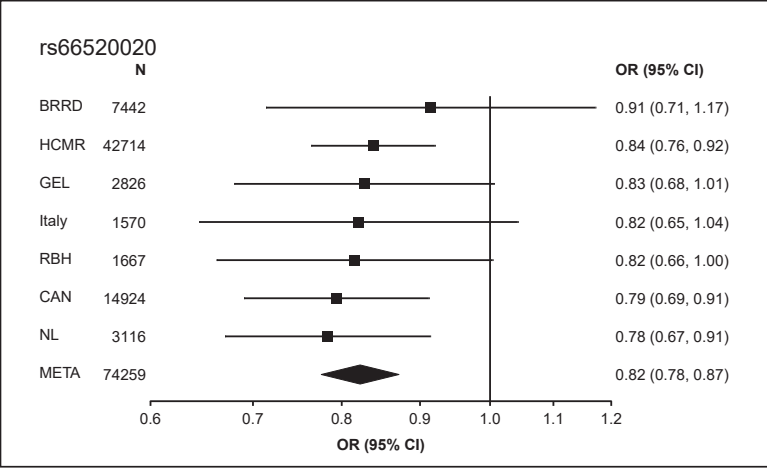
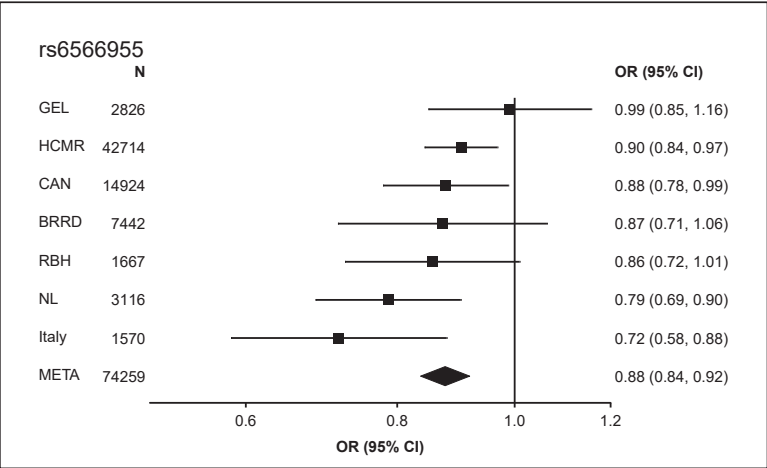
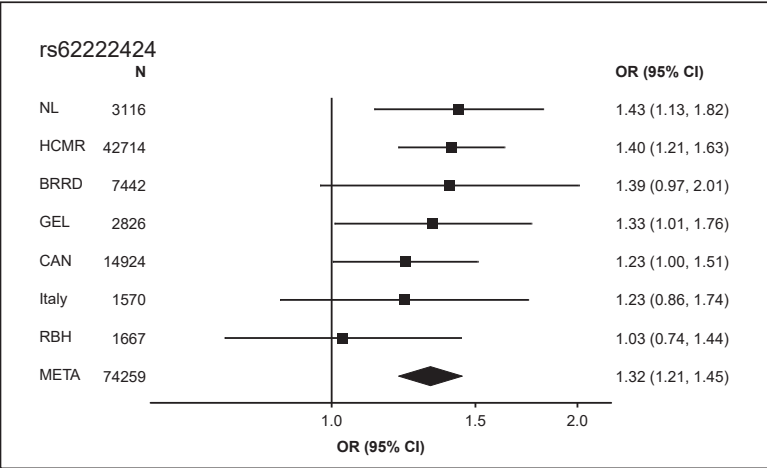


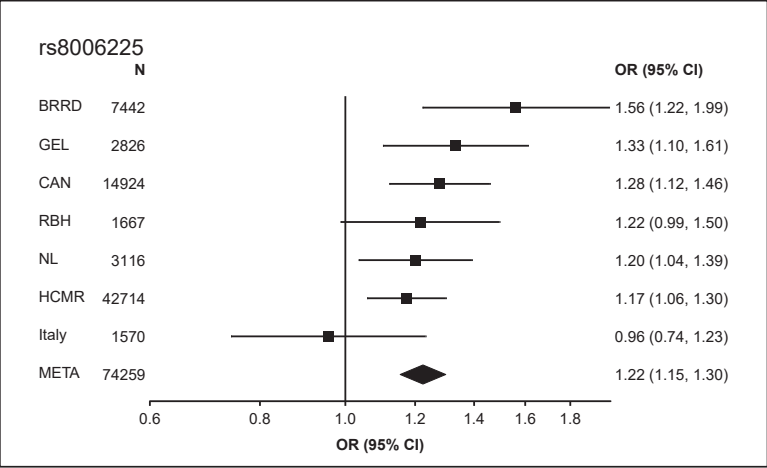
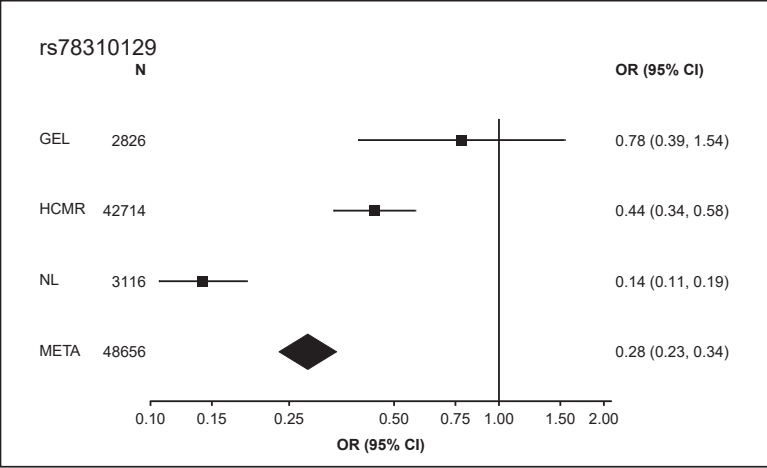
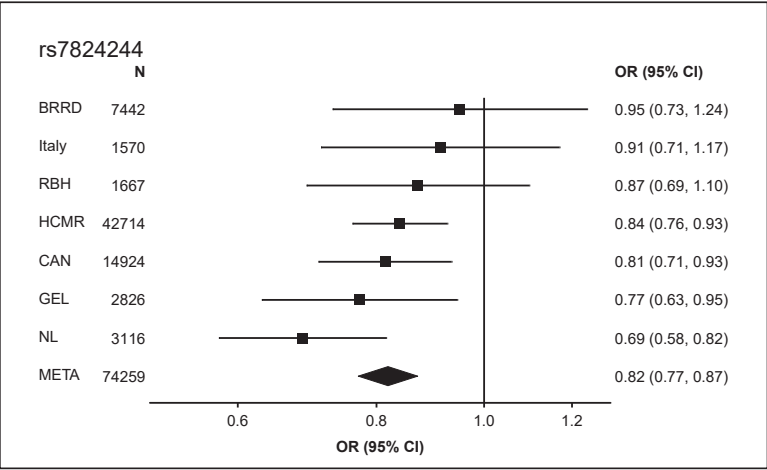
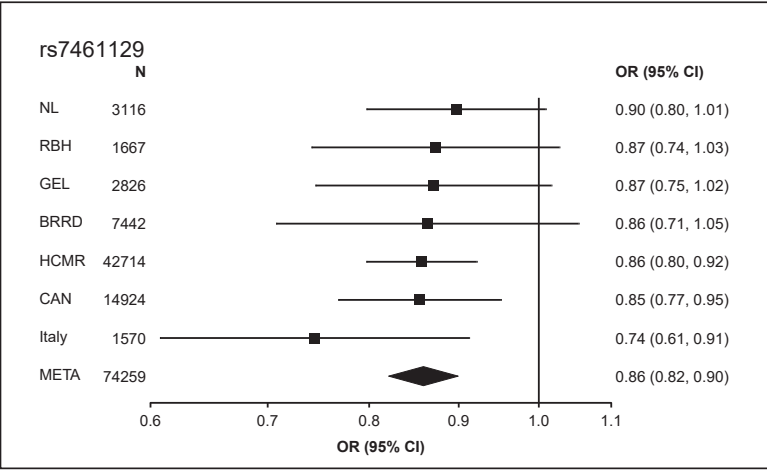


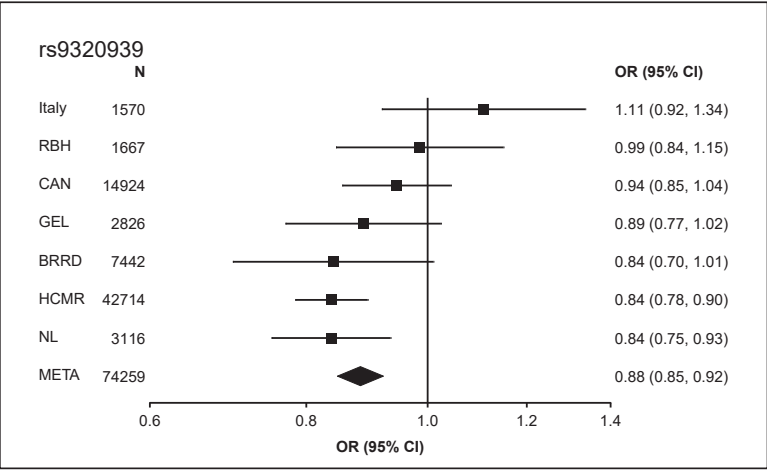
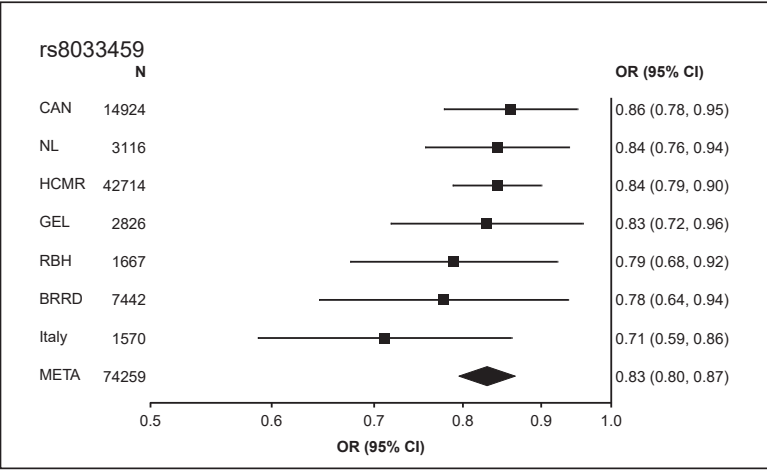


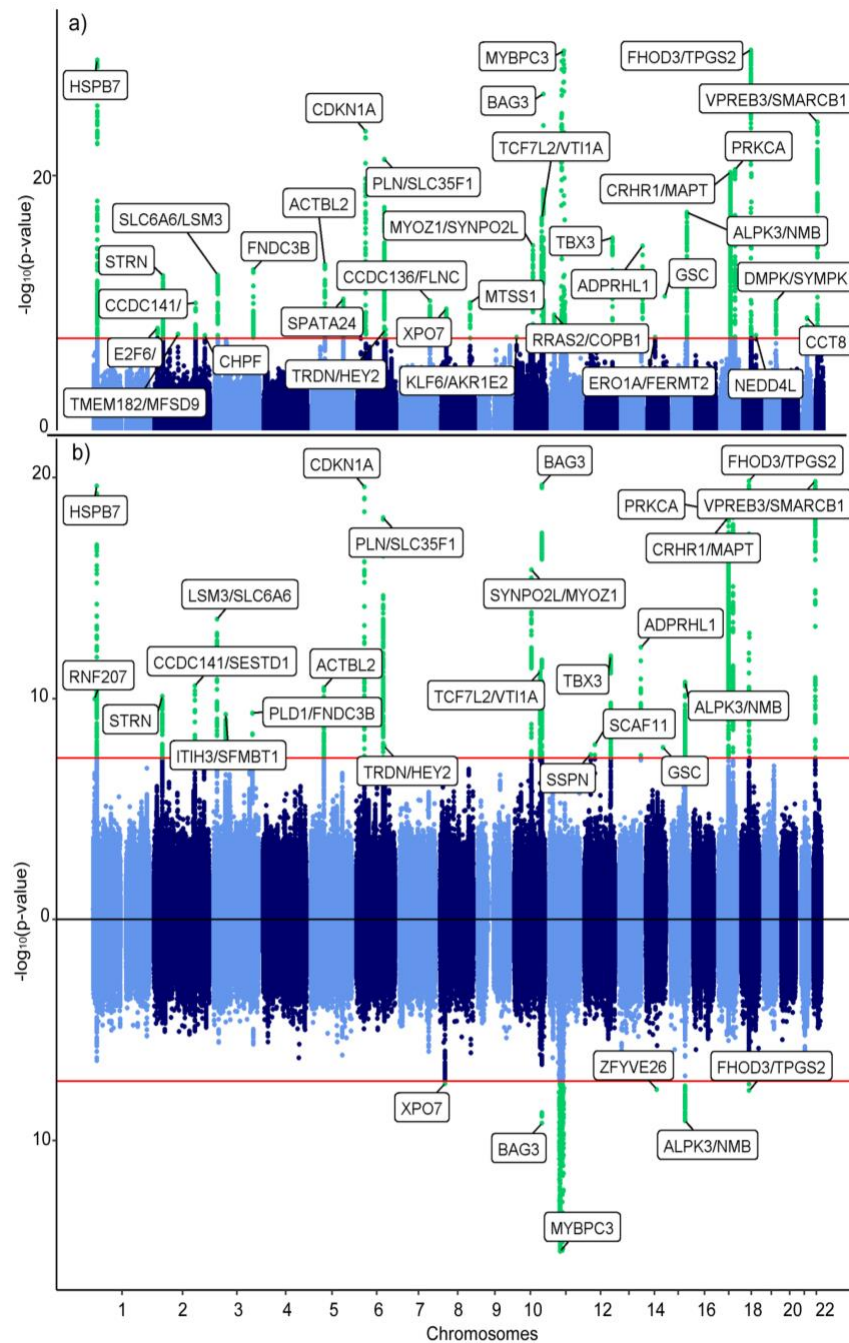




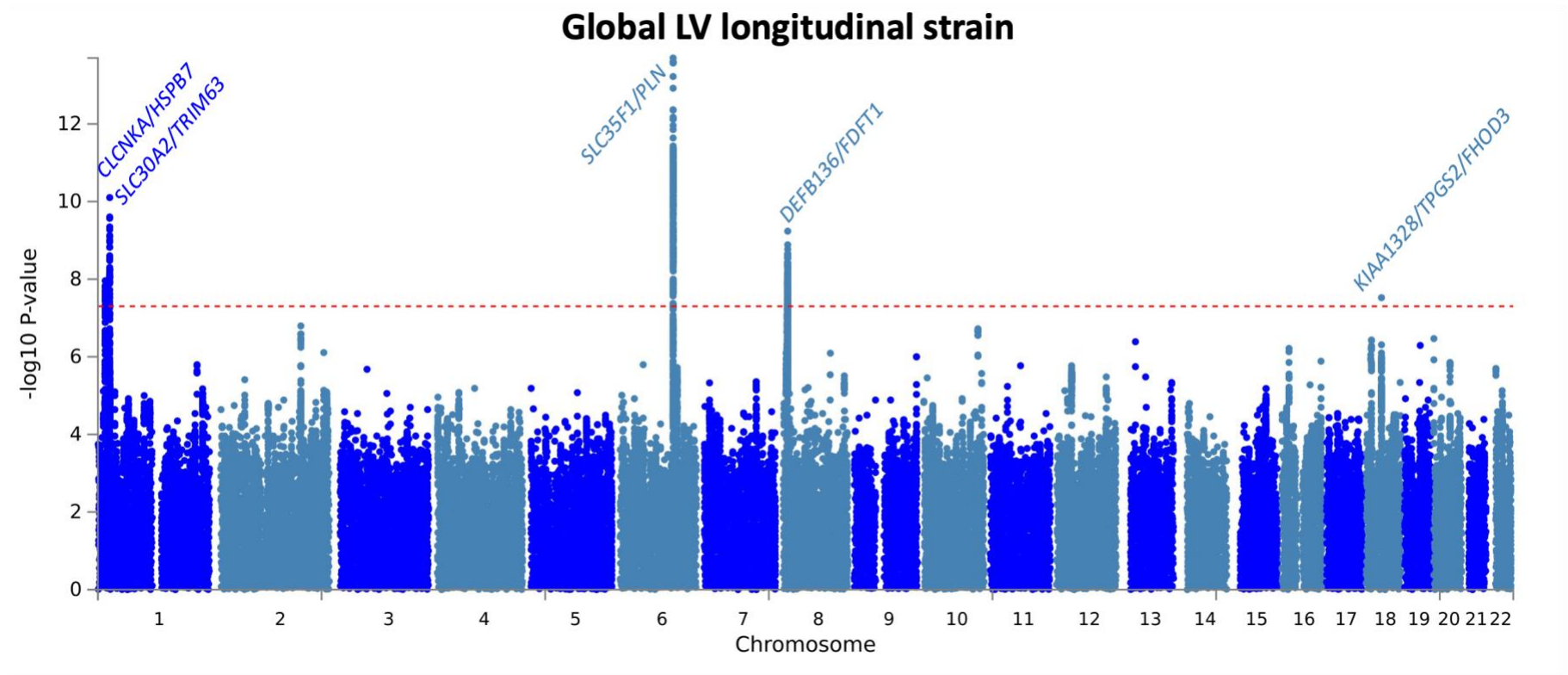




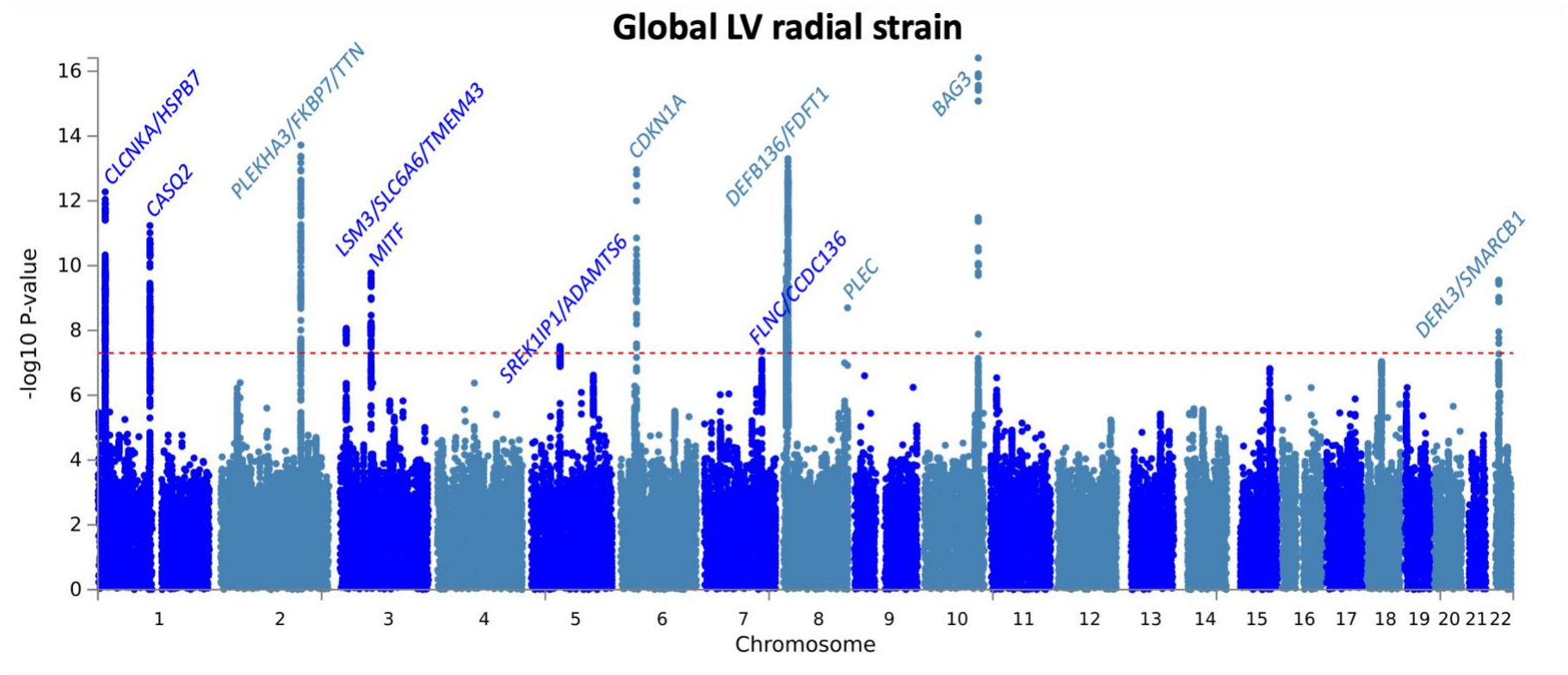




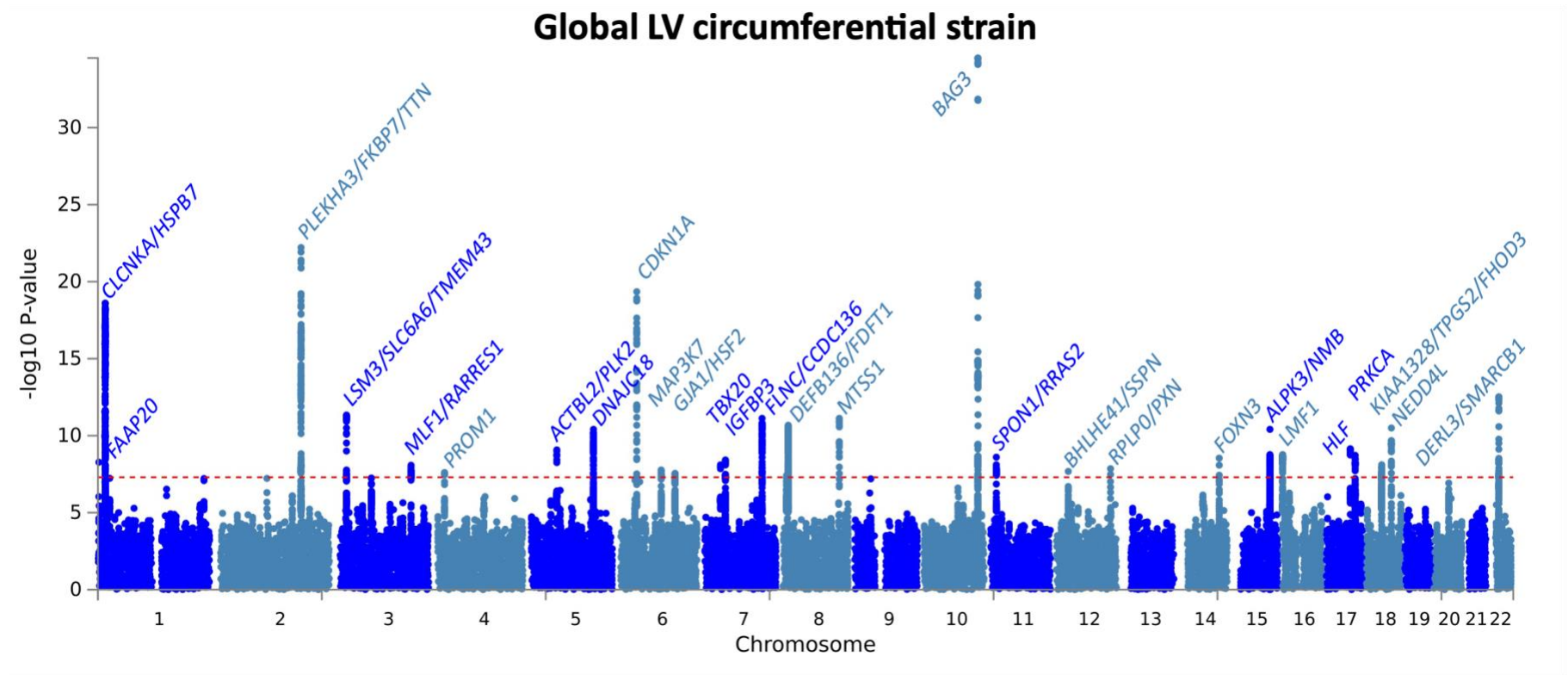
Supplementary Figure 2: Manhattan plots of GWAS of HCM for all comers (A) and sarcomere positive/negative cases (B). Sarcomere negative ($\text{HCM}_{\text{SARC-}}$) summary statistics are on the positive y-axis and sarcomere positive ($\text{HCM}_{\text{SARC+}}$) summary statistics on the negative y-axis in panel B. Red line denotes the significance threshold of $P=5 \times 10^{-8}$. Plots are truncated at $-\log_{10}(\text{p-value})$ of 30 (a), 20 (S-) and 15 (S+). P values correspond to the fixed-effects meta-analysis combining up to 7 case-control HCM GWAS and are not corrected for multiple testing. See Table 1 and Supplementary Table 2 for description of lead variants for the all-comer and stratified analyses ($\text{HCM}_{\text{SARC-}}$ and $\text{HCM}_{\text{SARC+}}$), respectively.



