nature genetics

Letter

https://doi.org/10.1038/s41588-024-02072-3

Meta-analysis of genome-wide associations and polygenic risk prediction for atrial fibrillation in more than 180,000 cases

Received: 12 March 2022

Accepted: 30 December 2024

Published online: 6 March 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Atrial fibrillation (AF) is the most common heart rhythm abnormality and is a leading cause of heart failure and stroke. This large-scale meta-analysis of genome-wide association studies increased the power to detect single-nucleotide variant associations and found more than 350 AF-associated genetic loci. We identified candidate genes related to muscle contractility, cardiac muscle development and cell–cell communication at 139 loci. Furthermore, we assayed chromatin accessibility using assay for transposase-accessible chromatin with sequencing and histone H3 lysine 4 trimethylation in stem cell-derived atrial cardiomyocytes. We observed a marked increase in chromatin accessibility for our sentinel variants and prioritized genes in atrial cardiomyocytes. Finally, a polygenic risk score (PRS) based on our updated effect estimates improved AF risk prediction compared to the CHARGE-AF clinical risk score and a previously reported PRS for AF. The doubling of known risk loci will facilitate a greater understanding of the pathways underlying AF.

The genetic basis of AF has been studied for over two decades, with approaches ranging from family-based studies[1–4] to larger case-control genome-wide association studies (GWAS)[5–13]. In this study, we analyzed the largest collection of AF GWAS so far, which included a total of 181,446 AF cases and 1,468,899 controls. We meta-analyzed 68 summary-level results from more than 40 primary cohorts (Supplementary Table 1) representing eight groups of ancestry and ethnicity, including non-Finnish, non-Icelandic European (n_{AF} = 128,044), Icelandic (n_{AF} = 20,953), Finnish (n_{AF} = 17,325), East Asian (n_{AF} = 11,350), admixed African and African American (n_{AF} = 1,782), Hispanic (n_{AF} = 1,203), Brazilian (n_{AF} = 571) and South Asian (n_{AF} = 218) (Supplementary Table 2). We analyzed a total of 29,789,980 single-nucleotide variants and insertions and deletions, with 8,272 variant associations exceeding the genome-wide significance level of $P < 5 \times 10^{-8}$. Conditional and joint analysis of the combined ancestry meta-analysis results revealed 403 independently associated signals (Supplementary Table 3).

We identified 354 genetic loci with minor allele frequency (MAF) ≥ 1% and at least 500 kb between sentinel variants (Fig. 1 and Supplementary Table 4). Of these 354 loci, 135 (38.1%) were farther than 1 Mb away from previously reported[8,14–17] sentinel variants for AF (Supplementary Table 5). A large subset (299 out of 354) of sentinel variants were available in the validation cohort from the Million Veteran Program (MVP). The effect estimates showed a positive correlation across the primary meta-analysis and the validation GWAS (Extended Data Fig. 1a,b). Nearly all sentinel variants (293 out of 299) demonstrated consistent directions of effect for the primary and validation analyses, with 215 variants exceeding a nominal significance level ($P < 0.05$) (Supplementary Table 4). Most (64%) of the common lead variants mapped to intronic regions of a gene, and 31% mapped to intergenic regions. The remaining 5% were coding variants (predominantly missense variants). An exploratory analysis of low-frequency variants (MAF < 1%) revealed 14 genetic loci (Fig. 1 and Supplementary Table 6), as described in the Supplementary Note, Supplementary Table 7 and Extended Data Fig. 1c,d. Seven of the common lead variants showed significant heterogeneity of effect estimates by ancestry (Supplementary Table 4). Further ancestry-specific loci are presented in the Supplementary Note.

Bridging the gap from single-variant associations and genetic loci to candidate genes remains a challenge with any GWAS. Ultimately,

e-mail: ellinor@mgh.harvard.edu

https://doi.org/10.1038/s41588-024-02072-3

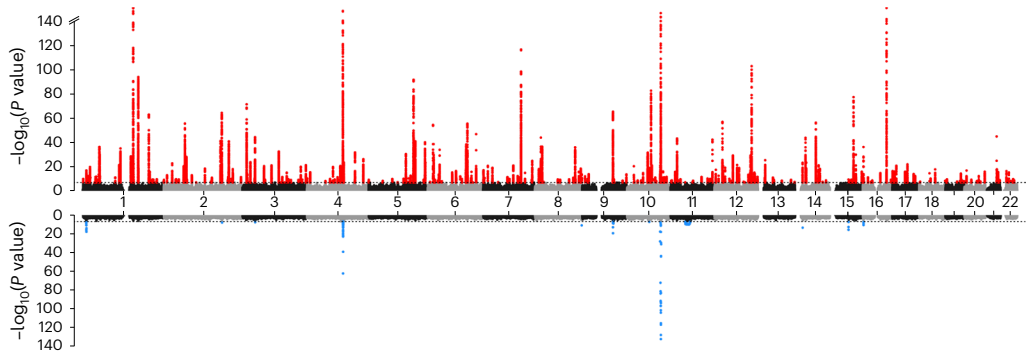


Fig. 1 | Miami plot of meta-analysis across 181,446 AF cases and 1,468,899 controls for common variants and low-frequency variants. The top panel depicts the common variant results with MAF ≥ 1%; significant loci are highlighted in red. The *y* axis for the common variant Miami plot is capped at $-\log_{10}(P) = 140$ for clarity. The signals that were capped are located on chromosome 1 (rs34515871, $P = 6.46 \times 10^{-154}$), chromosome 4 (rs6843082, $P = 1.74 \times 10^{-1141}$), chromosome 10 (rs11598047, $P = 1.95 \times 10^{-174}$) and chromosome 16 (rs2106261, $P = 7.24 \times 10^{-257}$). The bottom panel depicts the low-frequency results with MAF < 1%; significant loci are highlighted in blue. The meta-analysis included 181,446 AF cases and 1,468,899 controls. The genome-wide significance cutoff of $P < 5 \times 10^{-8}$ was applied to correct for multiple testing. The *P* values of the meta-analysis were calculated with the inverse-variance-weighted method.

functional effector genes, more so than AF-associated variants, have the potential to uncover novel biological mechanisms and identify drug targets for novel therapeutics. Therefore, we evaluated five different lines of evidence that prioritize candidate genes at GWAS loci: (1) a region-based method (multi-marker analysis of genomic annotation (MAGMA)[18]; Supplementary Table 8); (2) a similarity-based method (polygenic priority score (PoPS)[19]; Extended Data Fig. 2 and Supplementary Table 9); (3) cardiac gene expression (expression quantitative trait locus (eQTL); Supplementary Tables 10 and 11), (4) single-cell gene expression in cardiomyocytes (CMs) (single-nuclei RNA-sequencing (snRNA-seq), Extended Data Fig. 3a,b and Supplementary Table 12); and (5) predicted deleteriousness (Supplementary Tables 13 and 14). Findings from these five approaches are described in the Supplementary Note.

To integrate the results from each of the above modalities, we developed a simple approach called GenePrio (gene prioritization), which combines the five gene-level annotations to rank genes at each GWAS locus. The lines of evidence for each gene were summed (GenePrio sum) and the genes at each GWAS locus were ranked. Genes with at least two lines of evidence and the largest GenePrio sum at a locus are shown in Supplementary Table 15. The approach identified 139 GenePrio genes at AF loci. Most of the identified genes (56%) corresponded to the protein-coding gene closest to the sentinel variant (Extended Data Fig. 4). Ten of these genes had four supporting lines of evidence (Fig. 2a), and 40 genes were annotated with three lines of evidence (Fig. 2b). The remaining 89 genes were supported by two lines of evidence (Extended Data Fig. 5). Genes prioritized by GenePrio showed strong enrichment in rare variant association testing of loss-of-function and predicted damaging variants, based on recently analyzed whole-exome or whole-genome sequencing data[20] (Fig. 2c and Supplementary Table 16). Of note, five genes prioritized by GenePrio (*TTN*, *PKP2*, *FLNC*, *RBM20* and *CTNNA3*) showed Bonferroni significant (0.05/92 = 0.00054) associations in rare variant association testing (Fig. 2d and Supplementary Table 17).

We then ran colocalization analyses for all genes for which the top hit was an eQTL in atrial appendage, left ventricle, or left atrium tissue (Supplementary Tables 10 and 11). A gene set enrichment analysis of these 139 GenePrio genes revealed 164 enriched pathways (Extended Data Fig. 6) involving processes including muscle contractility, cardiac muscle development and cell–cell communication (Supplementary Table 18 and Extended Data Figs. 7 and 8). We additionally performed a Gene Ontology-based cluster analysis of the 139 genes after annotating them with cell-type-specific expression from the left atrium (Supplementary Table 19). The cluster analysis showed groups of Gene Ontology terms enriched for candidate genes related to actin binding, potassium channel activity and transcriptional regulation (Extended Data Fig. 9).

Our GenePrio approach identified recognized AF susceptibility genes, including *CAV1* (ref. 21), *SYNPO2L* (ref. 22), *TTN* (ref. 23), *CASQ2* (ref. 24), *CAMK2D* (ref. 25) and *MYH6* (ref. 26) (Fig. 2a). We also identified genes with less established links to AF. These genes, namely *FBXO32*, *IGF1R*, *HSPB7* and *ABLIM1*, are interesting candidates for functional follow-up and evaluation as novel therapeutic targets. *FBXO32* (F-box protein 32) encodes a protein known to be involved in muscle atrophy. eQTL data showed that increased expression was associated with a higher risk for AF. *IGF1R* (insulin-like growth factor 1 receptor) encodes a transmembrane receptor that binds insulin. Decreased expression was associated with increased risk for AF via eQTL. Mice that overexpress IGF1R develop cardiac hypertrophy[27]. *HSPB7* (heat shock protein family B (small) member 7) encodes a protein involved with the cardiac sarcomere and a binding partner of titin. Decreased expression of *HSPB7* was associated with an increased risk for AF in the eQTL results. Cardiac-specific knockdown of *Hspb7* in mice leads to embryonic lethality, implicating an important role in heart development[28]. *ABLIM1* (actin binding LIM protein 1) encodes a protein that binds to actin filaments and helps mediate interaction with targets in the cytoplasm. The eQTL results suggest that decreased expression of *ABLIM1* is associated with increased AF risk.

Next, we generated atrial CMs derived from induced pluripotent stem (iPS) cells and measured chromatin accessibility using assay for transposase-accessible chromatin with sequencing (ATAC–seq) and epigenetic modifications as assessed by histone H3 lysine 4 trimethylation (H3K4me3). At the transcription start sites of our 139 GenePrio genes, we found enhanced chromatin accessibility within the iPS-derived atrial CMs, compared to other tissues and cell types from the Encyclopedia of DNA Elements (ENCODE) (Fig. 2e). Moreover, we observed similar atrial CM enrichment when centering analyses around the 354 common AF sentinel variants (Fig. 2f). For example, in atrial iPS-CMs, we detected a marked enrichment in open chromatin and H3K4me3 peaks at several AF risk loci, including *TBX5*, *PITX2* and *HSPB7* (Extended Data Fig. 10).

Although most common genetic variants have modest effects on AF, the cumulative effect of risk variants across the genome can be harnessed through a PRS. To evaluate the gain of the increased sample size and the newly identified loci in this study, we created a polygenic score (details described in Methods) for AF (PRS_{AF}) from the GWAS presented here, excluding the Trøndelag Health Study (HUNT) and UK Biobank data ($n = 154{,}330$ cases). The PRS_{AF} was then evaluated using incident AF cases in both HUNT (2,474 incident AF cases; 50,283 controls) and UK Biobank (10,416 incident AF cases; 236,267 controls). To control

https://doi.org/10.1038/s41588-024-02072-3

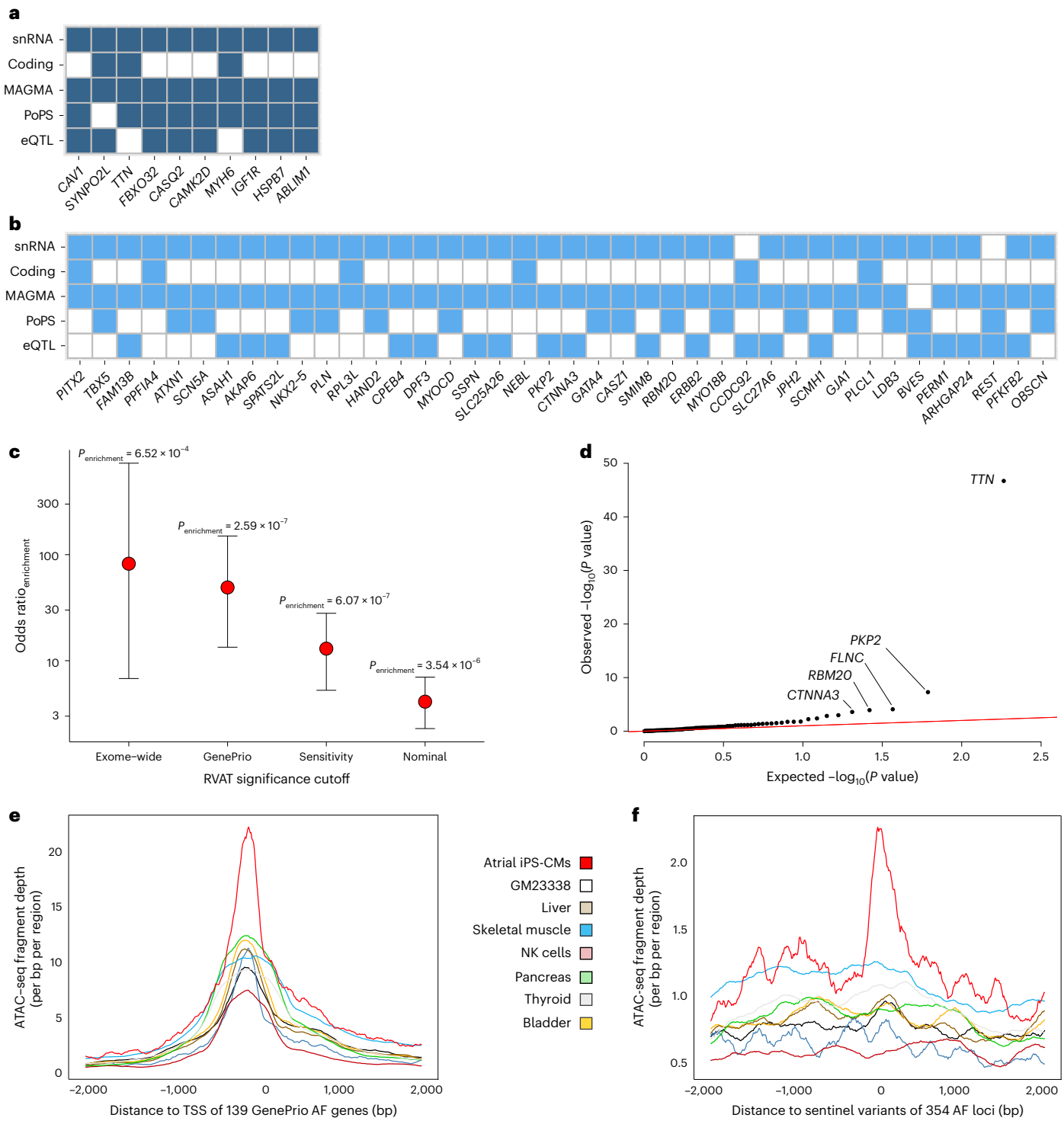


Fig. 2 | Prioritization of genes at common variant AF GWAS loci. a,**b**, Genes with four (**a**) and three (**b**) lines of evidence. Genes with two lines of evidence are shown in Extended Data Fig. 5. The five categories of evidence were snRNA-seq (marker gene for CMs in left atrium, labelled as snRNA); Coding (significant coding variant (loss-of-function or missense with high deleterious score)); MAGMA (significant in MAGMA analysis); PoPS (high PoPS score); and eQTL (significant eQTL for sentinel variant in cardiac tissue). **c**, Enrichment of GenePrio genes within four nested sets of rare variant association test (RVAT) genes for AF. The four *P* value thresholds used to determine the sets of RVAT genes for the enrichment analysis were defined as 'exome-wide' ($P < 0.05/11{,}242$), 'GenePrio' ($P < 0.05/92$), 'sensitivity' ($P < 0.005$) and 'nominal' ($P < 0.05$). The total number of genes tested in RVAT was $n = 11{,}242$ and the overlap of GenePrio genes with tested genes in RVAT was $n = 92$. The significance levels of 0.005 and 0.5 were included to show the trend of the enrichment analysis from higher to lower confidence RVAT results. A two-sided Fisher's exact test was performed. Shown as dots are the odds ratio and as error bars the 95% confidence interval. **d**, QQ plot for the 92 GenePrio genes available in the RVAT. Multiple testing was adjusted by Bonferroni ($P < 0.05/92$). The results are from a two-sided burden test in a logistic mixed-effects model. The sample size was 52,416 cases and 257,772 controls (for *TTN*, *CTNNA3*) and 51,019 cases and 253,267 controls (for *PKP2*, *FLNC*, *RBM20*). **e**,**f**, ATAC–seq fragment depth centered around the transcription start site (TSS) of the 139 GenePrio genes (**e**) as well as the 354 common sentinel variants (**f**). The different lines show fragment depth from ATAC–seq from our generated iPS cell-derived atrial CMs (atrial iPS-CMs) as well as seven publicly available cell lines and tissues from ENCODE. ATAC–seq fragment depth is elevated in atrial iPS-CMs compared to other tissues and cell lines. NK, natural killer.

https://doi.org/10.1038/s41588-024-02072-3

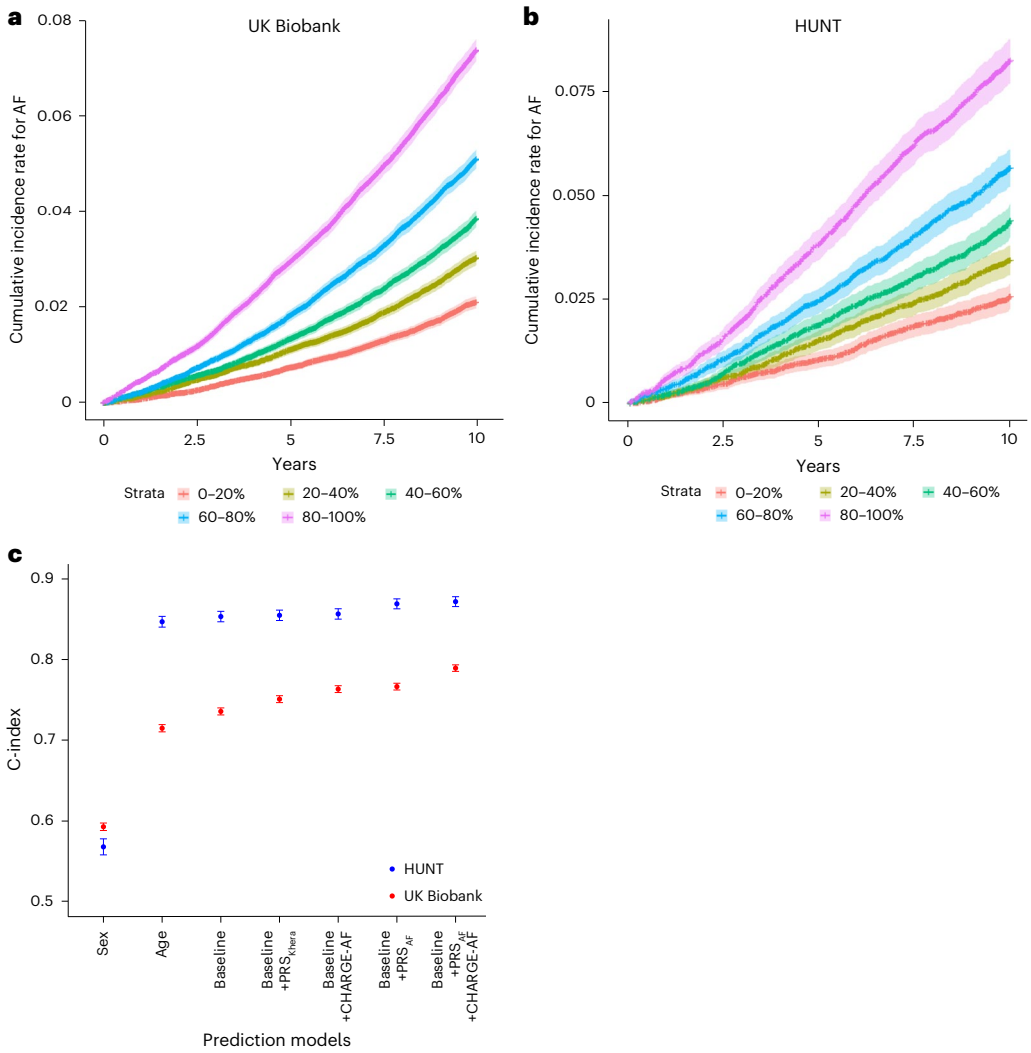


Fig. 3 | Polygenic risk prediction of AF with PRS in HUNT and UK Biobank. **a**,**b**, Plots showing the cumulative incidence rates of AF for the PRS score in the UK Biobank (**a**) and HUNT (**b**), separated by quintiles. The cumulative incidence plots include a lighter-shaded band showing the confidence intervals. The measures of center are the cumulative events, and the error bands are the 95% confidence intervals. **c**, Plot showing the model fit comparisons of several prediction models using the C-index with 95% confidence intervals. Each model was based on Cox proportional hazard regression. Model 1, 'sex' included sex. Model 2, 'age' included age and age^2. Model 3, 'baseline' included age, age^2 and sex. Model 4, 'baseline + PRS_{Khera}' included a previously published PRS[29] for AF on top of the baseline model. Model 5, 'baseline + CHARGE-AF' included the CHARGE-AF risk score[31] on top of the baseline model. Model 6, 'baseline + PRS_{AF}' included the newly generated PRS_{AF} on top of the baseline model. Model 7, 'baseline + PRS_{AF} + CHARGE-AF' included the newly generated PRS_{AF} and the CHARGE-AF clinical risk score on top of the baseline model. Both PRS_{AF} and PRS_{Khera} were inverse-normal-transformed and adjusted for the first ten principal components. HUNT included 52,757 samples with 2,474 incident AF cases. UK Biobank included 246,683 samples with 10,416 incident AF cases. The measure of center for the error bars is the C-index, and the error bars are the 95% confidence intervals. CHARGE, Cohorts for Heart and Aging Research in Genomic Epidemiology.

for ancestry-induced differences in the PRS_{AF} within European ancestry samples, we adjusted PRS_{AF} for genetically inferred principal components (Methods). We observed a clear stratification of the cumulative incidence rate for AF across quintiles of the score (Fig. 3a,b).

We created seven models for comparison of performance using Harrell's concordance index (C-index): model 1, sex; model 2, age; model 3, baseline; model 4, baseline + PRS_{Khera}; model 5, baseline + CHARGE-AF; model 6, baseline + PRS_{AF}; and model 7, baseline + PRS_{AF} + CHARGE-AF. Baseline covariates included sex, age, age^2 and study-specific variables. PRS_{Khera} is a published PRS for AF including 6.7 million variants[29] based on previously published AF GWAS results[7] and the LDPred algorithm[30]. CHARGE-AF is a published clinical risk score for AF[31]. PRS_{AF} is the newly generated PRS including 1.1 million variants, based on these GWAS results and the PRS-continuous shrinkage (PRS-CS) algorithm[32].

The predictive performance of each score was assessed over the first 10 years of follow-up with a Cox proportional hazards model. Each model showed an increased C-index over the previous model, from model 1 to model 7 (Fig. 3c and Supplementary Table 20). The model with the strongest predictive power (model 7) included both the CHARGE-AF clinical risk score and the PRS_{AF}, achieving a C-index of 0.872 (95% confidence interval, 0.866–0.878) in HUNT and a C-index of 0.790 (95% confidence interval, 0.786–0.794) in UK Biobank (Supplementary Tables 20 and 21). The incremental improvement of the best model with PRS relative to the model without was 0.0152 in HUNT and 0.026 in UK Biobank.

The new PRS_{AF} was used in the UK Biobank for a phenome-wide association analysis across a panel of 57 cardiometabolic and other diseases as well as 26 cardiometabolic traits (Fig. 4). Curated disease phenotypes were defined using reports from medical history interviews, inpatient and outpatient ICD-9 and ICD-10 codes, operation codes and death registry records[33]. The significance threshold was set at 6.02×10^{-4} to correct for multiple testing. Not surprisingly, the largest effect of the PRS_{AF} was for AF ($\beta = 0.557$, $P = 4.04 \times 10^{-1466}$).

https://doi.org/10.1038/s41588-024-02072-3

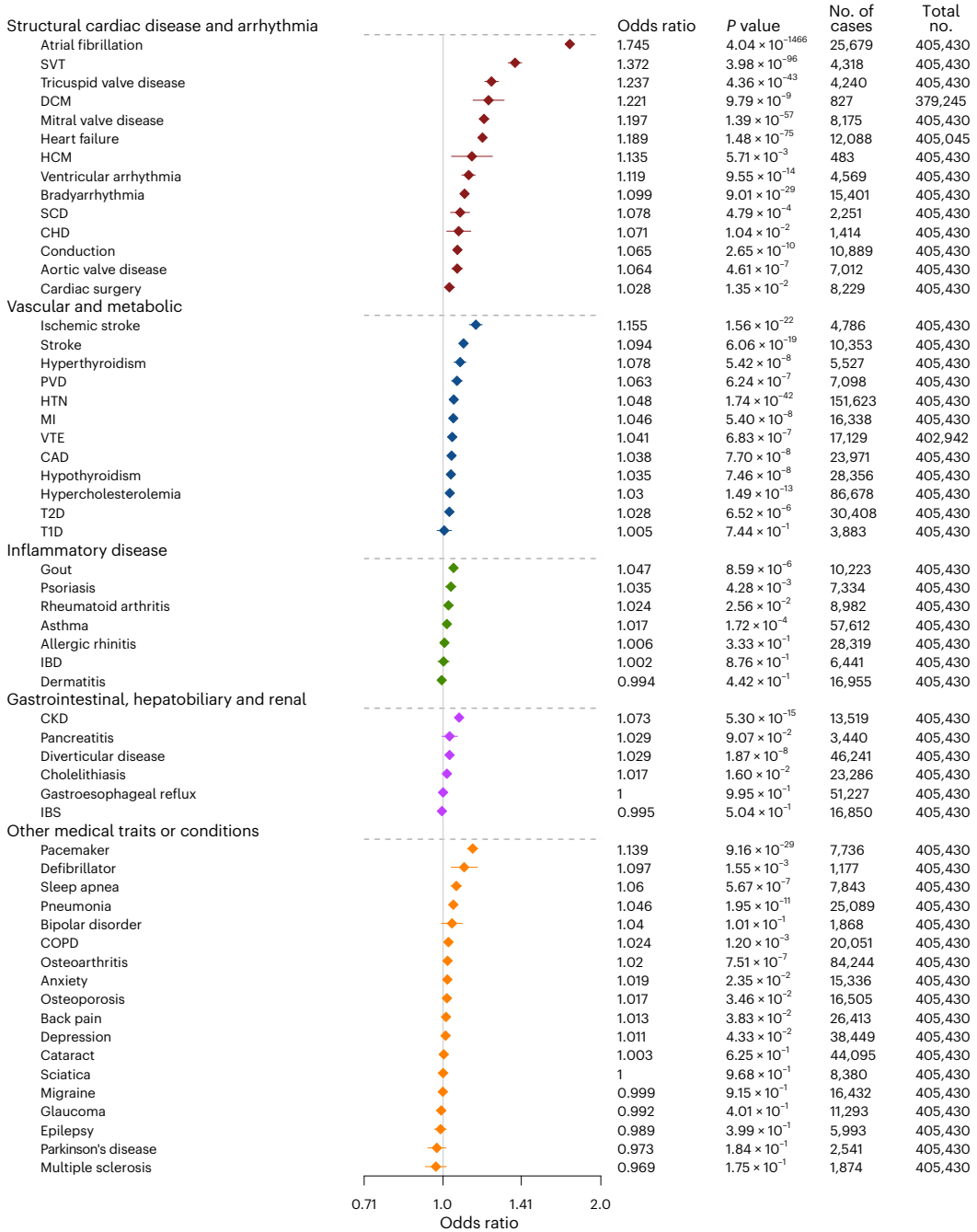


Fig. 4 | Phenome-wide associations of diseases and traits to the PRS_{AF} in the UK Biobank. Forest plot showing the associations of the PRS_{AF} with 57 cardiometabolic and other medical diseases and traits. The associations are sorted by odds ratio within each group. Significant associations are determined by Bonferroni correction as 0.05/83, resulting in a significance threshold of $P < 6.024 \times 10^{-4}$. The results are from a two-sided logistic regression model. The number of samples tested and the number of cases for each trait are included in the figure. The maximal available sample size was $n = 405{,}430$. The measure of center for the error bars is the odds ratio, and the error bars are the 95% confidence intervals. CAD, coronary artery disease; CHD, congenital heart disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; HTN, hypertension; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; MI, myocardial infarction; PVD, peripheral vascular disease; SCD, sudden cardiac death; SVT, supraventricular tachycardia; T1D; type 1 diabetes; T2D, type 2 diabetes; VTE, venous thromboembolism.