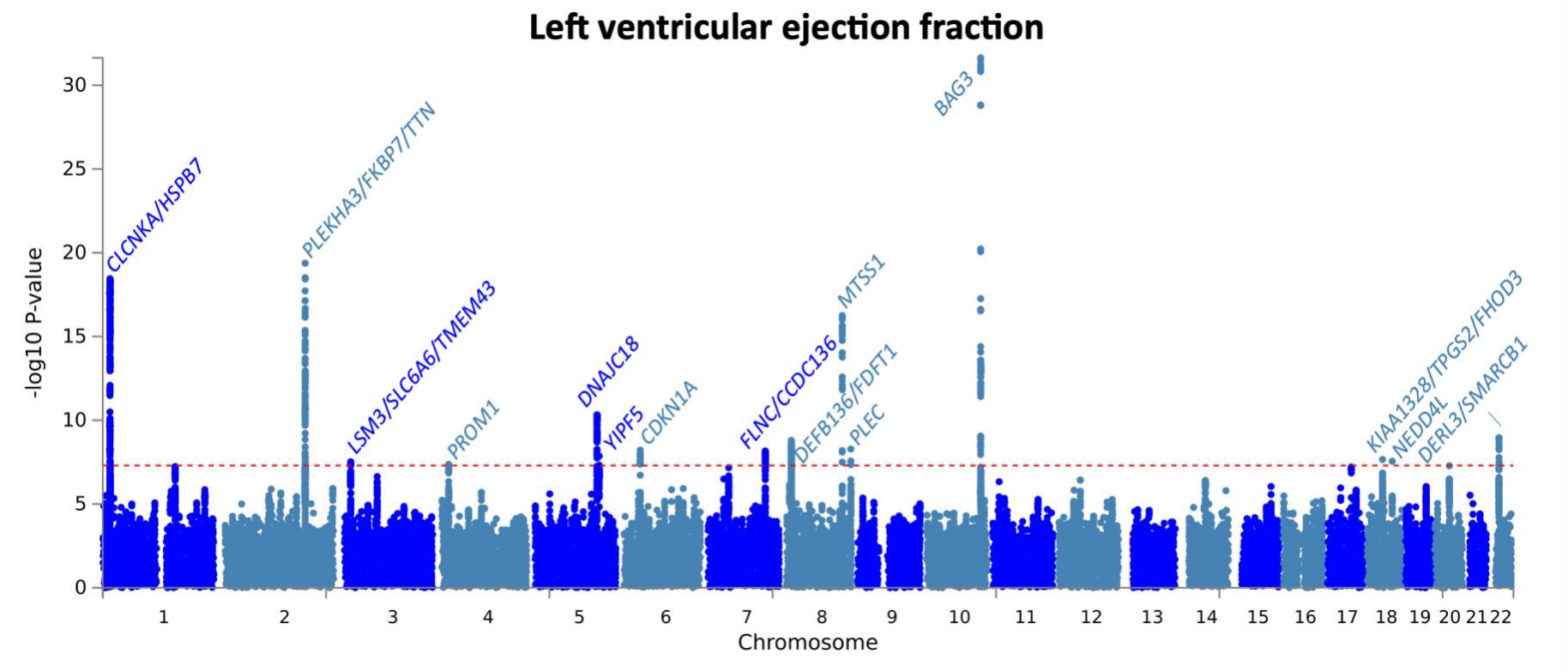
Supplementary Figure 3: Manhattan plot for GWAS of global left ventricular longitudinal strain. Based on an analysis of 35,052 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



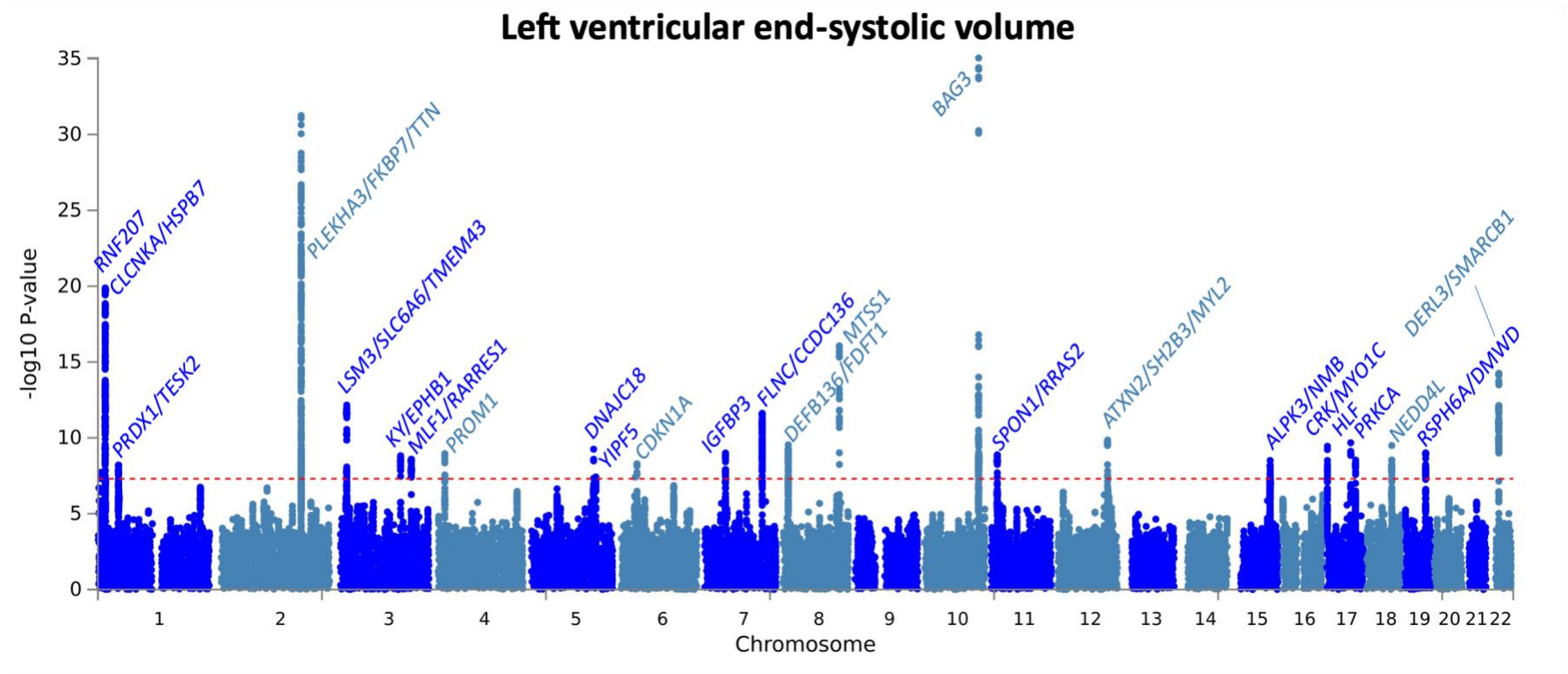
Supplementary Figure 4: Manhattan plot for GWAS of global left ventricular radial strain. Based on an analysis of 36,033 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



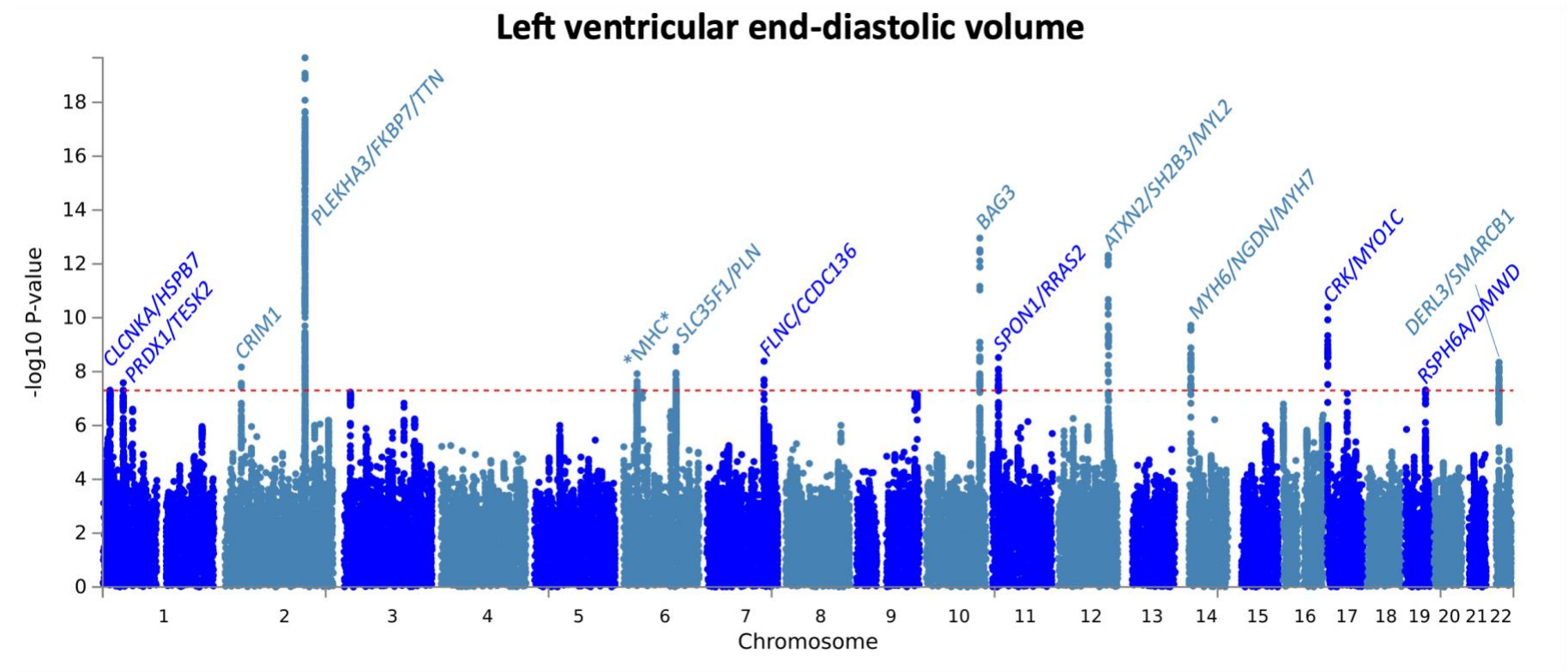
Supplementary Figure 5: Manhattan plot for GWAS of global left ventricular circumferential strain. Based on an analysis of 36,033 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



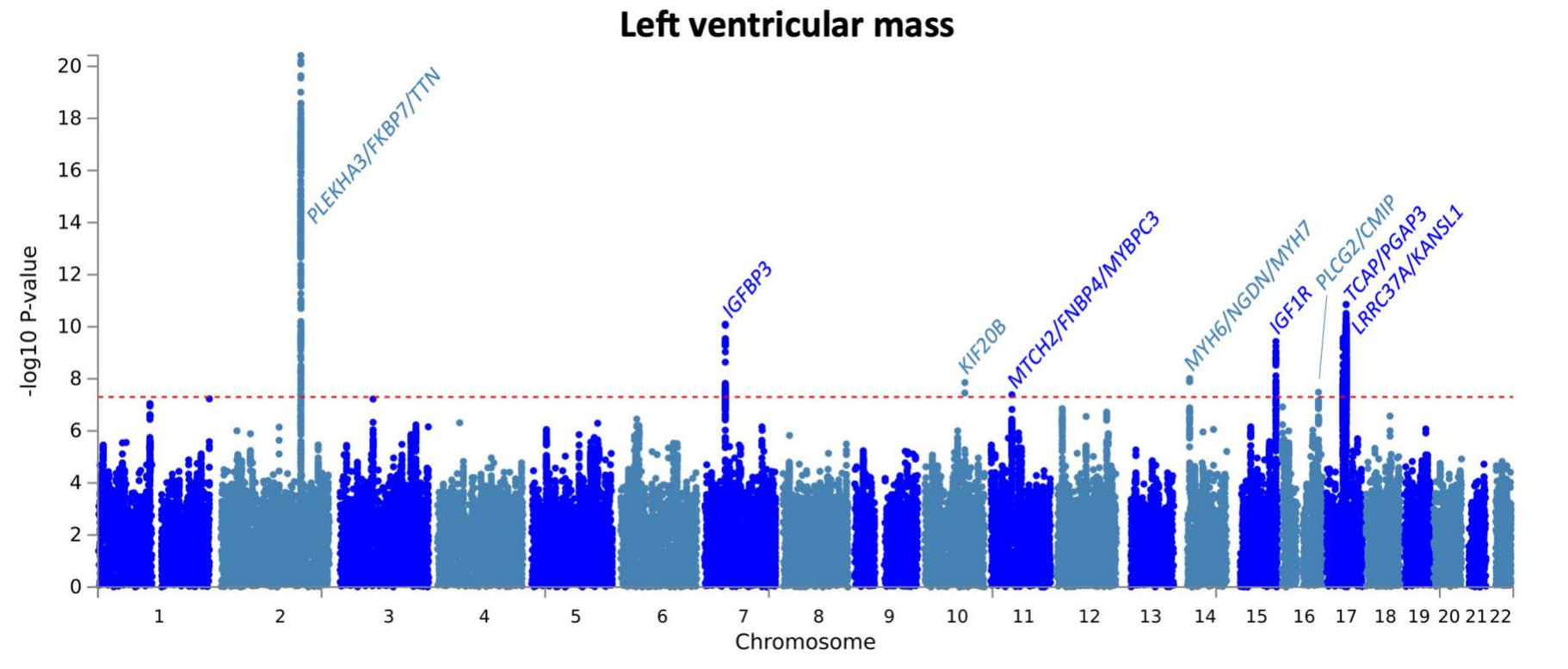
Supplementary Figure 6: Manhattan plot for GWAS of left ventricular ejection fraction. Based on an analysis of 36,083 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



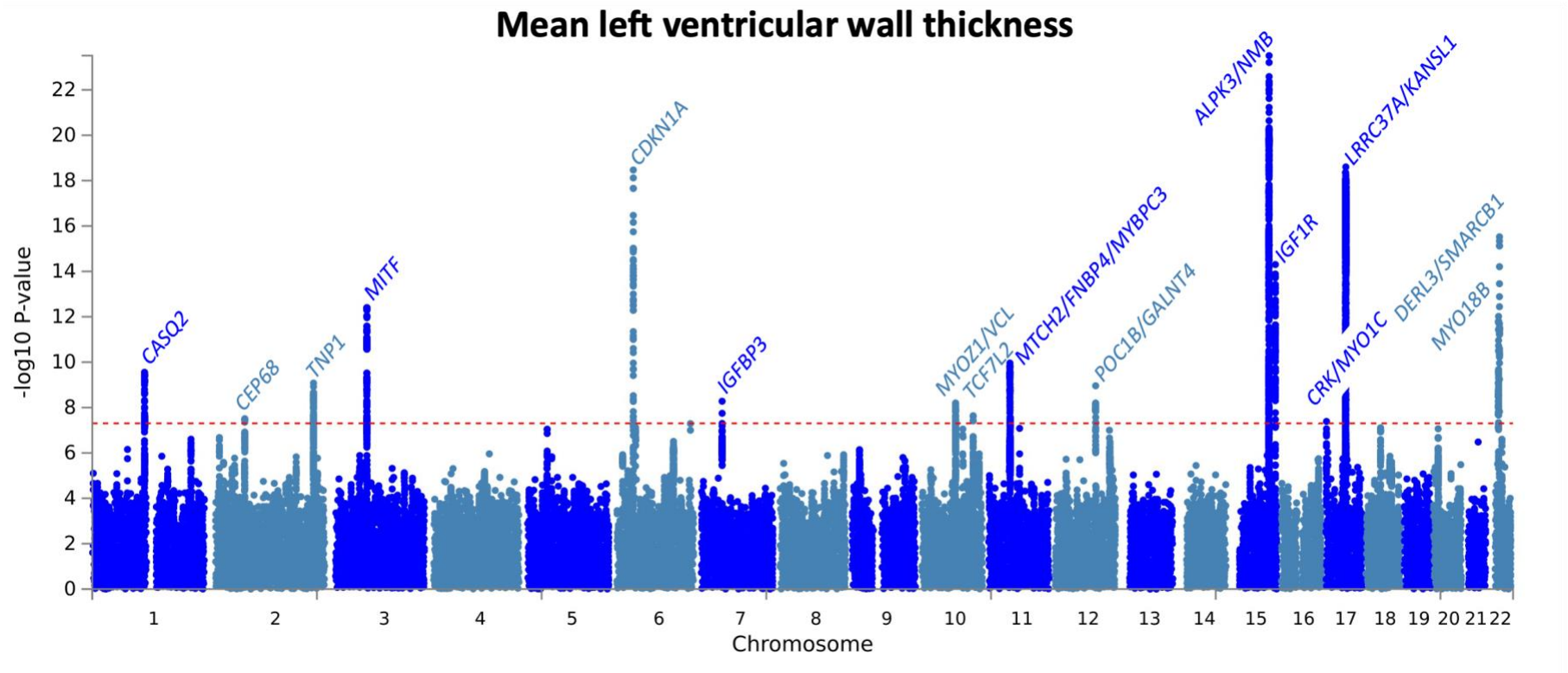
Supplementary Figure 7: Manhattan plot for GWAS of left ventricular end-systolic volume, indexed to the body surface area. GWAS based on an analysis of 36,083 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log$ (P-value) on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



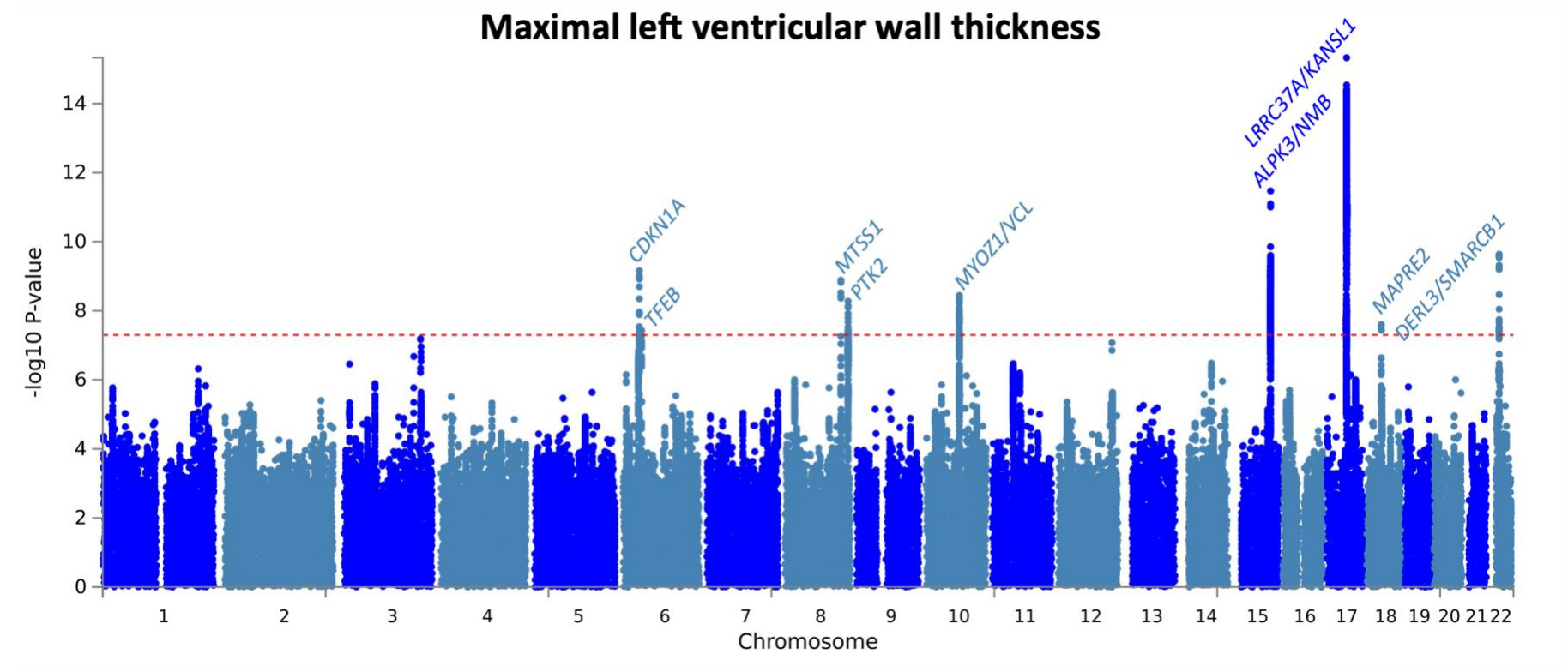
Supplementary Figure 8: Manhattan plot for GWAS of left ventricular end-diastolic volume, indexed to the body surface area. GWAS based on an analysis of 36,083 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. *MHC* refers to the major histocompatibility locus on chromosome 6. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



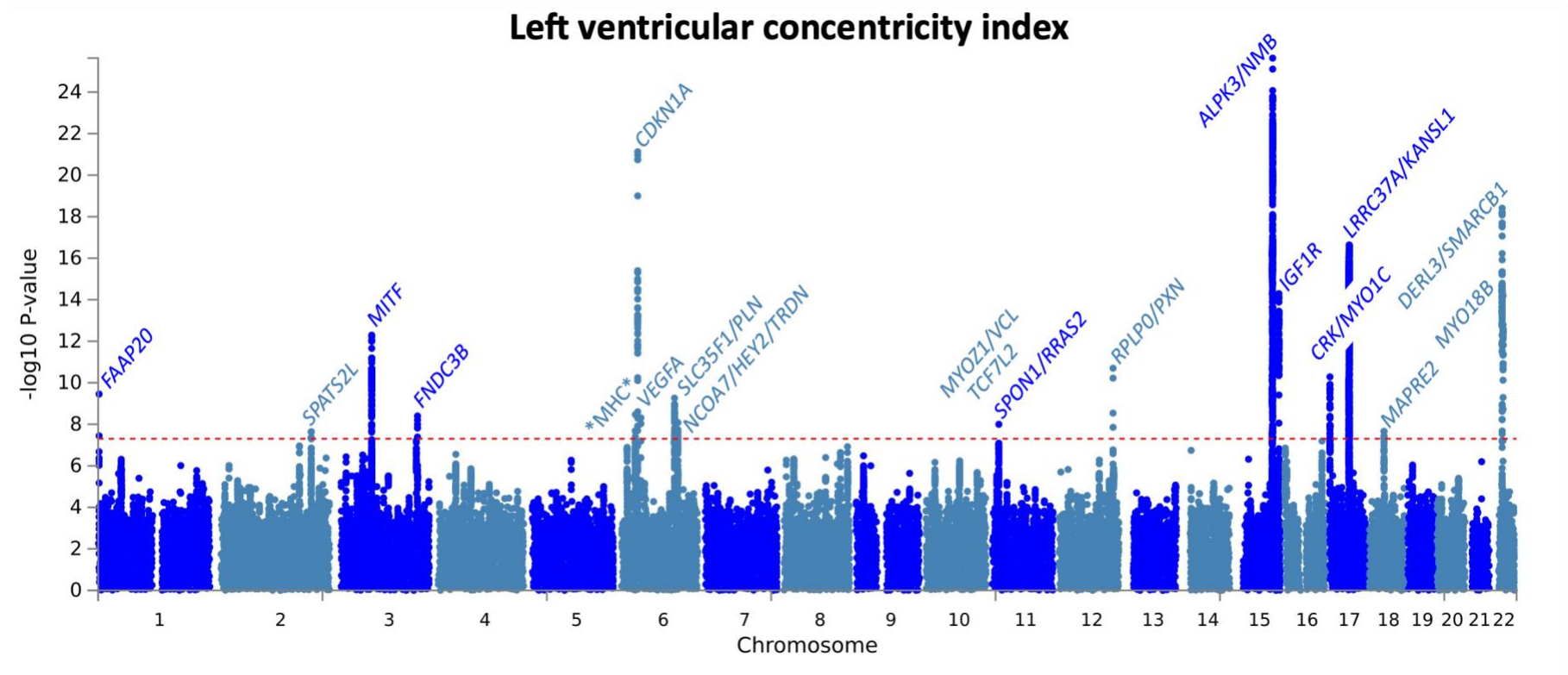
Supplementary Figure 9: Manhattan plot for GWAS of left ventricular mass, indexed to the body surface area. GWAS based on an analysis of 36,083 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



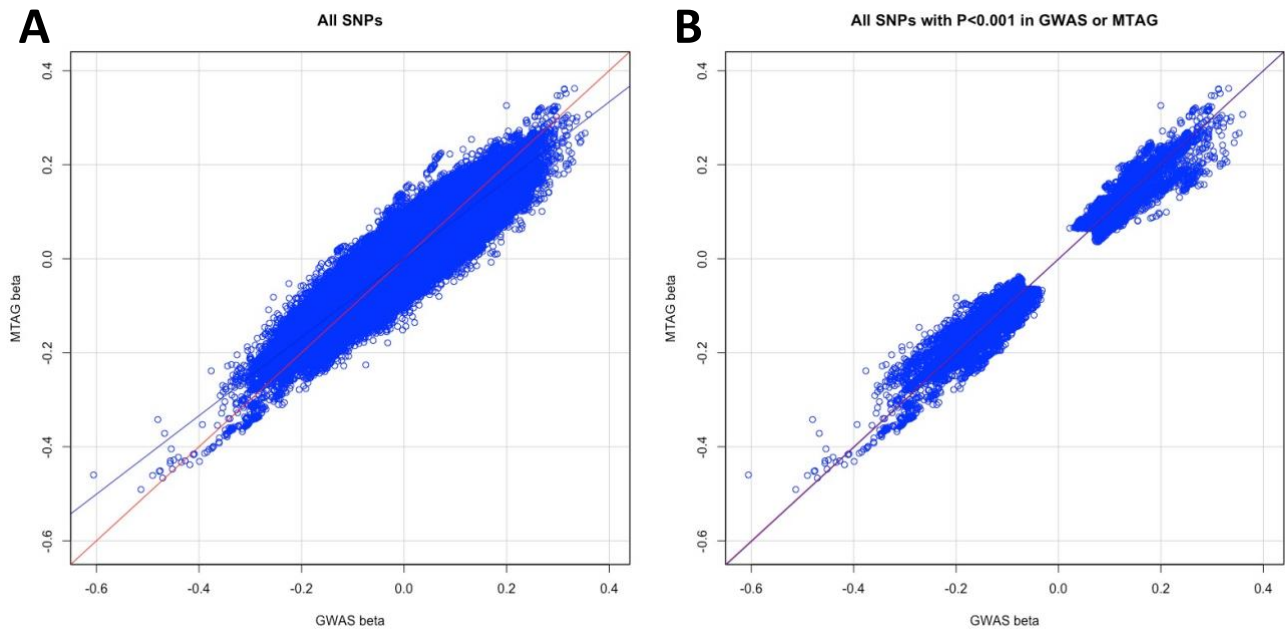
Supplementary Figure 10: Manhattan plot for GWAS of mean left ventricular wall thickness. Based on an analysis of 36,056 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.