Other associations with large effects include supraventricular tachycardia ($\beta = 0.316$, $P = 3.98 \times 10^{-96}$) and tricuspid valve disease ($\beta = 0.213$, $P = 4.36 \times 10^{-43}$; Supplementary Table 22). The results may indicate a shared genetic basis among several structural cardiac diseases and arrhythmias but may also reflect the co-morbidity of these diagnoses (Fig. 4). Among quantitative traits, some of the largest effects were observed for weight ($\beta = 0.504$, $P = 9.30 \times 10^{-111}$), P-wave duration ($\beta = -0.499$, $P = 5.52 \times 10^{-8}$), left ventricular diastolic volume ($\beta = 0.508$, $P = 2.01 \times 10^{-4}$) and left ventricular systolic volume ($\beta = 0.406$, $P = 6.25 \times 10^{-7}$; Fig. 5 and Supplementary Table 22). Surprisingly, PRS_{AF} was associated with shorter rather than longer P-wave and QRS duration in the population. This may reflect genetic differences in ion channel function or cell–cell connections rather than the conduction delays that occur as a result of fibrosis and dilation in AF.

In the present study, we performed the most comprehensive AF meta-analysis so far, including 68 genome-wide studies across

https://doi.org/10.1038/s41588-024-02072-3

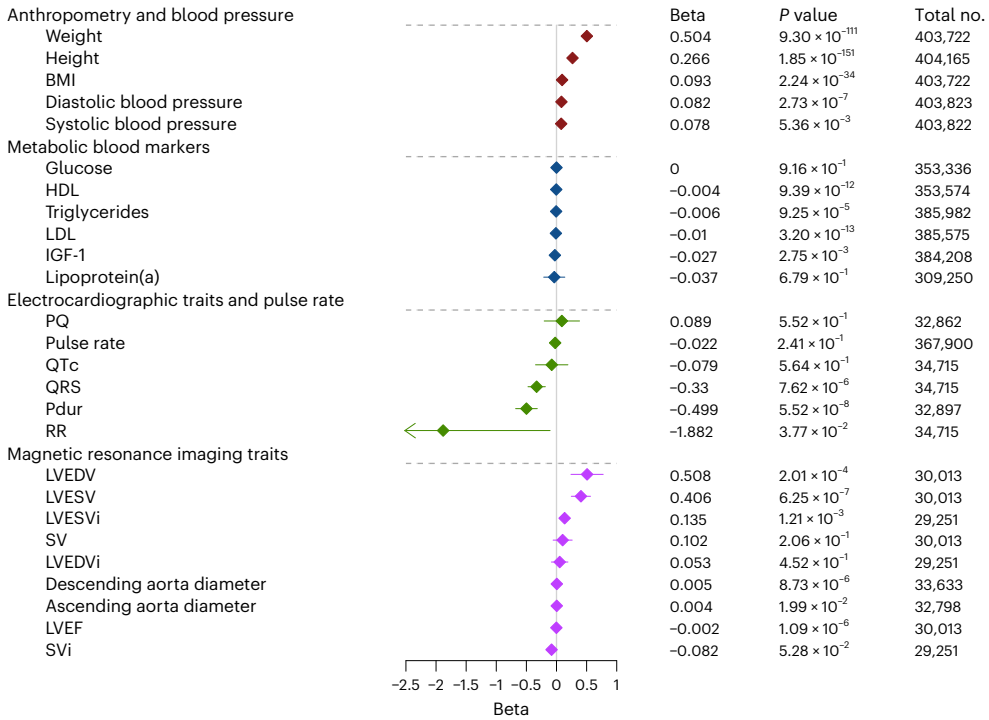


Fig. 5 | Phenome-wide associations of traits to the PRS_{AF} in the UK Biobank. Forest plot showing the associations of the PRS_{AF} with 26 quantitative traits that are grouped into four categories. The associations are sorted by beta within each group. Significant associations are determined by Bonferroni correction as 0.05/83, resulting in a significance threshold of $P < 6.024 \times 10^{-4}$. The results are from a two-sided linear regression model. The number of samples tested for each trait is included in the figure. The maximal available sample size was $n = 404{,}165$. The measure of center for the error bars is the beta, and the error bars are the 95% confidence interval. BMI, body mass index; HDL, high-density lipoprotein; IGF-1, insulin-like growth factor 1; LDL, low-density lipoprotein; LVEDV, left ventricular end-diastolic volume; LVEDVi, left ventricular end-diastolic volume indexed; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume; LVESVi, left ventricular end-systolic volume indexed; Pdur, P-wave duration; PQ, P–Q interval; QRS, QRS duration; QTc, corrected QT interval; RR, R–R interval; SV, systolic volume; SVi, systolic volume indexed.

40 primary cohorts and more than 180,000 AF cases (Supplementary Table 1). Most participants were of European ancestry, but we also incorporated two large-scale GWAS from European bottleneck populations (Iceland and Finland) as well as a large study sample from East Asia, predominantly represented by Biobank Japan[34] (Supplementary Table 2). We identified 354 AF-associated loci (Fig. 1), including five loci that were ancestry-enriched (Supplementary Tables 4 and 6). We found significant enrichment of overlap of chromatin accessibility and epigenetic modifications from iPS-derived atrial CMs with our GWAS loci. The novel PRS modestly improved the prediction of incident AF over the published PRS (Fig. 3c).

This work had several limitations. Most of the analyzed data were from individuals of European ancestry, and analyses that required a linkage disequilibrium (LD) reference panel were restricted to LD of European ancestry. The five lines of evidence used in gene prioritization can reflect shared information. For example, MAGMA results are included in the PoPS algorithm. There are many approaches to gene prioritization, but there is no gold standard for validation. A comparison of GenePrio to the contemporary locus-to-gene model[35] from OpenTargets can be found in the Supplementary Note.

In summary, we present results from the largest GWAS meta-analysis of AF so far and provide an improved PRS for the prediction of incident AF. The results implicate several candidate genes for AF that could serve as novel targets for therapeutics and may aid in determining the underlying biology of AF.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-02072-3.

References

1. Chen, Y.-H. et al. *KCNQ1* gain-of-function mutation in familial atrial fibrillation. *Science* **299**, 251–254 (2003).
2. Hodgson-Zingman, D. M. et al. Atrial natriuretic peptide frameshift mutation in familial atrial fibrillation. *N. Engl. J. Med.* **359**, 158–165 (2008).
3. Postma, A. V. et al. A gain-of-function *TBX5* mutation is associated with atypical Holt–Oram syndrome and paroxysmal atrial fibrillation. *Circ. Res.* **102**, 1433–1442 (2008).
4. Orr, N. et al. A mutation in the atrial-specific myosin light chain gene (*MYL4*) causes familial atrial fibrillation. *Nat. Commun.* **7**, 11303 (2016).
5. Gudbjartsson, D. F. et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
6. Ellinor, P. T. et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675 (2012).
7. Christophersen, I. E. et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952 (2017).
8. Thorolfsdottir, R. B. et al. Coding variants in *RPL3L* and *MYZAP* increase risk of atrial fibrillation. *Commun. Biol.* **1**, 68 (2018).
9. Sinner, M. F. et al. Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225–1235 (2014).

10. Nielsen, J. B. et al. Genome-wide study of atrial fibrillation identifies seven risk loci and highlights biological pathways and regulatory elements involved in cardiac development. *Am. J. Hum. Genet.* **102**, 103–115 (2018).
11. Gudbjartsson, D. F. et al. A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
12. Benjamin, E. J. et al. Variants in *ZFHX3* are associated with atrial fibrillation in individuals of European ancestry. *Nat. Genet.* **41**, 879–881 (2009).
13. Ellinor, P. T. et al. Common variants in *KCNN3* are associated with lone atrial fibrillation. *Nat. Genet.* **42**, 240–244 (2010).
14. Roselli, C. et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
15. Nielsen, J. B. et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
16. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
17. Miyazawa, K. et al. Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction. *Nat. Genet.* **55**, 187–197 (2023).
18. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
19. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
20. Choi, S. H. et al. Sequencing in over 50,000 cases identifies coding and structural variation underlying atrial fibrillation risk. *Nat. Genet.* https://doi.org/10.1038/s41588-025-02074-9 (2025).
21. van Ouwerkerk, A. F. et al. Identification of atrial fibrillation associated genes and functional non-coding variants. *Nat. Commun.* **10**, 4755 (2019).
22. Clausen, A. G., Vad, O. B., Andersen, J. H. & Olesen, M. S. Loss-of-function variants in the *SYNPO2L* gene are associated with atrial fibrillation. *Front. Cardiovasc. Med.* **8**, 650667 (2021).
23. Choi, S. H. et al. Association between Titin loss-of-function variants and early-onset atrial fibrillation. *JAMA* **320**, 2354 (2018).
24. Faggioni, M. et al. Suppression of spontaneous Ca elevations prevents atrial fibrillation in calsequestrin 2-null hearts. *Circ. Arrhythm. Electrophysiol.* **7**, 313–320 (2014).
25. Purohit, A. et al. Oxidized Ca^{2+}/calmodulin-dependent protein kinase II triggers atrial fibrillation. *Circulation* **128**, 1748–1757 (2013).
26. Holm, H. et al. A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nat. Genet.* **43**, 316–320 (2011).
27. McMullen, J. R. et al. The insulin-like growth factor 1 receptor induces physiological heart growth via the phosphoinositide 3-kinase(p110α) pathway. *J. Biol. Chem.* **279**, 4782–4793 (2004).
28. Wu, T. et al. HSPB7 is indispensable for heart development by modulating actin filament assembly. *Proc. Natl Acad. Sci. USA* **114**, 11956–11961 (2017).
29. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
30. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
31. Alonso, A. et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF Consortium. *J. Am. Heart Assoc.* **2**, e000102 (2013).
32. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
33. Jurgens, S. J. et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat. Genet.* **54**, 240–250 (2022).
34. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
35. 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 (2021).

Carolina Roselli[1,2,143], Ida Surakka[3,143], Morten S. Olesen[4,143], Gardar Sveinbjornsson[5,143], Nicholas A. Marston[6,7,143], Seung Hoan Choi[1], Hilma Holm[5], Mark Chaffin[1], Daniel Gudbjartsson[5,8], Matthew C. Hill[1,9], Hildur Aegisdottir[5], Christine M. Albert[10,11], Alvaro Alonso[12], Christopher D. Anderson[13,14], Dan E. Arking[15], David O. Arnar[5,16,17], John Barnard[18], Emelia J. Benjamin[19,20,21], Eugene Braunwald[6,7], Ben Brumpton[22,23], Archie Campbell[24], Nathalie Chami[25,26], Daniel I. Chasman[11,27], Kelly Cho[28,29], Eue-Keun Choi[30], Ingrid E. Christophersen[31,32], Mina K. Chung[33], David Conen[34], Harry J. Crijns[35], Michael J. Cutler[36], Tomasz Czuba[37,38], Scott M. Damrauer[39,40,41], Martin Dichgans[42,43], Marcus Dörr[44,45], Elton Dudink[35], ThuyVy Duong[15], Christian Erikstrup[46,47], Tõnu Esko[48], Diane Fatkin[49,50,51], Jessica D. Faul[52], Manuel Ferreira[53], Daniel F. Freitag[54], Santhi K. Ganesh[55,56], J. Michael Gaziano[57,58], Bastiaan Geelhoed[2], Jonas Ghouse[4,59], Christian Gieger[60], Franco Giulianini[11], Sarah E. Graham[3], Vilmundur Gudnason[16,61], Xiuqing Guo[62,63], Christopher Haggerty[64], Caroline Hayward[65], Susan R. Heckbert[66,67], Kristian Hveem[22], Kaoru Ito[68], Renee Johnson[49,50], J. Wouter Jukema[69,70], Sean J. Jurgens[1,71], Stefan Kääb[72,73], John P. Kane[74], Shinwan Kany[1,75,76], Sharon L. R. Kardia[77], Maryam Kavousi[78], Shaan Khurshid[1,79], Frederick K. Kamanu[6,7], Paulus Kirchhof[75,76,80,81], Marcus E. Kleber[82,83], Stacey Knight[36,84], Issei Komuro[85,86], Jose E. Krieger[87], Lenore J. Launer[88], Dadong Li[53], Honghuang Lin[89], Henry J. Lin[62,63], Ruth J. F. Loos[25,90], Luca Lotta[53], Steven A. Lubitz[1,79], Kathryn L. Lunetta[91], Peter W. Macfarlane[92,93], Patrik K. E. Magnusson[94], Rainer Malik[42,43], Helene Mantineo[95], Gregory M. Marcus[96], Winfried März[83,97], David D. McManus[98], Olle Melander[99], Giorgio E. M. Melloni[6], Pascal B. Meyre[100], Kazuo Miyazawa[68], Sanghamitra Mohanty[101], Laia M. Monfort[4,59],

Martina Müller-Nurasyid[102,103], Navid A. Nafissi[104], Andrea Natale[101,105], Saman Nazarian[106], Sisse R. Ostrowski[107,108], Hui-Nam Pak[109], Shichao Pang[110], Ole B. Pedersen[111], Nancy L. Pedersen[94], Alexandre C. Pereira[87], James P. Pirruccello[1], Michael Preuss[25], Bruce M. Psaty[112], Clive R. Pullinger[113], Daniel J. Rader[114], Joel T. Rämö[1,9,115], Paul M. Ridker[11,27], Michiel Rienstra[2], Lorenz Risch[116,117], Dan M. Roden[118], Jerome I. Rotter[62,119], Marc S. Sabatine[6,7], Heribert Schunkert[110,120], Svati H. Shah[104], Jaemin Shim[121], M. Benjamin Shoemaker[122], Bridget Simonson[1], Moritz F. Sinner[72,73], Roelof A. J. Smit[25], Jennifer A. Smith[52,77], Nicholas L. Smith[66,67,123], J. Gustav Smith[37,38], Elsayed Z. Soliman[124], Erik Sørensen[107], Nona Sotoodehnia[125], Daniel Strbian[126], Bruno H. Stricker[78], Maris Teder-Laving[48], Yan V. Sun[12,127], Sébastien Thériault[34,128], Rosa B. Thorolfsdottir[5], Unnur Thorsteinsdottir[5,16], Arnljot Tveit[31,129], Pim van der Harst[130], Joyce van Meurs[78,131], Biqi Wang[91], Stefan Weiss[45,132], Quinn S. Wells[118], Lu-Chen Weng[1,133,134], Peter W. Wilson[12,135,136], Ling Xiao[1,9], Pil-Sung Yang[137], Jie Yao[62], Zachary T. Yoneda[122], Tanja Zeller[76,138], Lingyao Zeng[54], Wei Zhao[52,77], Xiang Zhou[139], Sebastian Zöllner[140], The BioBank Japan Project*, Regeneron Genetics Center*, DBDS Genomic Consortium*, Christian T. Ruff[6,7,144], Henning Bundgaard[59,108,144], Cristen Willer[3,141,144], Kari Stefansson[5,16,144] & Patrick T. Ellinor[1,79,142,144]