Supplementary Figure 11: Manhattan plot for GWAS of maximal left ventricular wall thickness. Based on an analysis of 36,056 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. See Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.

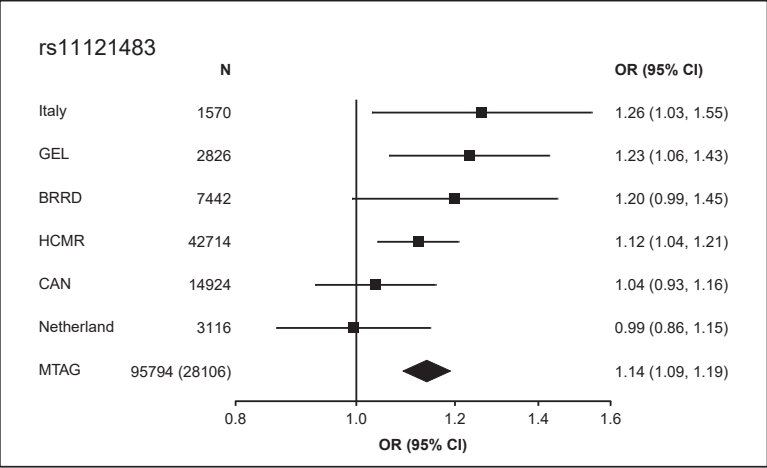
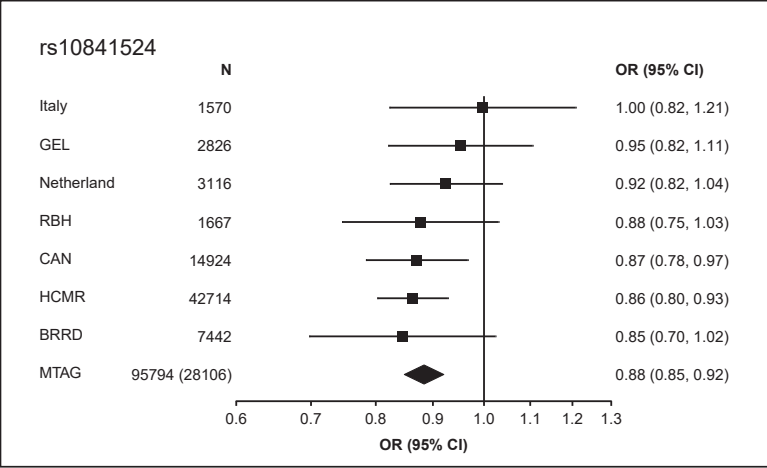
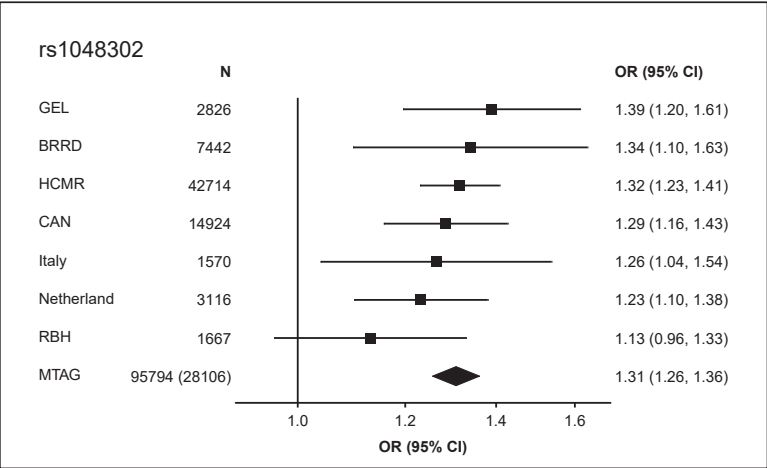
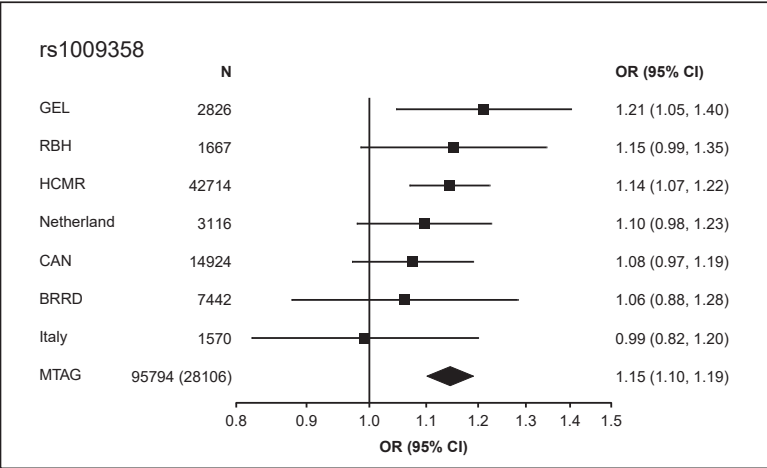


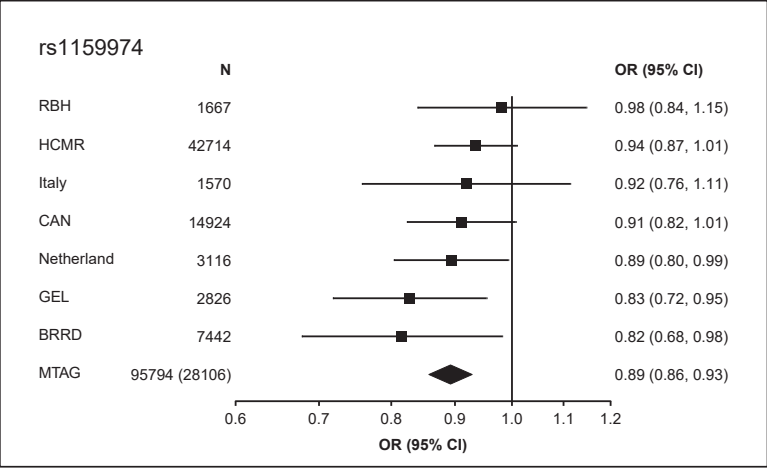
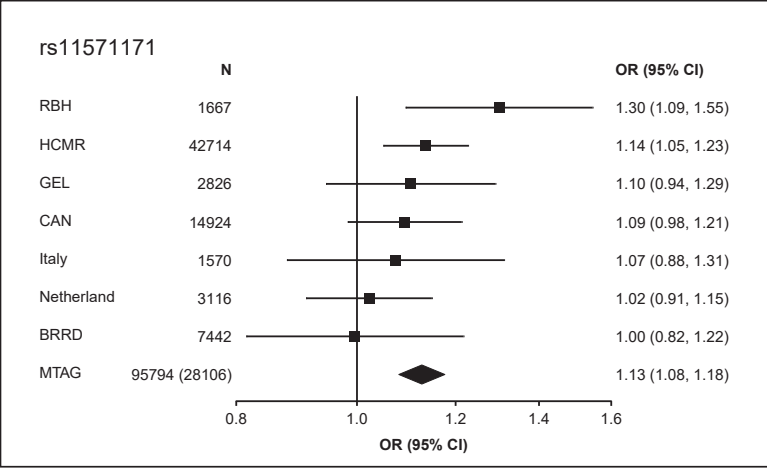
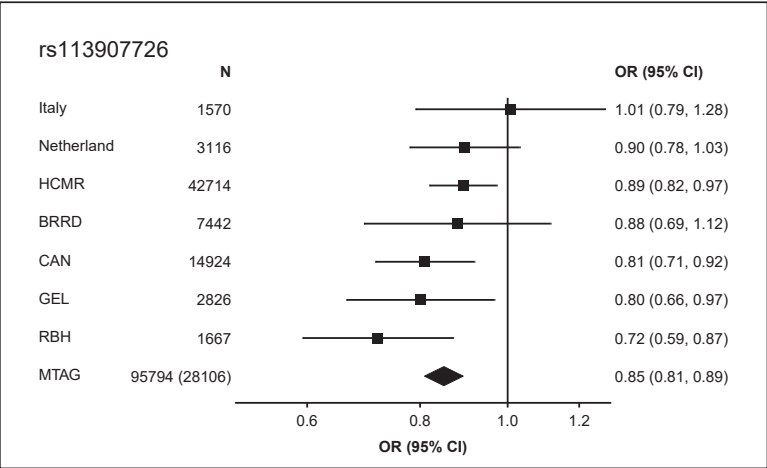
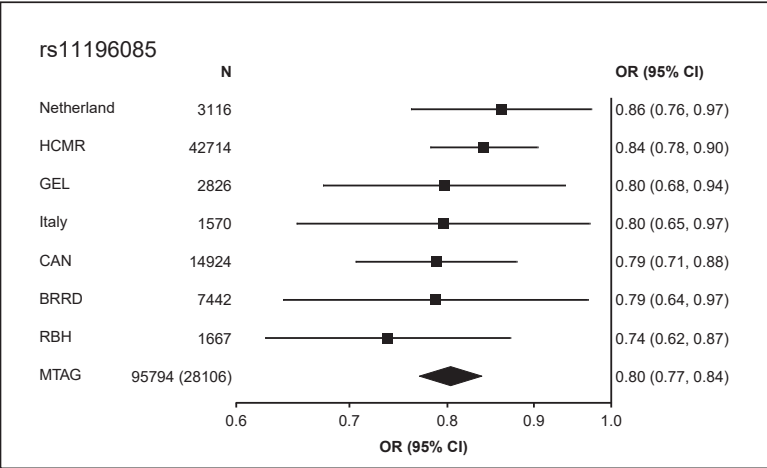
Supplementary Figure 12: Manhattan plot for GWAS of left ventricular concentricity index. The left ventricular concentricity index is the ratio of left ventricular mass/left ventricular end-diastolic volume. GWAS based on an analysis of 36,083 UK Biobank participants without cardiomyopathy and with available cardiac magnetic resonance data. Genomic position presented on the x-axis and $-\log(P\text{-value})$ on the y-axis. Manhattan plot generated using FUMA¹ and annotated with locus names based on nearest gene, top gene mapped using OpenTargets² variant to gene pipeline (25 February 2022 release), and any mendelian cardiomyopathy gene mapped using FUMA v1.3.8. Locus name and data point colors alternate by chromosome. “MHC” refers to the major histocompatibility locus on chromosome 6. See P values shown are calculated from a quantitative trait GWAS using BOLT-LMM and are not corrected for multiple testing. Supplementary Table 6 for description of the lead variants in the LV trait GWAS and lookup in the HCM GWAS.

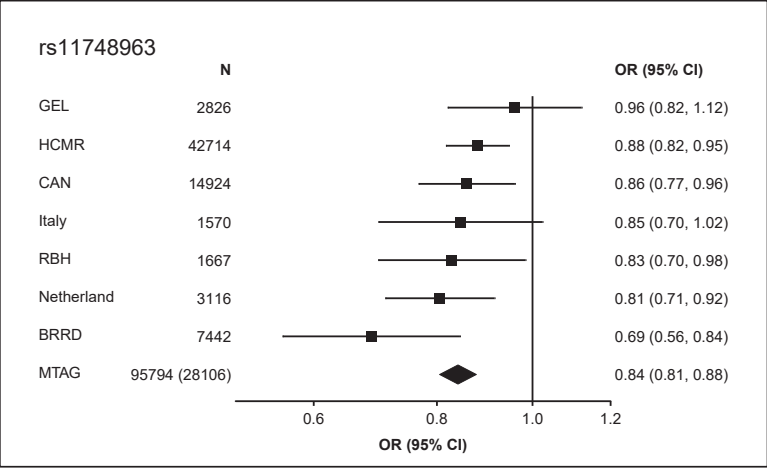
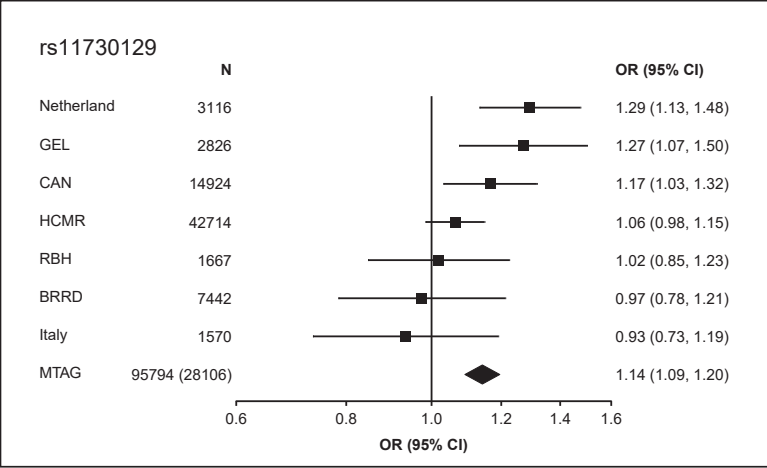
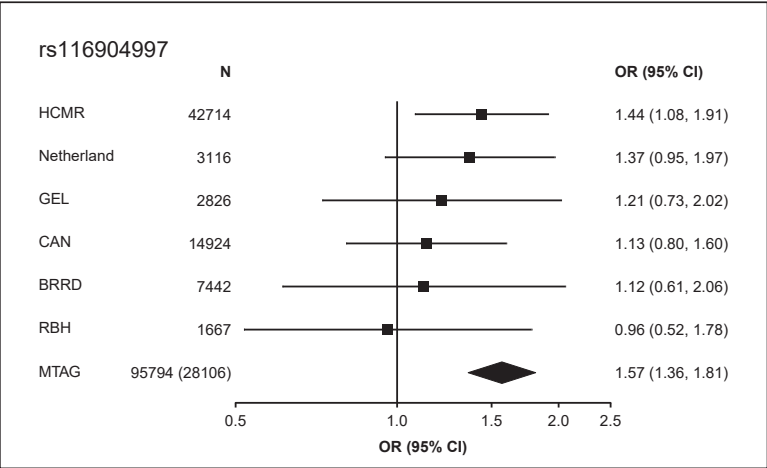
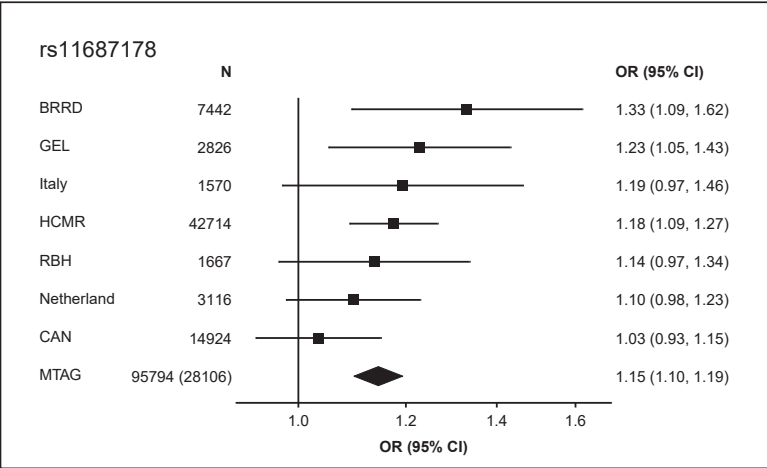


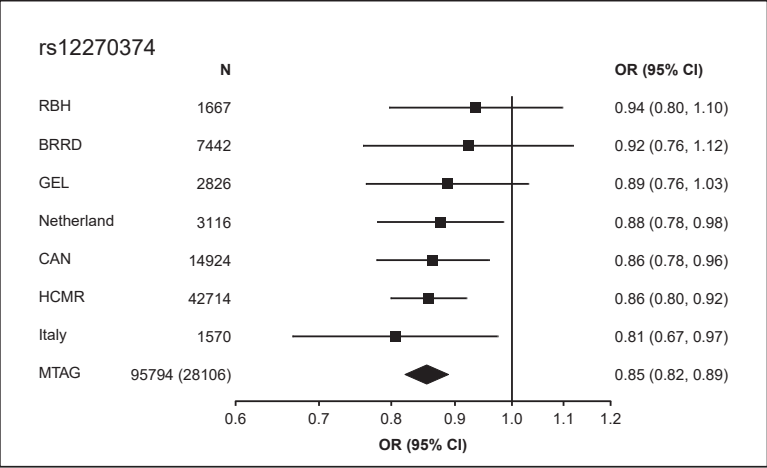
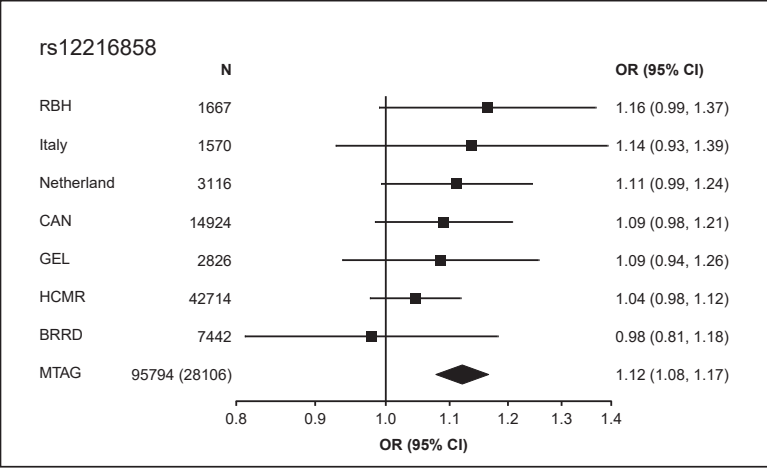
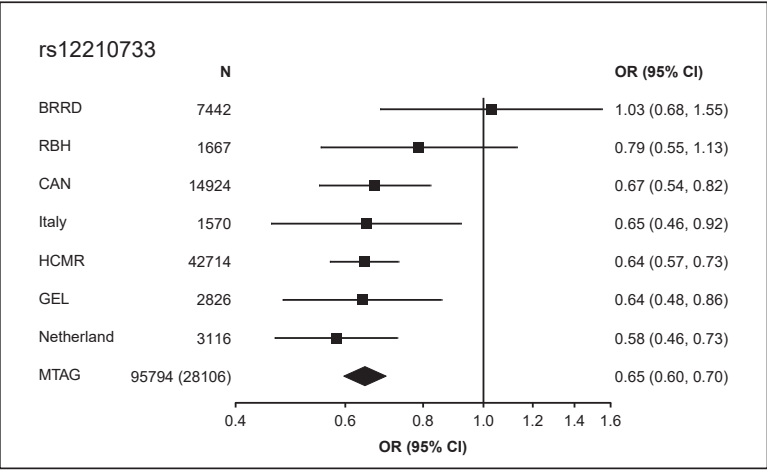
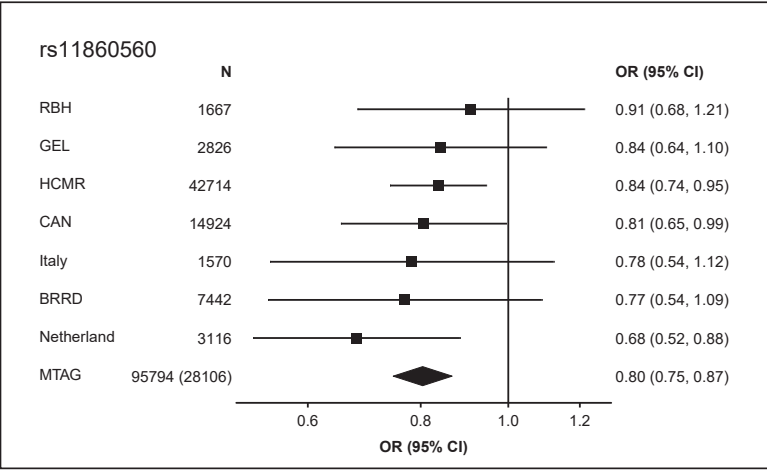
Supplementary Figure 13: Comparison of regression coefficients (beta) derived from HCM GWAS and MTAG. A) Analysis performed including all SNPs results in a Spearman's regression coefficient of 0.90 and a linear regression slope of 0.83. B) Analysis performed including all SNPs with $P < 0.001$ in either GWAS and/or MTAG results in a Spearman's regression coefficient of 0.96 and a linear regression slope of 1.00. Red line is the $x=y$ identity line, and Blue line is the linear regression fit. Both lines overlap in panel B. Data support an accurate estimation of SNP effects using MTAG.

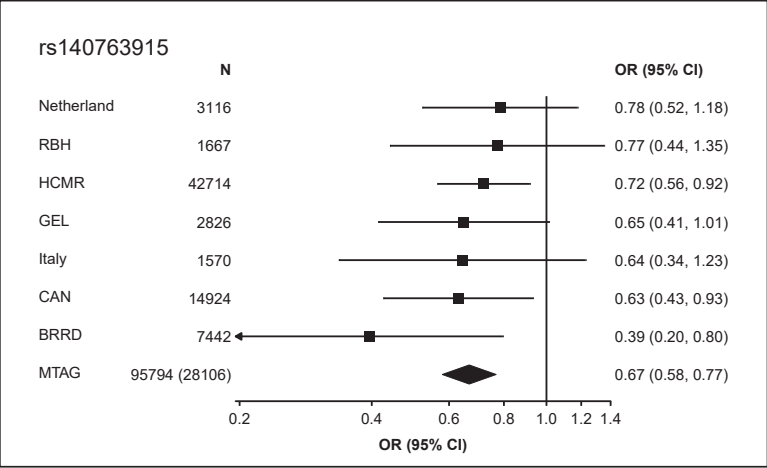
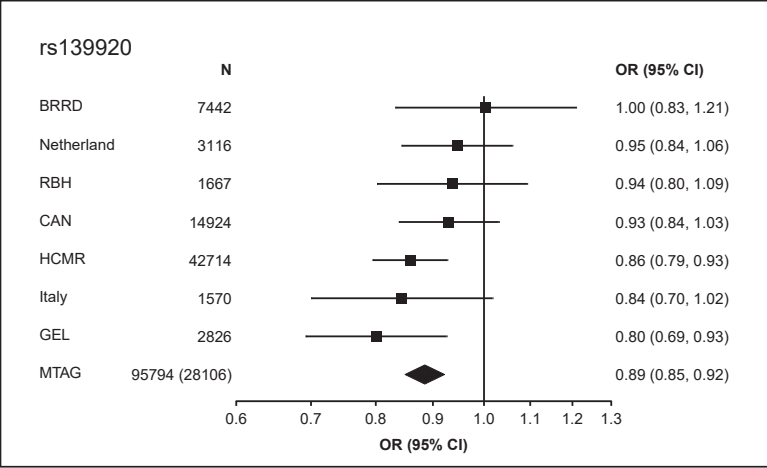
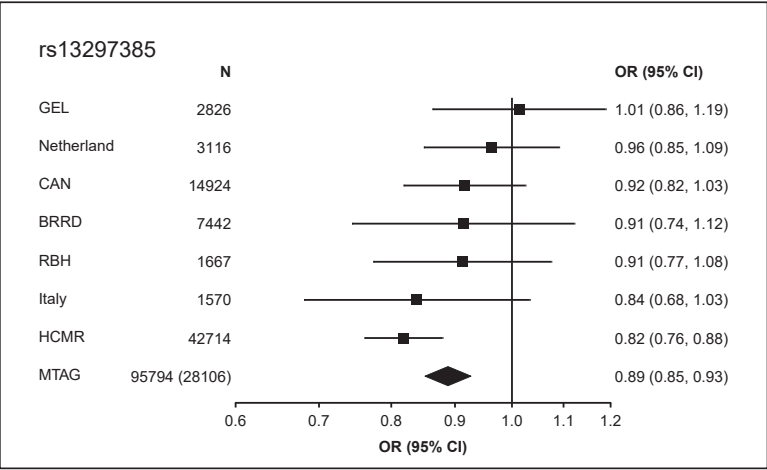
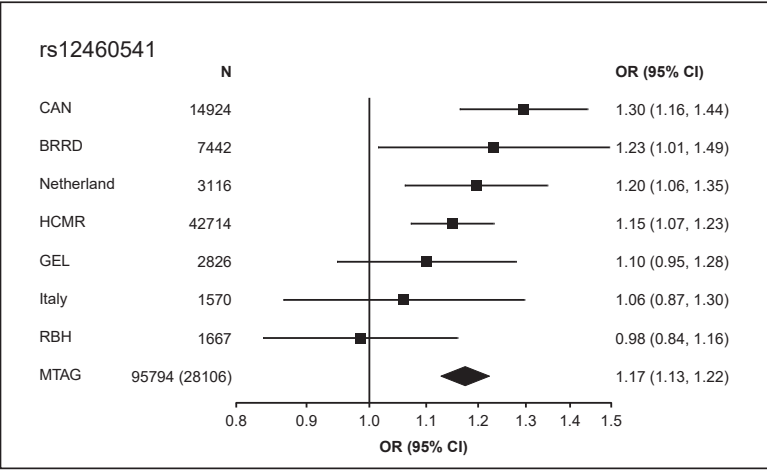
↓ **Supplementary Figure 14:** Forest plots of 68 MTAG lead variants showing the point estimate of the odds ratio (OR) and 95% confidence interval (CI) for each of the 7 case-control GWAS datasets (shown as filled squares and error bars, respectively), as well as OR and 95% CI from the HCM multitrait analysis of GWAS (MTAG) (shown as diamond). N represents the total sample size summing cases and controls for each dataset. For MTAG, N represents the total sample size $((5900 + 68359) \times 1.29 \text{ power increase})$, and the effective sample size between parentheses (see methods). BRRD, Bioresource Rare Disease; CAN, Canada; GEL, Genomics England; HCMR, Hypertrophic Cardiomyopathy Registry; META, meta-analysis; RBH, Royal Brompton Hospital.