[1]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [3]Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI, USA. [4]Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark. [5]deCODE genetics/Amgen Inc., Reykjavik, Iceland. [6]TIMI Study Group, Boston, MA, USA. [7]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [8]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [9]Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [10]Department of Cardiology, Cedars-Sinai, Los Angeles, CA, USA. [11]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [12]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [13]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA. [14]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [15]McKusick Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [16]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [17]Cardiovascular Centre, Landspitali—The National University Hospital of Iceland, Reykjavik, Iceland. [18]Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. [19]Department of Medicine, Boston Medical Center, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA. [20]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA. [21]Framingham Heart Study, Framingham, MA, USA. [22]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. [23]Clinic of Medicine, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway. [24]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [25]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [26]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [27]Harvard Medical School, Boston, MA, USA. [28]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA. [29]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [30]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [31]Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Gjettum, Norway. [32]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [33]Cardiovascular Medicine, Heart Vascular and Thoracic Institute and Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. [34]Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada. [35]Department of Cardiology and CARIM, Maastricht University Medical Centre, Maastricht, The Netherlands. [36]Intermountain Heart Institute, Intermountain Medical Center, Murray, UT, USA. [37]The Wallenberg Laboratory/Department of Molecular and Clinical Medicine, Institute of Medicine, Gothenburg University and the Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden. [38]Department of Cardiology, Lund University Diabetes Center, Wallenberg Center for Molecular Medicine, Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden. [39]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [40]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [41]Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA. [42]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. [43]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [44]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [45]German Center for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany. [46]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [47]Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark. [48]Institute of Genomics, University of Tartu, Tartu, Estonia. [49]Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia. [50]UNSW Sydney, Sydney, New South Wales, Australia. [51]St Vincent's Hospital, Sydney, New South Wales, Australia. [52]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [53]Regeneron Genetics Center, Tarrytown, NY, USA. [54]Bayer Pharmaceuticals, Wuppertal, Germany. [55]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA. [56]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA. [57]Million Veteran Program, VA Boston Healthcare System, Boston, MA, USA. [58]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [59]Department of Cardiology, Copenhagen University Hospital, Copenhagen, Denmark. [60]Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [61]Icelandic Heart Association, Kópavogur, Iceland. [62]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. [63]Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. [64]Department of Translational Data Science and Informatics, Geisinger, Danville, PA, USA. [65]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [66]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA. [67]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [68]Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [69]Department of Cardiology, Leiden University Medical Centre, Leiden, The Netherlands. [70]Netherlands Heart Institute, Utrecht, The Netherlands. [71]Department of Experimental Cardiology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands. [72]Department of Cardiology, University Hospital, LMU Munich, Munich, Germany. [73]German Centre for Cardiovascular Research (DZHK), Partner site: Munich Heart Alliance, Munich, Germany. [74]Cardiovascular Research Institute, University of California, San Francisco, CA, USA. [75]Department of Cardiology, University Heart and Vascular Center UKE Hamburg,

Hamburg, Germany. [76]German Center for Cardiovascular Research (DZHK), Hamburg, Germany. [77]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [78]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [79]Demoulas Center for Cardiac Arrhythmias and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. [80]Institute of Cardiovascular Sciences, Birmingham, UK. [81]AFNET, Münster, Germany. [82]SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany. [83]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany. [84]Department of Medicine, School of Medicine, University of Utah, Salt Lake City, UT, USA. [85]International University of Health and Welfare, Tokyo, Japan. [86]Department of Frontier Cardiovascular Science, The University of Tokyo, Tokyo, Japan. [87]Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of Sao Paulo, Sao Paulo, Brazil. [88]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, USA. [89]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA. [90]Novo Nordisk Foundation Center for Basic Metabolic Research, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [91]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [92]School of Health and Wellbeing, University of Glasgow, Glasgow, Scotland, UK. [93]Electrocardiology Group, New Lister Building, Royal Infirmary, Glasgow, Scotland, UK. [94]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [95]Precision Cardiology Laboratory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [96]Division of Cardiology, University of California, San Francisco, CA, USA. [97]Synlab Academy, SYNLAB Holding Deutschland GmbH, Mannheim, Germany. [98]University of Massachusetts Medical School, Worcester, MA, USA. [99]Department of Internal Medicine, Lund University and Skåne University Hospital, Malmö, Sweden. [100]Cardiovascular Research Institute Basel (CRIB), University Hospital Basel, Basel, Switzerland. [101]Texas Cardiac Arrhythmia Institute, Austin, TX, USA. [102]IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [103]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany. [104]Division of Cardiology, Duke University School of Medicine, Durham, NC, USA. [105]Department of Biomedicine and Prevention, Division of Cardiology, University of Tor Vergata, Rome, Italy. [106]Section of Cardiac Electrophysiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [107]Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark. [108]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [109]Yonsei University College of Medicine, Yonsei University Health System, Seoul, Republic of Korea. [110]Department of Cardiology, Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [111]Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark. [112]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Systems and Population Health, University of Washington, Seattle, WA, USA. [113]Cardiovascular Research Institute and Department of Physiological Nursing, University of California, San Francisco, CA, USA. [114]Division of Cardiovascular Medicine, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [115]Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland. [116]Institute of Laboratory Medicine, Faculty of Medical Sciences, Private University of the Principality of Liechtenstein, Triesen, Liechtenstein. [117]Center of Laboratory Medicine, University Institute of Clinical Chemistry, University of Bern, Inselspital, Bern, Switzerland. [118]Departments of Medicine Pharmacology and Biomedical Informatics, Divisions of Cardiovascular Medicine and Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA. [119]Departments of Pediatrics and Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. [120]Deutsches Zentrum für Herz- und Kreislauferkrankungen (DZHK), Partner site: Munich Heart Alliance, Munich, Germany. [121]Korea University Cardiovascular Center, Seoul, Republic of Korea. [122]Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [123]Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, WA, USA. [124]Epidemiological Cardiology Research Center, Section on Cardiovascular Medicine, Department of Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [125]Division of Cardiology, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA, USA. [126]Department of Neurology and Neurosciences, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [127]Atlanta VA Health Care System, Decatur, GA, USA. [128]Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Québec City, Québec, Canada. [129]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [130]Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands. [131]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [132]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [133]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. [134]VA Boston Healthcare System, Boston, MA, USA. [135]Atlanta VA Medical Center, Atlanta, GA, USA. [136]Division of Cardiology, Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA. [137]Cha University College of Medicine, Seoul, Republic of Korea. [138]University Center of Cardiovascular Sciences, Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [139]Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA. [140]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [141]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [142]Heart and Vascular Institute, Mass General Brigham, Boston, MA, USA. [143]These authors contributed equally: Carolina Roselli, Ida Surakka, Morten S. Olesen, Gardar Sveinbjornsson, Nicholas A. Marston. [144]These authors jointly supervised this work: Christian T. Ruff, Henning Bundgaard, Cristen Willer, Kari Stefansson, Patrick T. Ellinor. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: ellinor@mgh.harvard.edu

The BioBank Japan Project*

Kaoru Ito[68], Issei Komuro[85,86] & Kazuo Miyazawa[68]

A full list of members and their affiliations appears in the Supplementary Information.

Regeneron Genetics Center*

Manuel Ferreira[53], Dadong Li[53] & Luca Lotta[53]

A full list of members and their affiliations appears in the Supplementary Information.

DBDS Genomic Consortium*

Christian Erikstrup[46,47], Ole B. Pedersen[111], Erik Sørensen[107], Sisse R. Ostrowski[107,108], Daniel Gudbjartsson[5,8], Kari Stefansson[5,16,144] & Unnur Thorsteinsdottir[5,16]

A full list of members and their affiliations appears in the Supplementary Information.

Methods

Sample

This research complies with all relevant ethical regulations. Written informed consent was obtained from all participants in this study. The UK Biobank resource was approved by the UK Biobank Research Ethics Committee, and all participants provided written informed consent to participate. The use of UK Biobank data was authorized under application number 17488 and was approved by the local Massachusetts General Brigham Institutional Review Board. The Institutional Review Board at Massachusetts General Hospital reviewed and approved the overall study. Our sample included a total of 68 summary-level result files (Supplementary Table 1). Of those, 26 were previously published and unchanged since the last publication[14] (Supplementary Table 23), whereas 42 non-overlapping summary-level results were novel or updated. The baseline information for these 42 novel or updated studies is reported in Supplementary Table 24. The study description of the novel or updated summary-level results can be found in the Supplementary Note. The baseline and study information for the 26 previously published study-based AF GWAS results are described elsewhere[14]. Across the entirety of the 68 study results, we included samples of European ancestry (n_{case} = 166,322; $n_{control}$ = 1,313,950), East Asian ancestry (n_{case} = 11,350; $n_{control}$ = 137,515), admixed African and African American ancestry (n_{case} = 1,782; $n_{control}$ = 9,356), Hispanic ancestry (n_{case} = 1,203; $n_{control}$ = 6,569), Brazilian ancestry (n_{case} = 571; $n_{control}$ = 1,096) and South Asian ancestry (n_{case} = 218; $n_{control}$ = 413). The European subset furthermore incorporates samples from the two bottleneck populations of Finland (n_{case} = 17,325; $n_{control}$ = 97,214) and Iceland (n_{case} = 20,953; $n_{control}$ = 353,822) (Supplementary Table 2). AF was broadly defined as paroxysmal or permanent AF or atrial flutter. Controls did not present with these diagnoses.

Genotyping, pre-imputation quality control and imputation

The genotyping, pre-imputation quality control procedure and imputation of the 26 previously published summary-level results and the UK Biobank has been described elsewhere[14,36]. For novel and updated studies, genotyping arrays, calling algorithms and pre-imputation quality control steps by study are listed in Supplementary Table 25. In short, the pre-imputation quality control consisted of sample-level filtering (low call rate, excess heterozygosity, relatedness) and variant-level filtering (low call rate, deviation from Hardy–Weinberg equilibrium, excess heterozygosity, low MAF). In total, 32 of the novel or updated studies were imputed to the TOPMed reference panel[37] and six to the Haplotype Reference Consortium panel[38]. The studies FinnGen and deCODE were imputed to population-specific reference panels based on whole-genome sequence (FinnGen, n = 3,775 Finnish individuals from Sequencing Initiative Suomi (SISu v.3); deCODE, n = 49,708 Icelandic individuals). Biobank Japan was imputed to a merged reference panel, including 1000 Genomes phase 3 (v.5) and a population-specific reference panel consisting of whole-genome sequences from 1,037 Japanese individuals[39].

Single-variant GWAS

Single-variant testing was performed using an additive genetic model and genotype dosages. Most studies included prevalent AF cases, and a subset combined prevalent and incident cases, in either case applying a logistic regression model. Studies with separate results for incident and prevalent cases used the Cox proportional hazards model and logistic regression, respectively. The analytical details and tools used by each study are listed in Supplementary Table 25 or have previously been described (Supplementary Table 23). In general, studies were advised to include the covariates of age, sex and genetic principal components in their models. Studies and biobanks with large case-control imbalances used appropriate tools, such as SAIGE, to account for the imbalance. Each summary-level results file was evaluated based on the following quality control criteria: inspection of Manhattan plots, quantile–quantile plots and PZ plots (P value reported versus P value from Z-score), a reasonable genomic inflation parameter (λ_{GC}), consistent direction of effect for known AF-associated variants, distribution of effect estimates and allele frequencies. We furthermore compared reported allele frequencies to those of the Haplotype Reference Consortium or TOPMed freeze 5 reference panel. Variants were filtered before the meta-analysis for imputation quality of >0.3 and MAF × imputation quality × n events of ≥10.

Meta-analysis

The tool used for the meta-analysis was METAL (version released on 2018-08-28), with the inverse-variance-weighted approach. Standard errors were adjusted according to the genomic inflation for each study if λ_{GC} was >1. The meta-analysis was first conducted by each one of the following eight groups: (1) European (non-Finnish and non-Icelandic); (2) Finnish; (3) Icelandic; (4) African and African American; (5) Brazilian; (6) South Asian; (7) East Asian; and (8) Hispanic. Subsequently, these eight result files were meta-analyzed and heterogeneity was calculated. We applied the commonly used genome-wide significance cutoff of $P < 5 \times 10^{-8}$. We evaluated our results separated by allele frequency. A total of 18,842,305 variants were common with MAF ≥ 1%. Low-frequency variants with MAF < 1% were more stringently filtered for a mean imputation quality of ≥0.8. A total of 10,947,675 variants with low frequency were assessed. Genetic loci were identified in a region-based approach, defining a locus to a 1 Mb region around a sentinel variant. The sentinel variant, or lead variant, was the variant with the lowest P value in that 1 Mb region. The variants were annotated for functional consequence and overlapping genes with the VEP tool (97 version)[40].

Conditional and joint analysis

Conditional and joint analysis of all ancestry meta-analysis results was performed with the tool GCTA (Genome-wide Complex Trait Analysis)[41] and the stepwise model selection procedure. Owing to computational limitations, the P values were capped at 1×10^{-300}. The threshold for significance was defined as 5×10^{-8}. As an LD reference, the TOPMed imputed genotype data from the European ancestry participants of the Broad Cardiovascular Disease Initiative (Broad CVDi) study was used (n = 32,715). The LD references included variants with imputation quality scores of >0.3 and MAF > 0.01% and were converted to hard calls with a threshold of 80%.

Validation

We sought to validate the sentinel variants in the MVP using genotypes (release 3). The MVP is an independent cohort with 30,831 AF cases and 268,407 controls of European ancestry, 4,539 AF cases and 76,046 controls of African American ancestry, 1,428 AF cases and 30,507 controls of Hispanic descent, and 163 AF cases and 4,165 controls of Asian descent. The AF cases and controls were analyzed in a combined ancestry analysis, including age, sex and AF-related principal components as covariates.

Ancestry-specific effects

Ancestry-specific effects were evaluated in a heterogeneity analysis of sentinel variants across the eight listed ancestry groups reported above. We applied a Bonferroni correction for multiple testing. For the common lead variants, the threshold was $0.05/354 = 1.41 \times 10^{-4}$. For low-frequency lead variants, the threshold was $0.05/39 = 1.28 \times 10^{-3}$. We report the I^2 statistic and the P value for the heterogeneity test across ancestries. Additionally, we annotated each lead variant with the direction of effect by ancestry.

Gene ranking at GWAS loci with GenePrio annotation

Genes at the common variant GWAS loci were identified by intersecting a 1 Mb region around each sentinel variant with the GENCODE[42] gene

reference (v.26). Any gene with a transcription start position (txStart) within the 1 Mb region of the sentinel variant was assigned to the respective locus. In the case of overlapping loci, the gene was associated with the locus with the shorter distance between the txStart and sentinel variant. The nearest gene and the nearest protein-coding gene were identified as the gene or protein-coding gene with the closest proximity of its txStart to the coordinates of the sentinel variant.

Subsequently, each gene was annotated across five categories: (1) eQTL; the gene was a significant eQTL in either cardiac tissue from the Genotype-Tissue Expression (GTEx) or left atrial tissue from MAGNet; (2) PoPS; the gene was within the top-ranked genes (>mean +3 s.d.) of the similarity-based PoPS analysis; (3) MAGMA; the gene was significantly associated in the region-based MAGMA analysis; (4) coding; the gene had a significant coding variant that was either annotated as loss-of-function (high confidence) or missense with a deleterious score greater than 0.3; and (5) snRNA-seq; the gene was within top 10% of specific genes for CMs in left atrial single-nuclei tissue from the human heart.

Each annotation had a binary flag (0, no; 1, yes). The five lines of evidence were summed for each gene, and the genes were subsequently ranked by locus based on that sum. Tied ranks were assigned as the average. We refer to this gene-based annotation on a locus level as GenePrio. At each locus, a gene was selected if it presented with a GenePrio rank of 1 as well as a minimum of two lines of evidence.

Gene-level analyses

MAGMA. The proximity-based gene-level analysis of our GWAS data was performed with the MAGMA[18] (v.1.09) tool. European ancestry samples from the 1000 Genomes project were used as an LD reference, and 9,793,179 variants available in the LD reference were included. Gene-level annotations from the file *magma_0kb.genes.annot* generated by the PoPS tool were used for this analysis. This annotation file contains 18,383 protein-coding genes with no added window. LD reference and MAGMA gene annotations were downloaded on 5 April 2021 (https://www.dropbox.com/sh/o6t5jprvxb8b500/AADZ8qD6Rpz4uvCk0b5nUnPaa/data?dl=0). The SNP-wise mean model was applied. Owing to computational constraints, very low *P* values of variants were capped at 5×10^{-324} for the input data. A total of 18,116 genes from the gene annotations contained single-nucleotide polymorphisms (SNPs) in the genotype data. We used a Bonferroni correction to account for multiple testing with a significance threshold of $0.05/18{,}116 = 2.76 \times 10^{-6}$.

PoPS. The similarity-based gene-level analysis of our GWAS data was performed with the PoPS[19] (v.0.1) algorithm. Step 1 contained the generation of MAGMA results, as previously described. Step 2 aimed to select features using a predefined set of features *PoPS.features.txt.gz*. Those 57,543 gene features were derived from gene expression (40,546), protein–protein interaction (8,718) and pathway membership (8,479). A total of 22,017 features were selected in step 2. Step 3 calculated the PoPS score using the selected features, a predefined set of control features from the file *control.features* as well as the gene locations from *gene_loc.txt*. The precomputed files *PoPS.features.txt.gz*, *control.features* and *gene_loc.txt* were downloaded on 5 April 2021 from (https://www.dropbox.com/sh/o6t5jprvxb8b500/AADZ8qD6Rpz4uvCk0b5nUnPaa/data?dl=0). The final results included a PoPS score for a total of 18,383 genes. The top-ranking PoPS genes were selected by subsetting to genes with a score greater than mean score + 3 s.d. of the score (PoPS score > 0.855) and additionally being located within 500 kb of a genome-wide significant variant.

Transcriptional profiling

LD-score regression analysis with snRNA-seq data. LD-score regression analyses were performed using annotations based on snRNA-seq data from the healthy human heart[43]. The data were obtained from the four cardiac chambers: left atrium, left ventricle, right atrium and right ventricle. The data were subsetted to genes from the autosomal chromosomes with a total read count of >10. For the analysis, the annotation file was built using 1000 Genomes Europeans LD reference files and a window size of 100 kb for SNPs that were used in the baseline model. The LD-score regression was performed on high-quality SNPs from the HapMap3 reference and included the baseline annotations from a previous work[44].

The first analysis (CM by chamber) was conducted with gene expression profiles of CMs across the four cardiac chambers. For each chamber, the top 10% of upregulated genes in that chamber were chosen. The second analysis (left atrium by cell type) was conducted with gene expression profiles in the left atrium only of the cell types CMs, fibroblasts, endothelial cells, pericytes, macrophages, vascular smooth muscle cells, adipocytes, neuronal cells and lymphocytes. For each cell type, the top 10% of upregulated genes in that cell type were chosen. In both analyses, all considered genes were included in the control group. The significance threshold for each analysis was determined using Bonferroni correction: the CM by chamber analysis had a cutoff of 0.05/4 = 0.0125 and the left atrium by cell type analysis had a cutoff of 0.05/9 = 0.0056.

Intersection with eQTLs. Sentinel variants were intersected with significant eQTLs from two sources: cardiac tissues (left ventricle and atrial appendage) from GTEx[45] (v.8) and left atrial tissue from MAGNet[14]. Only significant *cis*-eQTLs as defined in the primary analysis of the GTEx (v.8) and MAGNet eQTL datasets were included. We performed colocalization analyses with the R package *coloc*[46]. We used default prior probabilities $p1 = 1 \times 10^{-4}$, $p2 = 1 \times 10^{-4}$ and $p12 = 1 \times 10^{-5}$.

Variant consequences

We annotated variants with $P < 0.05$ using VEP (v.97)[40]. The most severe consequence of a variant was selected with the *pick_allele* option. The variants in the coding region were also annotated with the dbNSFP4.1a database[47], and we created a deleterious score using 30 in silico prediction tools. The deleterious score was calculated if a variant had more than seven in silico prediction tools available.

Rare variant association testing analysis

We queried all genes prioritized with GenePrio against recently analyzed whole-genome sequencing and whole-exome sequencing rare variant analysis for AF. In brief, predicted loss-of-function and predicted damaging mutations were aggregated across genes and tested for an association with AF using burden tests in a logistic mixed-effects model (for details, see ref. 20). Enrichment for GenePrio genes among rare variant signals was performed at multiple *P* value cutoffs using a Fisher's exact test. For GenePrio and nearby gene categorizations from GWAS, only protein-coding genes were considered.

Enrichment analysis

We performed a gene set enrichment analysis of the GenePrio genes with the web-based tool g:Profiler[48] (https://biit.cs.ut.ee/gprofiler/gost). Data sources for pathways included Gene Ontology[49], Kyoto Encyclopedia of Genes and Genomes[50], Reactome[51], Human Phenotype Ontology[52], the Human Protein Atlas[53] and WikiPathways[54]. We applied the g:SCS algorithm for multiple testing.

Chromatin accessibility and epigenetic modifications

The cell line HUES8 was obtained from Memorial Sloan Kettering Cancer Center. Human pluripotent stem cells were maintained in feeder-free culture until 90% confluence in 5% CO_2 at 37 °C. They were then dissociated into single-cell suspension and cultured with constant shaking to form spheroids, followed by manipulation of activin A/BMP4 signaling and biphasic control of the WNT pathway with activation of retinoic acid signaling to generate atrial CMs.

A total of 50,000 atrial iPS-CMs were used as input for ATAC–seq, following the OMNI-ATAC–seq protocol[55]. Transposed DNA was purified with a Qiagen PCR MinElute kit (Qiagen 28004), and final ATAC–seq libraries were purified with a 1.8× SPRI purification using SPRISelect beads (Beckman Coulter) following PCR amplification. Libraries were sequenced on an Illumina NextSeq 500. Reads were mapped to the human genome (GRCh38) using Bowtie2 (ref. 56) with default paired-end settings and all non-nuclear and unmapped paired reads were discarded. Duplicated reads were removed with the Picard MarkDuplicates function, using default settings. Visualization of ATAC–seq signals was done with HOMER, and all reads were normalized by read count in which scores represent read count per bp per 1×10^7 reads.

The following ENCODE ATAC–seq datasets were used in this study: ENCFF676KNV (skeletal muscle), ENCFF552GSS (natural killer cell), ENCFF246NCJ (pancreas), ENCFF466EAM (thyroid gland), ENCFF428WFH (urinary bladder), ENCFF348MQI (GM23338) and ENCFF258LP (liver).

The ATAC–seq and histone modification tracks were generated with the USCS browser[57]. The commercially available Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb (no. 9751) was used. The H3K4me3 sites in our atrial iPS-CMs were produced with the Cleavage Under Targets & Release Using Nuclease (CUT&RUN)[58] protocol. The layered track of histone H3 Lysine 27 acetylation (H3K27ac) was combined from seven published ENCODE H3K27ac cell lines: GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK and NHLF.

Cluster analysis

We performed a cluster analysis of the GenePrio genes with the R (v.4.1) package *clusterProfiler* (v.4.0.5)[59]. The function *compareCluster* was used with the settings fun = 'enrichGO', OrgDb = 'org.Hs.eg.db' and pvalueCutoff = 0.05. The input lists of genes were created by intersecting the GenePrio genes with the top 10% of specific genes for each cell type in the left atrium. A gene could be assigned to multiple cell types. In total, 112 out of 139 genes were assigned to one or more cell types and included in the analysis. The plot was created with the R library *enrichplot* (v.1.12.2) using the functions *pairwise_termsim* and *emapplot*.

Polygenic risk scores

Derivation. To create a PRS for AF, we ran a meta-analysis, leaving out HUNT and UK Biobank, including a total of 154,330 cases and 999,609 controls. We selected variants present in UK Biobank and HUNT for the analysis. We used the PRS-CS[32] method to calculate the weights for the score with 1000 Genomes Europeans as the LD reference panel. The resulting weights (w_m) for 1,113,668 genetic variants were then used to calculate the raw PRS_{AF} in both HUNT and UK Biobank using the following formula:

$$PRS_{AF_i} = \sum_m w_m G_{m,i}$$

where $G_{m,i}$ is the dosage of effect alleles of individual i for marker m. For HUNT and UK Biobank, the raw PRS_{AF} was then adjusted for the first ten genetic principal components. The adjusted score was further inverse-normal-transformed for the analyses to have the hazard ratios on the s.d. scale.

PRS testing in HUNT. HUNT[60] samples were collected over three different time periods: HUNT1 (1984–1986), HUNT2 (1995–1997) and HUNT3 (2006–2008). HUNT1 was excluded, as genotypes were not available for this collection. HUNT2 and HUNT3 comprised the primary dataset for this analysis, and the baseline for each individual was set as the earliest time point with all clinical risk factors used in the CHARGE-AF risk score calculation available. Individuals with prevalent AF at baseline were excluded. The final dataset consisted of 52,757 individuals between the ages of 19 and 94 years (median follow-up, 21.2 years), including 2,474 incident AF cases. The analyses were restricted to the first 10 years of follow-up. All AF prediction analyses were performed in R (v.3.6.3) with the library *survival*. We adjusted all models for the collection period (HUNT2 vs HUNT3). Baseline information for the components of the CHARGE-AF risk score in the HUNT cohort is provided in Supplementary Table 26.

PRS testing in UK Biobank. The PRS testing in the UK Biobank was performed on data from a total of 246,683 individuals with European ancestry. We excluded the following participants from this analysis: UK Biobank phase 1 participants, prevalent AF at enrollment, withdrew consent ($n = 127$), up to third-degree relatives (kinship coefficient of >0.0442, $n = 81{,}849$) and missing data for covariates. Our sample included a total of 10,416 incident AF cases within a 10-year follow-up. We adjusted all models for the genotyping array. We used R (v.3.6.0) and the packages survival (v2.44-1.1) and survminer (v0.4.3) for the statistical analyses and visualization. Baseline information for the components of the CHARGE-AF risk score in the UK Biobank cohort is provided in Supplementary Table 26. The variables for the CHARGE-AF risk score were taken at enrollment.

AF risk prediction with PRS. We compared seven different models. Model 1, 'sex' included sex. Model 2, 'age' included age and age^2. The age^2 term was included to account for the nonlinear relationship between AF risk and age[61]. Model 3, 'baseline' included age, age^2 and sex. Model 4, 'baseline + PRS_{Khera}' added a previously published PRS[29] for AF. Model 5, 'baseline + CHARGE-AF' added the CHARGE-AF clinical risk score[31]. Model 6, 'baseline + PRS_{AF}' added the newly generated PRS_{AF}. Model 7, 'baseline + PRS_{AF} + CHARGE-AF' added the newly generated PRS_{AF} and the CHARGE-AF clinical risk score. All scores were inverse-normal-transformed. We used Cox proportional hazards models with follow-up time as the timescale to test the performance of different AF predictors. We compared the model fit using the C-index, a model fit statistic for survival models that is a generalization of the receiver operating characteristic curve that also handles censored data.

PRS phenome-wide association in UK Biobank. There were 488,374 samples with genetic data available. The participants who withdrew their consent ($n = 127$) and within third-degree relatives (kinship coefficient of >0.0442, $n = 81{,}849$) were removed. We performed association tests between 83 cardiometabolic traits (57 diseases and 26 quantitative traits) and PRS_{AF}. The 57 diseases included prevalent and incident cases; the definitions are listed in Supplementary Table 27. The quantitative traits were measured at the time of enrollment. We used logistic regression for binary traits and linear regression for quantitative traits, adjusting for age, sex, genotyping array and the first five genetic principal components. The quantitative traits were inverse-rank normalized for analyses. The significance was determined at 6.02×10^{-4} (0.05/83 traits).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The summary-level results file as well as the weights file for the PRS_{AF} are available for download at the Cardiovascular Disease Knowledge Portal under the weblinks https://cvd.hugeamp.org/downloads.html#polygenic and https://cvd.hugeamp.org/downloads.html#summary. The raw and processed ATAC–seq and H3K4me3 data have been deposited at the NCBI Gene Expression Omnibus under accession number GSE225293. The following datasets were used in this study and are publicly available under the listed weblinks: GENCODE: https://www.gencodegenes.org; 1000G LD reference, MAGMA gene

annotations and precomputed files for PoPS algorithm: https://www.dropbox.com/sh/o6t5jprvxb8b500/AADZ8qD6Rpz4uvCk0b5nUnPaa/data?dl=0; GTEx: https://www.gtexportal.org/home; ENCODE: https://www.encodeproject.org; OpenTargets: https://www.opentargets.org.

Code availability

All software programs used in the study are publicly available and described in the Methods and Reporting Summary.

References

36. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
37. Kowalski, M. H. et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, e1008500 (2019).
38. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
39. Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).
40. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
41. 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).
42. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
43. Tucker, N. R. et al. Transcriptional and cellular diversity of the human heart. *Circulation* **142**, 466–482 (2020).
44. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
45. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
46. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
47. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **12**, 103 (2020).
48. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
49. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
50. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
51. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
52. Köhler, S. et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).
53. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
54. Martens, M. et al. WikiPathways: connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).
55. Corces, M. R. et al. An improved ATAC–seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Kent, W. J. et al. The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
58. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
59. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).
60. Krokstad, S. et al. Cohort profile: the HUNT study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
61. Staerk, L. et al. Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. *Br. Med. J.* **361**, k1453 (2018).