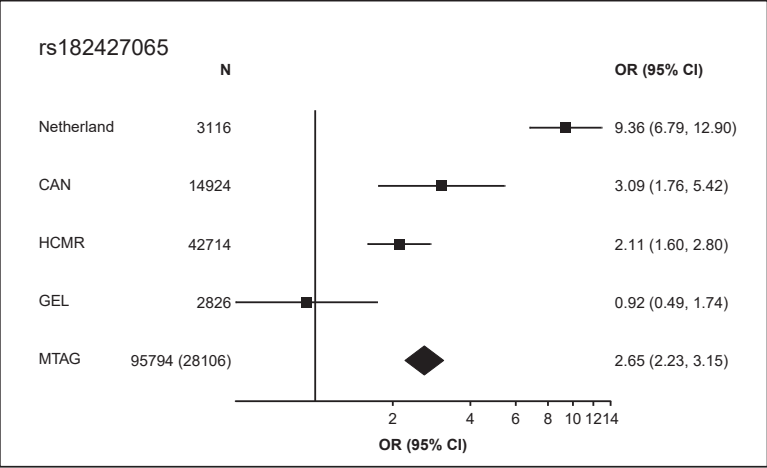
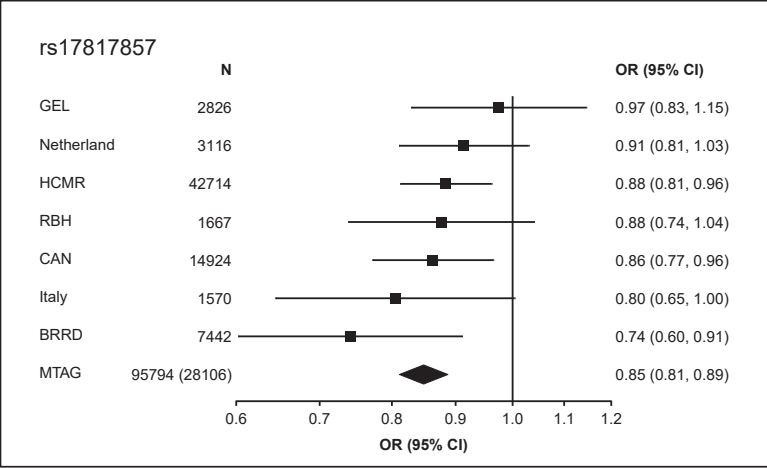
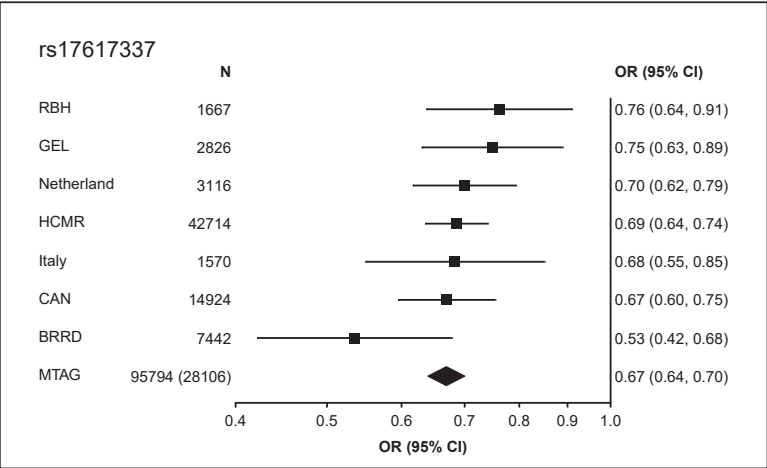
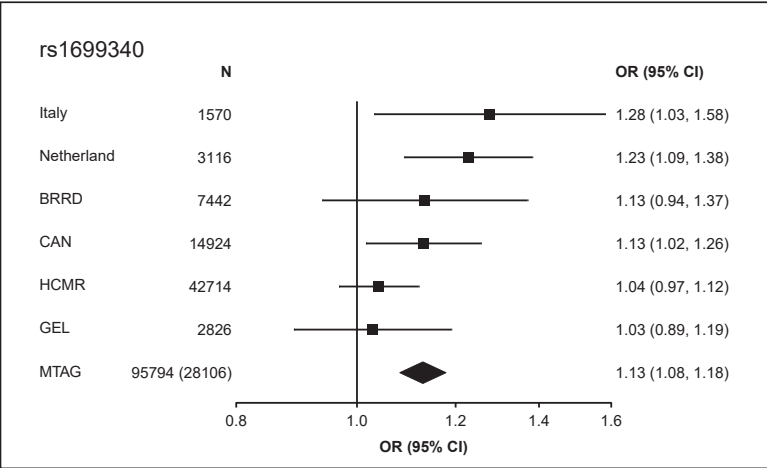


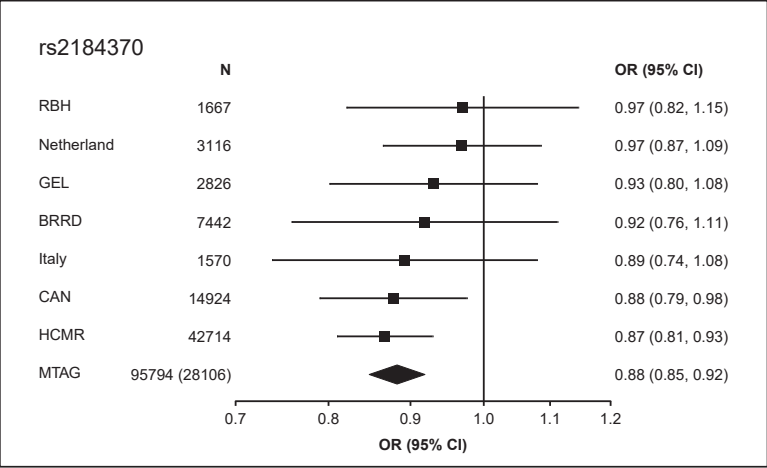
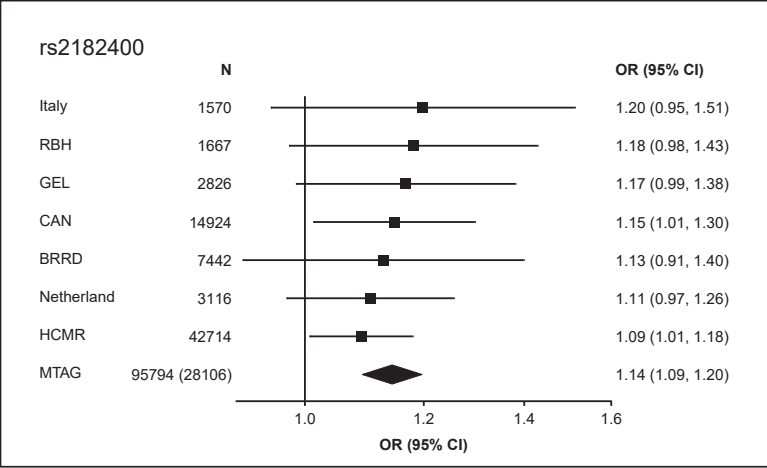
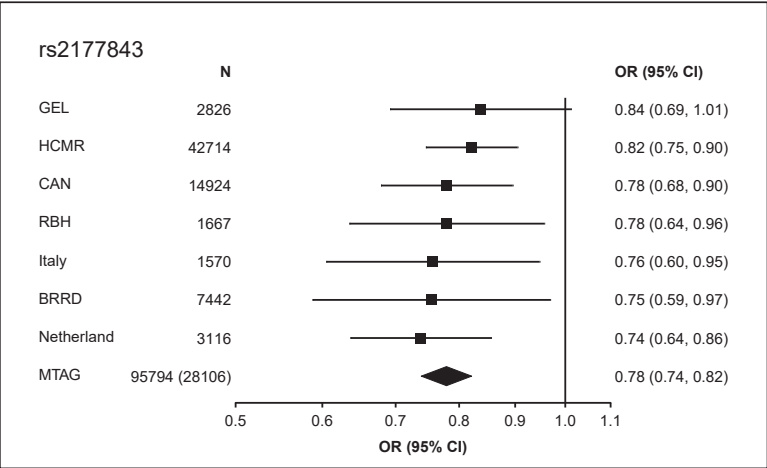
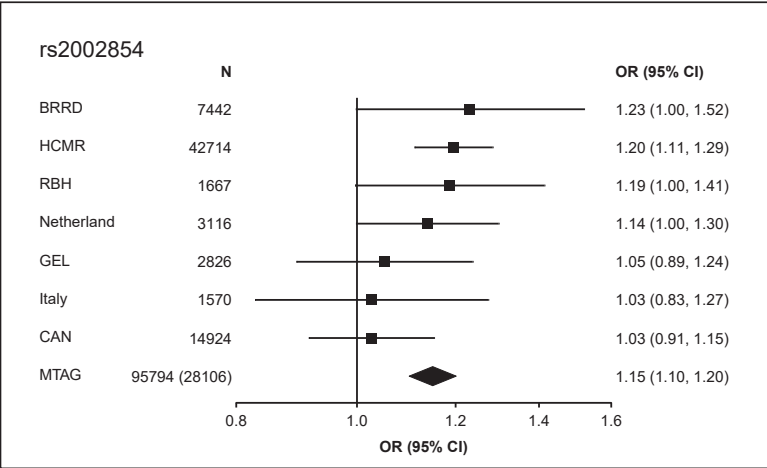


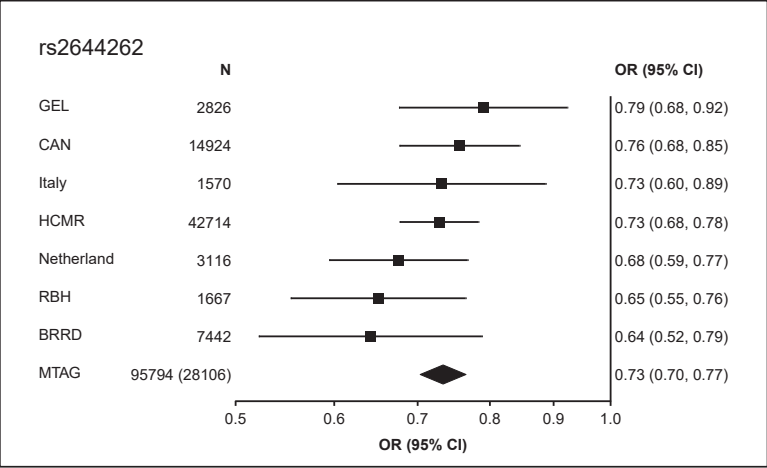
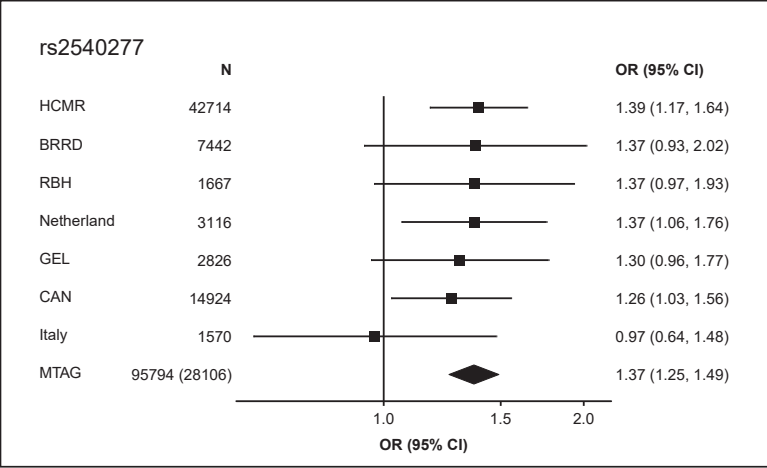
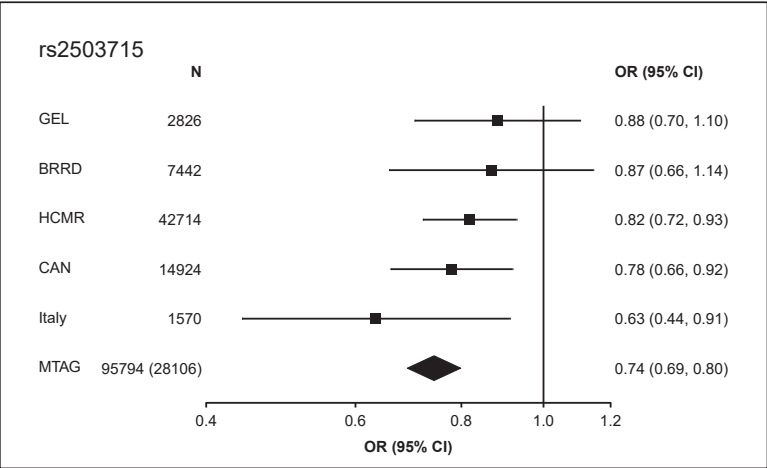
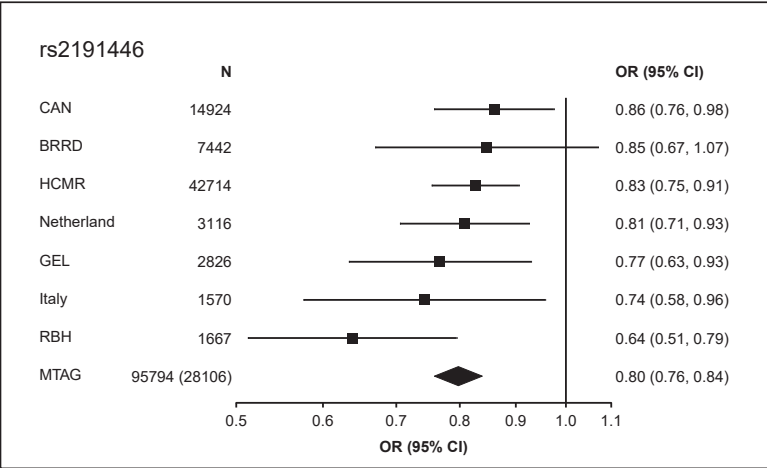


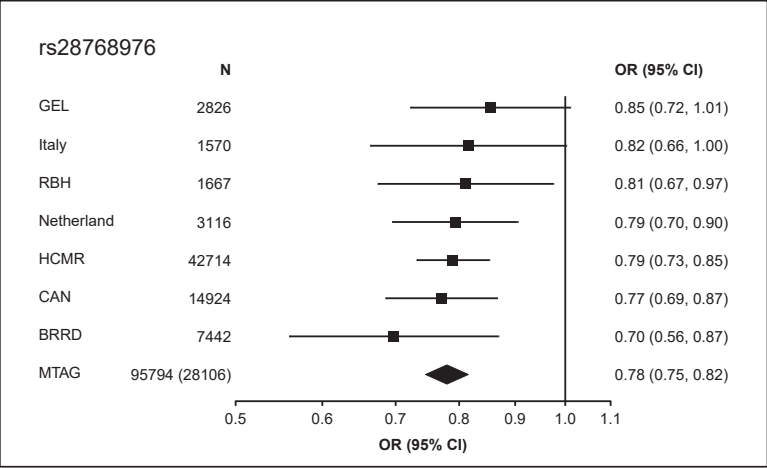
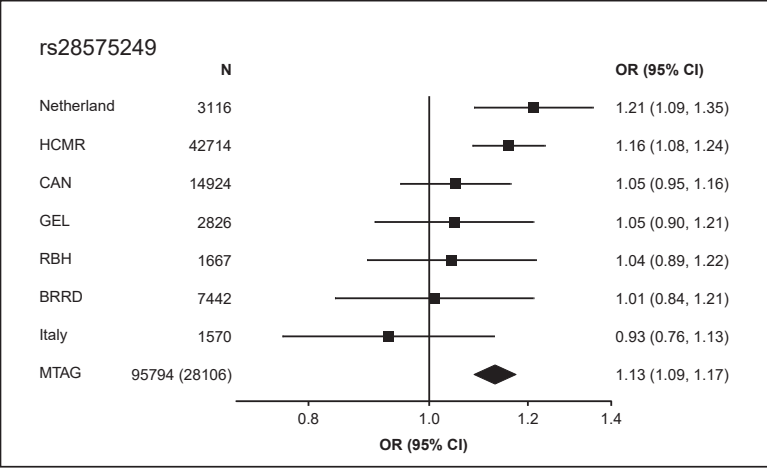
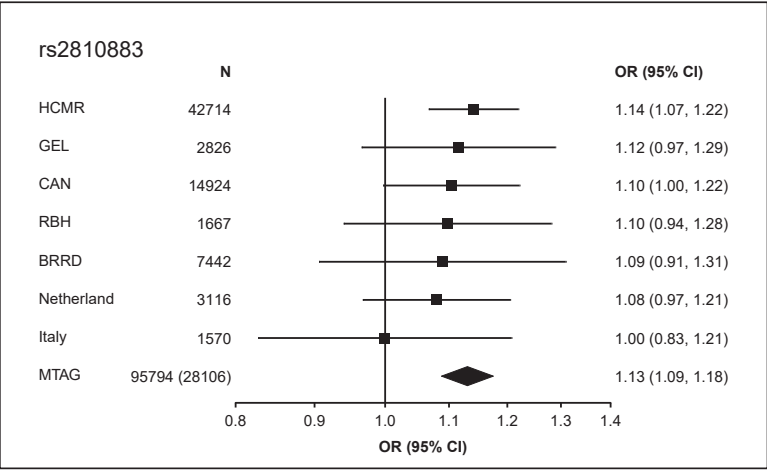
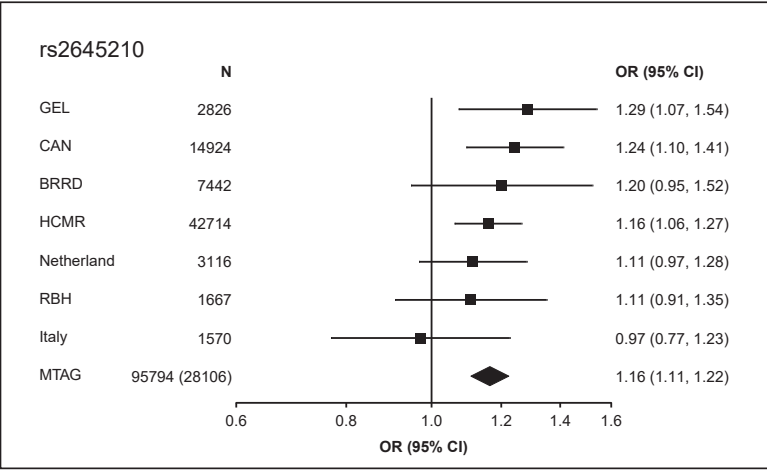


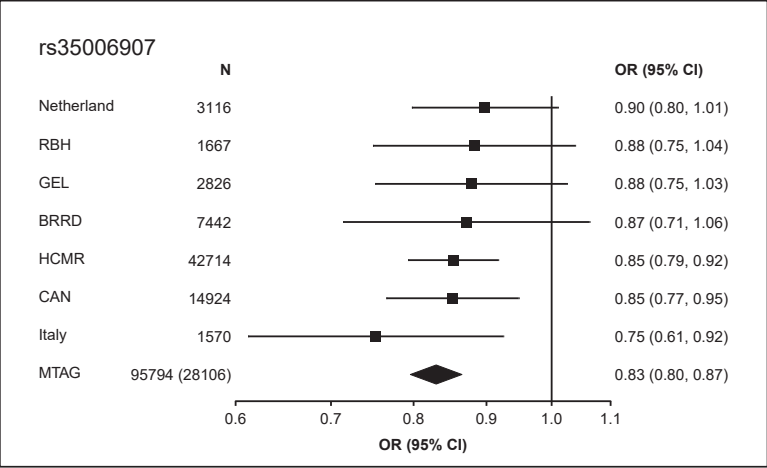
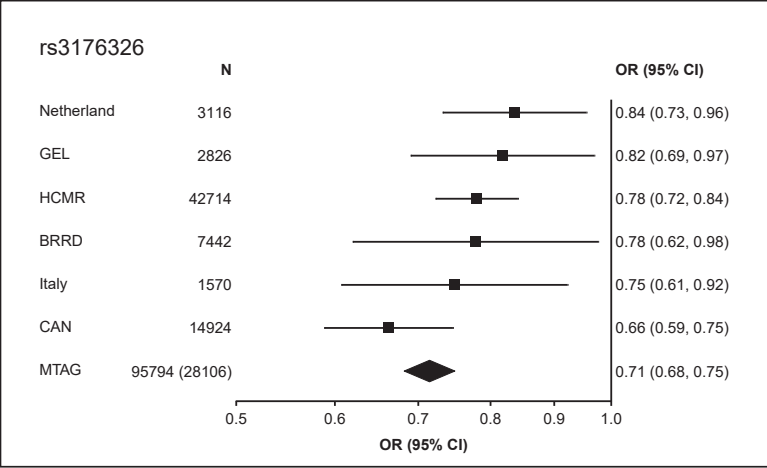
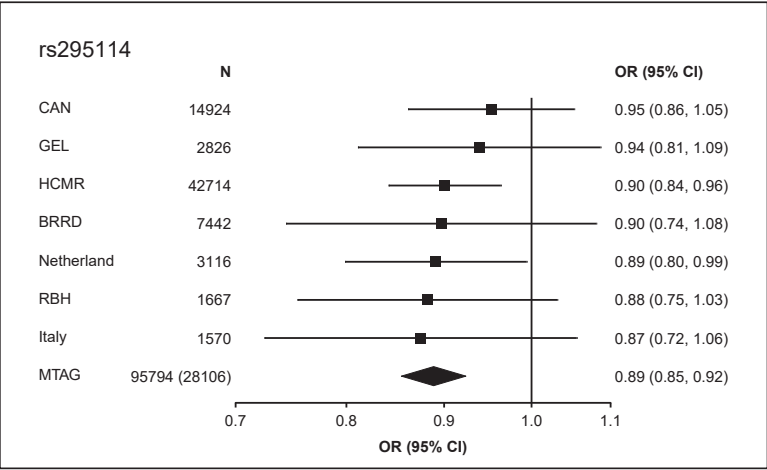
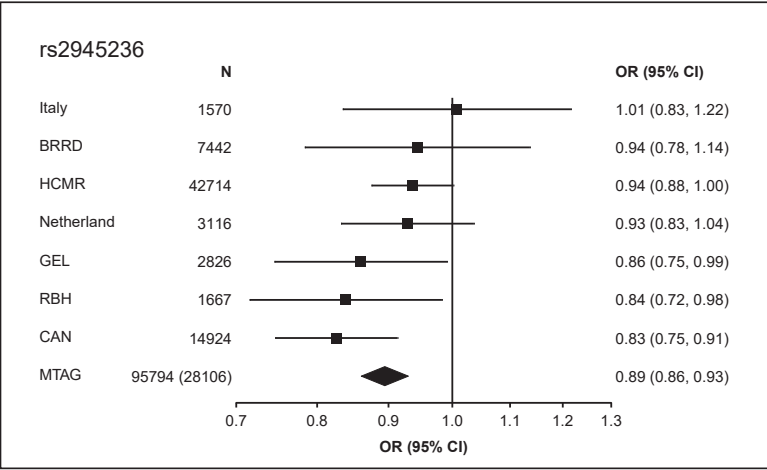


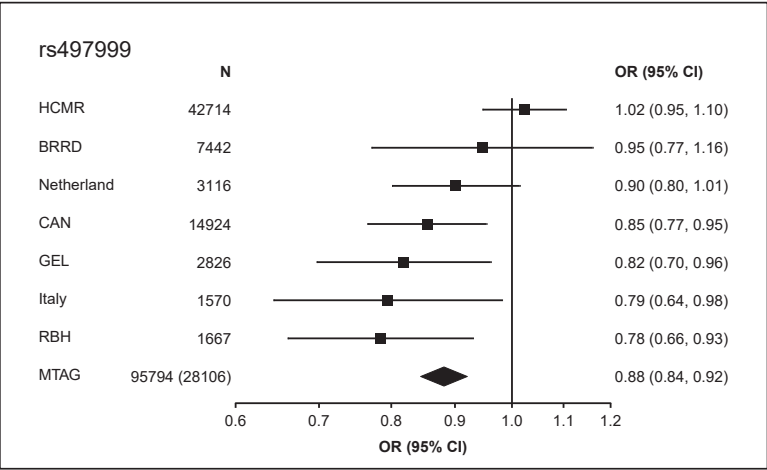
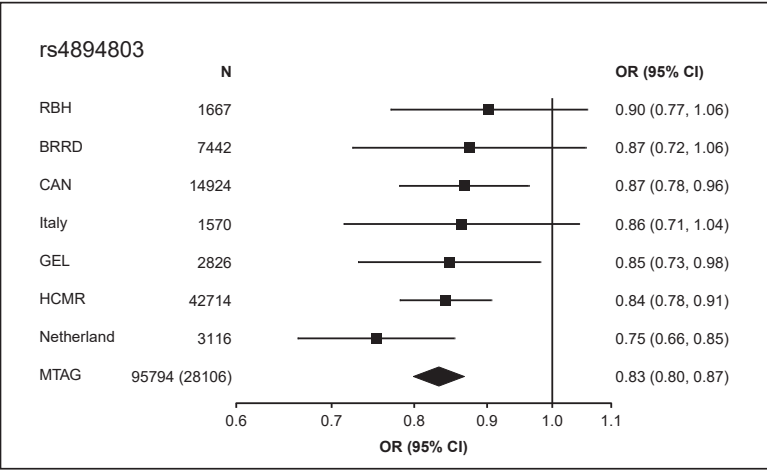
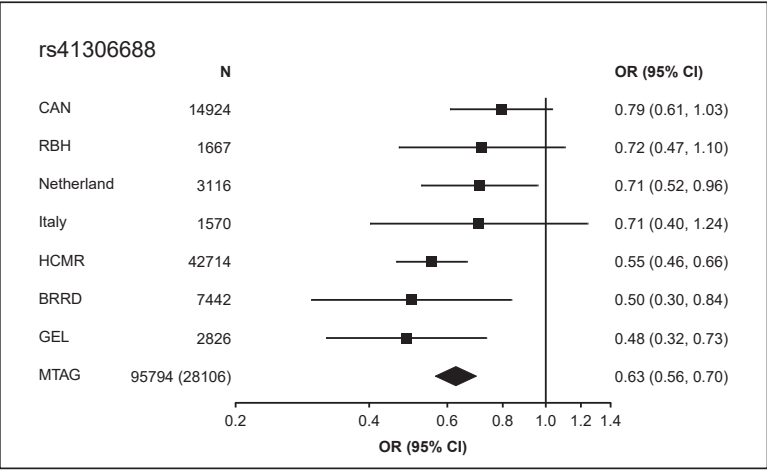
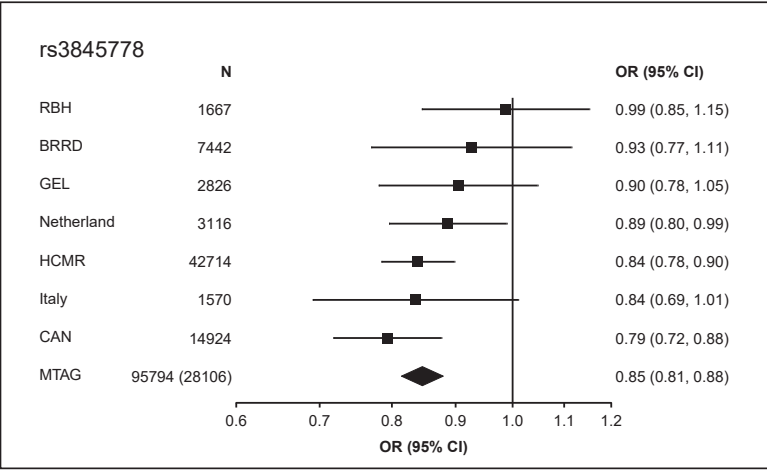


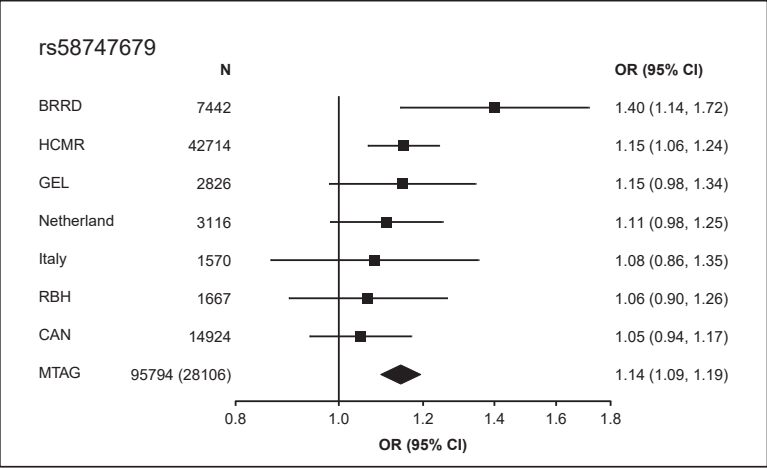
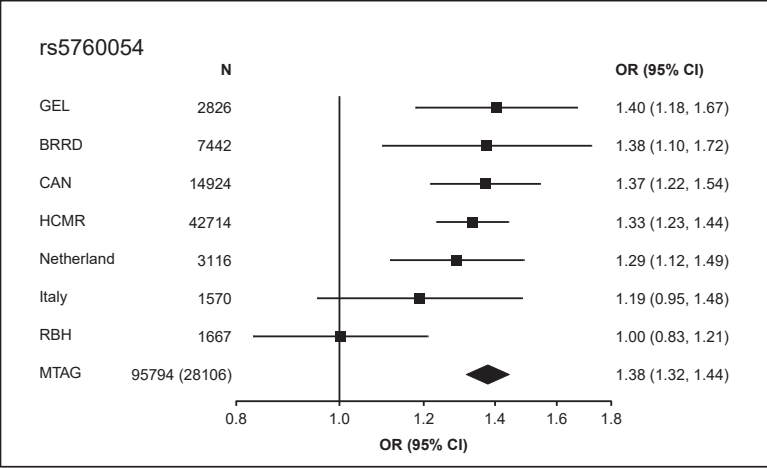
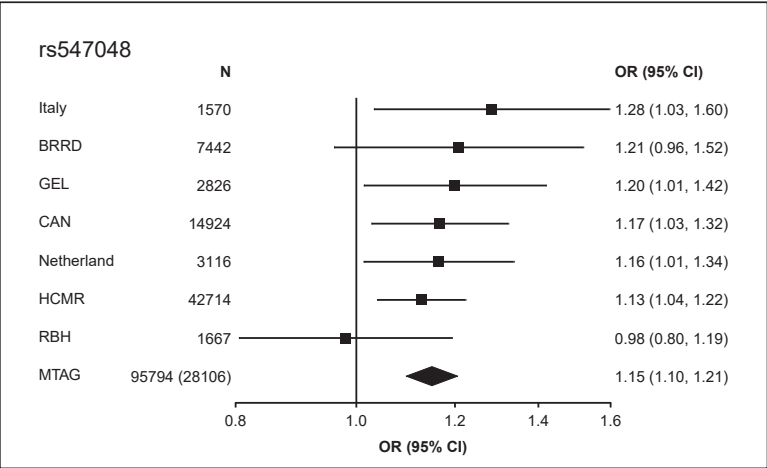
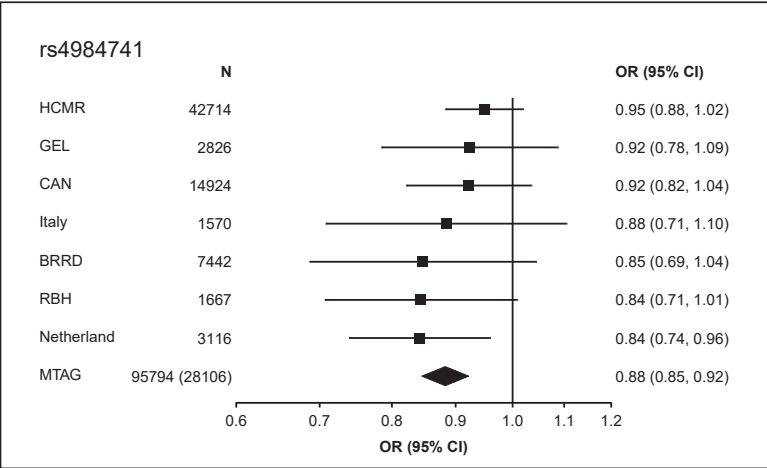


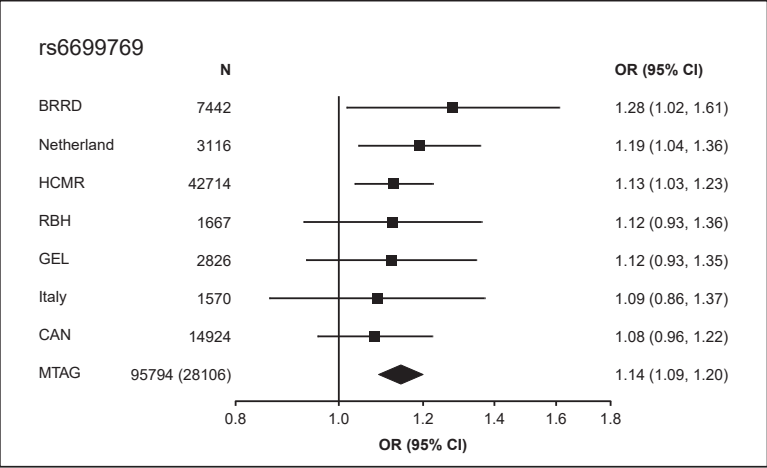
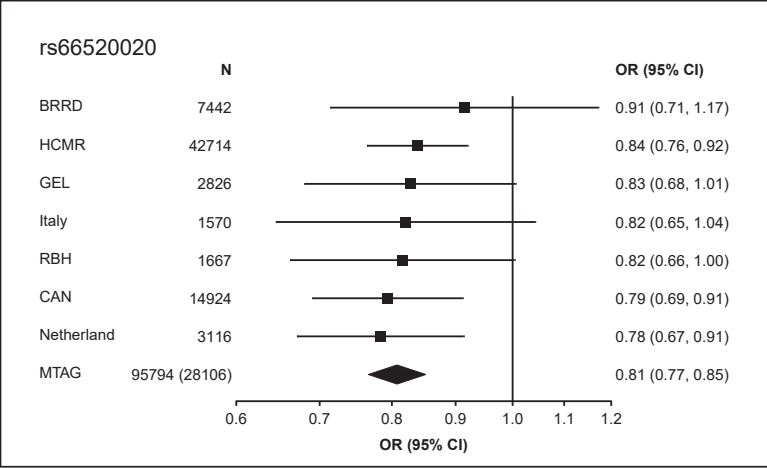
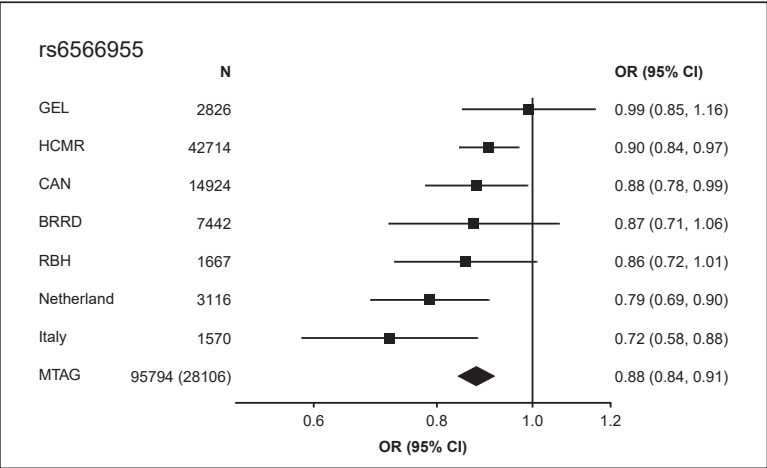
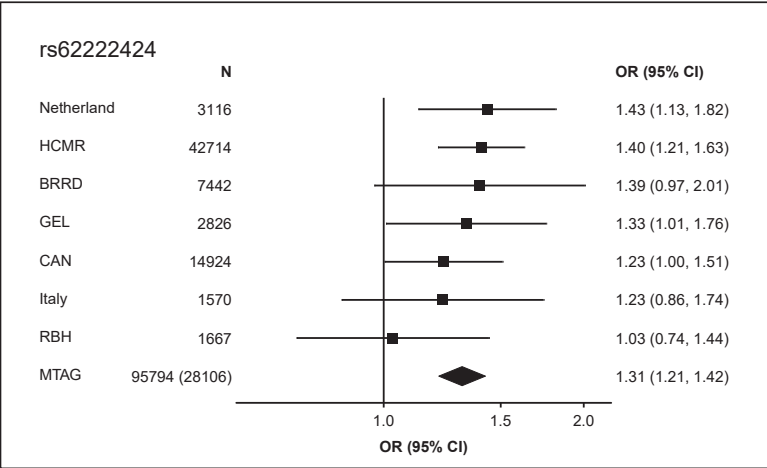


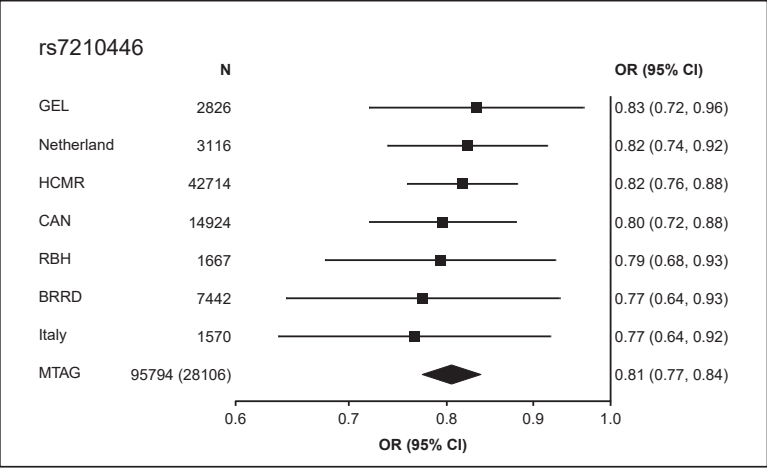
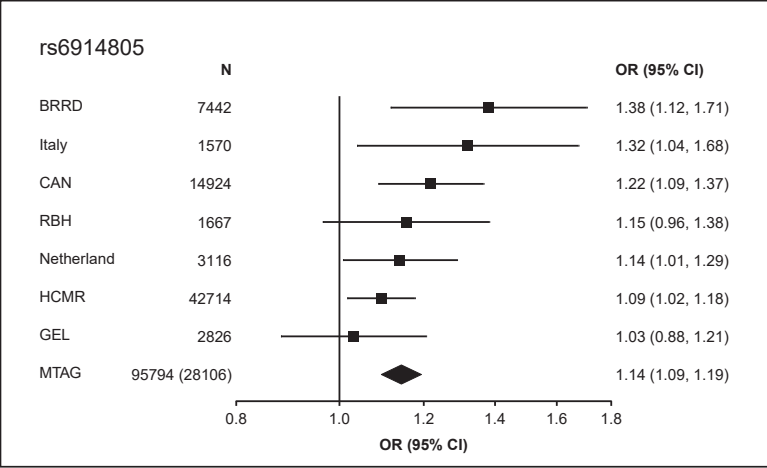
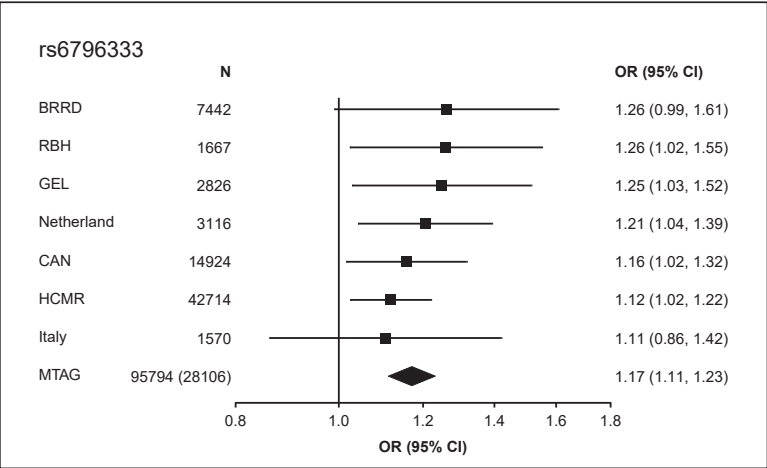
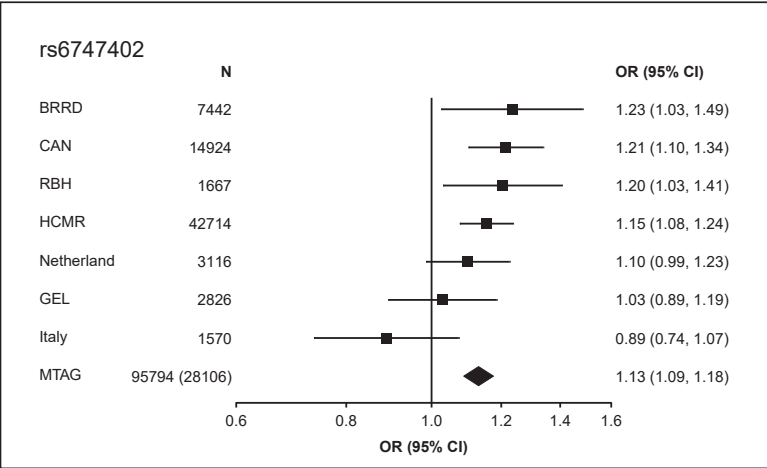


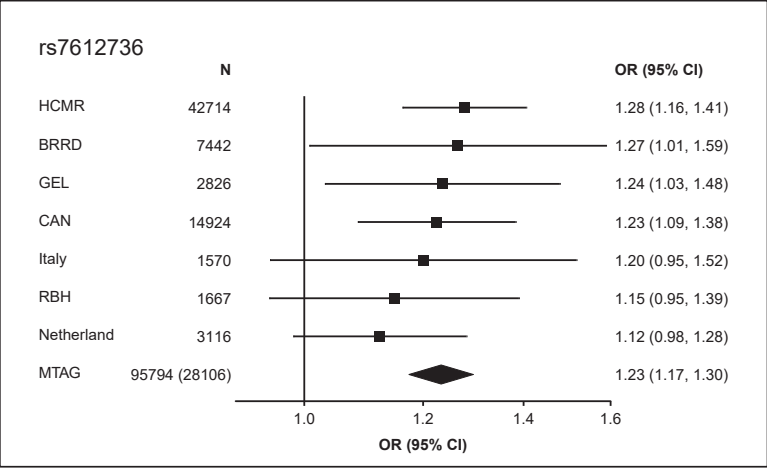
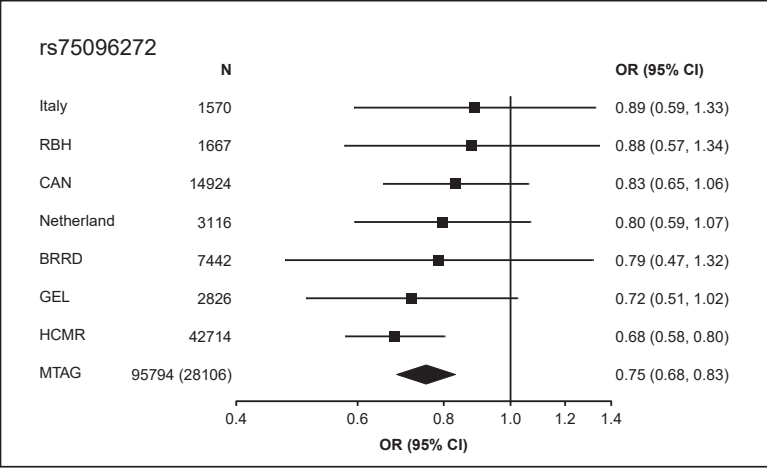
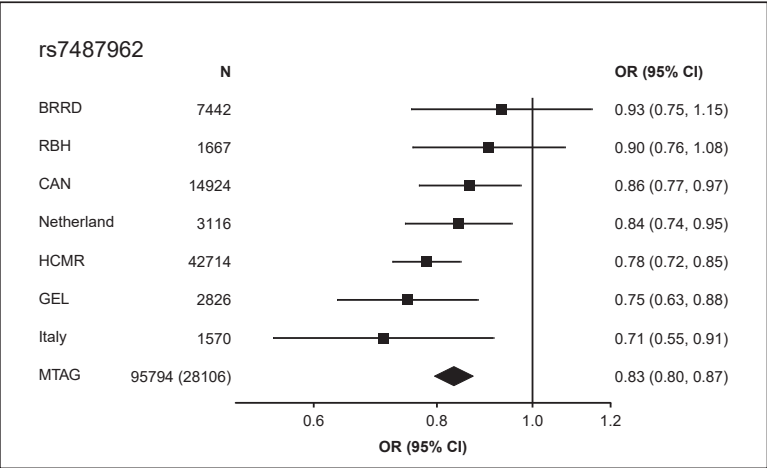
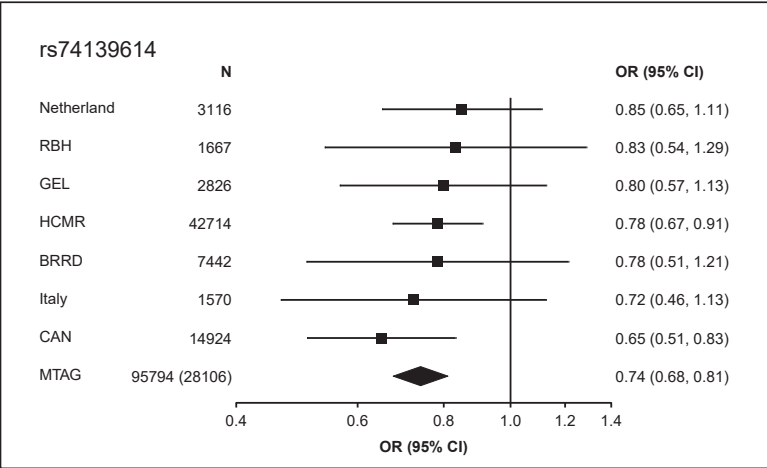


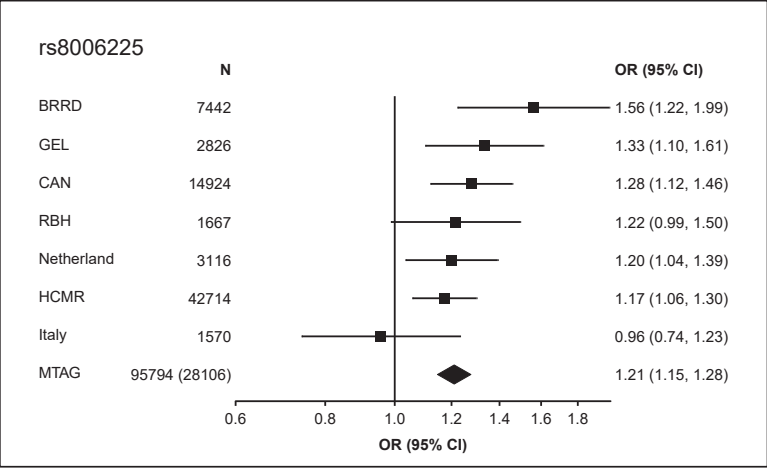
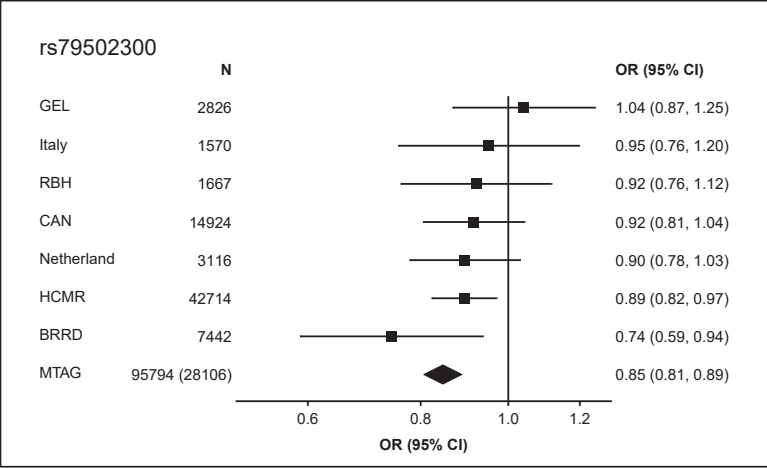
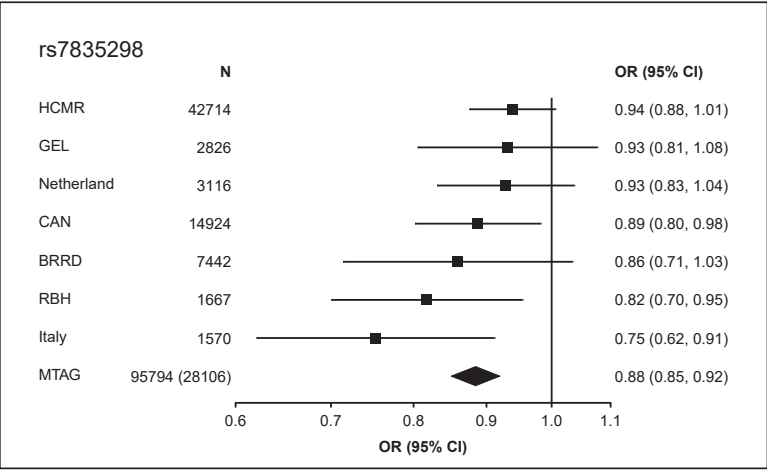
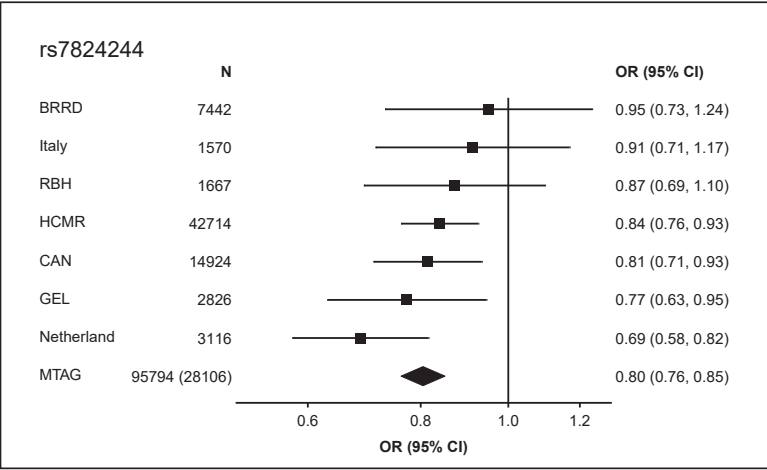