Acknowledgements

The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 02/10/2021. This research is based on data from the MVP, Office of Research and Development, Veterans Health Administration and was supported by award no. I01-BX004821. This publication does not necessarily represent the views of the Department of Veteran Affairs or the United States Government. Study-specific acknowledgements, including grants, are listed in the Supplementary Note.

Author contributions

C.R., C.W., D.G., G.S., H.H., I.S., K.S., M.C., M.C.H., M.S.O., N.A.M. and P.T.E. crafted and finalized the manuscript. A.A., A.C., A.C.P., A.N., A.T., B.B., B.G., B.H.S., B.M.P., B.S., C.D.A., C.E., C.G., C. Haggerty, C. Hayward, C.R., C.R.P., C.T.R., C.W., D.O.A., D.C., D.D.M., D.E.A., D.F., D.F.F., D.G., D.I.C., D.J.R., D.L., D.S., E.-K.C., E.J.B., E.D., E.S., E.Z.S., G.M.M., J.G.S., G.S., H.-N.P., H.A., H.B., H.H., H.J.C., H.J.L., H.M., H.S., I.E.C., I.S., J.B., J.E.K., J.G., J.I.R., J.P.K., J.P.P., J.S., J.T.R., J.v.M., K.C., K.H., K.I., K.L.L., K.S., L.-C.W., L.J.L., L.L., L.M.M., L.R., L.X., L.Z., M.C., M. Dichgans, M. Dörr, M.E.K., M.F., M.F.S., J.M.G., M.C.H., M.J.C., M.K., M.K.C., M.M.-N., M.P., M.R., M.S.O., M.S.S., M.T.-L., N.A.M., N.A.N., N.C., N.L.P., N.L.S., N.S., O.B.P., O.M., P.-S.Y., P.K., P.K.E.M., P.B.M., P.T.E., P.v.d.H., P.W.W., P.W.M., R.A.J.S., R.J., R.J.F.L., R.M., R.B.T., S.K.G., S.H.C., S.H.S., S.J.J., S. Kääb, S. Kany, S. Khurshid, S. Knight, S.A.L., S.M., S.M.D., S.N., S.R.O., S.P., S.R.H., S.W., S.Z., T.C., T.D., T.E., T.Z., U.T., V.G., J.W.J., W.M., X.G., X.Z. and Y.V.S. contributed to or revised the manuscript. B.W., C.R., D.G., E.-K.C., F.G., F.K.K., G.E.M.M., G.S., H.L., I.S., J.A.S., J.Y., K.M., L.-C.W., M.C., M.F., M.P., S.E.G., S.J.J., S.T., T.D. and W.Z. performed statistical analyses. A.A., C.D.A., C.G., C. Haggerty, C.M.A., C.T.R., C.W., D.C., D.I.C., D.L., D.M.R., D.S., E.J.B., E.B., F.G., H.-N.P., H.B., I.K., J.D.F., K.I., K.L.L., K.S., L.L., L.R., M.B.S., M.F., M.M.-N., M.P., M.S.S., N.A.M., P.M.R., P.T.E., Q.S.W., R.J.F.L., S.A.L., S.L.R.K. and Z.T.Y. contributed or managed samples and phenotype data. C.M.A., C.T.R., C.W., D.G., D.I.C., H.B., H.H., K.S. and P.T.E. conceived, designed and supervised the overall project.

Competing interests

B.M.P. serves on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. C.D.A. receives sponsored research support from Bayer AG and has consulted for ApoPharma. C. Haggerty receives research support from Tempus Labs, outside the scope of the present work. C.M.A. receives sponsored research support from St. Jude, Abbott and Roche. C.R. is supported by a grant from Bayer AG to the Broad Institute focused on the development of therapeutics for cardiovascular disease. C.R. is a full-time employee of GSK as of 1 July 2024. C.T.R. reports research grants through Brigham and Women's Hospital from Amgen, Anthos, AstraZeneca, Daiichi Sankyo, Janssen, Merck and Novartis and has received honoraria for scientific advisory boards and consulting from Anthos, Bayer,

Bristol Myers Squibb, Daiichi Sankyo, Janssen, Pfizer, Regeneron and Sirius. The spouse of C.W. works at Regeneron Pharmaceuticals. D.O.A., D.G., G. Sveinbjornsson, H.A., H.H., K.S., R.B.T. and U.T. are employees of deCODE genetics/Amgen. D.C. has received consultancy fees from Roche Diagnostics and Trimedics, and speaker's fees from BMS/Pfizer and Servier, all outside of the current work. D.F.F. is a full-time employee of Bayer. E.B. performs uncompensated consultancies and lectures with The Medicines Company. E.-K.C. reports research grants or speaking fees from Abbott, Bayer, BMS/Pfizer, Biosense Webster, Chong Kun Dang, Daewoong Pharmaceutical, Daiichi Sankyo, DeepQure, Dreamtech, Jeil Pharmaceutical, Medtronic, Samjinpharm, Seers Technology and Skylabs. L.-C.W. receives sponsored research support from IBM to the Broad Institute. L.Z. is a full-time employee of Bayer AG. M.E.K. is employed by SYNLAB Holding Deutschland. M.S.S. receives research grant support through Brigham and Women's Hospital from Abbott, Amgen, Anthos Therapeutics, AstraZeneca, Bayer, Daiichi Sankyo, Eisai, Intarcia, Ionis, Medicines Company, MedImmune, Merck, Novartis, Pfizer and Quark Pharmaceuticals and consults for Althera, Amgen, Anthos Therapeutics, AstraZeneca, Beren Therapeutics, Bristol Myers Squibb, DalCor, Dr. Reddy's Laboratories, Fibrogen, Intarcia, Merck, Moderna, Novo Nordisk and Silence Therapeutics; additionally, M.S.S. is a member of the TIMI Study Group, which has also received institutional research grant support through Brigham and Women's Hospital from ARCA Biopharma, Janssen Research and Development, Siemens Healthcare Diagnostics, Softcell Medical Limited, Regeneron, Roche and Zora Biosciences. N.A.M. reports involvement in clinical trials with Amgen, Pfizer, Ionis, Novartis and AstraZeneca without personal fees, payments or increase in salary. P.M.R. has received investigator-initiated research grant support for unrelated projects from NHLBI, Operation Warp Speed, Novartis, Kowa, Amarin and Pfizer and has served as a consultant on unrelated issues to Novo Nordisk, Flame, Agepha, Uppton, Novartis, Jansen, Health Outlook, Civi Biopharm, Alnylam and SOCAR. P.T.E. receives sponsored research support from Bayer AG, Bristol Myers Squibb, Pfizer and Novo Nordisk; he has also served on advisory boards or consulted for Bayer AG. S.A.L. is a full-time employee of Novartis Institutes for BioMedical Research as of 18 July 2022. Previously, S.A.L. received sponsored research support from Bristol Myers Squibb/Pfizer, Bayer, Boehringer Ingelheim, Fitbit, IBM, Medtronic and Premier, and consulted for Bristol Myers Squibb/Pfizer, Bayer, Blackstone Life Sciences and Invitae. S.M.D. receives research support from RenalytixAI and personal consulting fees from Calico Labs, outside the scope of the current research. W.M. reports grants and/or personal fees from Siemens Healthineers, Aegerion Pharmaceuticals, AMGEN, AstraZeneca, Sanofi, Alexion Pharmaceuticals, BASF, Abbott Diagnostics, Numares, Berlin-Chemie, Akzea Therapeutics, Bayer Vital, bestbion dx, Boehringer Ingelheim Pharma, Immundiagnostik, Merck Chemicals, MSD Sharp and Dohme, Novartis Pharma and Olink Proteomics, and other from Synlab Holding Deutschland, all outside the submitted work. A.N. is a consultant for Abbott, Biosrnse Webster, Biotronik, Boston Sci, iRhythm, Field Medical, Pulse Bioselect and Medtronic. S. Khurshid receives sponsored research support from Bayer AG. All other authors declare no competing interests.

Additional information

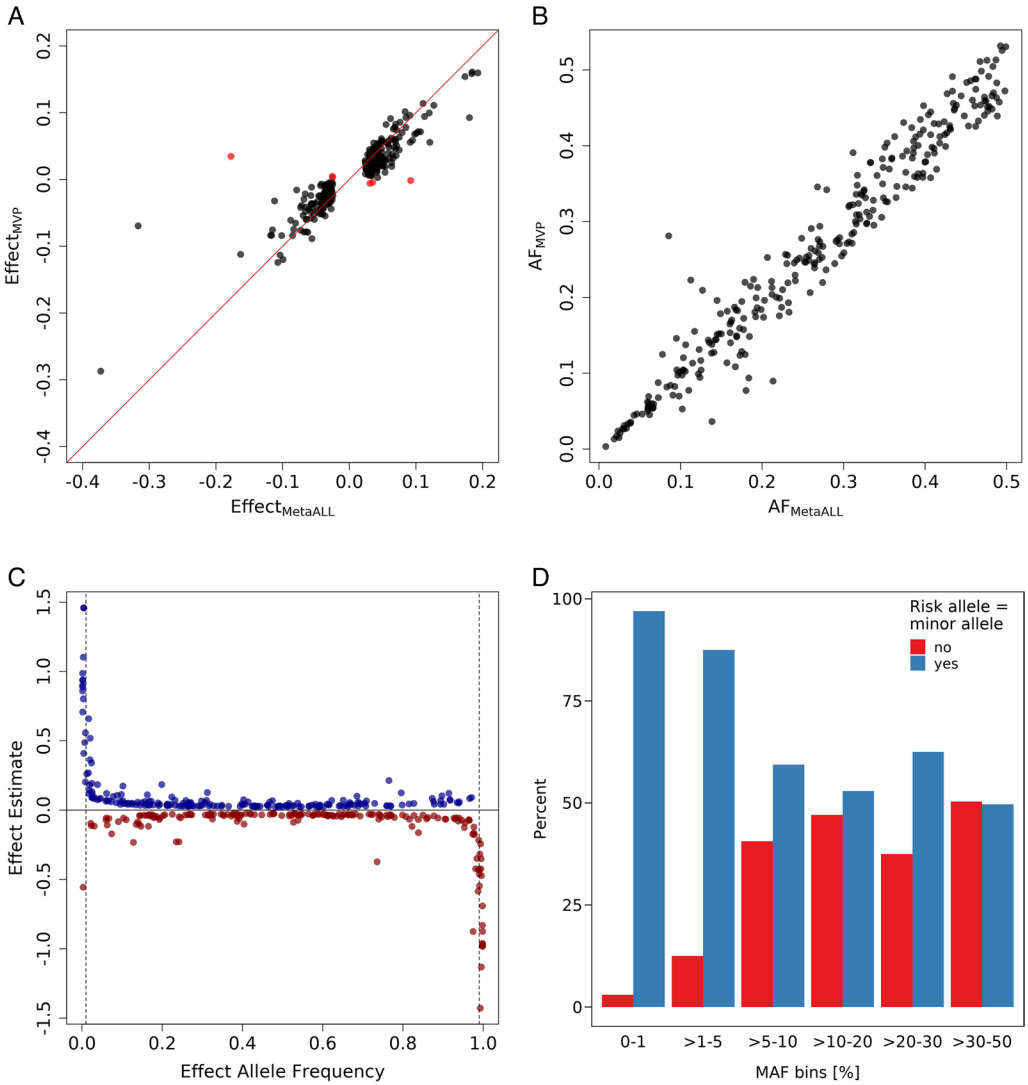
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-02072-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-02072-3.

Correspondence and requests for materials should be addressed to Patrick T. Ellinor.

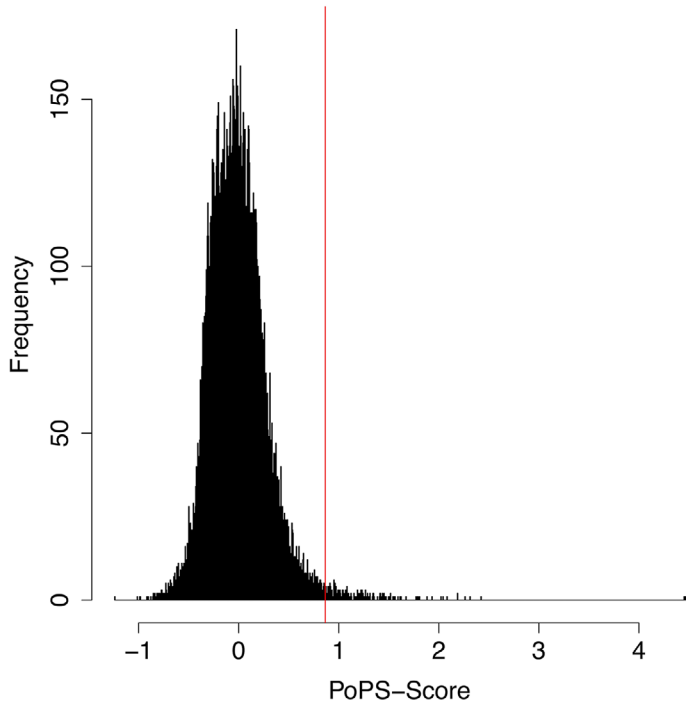
Peer review information *Nature Genetics* thanks Rafik Tadros and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41588-024-02072-3