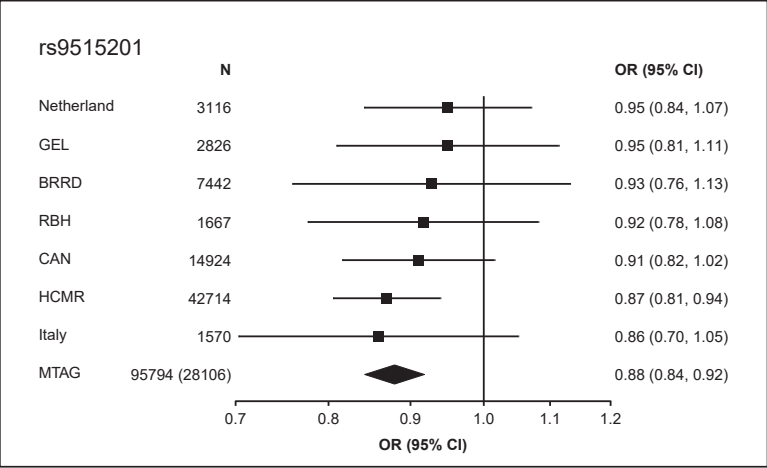
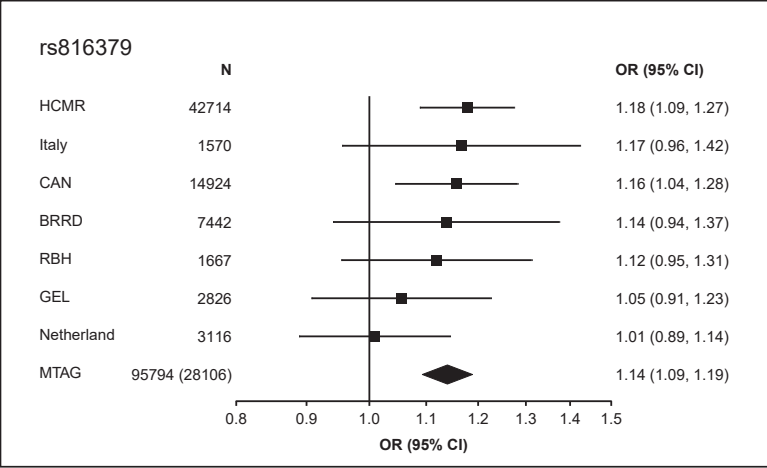
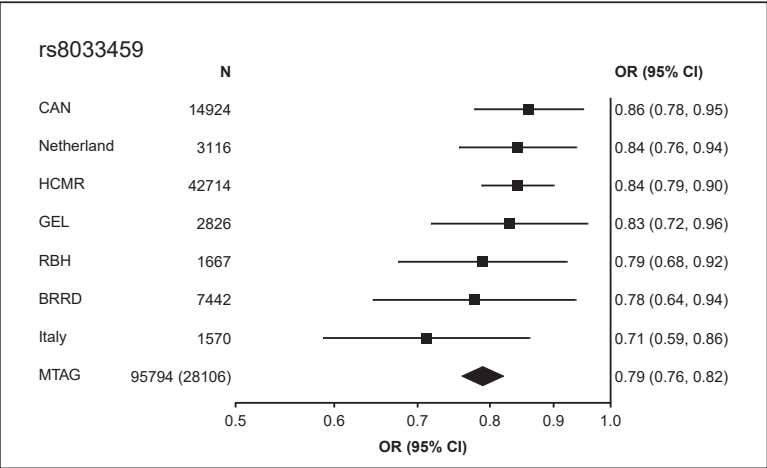
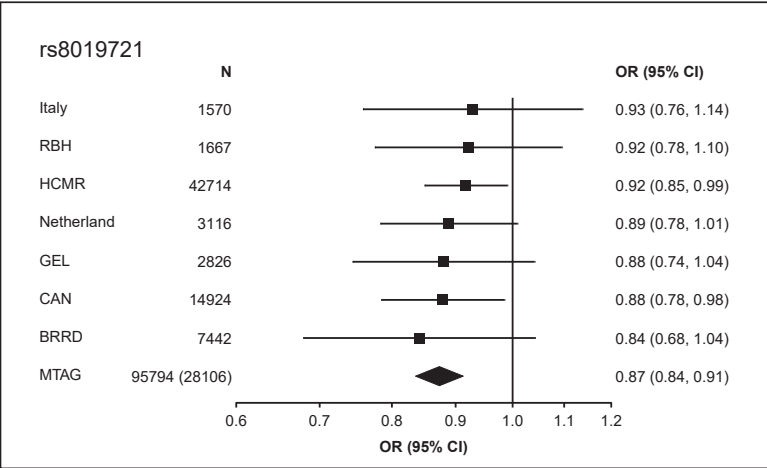


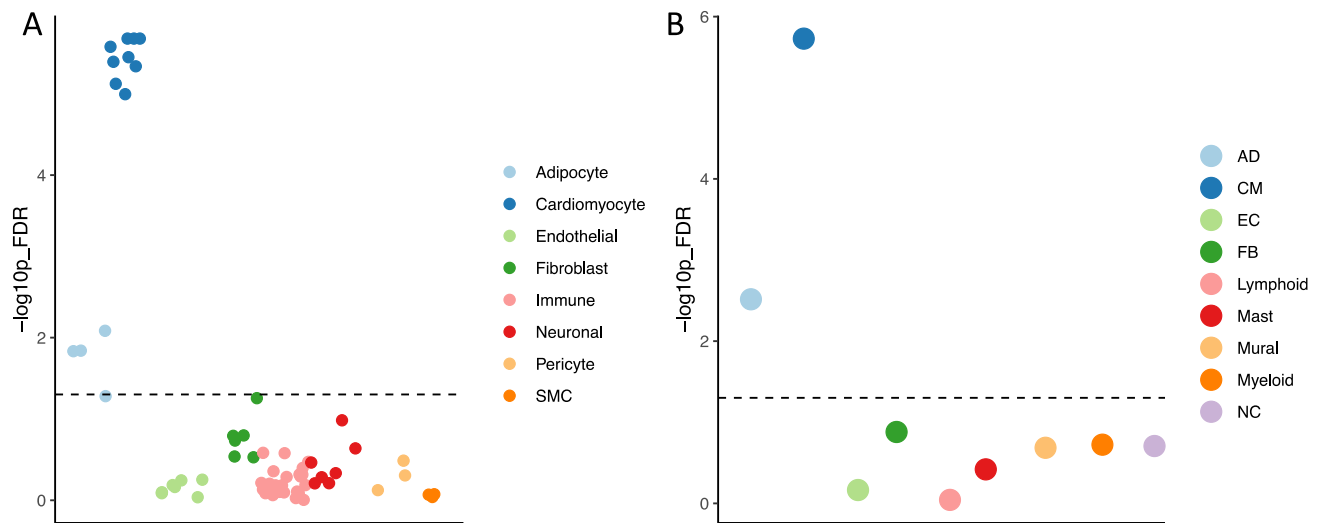




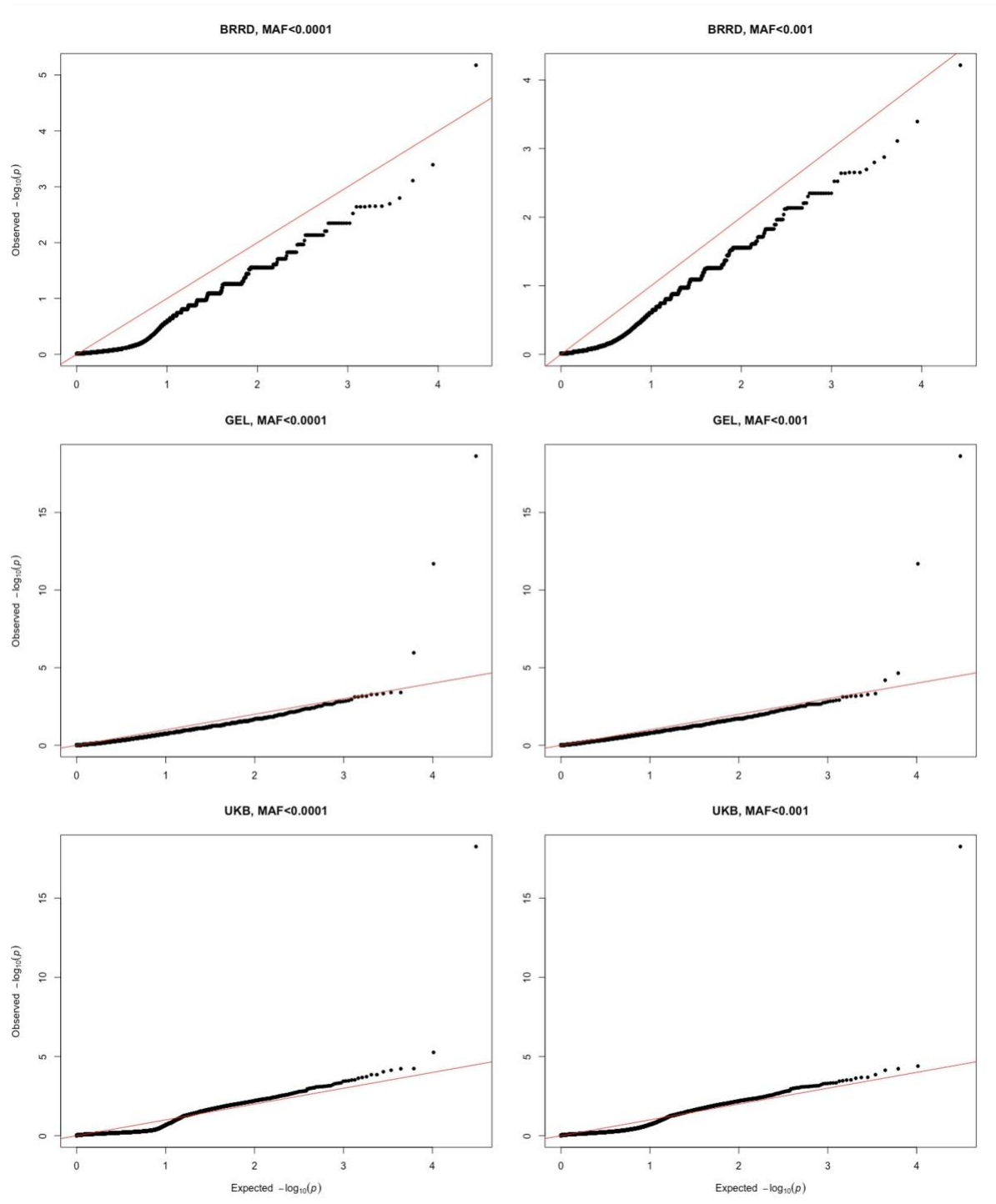




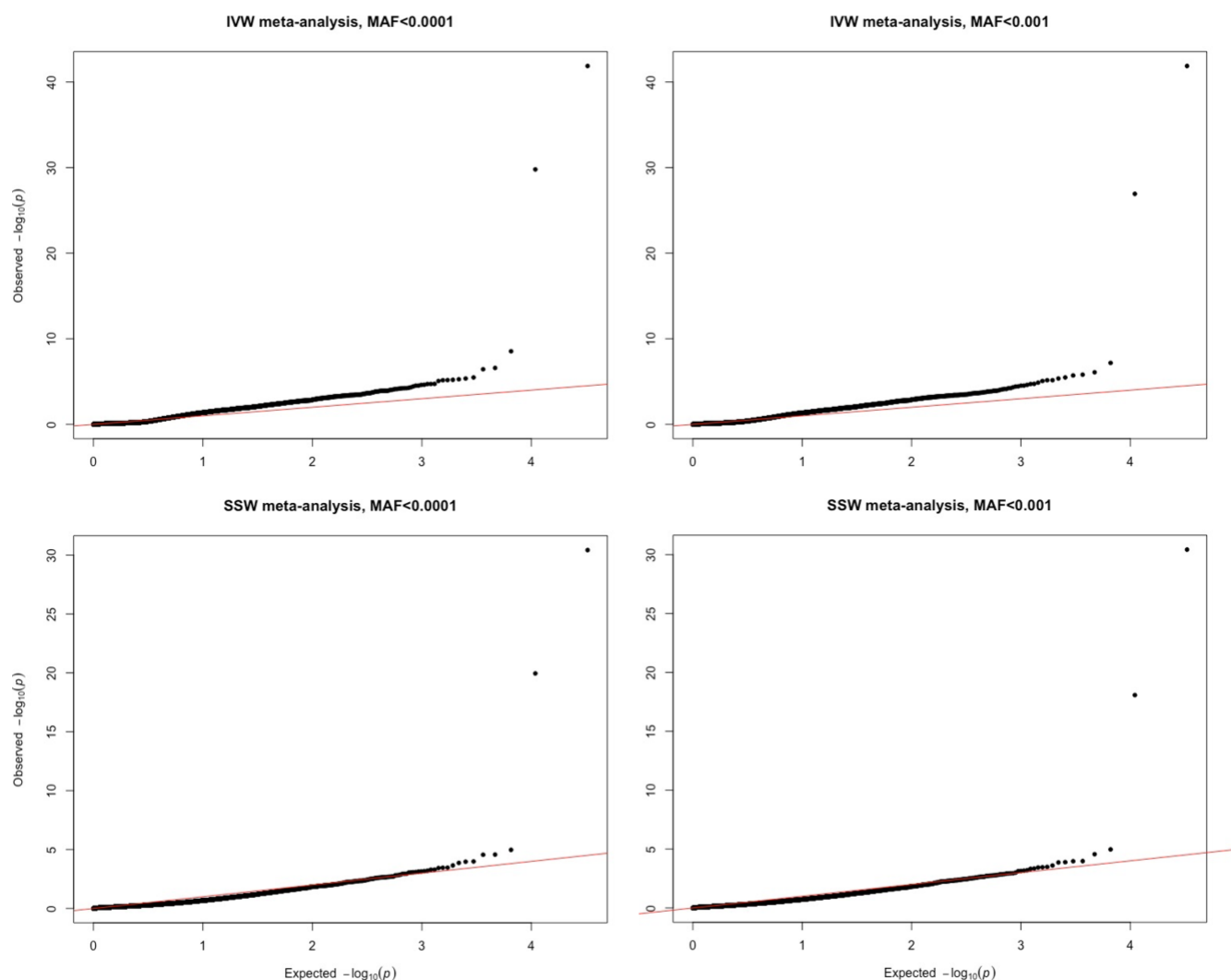




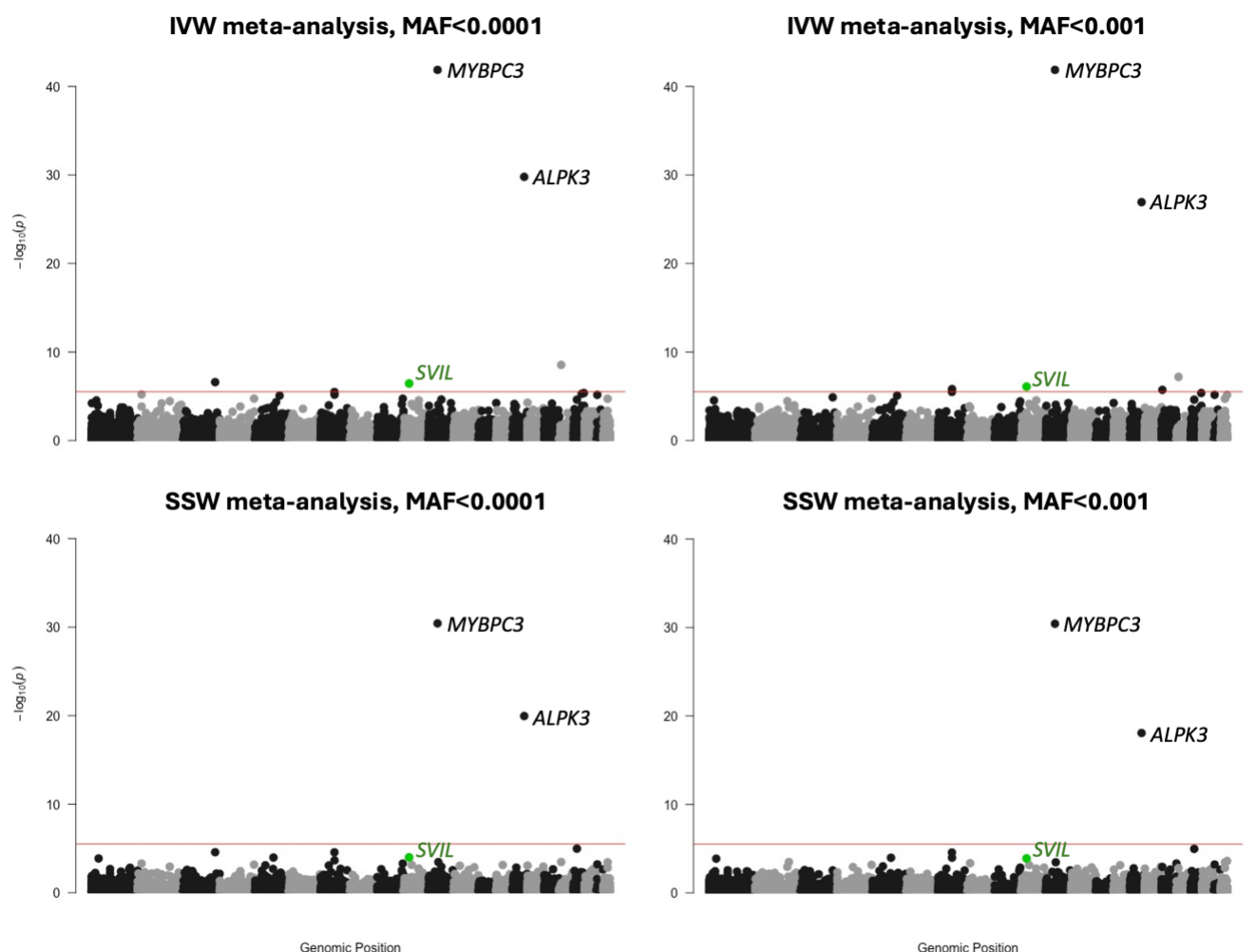
Supplementary Figure 15: Enrichment of 68 cell-state (A) and 9 cell-type (B) gene programs from single nuclei data of control heart samples from Reichart D et al³ using the sc-linker approach and ABC-model enhancer-gene linking strategy. Each point represents a cell state (A) or type (B) with $-\log_{10}(P)$ FDR of the heritability enrichment using S-LDSC. Dashed line represents $FDR=0.05$ for significance. FDR, false discovery rate.



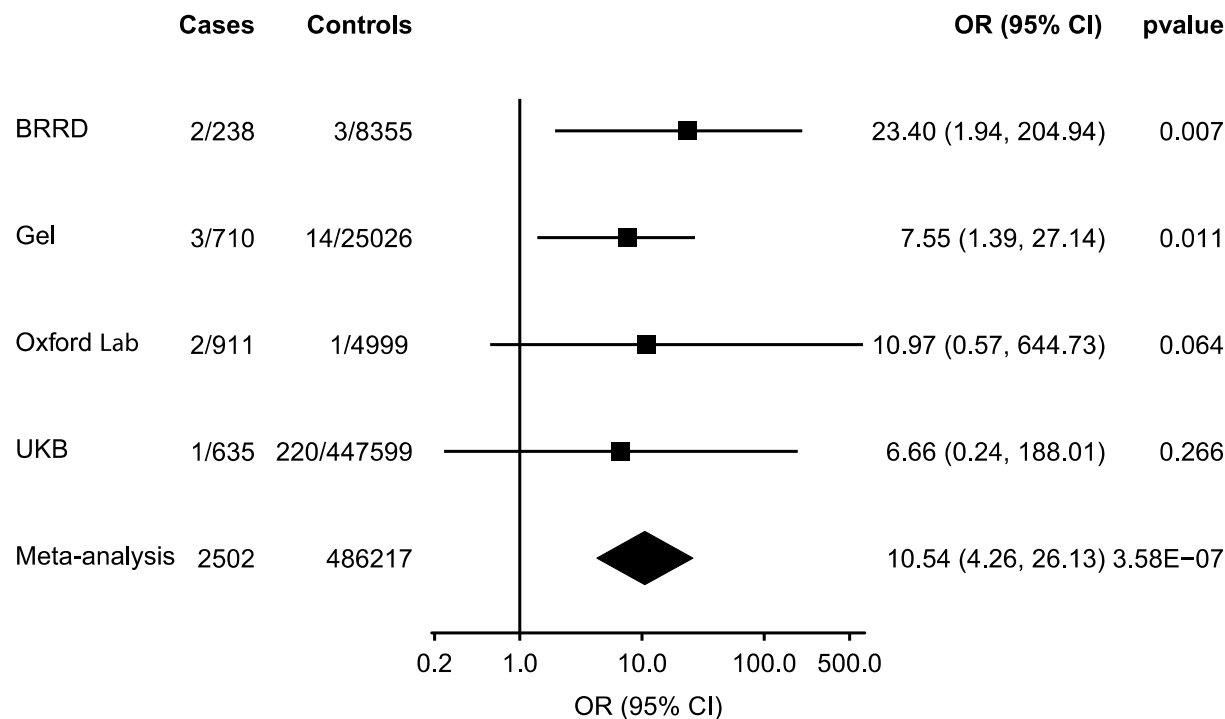
Supplementary Figure 16: Cohort level quantile-quantile (QQ) plots for the exploratory exome-wide rare loss of function variant gene-based burden testing. Analyses performed using 2 minor allele frequency thresholds (MAF<0.001 and MAF<0.0001) in 3 datasets: BioResource for Rare Disease (BRRD; 240 cases and 8,358 controls), Genomics England (GEL; 713 cases and 25,040 controls), and UK Biobank (UKB; 637 cases and 452,818 controls). Association of predicted rare loss-of-function variants with HCM was performed using Fisher Exact Test in GEL and BRRD, and using REGENIE for UKB.



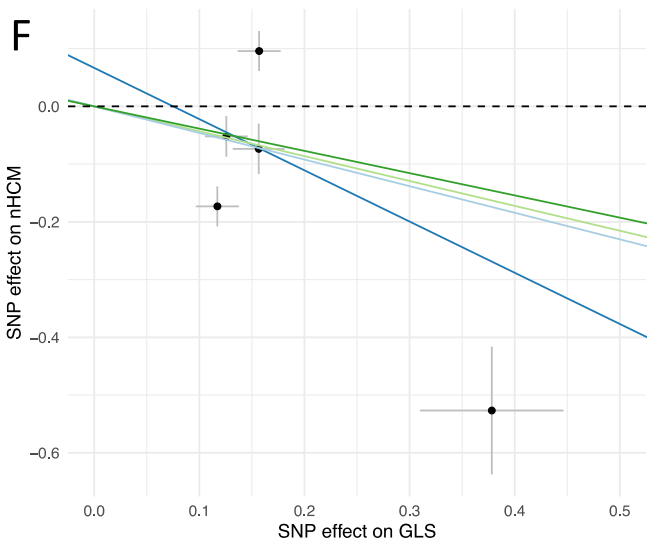
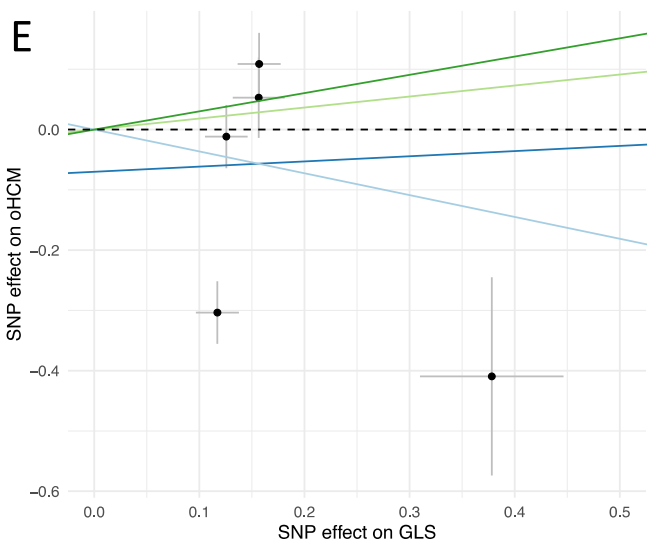
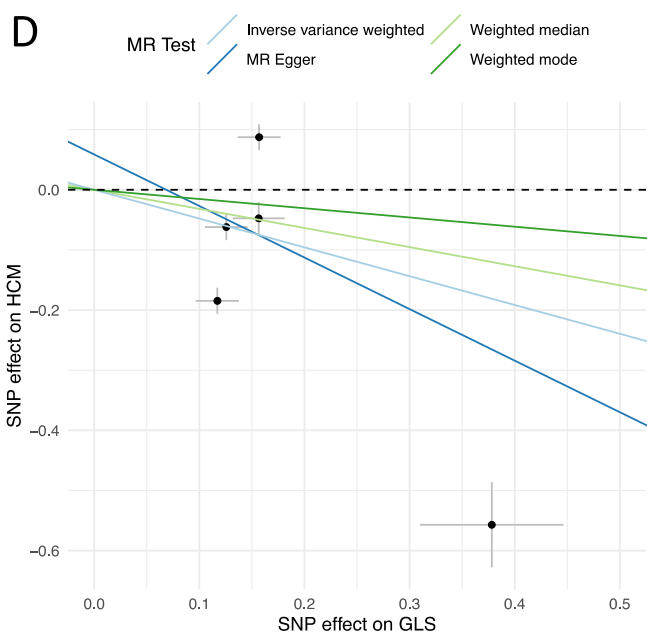
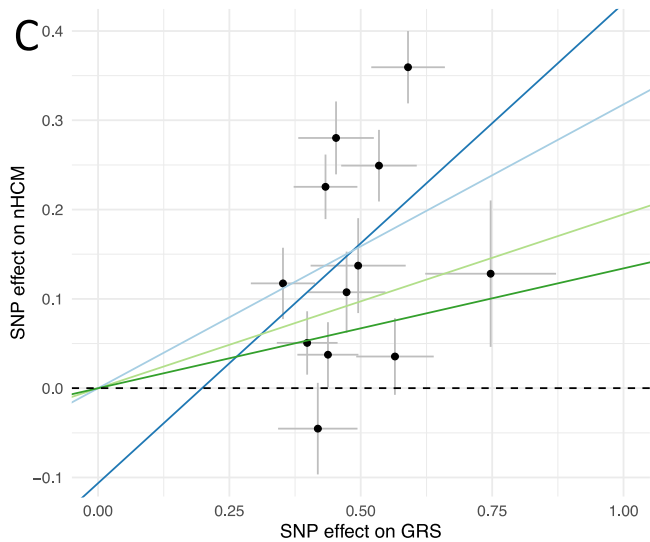
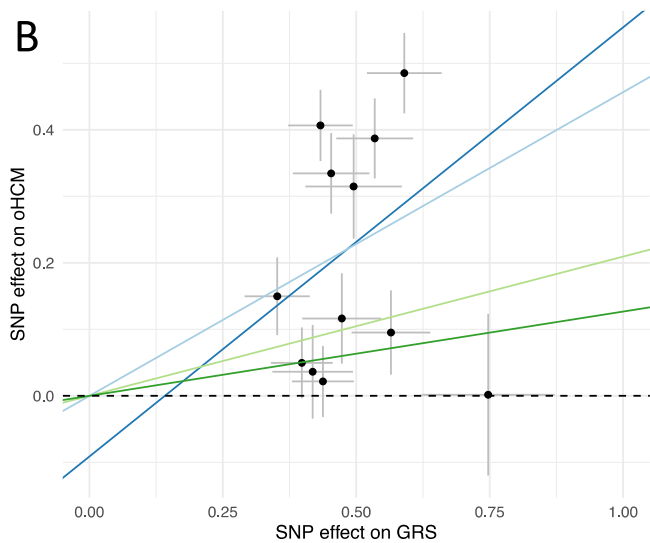
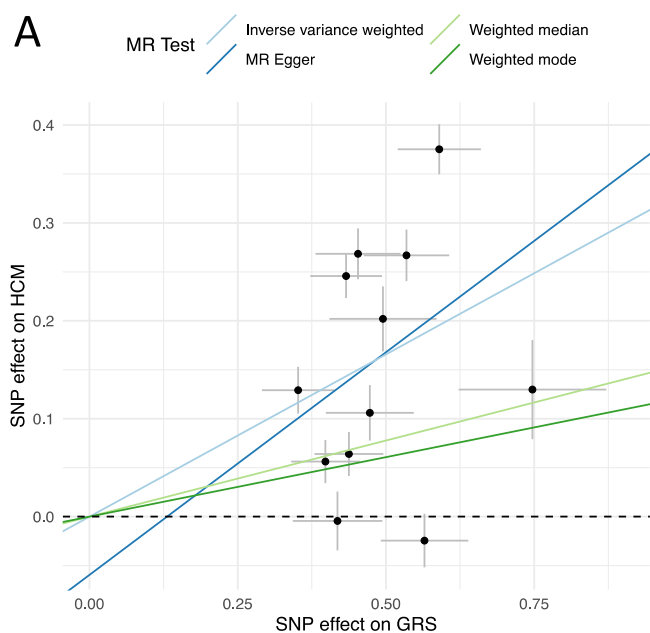
Supplementary Figure 17: Meta-analysis quantile-quantile (QQ) plots for the exploratory exome-wide rare loss of function variant gene-based burden testing. Analyses performed using 2 minor allele frequency thresholds (MAF<0.001 and MAF<0.0001) exome-wide in 3 datasets: BioResource for Rare Disease (BRRD; 240 cases and 8,358 controls), Genomics England (GEL; 713 cases and 25,040 controls), and UK Biobank (UKB; 637 cases and 452,818 controls). A 4th dataset (Oxford Medical Genetics Laboratory, OMGL; 913 cases and 5000 controls) was also used only for *SVIL*. Meta-analyses were first performed using inverse variance weighting (IVW). In presence of apparent test statistic inflation when using IVW, meta-analyses were also performed using effective sample size (N_{eff}) weighting (SSW), where $N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$. All protein-coding genes with at least 1 loss-of-function variant carrier in UKB, or at least one case and one control carrier in any other dataset were included in the meta-analyses (16,312 genes for MAF<0.0001, and 16,508 for MAF<0.001).

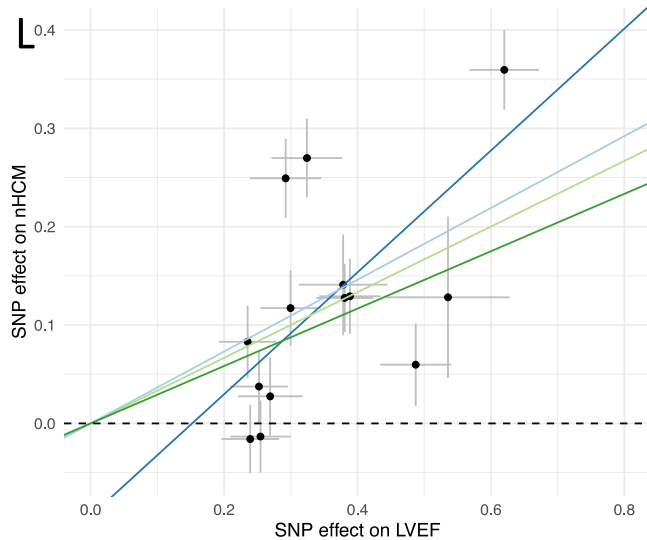
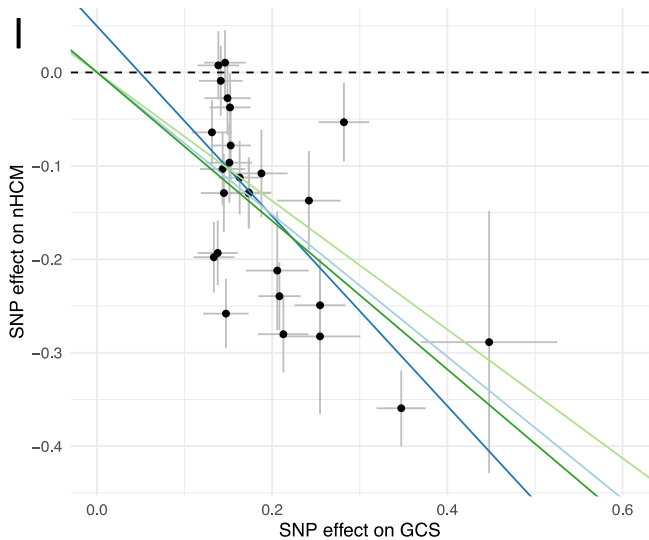
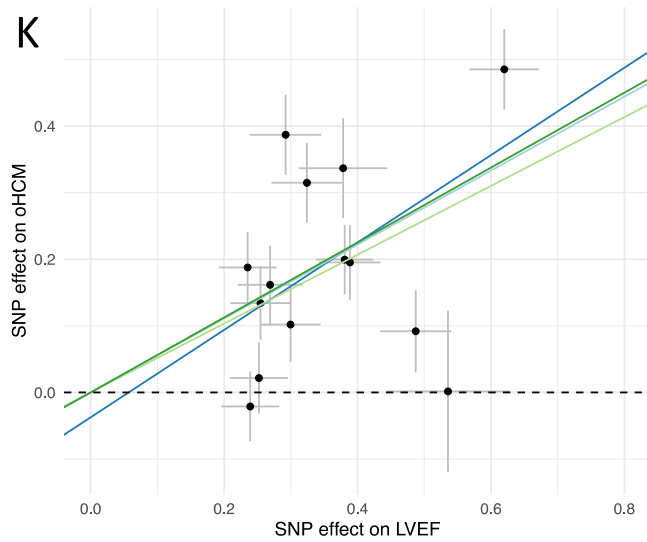
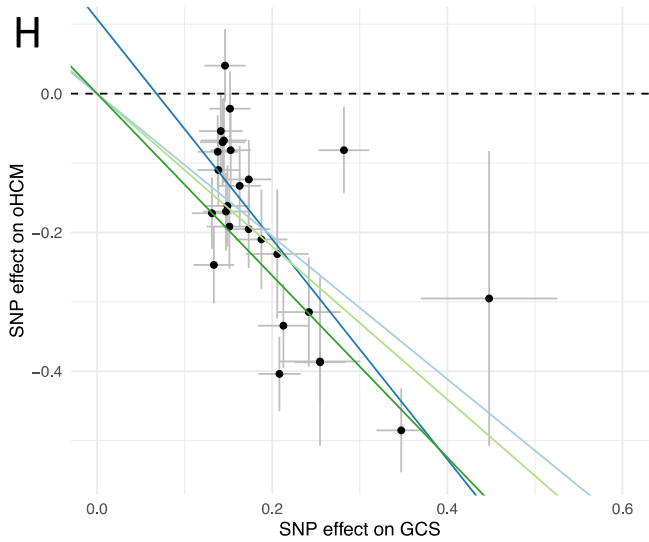
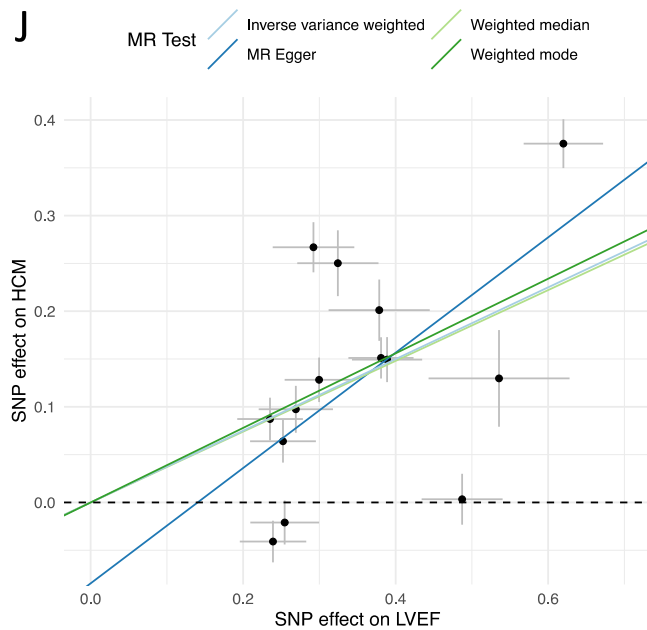
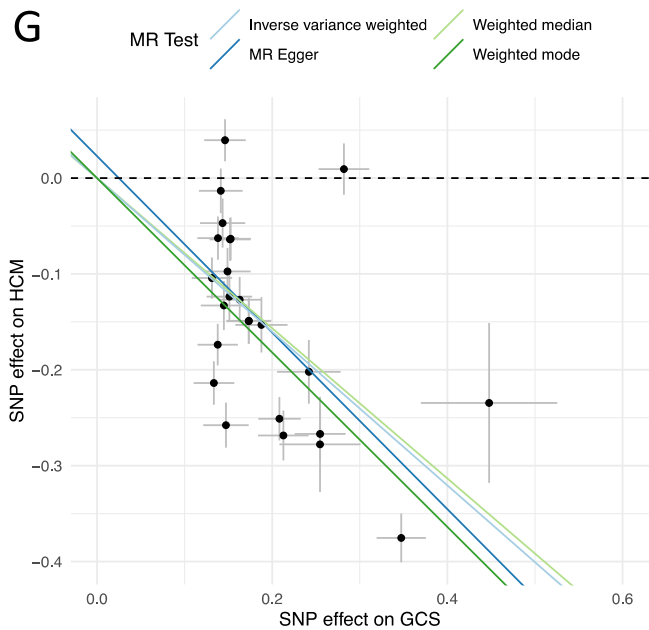


Supplementary Figure 18: Meta-analysis Manhattan plots for the exploratory exome-wide rare loss of function variant gene-based burden testing. Analyses performed with a Fisher Exact Test (or REGENIE for UK Biobank) using 2 minor allele frequency thresholds (MAF<0.001 and MAF<0.0001) in 4 datasets, followed by inverse variance weighting (IVW) and effective sample size weighting (SSW) meta-analyses. All protein-coding genes with at least 1 loss-of-function variant carrier in UKB, or at least one case and one control carrier in any other dataset were included in the meta-analyses (16,312 genes for MAF<0.0001, and 16,508 for MAF<0.001). Genes showing consistent association with HCM (*MYBPC3* and *ALPK3*) at the exome-wide threshold ($3 \times 10^{-6} = 0.05/16,508$) are flagged in black. *SVIL* is also flagged in green. P values shown are not corrected for multiple testing. See Supplementary Tables 17 and 18 for full summary statistics.

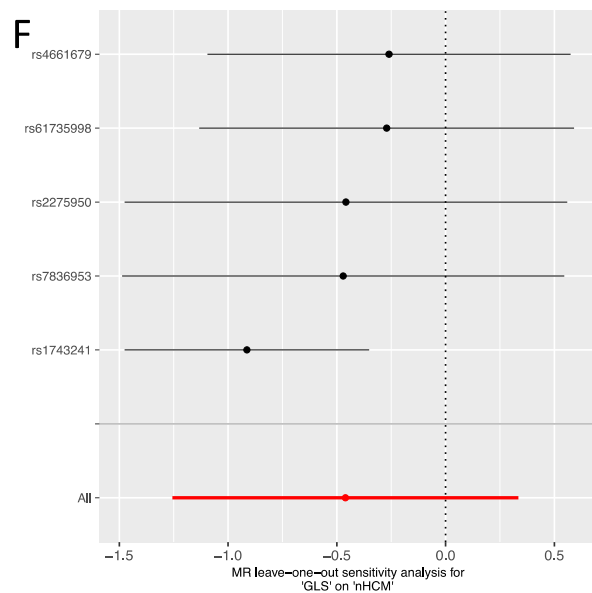
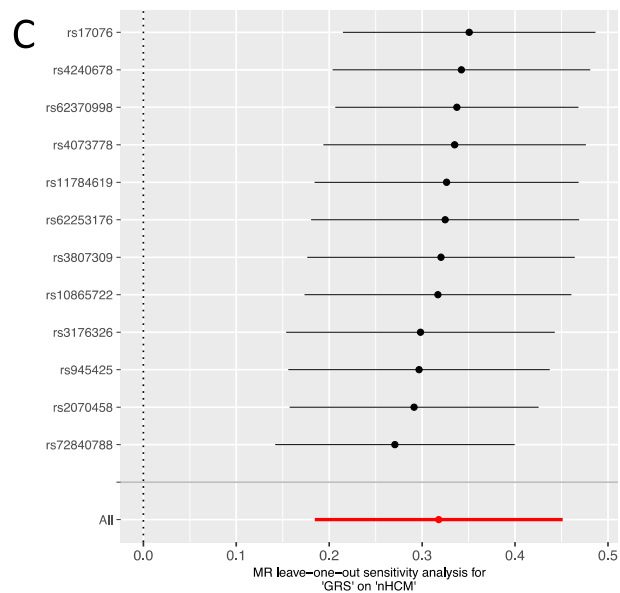
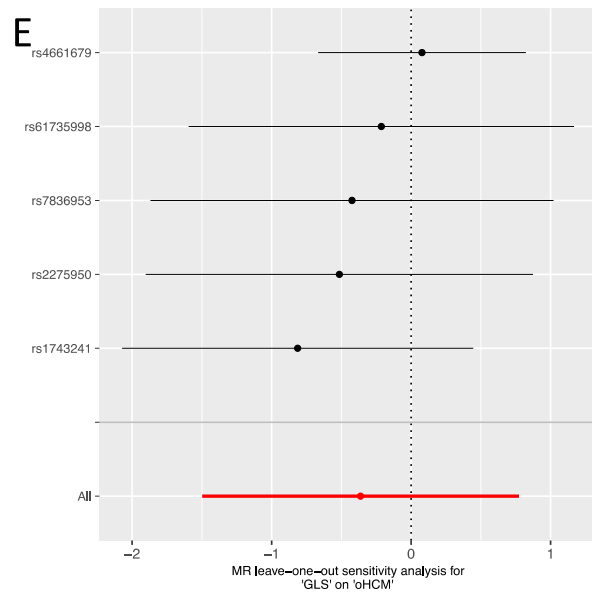
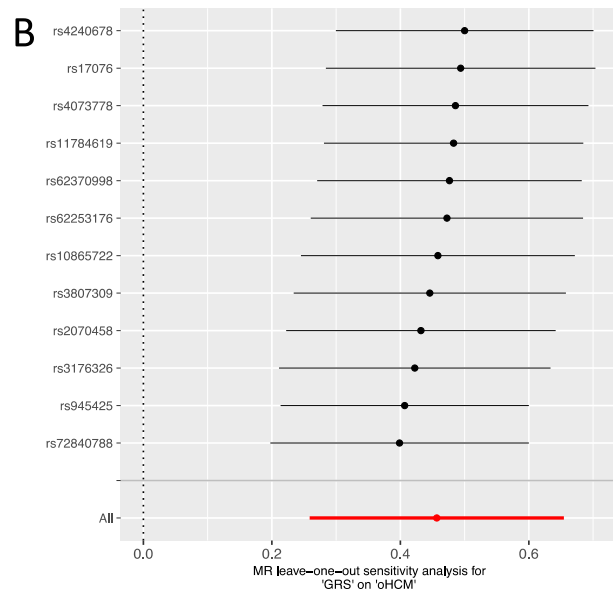
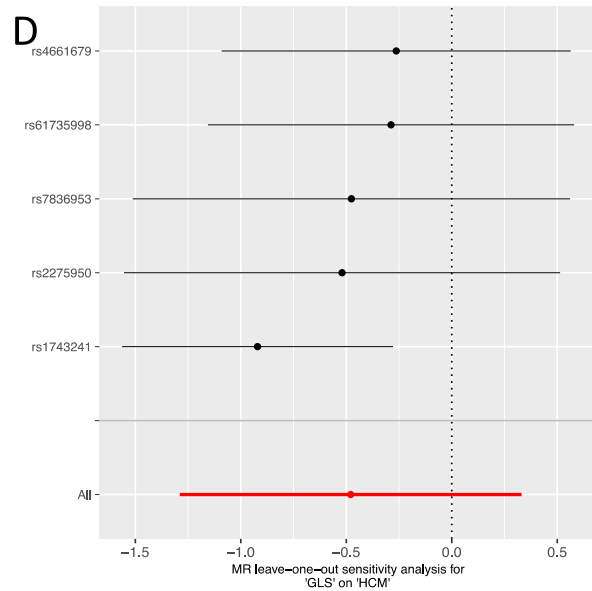
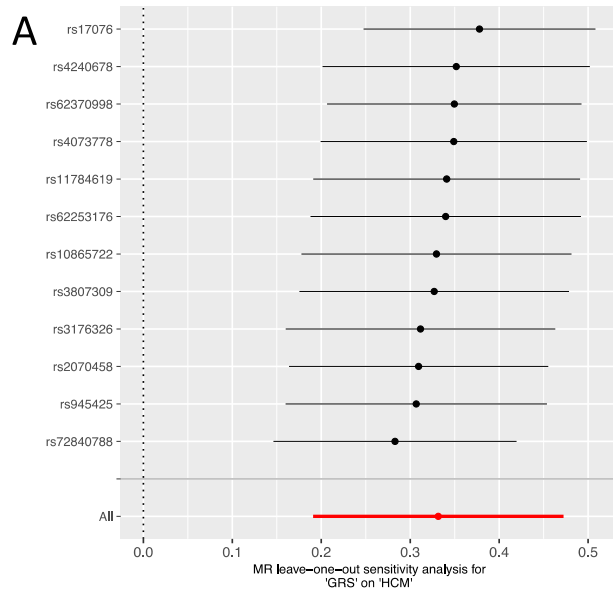


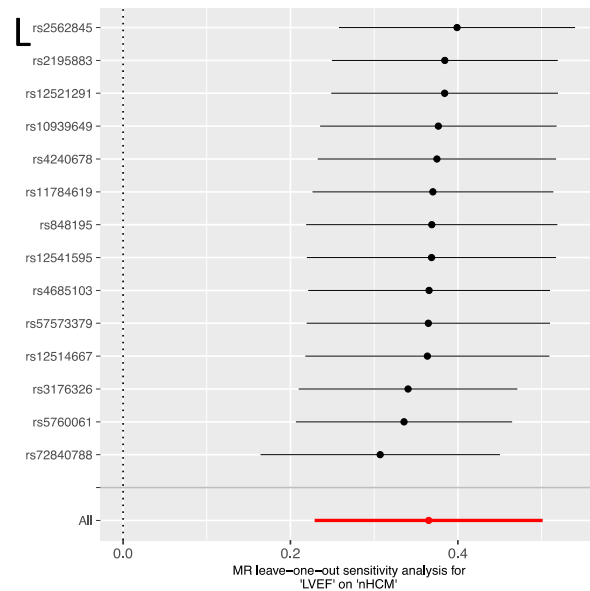
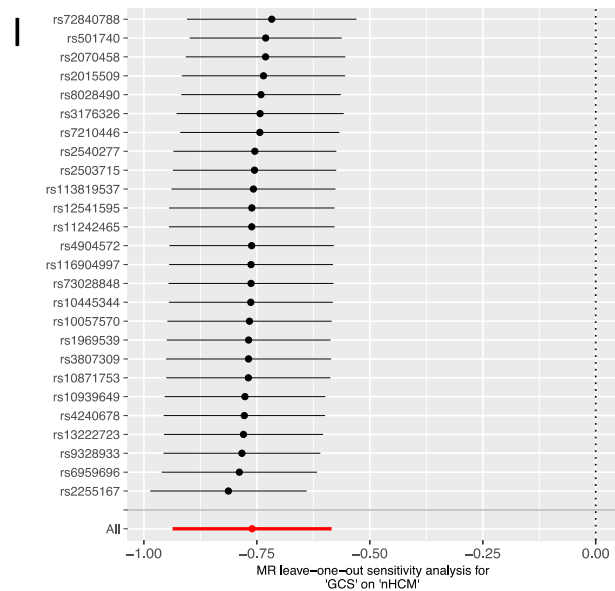
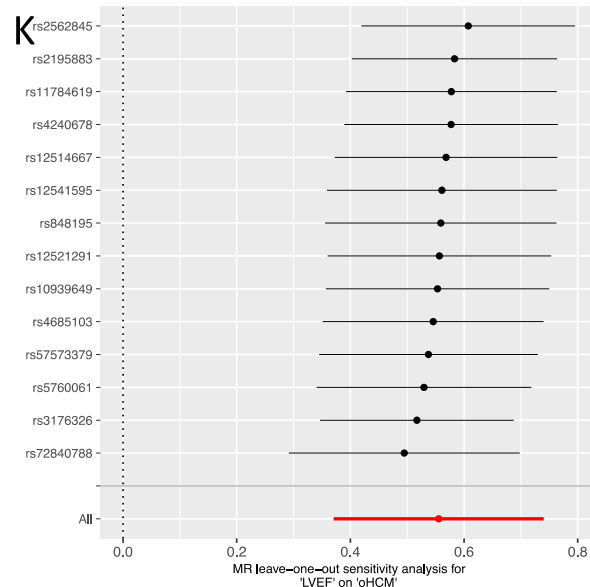
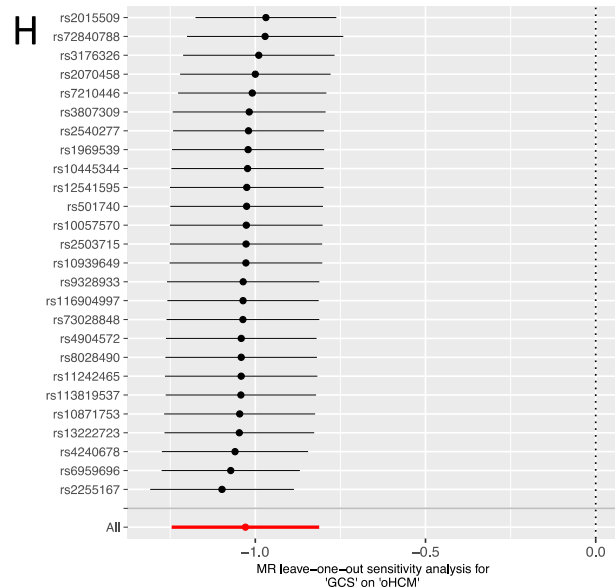
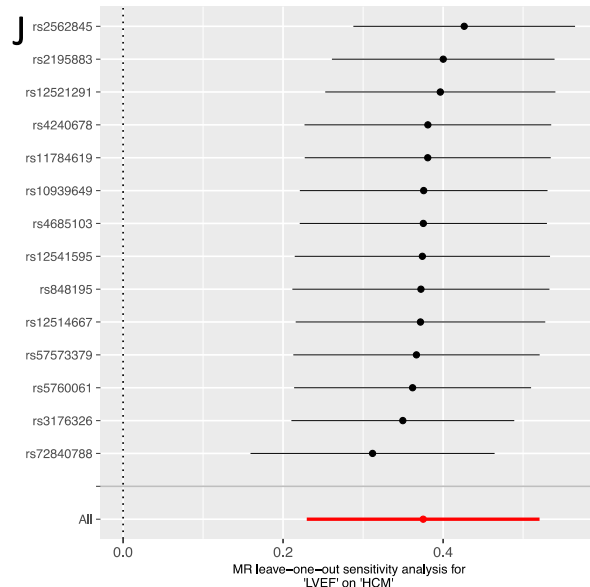
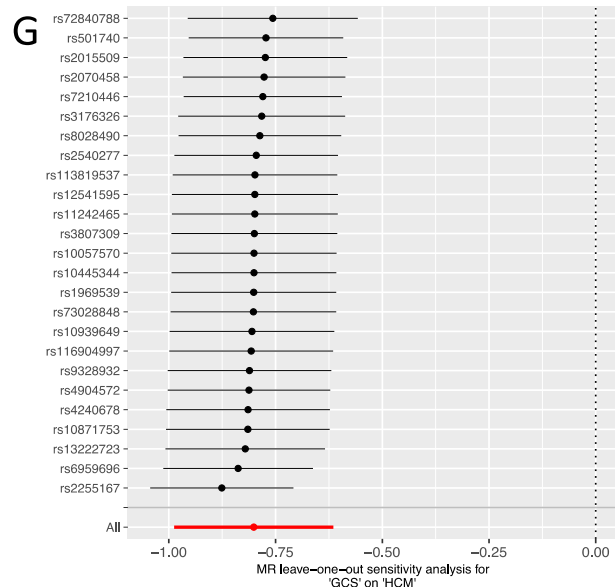
Supplementary Figure 19: Forest plot of rare *SVIL* loss-of-function variants association with HCM, including odds ratio (OR), 95% confidence interval (CI), and P value for each of the 4 datasets (squares representing point estimate of OR, and error bars represent the 95%CI of the OR) and an inverse variance weighted meta-analysis (diamond representing the extent of the 95%CI of the OR). Counts shown at the left side of the forest plot refer to carriers/non-carriers. Total sample size included 2,502 cases and 486,217 controls. OR and P value derived using a Fisher Exact Test for BRRD, GeL and Oxford Lab, and using REGENIE for UKB. BRRD, BioResource for Rare Disease; GeL, Genomics England; UKB, UK Biobank.