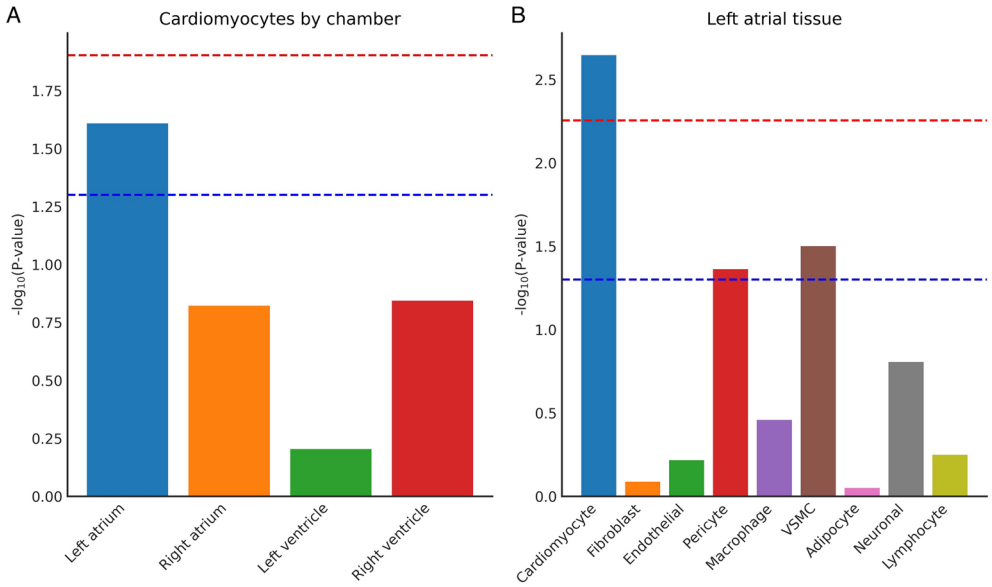
Extended Data Fig. 1 | Evaluation of effect estimates and allele frequencies for main GWAS meta-analysis and, in comparison, to the validation analysis. **a**,**b**, Plots showing data for the 299 sentinel variants from the common variant analysis (MAF ≥ 1%) that were also available in the validation set. **a**, Correlation of allele frequencies between the meta-analysis and the validation cohort from MVP. **b**, Correlation of effect estimates between meta-analysis and the validation cohort from MVP. The red line is the identity line ($x = y$). Labelled in red are the variants with discordant direction of effect between meta-analysis and validation. **c**, Plot showing the relationship between effect allele frequency and strength of effect for sentinel variants of the meta-analysis. The effect estimates are from the inverse variance weighted method for meta-analysis. The dotted vertical lines show the cutoff for rare variants with MAF < 1%. The genome-wide significance cut-off of $P < 5 \times 10^{-8}$ was applied to correct for multiple testing. **d**, Plot showing co-occurrence of risk allele for atrial fibrillation and minor allele in blue and the inverse in red for sentinel variants of the meta-analysis. AF, allele frequency; ALL, all-ancestry; MAF, minor allele frequency; Meta, meta-analysis; MVP, Million Veteran Program.

https://doi.org/10.1038/s41588-024-02072-3



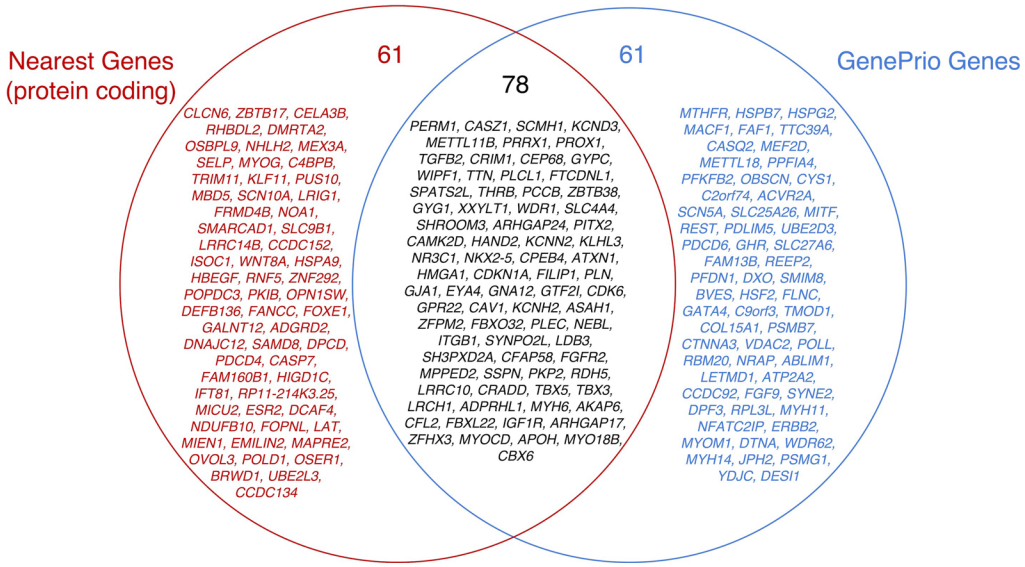
Extended Data Fig. 2 | Distribution of PoPS-score. Red line shows the cutoff for mean + 3 standard deviations of the score (cutoff = 0.8548401). There were 205 genes with a PoPS-score higher than the cutoff.

https://doi.org/10.1038/s41588-024-02072-3



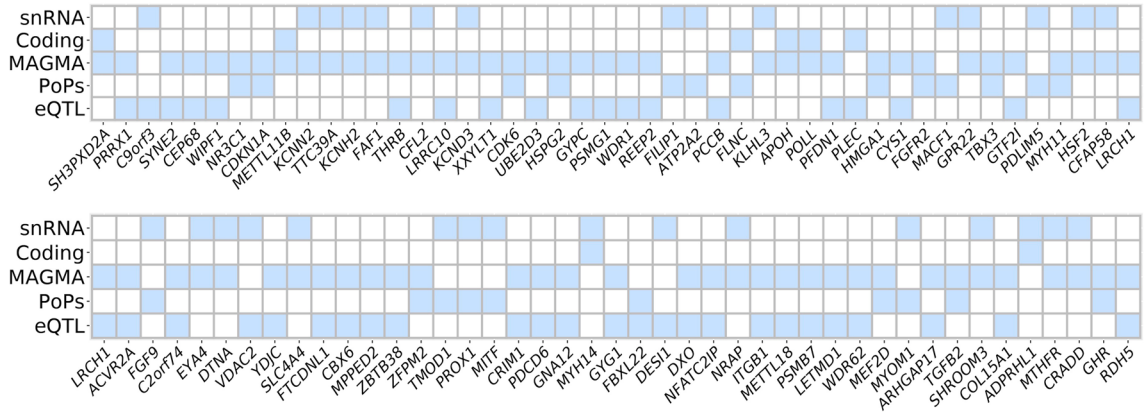
Extended Data Fig. 3 | Partitioned heritability analyses with annotations from human cardiac single-nuclei RNA-sequencing expression data. a, Plot showing the partitioned heritability results for gene expression of cardiomyocytes by heart chamber ($P_{LeftAtrium} = 2.46 \times 10^{-2}$, $P_{RightAtrium} = 1.51 \times 10^{-1}$, $P_{LeftVentricle} = 6.25 \times 10^{-1}$, $P_{RightVentricle} = 1.43 \times 10^{-1}$). The blue dotted line shows the cutoff for nominal significance $P < 0.05$. The red dotted line shows the Bonferroni corrected significance cut off 0.05/4. **b**, Plot showing the partitioned heritability results for gene expression in left atrial tissue by cell type ($P_{Cardiomyocyte} = 2.26 \times 10^{-3}$, $P_{Fibroblast} = 8.18 \times 10^{-1}$, $P_{Endothelial} = 6.07 \times 10^{-1}$, $P_{Pericyte} = 4.34 \times 10^{-2}$, $P_{Macrophage} = 3.48 \times 10^{-1}$, $P_{VSMC} = 3.15 \times 10^{-2}$, $P_{Adipocyte} = 8.90 \times 10^{-1}$, $P_{Neuronal} = 1.56 \times 10^{-1}$, $P_{Lymphocyte} = 5.63 \times 10^{-1}$). The blue dotted line shows the cutoff for nominal significance $P < 0.05$. The red dotted line shows the Bonferroni corrected significance cut off 0.05/9. VSMC, vascular smooth muscle cells.

https://doi.org/10.1038/s41588-024-02072-3



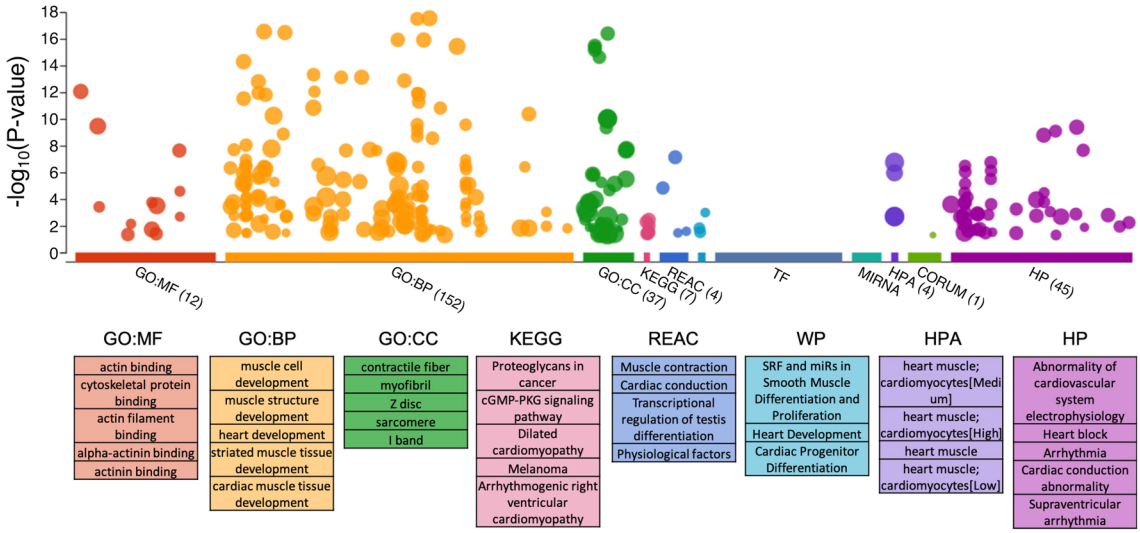
Extended Data Fig. 4 | Venn diagram of consensus genes between GenePrio and nearest protein coding genes. Venn diagram for the overlap of GenePrio vs. nearest gene (protein coding, in relation to the transcription start position) at 139 loci. In red are the genes identified as the nearest, in blue are the GenePrio genes, and in black are the genes that overlap between the two groups. 56% of genes overlap.

https://doi.org/10.1038/s41588-024-02072-3



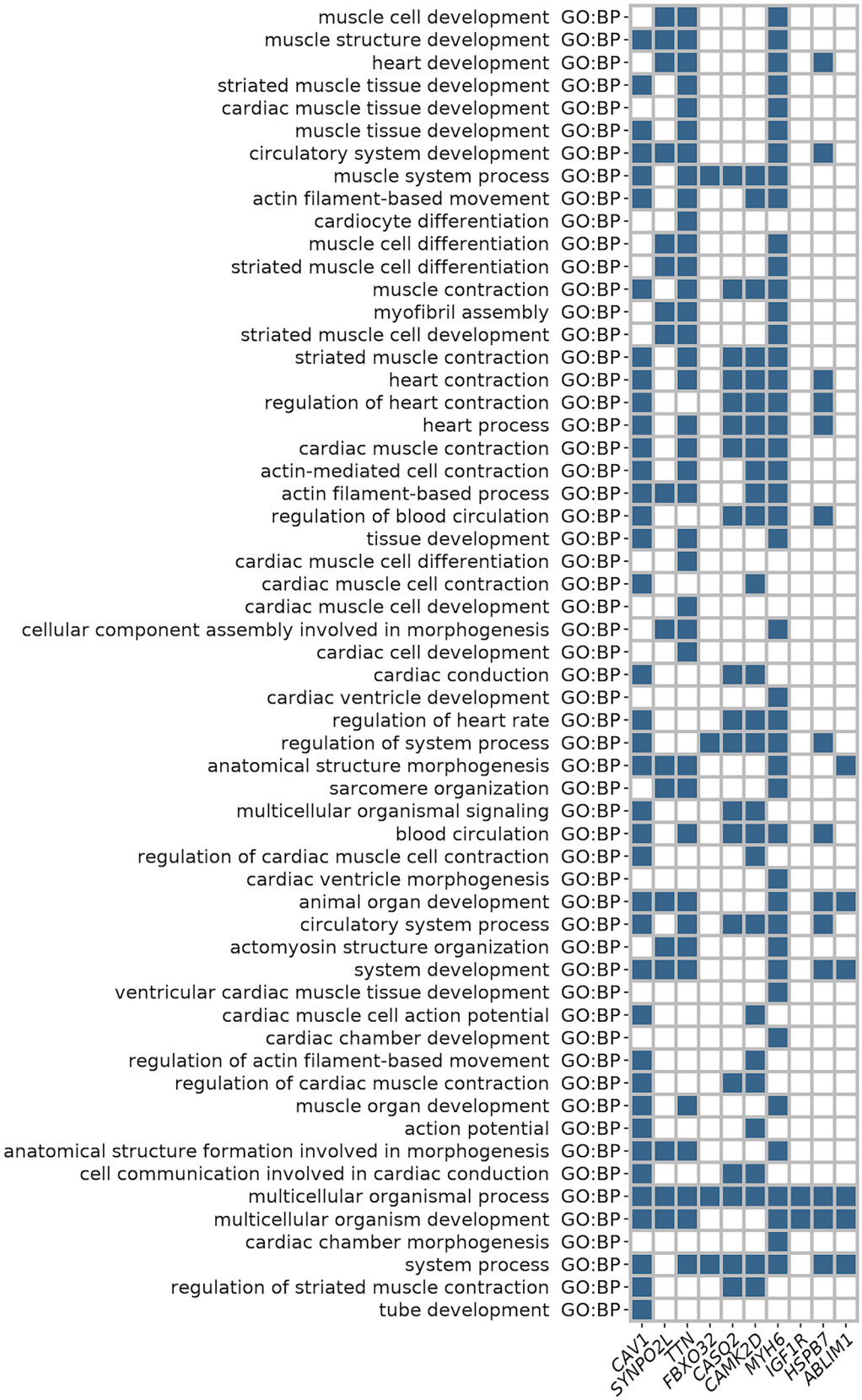
Extended Data Fig. 5 | Heatmap for GenePrio genes with two lines of evidence. GenePrio genes with two lines of evidence. The five categories of evidence that were assessed to prioritize genes at GWAS loci: snRNA-seq (labelled as snRNA), gene was a top 10% marker gene for cardiomyocytes in left atrial tissue; Coding, gene had genome-wide significant loss-of-function variant or missense variant with predicted to be damaging effect; MAGMA, significant result for the gene in MAGMA analysis; PoPs, gene had a high PoPs score; eQTL, sentinel variant at locus had a significant eQTL to that gene in cardiac tissue. The genes are sorted from lowest to highest *P*-value at the sentinel variant of the locus. AF, atrial fibrillation; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; MAF, minor allele frequency; MAGMA, Multi-marker Analysis of GenoMic Annotation; PoPS, polygenic priority score; snRNA-seq, single-nuclei RNA-sequencing.