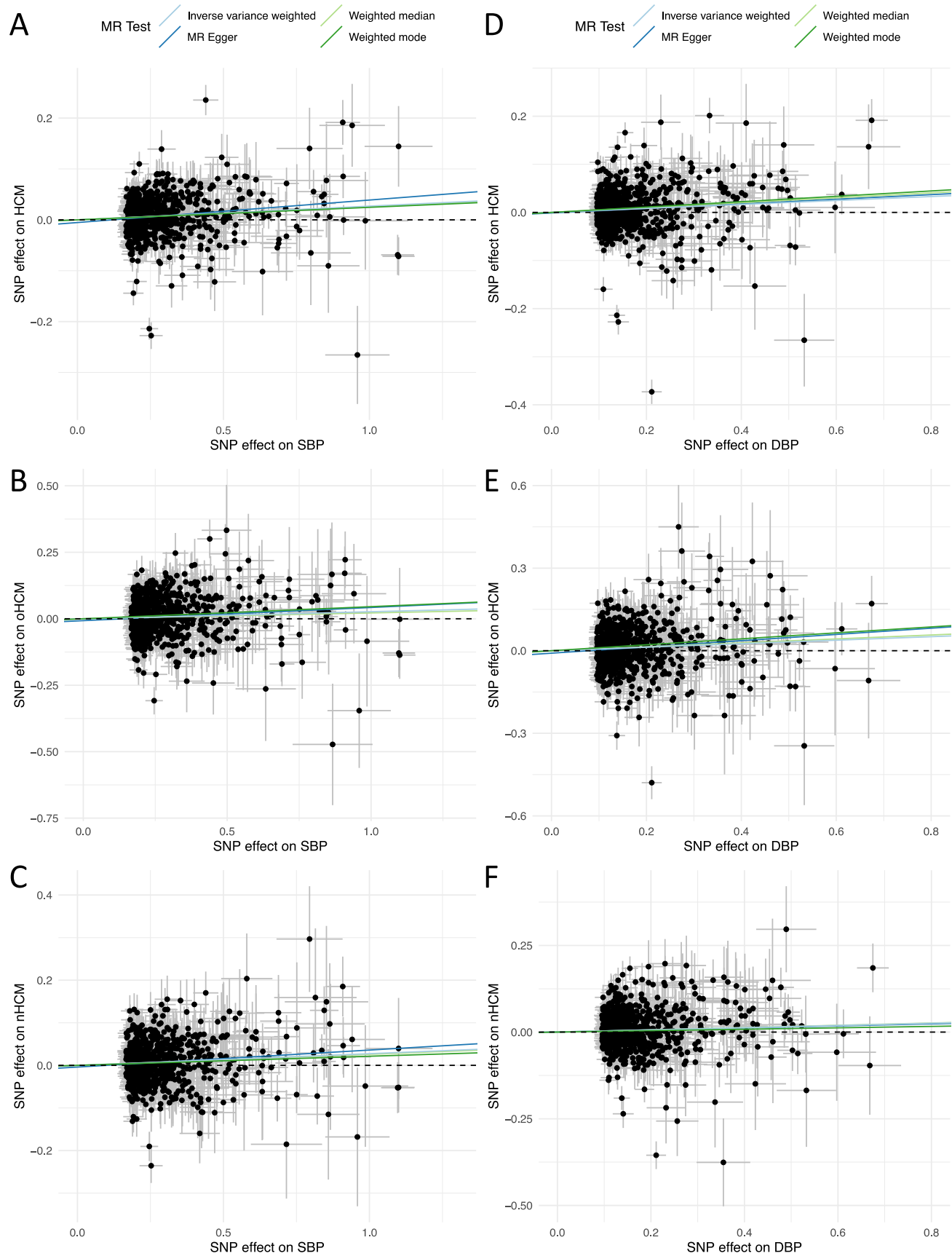


↑ **Supplementary Figure 20:** SNP effects of contractility on risk of HCM, oHCM and nHCM. Mendelian randomization (MR) analysis of LV contractility (exposure) on risk of hypertrophic cardiomyopathy (HCM, top panels) and its obstructive (oHCM, middle panels) and non-obstructive (nHCM, bottom panels) forms (outcomes). Genetic instruments for LV contractility were selected from the present GWAS of global strain in the radial (GRS, panels A-C), longitudinal (GLS, panels D-F) and circumferential (GCS, panels G-I) directions, and left ventricular ejection fraction (LVEF, panels J-L), in up to 36,083 participants of the UKB without cardiomyopathy and with available CMR. The outcome HCM GWAS included 5,900 HCM cases vs. 68,359 controls. Of those, 964 cases and 27,163 controls were included in the oHCM GWAS, and 2,491 cases and 27,109 were included in the nHCM GWAS. The effects (β) of the exposure variable-increasing allele at all independent SNPs reaching $P < 5 \times 10^{-8}$ are plotted as datapoints and associated standard errors are represented as lines extending from the datapoints. The x-axis represents the β for the exposure variable (contractility: GRS, GLS, GCS, or LVEF) and the y-axis represents the β for the outcome variable (HCM, oHCM, or nHCM). Lines represent the causal effects using 4 MR models (legend on top applicable for all panels). Each marker represents a single SNP, where the center and bars represent the regression coefficient and standard errors, respectively. Note that a positive β for LVEF and GRS represents *increased* contractility, while for GCS and GLS it reflects *decreased* contractility since GCS and GLS are negative measures where increasingly negative values reflect increase in contractility. See Figure 5 and Supplementary Table 21 for full MR results.





↑ **Supplementary Figure 21:** Leave-one-out analyses for Mendelian randomization (MR) of LV contractility on risk of hypertrophic cardiomyopathy (HCM, top panels) and its obstructive (oHCM, middle panels) and non-obstructive (nHCM, bottom panels) forms. Genetic instruments for LV contractility were selected from the present GWAS of global strain in the radial (GRS, panels A-C), longitudinal (GLS, panels D-F) and circumferential (GCS, panels G-I) directions, and left ventricular ejection fraction (LVEF, panels J-L). The IVW MR method is used to estimate the causal effect of contractility on HCM (x-axis) leaving out one SNP at a time (y-axis) to assess whether the association is driven by this single SNP. The IVW effect estimate (center mark) is shown along with the standard error (bars). The overall effect estimate for each contractility measure is shown in the corresponding panel in red. All associations are significant overall and when excluding each SNP at a time, except for GLS.



↑ **Supplementary Figure 22:** SNP effects of blood pressure on risk of HCM, oHCM and nHCM.

Mendelian randomization (MR) analysis of systolic (SBP, panels A-C) and diastolic (DBP, panels D-F) blood pressure (exposure) on risk of hypertrophic cardiomyopathy (HCM, top panels A and D) and its obstructive (oHCM, middle panels B and E) and non-obstructive (nHCM, bottom panels C and F) forms (outcomes). Genetic instruments for SBP and DBP were selected from a published GWAS of blood pressure including up to 801,644 individuals.⁴ The outcome HCM GWAS included 5,900 HCM cases vs. 68,359 controls. Of those, 964 cases and 27,163 controls were included in the oHCM GWAS, and 2,491 cases and 27,109 were included in the nHCM GWAS. The effects (β) of the exposure variable-increasing allele at all independent SNPs reaching $P < 5 \times 10^{-8}$ are plotted as datapoints and associated standard errors are represented as lines extending from the datapoints. The x-axis represents the β for the exposure variable (SBP or DBP) and the y-axis represents the β for the outcome variable (HCM, oHCM, or nHCM). Lines represent the causal effects using 4 MR models (legend on top applicable for all panels). Each marker represents a single SNP, where the center and bars represent the regression coefficient and standard errors, respectively. See Figure 5 and Supplementary Table 21 for full MR results.

Supplementary Note

HCM GWAS study cohorts

The HCM meta-analysis included 7 cohorts, described below and in Supplementary Table 1.

The Hypertrophic Cardiomyopathy Registry (HCMR). Details on the HCMR cohort are described in Harper et al⁵ and Neubauer et al⁶ Briefly, HCMR recruited 2755 incident HCM cases with evidence of unexplained left ventricular hypertrophy (LV wall thickness, LVWT > 15mm). After QC and removal of cases common to both HCMR and other cohorts described below, 2431 HCM cases were available for analysis. HCM cases underwent gene panel sequencing; variants identified within 8 core sarcomere genes (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1* and *TPM1*) were classified using the American College of Medical Genetics and Genomics (ACMG) guidelines.⁷ HCM cases were dichotomised into sarcomere-positive (n=850) and sarcomere-negative (N=1,546) groups using a classification framework previously reported in Neubauer et al.⁶ An all-comer analysis evaluated 2,431 HCM cases and 40,283, age- and sex-matched, UK Biobank (UKB) participants (application 11223) as controls. Separate sarcomere-positive and sarcomere-negative HCM analyses were performed using randomly allocated non-overlapping 20K UKB controls. Analyses were performed with logistic regression using SNPTEST⁸ v2.5.4-beta3 with the Newton-Raphson method adjusting for the first ten ancestry informative principal components (PCs). LD score regression⁹ intercept was used to adjust the standard errors for all comer and sarcomere-negative analyses. GC Lambda after correction for both all-comer and sarcomere-negative datasets was 1.05.

Canadian HCM cohort. Unrelated cases diagnosed with HCM with included from the Toronto General Hospital HCM clinic, the Montreal Heart Institute (MHI) Cardiovascular Genetics Clinic, the MHI Biobank and the London Health Sciences Centre. All cases had a clinical diagnosis of HCM (LVWT>15mm or LVWT>13mm in presence of family history of HCM or in presence of a pathogenic or likely pathogenic variant in the 8 core sarcomere genes). Cases were excluded if they had syndromic HCM, metabolic disease, or had >1 sarcomeric pathogenic or likely pathogenic variants (homozygous, compound heterozygous or digenic). Cases underwent targeted sequencing of genes associated with HCM, as per local clinical practice at the time of genetic testing. Rare variants detected through sequencing in each of the contributing cohorts of this study were assessed centrally for pathogenicity by the Oxford laboratory using the classification framework previously reported in Neubauer et al.⁶ Controls without a diagnosis of HCM were included from the MHI Biobank and the London Health

Sciences Centre. All cases and controls were genotyped on the Illumina Global Screening Array (GSA). QC was performed as previously described¹⁰, followed by imputation with Minimac4 on the Michigan Imputation server using the Haplotype Reference Consortium r1.1.¹¹ A total of 1,035 cases and 13,889 controls were then included in the association analysis using SNPTEST v2.5 with the score method adjusting for the first 20 genotypic PCs. In addition to the primary all-comer analysis, a case-control analysis of sarcomere-positive (N=261) and sarcomere-negative (N=582) was performed using random allocation of controls.

Netherlands HCM cohort. The Netherlands cohort was previously described.¹⁰ Cases with HCM were included from Amsterdam University Medical Center, Erasmus Medical Center and the University Medical Center Groningen. The control group consisted of a previously genotyped Dutch control population from project Mine (cohort NL4, as described by van Rheenen et al¹²). All cases and controls underwent genome-wide array genotyping on an Illumina Infinium BeadChip. Following QC and imputation, 999 unrelated HCM cases and 2,117 controls were included in the association analysis using SNPTEST v.2.5.2 with the score method correcting for the first 3 PCs. HCM cases were dichotomized into sarcomere-positive (N=436) and sarcomere-negative (N=538) as described for the HCMR cohort and stratified cases control analyses were performed using randomly allocated controls.

Genomics England 100K Genome Project (GEL). A GWAS was performed for HCM cases and healthy unrelated controls recruited to the 100K Genome Project (GEL)¹³ using whole-genome sequencing data. Cases were defined as individuals recruited to GEL with a primary clinical diagnosis of hypertrophic cardiomyopathy. Controls were selected from healthy relatives recruited to GEL, in whom the proband did not have any cardiac phenotype and matched 5:1 with cases for sex. SNP-level QC excluded variants with a sequencing depth <10, median GQ<15, AB ratio <0.25, missingness >0.05, MAF <0.01, differential missingness between cases and controls $P < 10^{-5}$, and Hardy-Weinberg equilibrium test in unrelated controls $P < 10^{-5}$. A total of 471 cases and 2,355 controls were included in the association analysis. A mixed-model analysis adjusted for age, sex and the first 10 genotypic PCs was used to account for distant relatedness and population stratification using SAIGE version 0.42.1.¹⁴ All analyses were performed in the Genomics England Research Environment under project ID 291. Since GEL predominantly included sarcomere-negative HCM cases (442/471 cases), no sarcomere-positive analysis was performed.

Royal Brompton Hospital HCM cohort. Unrelated British HCM cases from the Royal Brompton & Harefield Hospitals NHS Trust Cardiovascular Research Biobank, and healthy controls from the UK Digital Heart Project¹⁵ were included. The presence of HCM was excluded in all controls using cardiac magnetic resonance (CMR) imaging. Genotyping was performed in 4 batches at 3 centres (the Sanger Institute, London; Duke-NUS in Singapore; and King's college London) using the Illumina Human OmniExpress Beadchip. All downstream QC and data analyses were performed centrally. Data quality control was performed using Plink v1.9 and in-house scripts. We first mapped all SNPs to the positive strand of GRCh37 build. Pre-imputation SNP-level QC excluded SNPs with MAF<0.01, and Hardy-Weinberg equilibrium test $P<10^{-7}$. Sample QC excluded samples with sex mismatch, heterozygosity rate >3 standard deviations (SD) from the mean, and missingness rate >0.05. We aligned study genotypes with those from the HapMap3 cohort, and used PC analysis to identify genotypically Caucasian samples to take forward for imputation, based on cut-offs defined by the mean and SD of each HapMap ethnicity cohort. Batches were merged pre-imputation, excluding SNPs which failed QC checks in any batch. We also checked for differential missingness and excluded any SNPs with differential missingness test $P<1\times 10^{-7}$. We used the combined UK10K + 1000 Genomes Project dataset as the imputation reference panel. This has a proven record of accuracy for UK study cohorts, across a range of allele frequencies.¹⁶ We pre-phased study genotypes using SHAPEIT¹⁷ (v2.r790) and imputed the genotype data using IMPUTE2⁸ (v2.3.2). We finally performed several post-imputation QC steps per-SNP and per-sample, excluding SNPs with an INFO criterion score <0.4, MAF<1%, Hardy-Weinberg equilibrium $P<10^{-7}$. We then removed related individuals from the dataset, using a PI-HAT cut-off of 0.187. After the QC steps, we performed another PC analysis to generate the eigenvectors used as covariates in the regression analysis. Association test was performed by SNPTEST v2.5.4 with the Newton-Raphson method adjusted by sex, age, and top 3 PCs. Stratified analyses for sarcomere-positive (N=118) and sarcomere-negative (N=330) HCM cases using randomly allocated non-overlapping controls were also performed.

Italian HCM cohort. Individuals with HCM were diagnosed or referred to the Cardiomyopathy Unit of the Careggi University Hospital (Florence, Italy). HCM diagnosis was defined as unexplained LV hypertrophy with a maximal LV wall thickness >13mm on echocardiography or CMR, complemented by family history and genotype to enable an informed diagnosis. Unrelated healthy Italian controls of European ancestry were obtained from the multicentre international HYPERGENES study.¹⁸ A total of

1,293 control samples were selected, comprising 561 hypertensives (defined as diastolic blood pressure ≥ 90 mmHg, systolic blood pressure ≥ 140 mmHg, or on antihypertensive treatment before the age of 50) and 732 normotensives. As expected by the inclusion criteria, the mean age was significantly higher in normotensives than in hypertensives. Since aging is associated with an increased prevalence of hypertension, we selected exclusively hyper-normal controls older than 50-year-old allowing for the exclusion of subjects that developed hypertension at a later age. Cases were genotyped at King's College London by Illumina OmniExpress Beadchip and controls were genotyped by Human1M-Duo v3.0 array. All downstream QC and data analyses were performed centrally. Data quality control was performed using Plink v1.9 and in-house scripts. We first mapped all SNPs to the positive strand of GRCh37 build. Pre-imputation SNP-level QC excluded SNPs with $MAF < 0.01$, and Hardy-Weinberg equilibrium test $P < 10^{-7}$. Sample QC excluded samples with sex mismatch, heterozygosity rate > 3 standard deviations (SD) from the mean, and missingness rate > 0.05 . We aligned study genotypes with those from the HapMap3 cohort, and used PC analysis to identify genotypically Caucasian samples to take forward for imputation, based on cut-offs defined by the mean and SD of each HapMap ethnicity cohort. Case and control batches were combined after pre-imputation steps. We also checked for differential missingness and excluded any SNPs with differential missingness test $P < 1 \times 10^{-7}$. Genome wide imputation was performed using eagle v2.4 phasing, Minimac4 1.5.7 and HRC r1.1 reference panel implemented on the Michigan Imputation Server v1.2.4.¹¹ We finally performed a number of post-imputation QC steps per-SNP and per-sample, excluding SNPs with a Minimac $R^2 < 0.5$, $MAF < 1\%$, Hardy-Weinberg equilibrium $P < 10^{-7}$. We then removed related individuals from the dataset, using a PI-HAT cut-off of 0.187. After the QC steps, we performed another PC analysis to generate the eigenvectors used as covariates in the regression analysis. Association test was performed by SNPTEST version 2.5.4 with the expected method adjusted by age, sex and top 10 PCs. Stratified analyses for sarcomere-positive ($N=111$) and sarcomere-negative ($N=183$) HCM cases using randomly allocated non-overlapping controls were also performed. Note that the total number of HCM cases included in the sarcomere status stratified analyses was higher than the number of cases included in the all-comer analysis which was completed at an earlier phase.

The BioResource for Rare Disease (BRRD). The BRRD cohort, a pilot study of the Genomics England 100,000 Genomes Project (GEL), has been described elsewhere.^{5,19} In brief, following QC, 239

sarcomere negative HCM cases and 7,203 controls were available for analyses. Logistic regression analysis was performed using SAIGE (v0.29.4.2)¹⁴ with correction for the first three genotypic PCs.

Automated cardiac magnetic resonance (CMR) image analysis

The LV traits GWAS was performed using CMR data from 39,559 UK Biobank participants (36,083 following genotypic and imaging QC and exclusions for co-morbidities). LV trait data were computed using a machine learning analysis of CMR images as described below.

We segmented short and long axis cine images using a fully convolutional network²⁰, trained on manual annotations of 3,975 subjects, as previously reported.¹⁰ All image segmentations were manually quality controlled by an experienced cardiologist. The segmentation screenshots for short-axis and long-axis end-diastolic and end-systolic frames were visually inspected. Bad segmentations, images with insufficient coverage of the LV or missing anatomical structures were discarded. For motion tracking, subjects with failed image registration or outlier peak global strain values (positive circumferential or longitudinal strain, or negative radial strain) were discarded.

LV end-diastolic (LVEDV) and end-systolic (LVESV) volumes and ejection fraction (LVEF, defined as $[LVEDV - LVESV] / LVEDV$) were derived from these segmentations. The LV myocardial mass (LVM) was calculated from the myocardial volume using a density of 1.05 g/mL. LV concentricity (LVconc) was defined as $LVM / LVEDV$. The LV wall thickness (WT) was measured at end-diastole. The myocardium was divided into 16 segments, according to the American Heart Association nomenclature.²¹ Maximum and mean wall thickness was derived for each AHA segment. Overall mean wall thickness (meanWT) per subject was calculated from the mean of each of the 16 mean wall thickness measurements. Overall maximum wall thickness (maxWT) was the maximum wall thickness measurement from any segment.

Motion tracking was performed using non-rigid image registration between successive timeframes, using the MIRTk toolkit.²² Inter-frame displacement fields were composed to obtain the displacement with respect to a reference frame (the end-diastolic frame, or frame 0). To avoid drift effect due to accumulation of registration errors²³, motion tracking is performed twice – along the forward direction (tracking starting from frame 0 to frames 1, 2, 3,...) and backward direction (tracking from frame 0 to frames $T-1$, $T-2$, $T-3$, ...) where T denotes the total number of frames per cardiac cycle). The average displacement field is calculated by weighted averaging of the forward and backward

displacement field such that, for a frame at the start of the cardiac cycle, the forward displacement field will have a higher weight, whereas for a frame at towards the end of the cardiac cycle, the backward displacement field will have a higher weight. Four image slices were used for this: in the short axis, a basal slice at 75% LV location, a mid-cavity location and an apical slice at 25% LV location, and in the long axis, a 4-chamber view as previously described.^{24,25} The myocardial contours on the three slices were divided into the 16 segments. Based on the displacement field from motion tracking, myocardial contours at the end-diastolic frame were warped onto each time frame of the cardiac cycle. Circumferential, radial and longitudinal strains were calculated for each time frame based on the change of length for each line segment²⁶, using the equation $E = \frac{\Delta L}{L}$, where E is the strain, ΔL is the change in length of the line segment, and L is the starting (end-diastolic) length of the line segment. Peak strain for each AHA segment, and *global* peak strain were calculated in radial (strain^{rad}), longitudinal (strain^{long}) and circumferential (strain^{circ}) directions.²⁶ Note that in contrast to strain^{rad}, both strain^{long} and strain^{circ} are negative values where more negative values reflect higher contractility. As such, we will sometimes present results for -strain^{long} and -strain^{circ} to facilitate interpretation of directionality.

LV traits GWAS: genotyping and imputation

Details of the genotyping and QC strategy employed by UKB were previously published.²⁷ Genotypes were called from 2 purpose-built arrays: the UK Biobank Axiom Array (825,927 markers) and UK BiLEVE Axiom Array (807,411 markers). These directly called genotypes were imputed to more than 90 million variants using the haplotype reference consortium (HRC) and the UK10K reference panels. See UKB website for details of imputation and genotype quality control performed by a collaborative group headed by the Wellcome Trust Centre for Human Genetics. We excluded samples with outlying heterozygosity or missingness rates, as defined by UKB and those with mismatches between the genotypic and recorded sex, as well as those with aneuploidy or excess kinship, as defined by UKB. We excluded SNPs failing UKB protocols (filtered per batch by Hardy-Weinberg equilibrium and missingness)²⁷, those with imputation INFO score <0.3 or MAF <0.01, or with missingness >0.1.

References for Supplementary Figures and Note

- 1 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
- 2 Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet* **53**, 1527-1533, doi:10.1038/s41588-021-00945-5 (2021).
- 3 Reichart, D. *et al.* Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science* **377**, eabo1984, doi:10.1126/science.abo1984 (2022).
- 4 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, doi:10.1038/s41588-018-0205-x (2018).
- 5 Harper, A. R. *et al.* Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat Genet* **53**, 135-142, doi:10.1038/s41588-020-00764-0 (2021).
- 6 Neubauer, S. *et al.* Distinct Subgroups in Hypertrophic Cardiomyopathy in the NHLBI HCM Registry. *J Am Coll Cardiol* **74**, 2333-2345, doi:10.1016/j.jacc.2019.08.1057 (2019).
- 7 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, doi:10.1038/gim.2015.30 (2015).
- 8 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913, doi:10.1038/ng2088 (2007).
- 9 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 10 Tadros, R. *et al.* Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat Genet* **53**, 128-134, doi:10.1038/s41588-020-00762-2 (2021).
- 11 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
- 12 van Rheenen, W. *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* **48**, 1043-1048, doi:10.1038/ng.3622 (2016).
- 13 Genomics England: The National Genomics Research and Healthcare Knowledgebase v5. (2019). <doi:10.6084/m9.figshare.4530893.v5>.
- 14 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341, doi:10.1038/s41588-018-0184-y (2018).
- 15 Schafer, S. *et al.* Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet* **49**, 46-53, doi:10.1038/ng.3719 (2017).
- 16 Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111, doi:10.1038/ncomms9111 (2015).
- 17 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).

- 18 Salvi, E. *et al.* Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension* **59**, 248-255, doi:10.1161/HYPERTENSIONAHA.111.181990 (2012).
- 19 Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102, doi:10.1038/s41586-020-2434-2 (2020).
- 20 Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* **20**, 65, doi:10.1186/s12968-018-0471-x (2018).
- 21 Cerqueira, M. D. *et al.* Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **105**, 539-542, doi:10.1161/hc0402.102975 (2002).
- 22 Rueckert, D. *et al.* Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* **18**, 712-721, doi:10.1109/42.796284 (1999).
- 23 Tobon-Gomez, C. *et al.* Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Med Image Anal* **17**, 632-648, doi:10.1016/j.media.2013.03.008 (2013).
- 24 Schuster, A. *et al.* Cardiovascular magnetic resonance feature-tracking assessment of myocardial mechanics: Intervendor agreement and considerations regarding reproducibility. *Clin Radiol* **70**, 989-998, doi:10.1016/j.crad.2015.05.006 (2015).
- 25 Taylor, R. J. *et al.* Myocardial strain measurement with feature-tracking cardiovascular magnetic resonance: normal values. *Eur Heart J Cardiovasc Imaging* **16**, 871-881, doi:10.1093/ehjci/jev006 (2015).
- 26 Puyol-Antón, E. *et al.* in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 1139-1143.
- 27 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).