https://doi.org/10.1038/s41588-024-02072-3



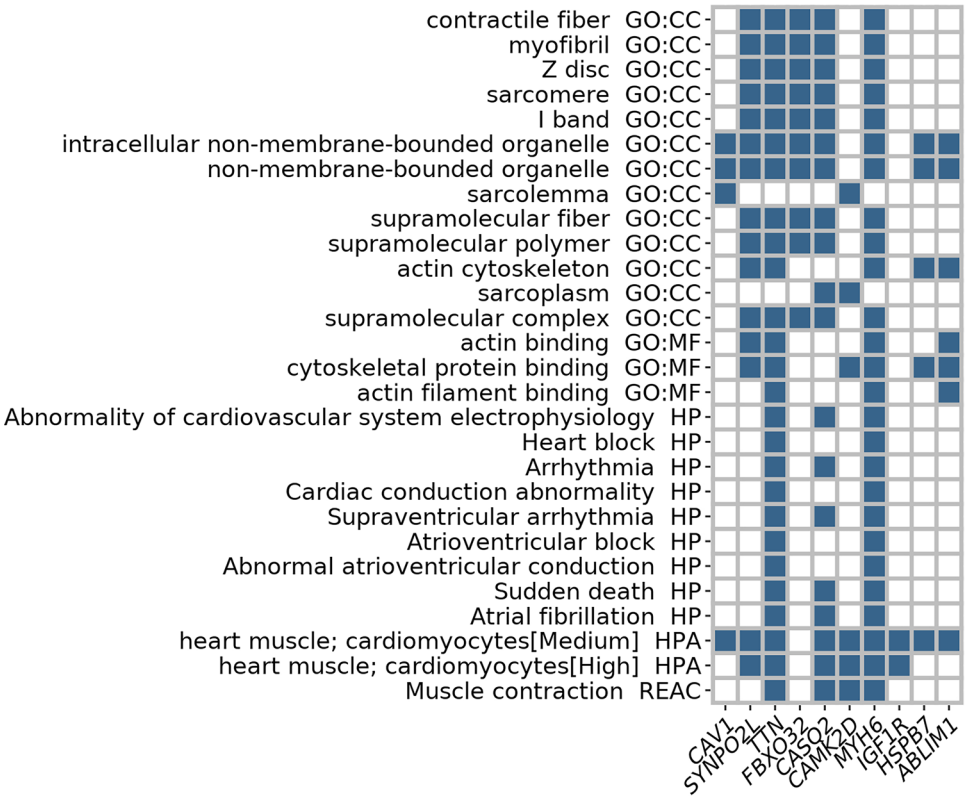
Extended Data Fig. 6 | Gene set enrichment analysis for all GenePrio genes. Results of the gene set enrichment analysis for all 139 GenePrio genes across several databases. The $-\log_{10}$(*P*-values) are plotted sorted by gene set category. The top 5 gene sets by *P*-value are listed for each category. The size of each dot is proportional to the term size (*n* genes) of the gene set, that is larger terms have larger dots. The enrichment testing is done using a Fisher's one-sided test (cumulative hypergeometric probability). *P*-values were adjusted for multiple testing using the g:SCS algorithm from the g:Profiler tool. BP, biological process; CC, cellular component; GO, gene ontology; HP, human phenotype ontology; HPA, human protein atlas; MF, molecular function; REAC, reactome; WP, wiki pathways.

https://doi.org/10.1038/s41588-024-02072-3



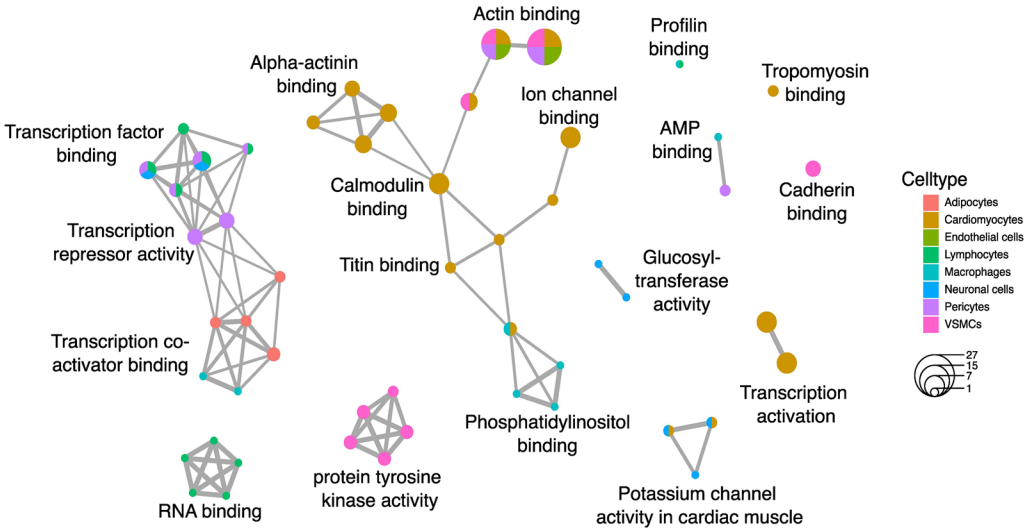
Extended Data Fig. 7 | Gene set enrichment results for top 10 GenePrio genes and GO:BP. Heatmap of significant (adjusted $P < 5 \times 10^{-6}$) gene sets for GO:BP, showing top 10 GenePrio genes (GenePrio sum = 4) and their affiliation to each set. The enrichment testing is done using a Fisher's one-sided test (cumulative hypergeometric probability). *P*-values were adjusted for multiple testing using the g:SCS algorithm from the g:Profiler tool. BP, biological process; GO, gene ontology.

https://doi.org/10.1038/s41588-024-02072-3



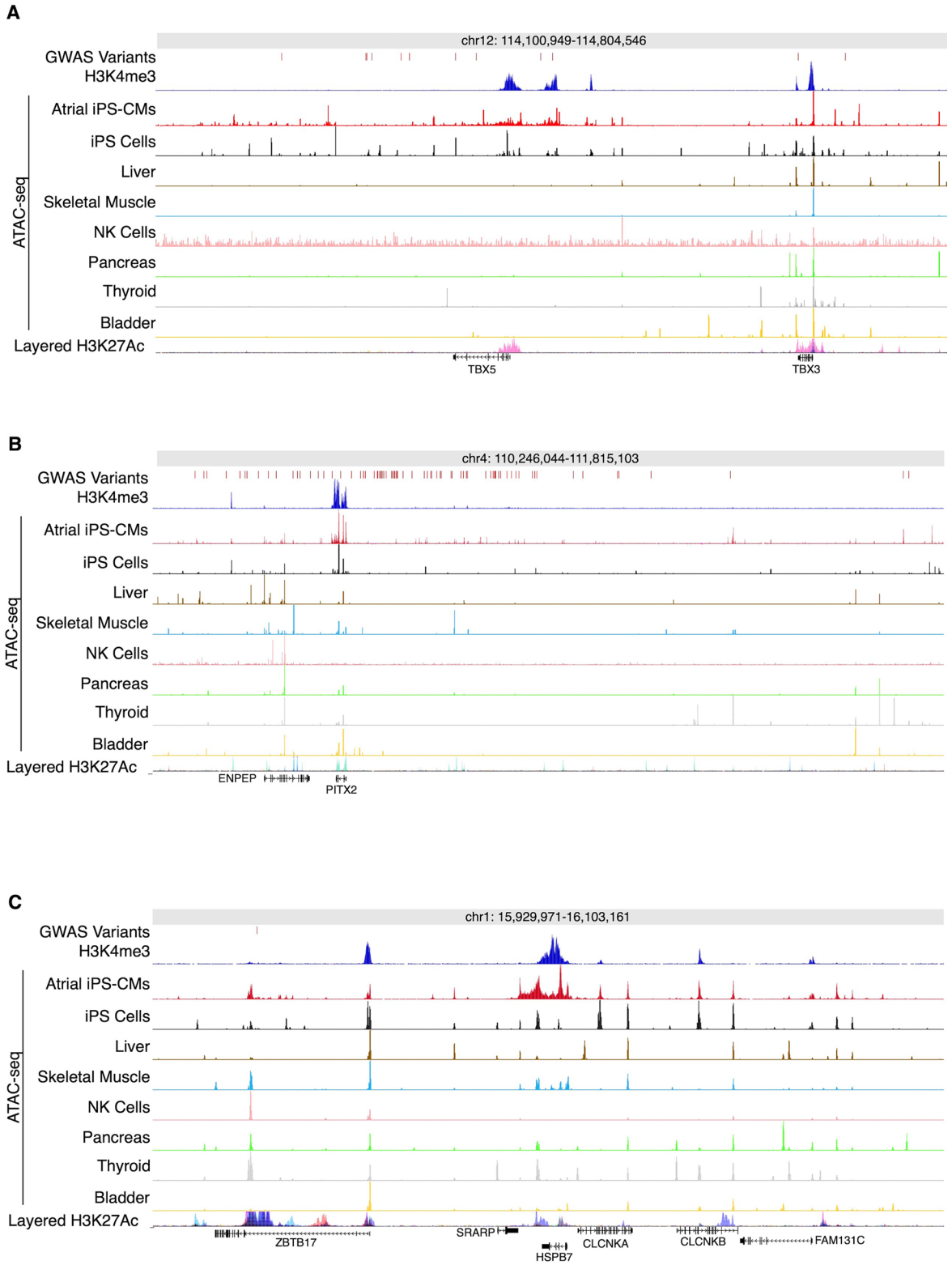
Extended Data Fig. 8 | Gene set enrichment results for top 10 GenePrio genes and GO:MF, GO:CC, HP, HPA and REAC. Heatmap of significant (adjusted $P < 5 \times 10^{-6}$) gene sets for GO:MF, GO:CC, HP, HPA and REAC, showing top 10 GenePrio genes (GenePrio sum = 4) and their affiliation to each set. The enrichment testing is done using a Fisher's one-sided test (cumulative hypergeometric probability). *P*-values were adjusted for multiple testing using the g:SCS algorithm from the g:Profiler tool. CC, cellular component; HP, human phenotype ontology; HPA, human protein atlas; MF, molecular function; REAC, Reactome.

https://doi.org/10.1038/s41588-024-02072-3



Extended Data Fig. 9 | Cluster analysis of the GenePrio genes based on cell type specific expression and Gene Ontology. Results from a cluster analysis of 112 of the 139 GenePrio genes for Gene Ontology (GO) with *clusterProfiler*. Genes were annotated to cell types (one or more) based on the top 10% specific genes for each cell type. The over representation analysis uses a one-sided Fisher's exact test. Multiple testing adjustment is performed with the Benjamini-Hochberg method. GO, gene ontology; VSMCs, vascular smooth muscle cells.

https://doi.org/10.1038/s41588-024-02072-3



Extended Data Fig. 10 | ATAC-seq tracks for *TBX5*, *PITX2* and *HSPB7* loci. **a**, *TBX5* locus. **b**, *PITX2* locus. **c**, *HSPB7* locus. Tracks for ATAC-seq from our iPSC-derived atrial cardiomyocytes and seven publicly available ENCODE ATAC-seq datasets (GM23338, liver, skeletal muscle, NK cells, pancreas, thyroid and bladder), as well as two histone modification tracks: H3K4me3 from a CUT&RUN experiment on our iPSC-derived atrial cardiomyocytes and layered H3K27ac from seven published ENCODE cell lines. The histone modifications H3K27ac and H3K4me3 are both associated with active regions in the genome, and are often located at promoters, enhancers or transcription start sites. GWAS variants indicate the location of common variants with genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-analysis. Coordinates are in build GRCh38. ATAC-seq, Assay for Transposase-Accessible Chromatin using sequencing; CMs, cardiomyocytes; CUT&RUN, Cleavage Under Targets & Release Using Nuclease; ENCODE, Encyclopedia of DNA Elements; H3K27ac, histone H3 Lysine 27 acetylation; GWAS, genome-wide association study, H3K4me3, histone H3 lysine 4 trimethylation; iPS cells, induced pluripotent stem cells.

nature portfolio

Corresponding author(s): Patrick T. Ellinor

Last updated by author(s): 07/14/2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not use any commercial software to collect data.
Data analysis	Beagle 4.1; BOLT-LMM v2.3.4; Bowtie2; Eagle v2.3, v2.4; FAST v2.4; g:Profiler ve104_eg51_p15; GCTA v1.93.2beta; GraphTyper v2; gwasurvivr v1.2.0; LDSC v1.0.1; MAGMA v1.09; METAL (version released on 2018-08-28); Minimac v3, v4; PLINK v2.00a1LM, v2.00a2LM; PoPs v0.1; ProbAbel v0.5.0; PRS-CS v1.0.0; R v4.1, v3.5, v3.6.0, v3.6.3; R-package clusterProfiler v4.0.5; R-package coloc v5.1.0; R-package enrichplot v1.12.2 ; R-package survival v2.44-1.1; R-package survminer v0.4.3; rvtest v 2.1.0; SAIGE v0.36.3.2, v0.6; ShapeIT v2.r837; SNPtest v2.5.2; VEP tool v97

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The summary level results file as well as the weights file for the PRSAF are available for download at the Cardiovascular Disease Knowledge Portal under the weblinks https://cvd.hugeamp.org/downloads.html#polygenic and https://cvd.hugeamp.org/downloads.html#summary. The raw and processed ATAC-seq and H3K4me3 data has been deposited at the NCBI Gene Expression Omnibus under accession number GSE225293.

The following datasets were utilized in this study and are publicly available under the listed weblinks:
GENCODE: https://www.gencodegenes.org/
1000G LD reference, MAGMA gene annotations and precomputed files for PoPs algorithm: https://www.dropbox.com/sh/o6t5jprvxb8b500/AADZ8qD6Rpz4uvCk0b5nUnPaa/data?dl=0
GTEx: https://www.gtexportal.org/home/
ENCODE: https://www.encodeproject.org/
OpenTargets: https://www.opentargets.org/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were based on the maximum number of cohorts that were available and could contribute to the meta-analysis at the time. No power calculations were performed to pre-determine the required sample size.
Data exclusions	The pre-imputation quality control consisted of sample level filtering (low call rate, excess heterozygosity, relatedness) and variant level filtering (low call rate, deviation from Hardy-Weinberg Equilibrium, excess heterozygosity, low minor allele frequency). Variants were filtered prior to the meta-analysis for imputation quality > 0.3 and MAF * imputation quality * N events ≥ 10.
Replication	We performed one validation analysis of the common and low frequency sentinel variants in the cohort MVP (Million Veteran Program). Nearly all available sentinel variants (293 of 299) demonstrated consistent directions of effect for the primary and validation analyses, with 215 of the validated variants exceeding a nominal significance level (P-value < 0.05). Additionally we attempted an exploratory analysis for the 14 low frequency sentinel variants based on sequencing data. The sequencing sample had partial overlap with the primary GWAS meta-analysis sample, and can therefore not be considered an independent replication. We could test 11 loci and saw consistent directionality for 10 of them, 5 reached nominal significance.
Randomization	Samples were not experimentally randomized, given that the exposure in our analysis is genetic variation.
Blinding	This is a GWAS meta-analysis of summary level GWAS datasets. No group assigning is performed that would require blinding for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb #9751 was sourced from Cell Signaling Technologies (https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys4-c42d8-rabbit-mab/9751).
Validation	Received directly from the company Cell Signaling Technology: https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys4-c42d8-rabbit-mab/9751 The antibody is commercially available, and has been verified by Western blotting or by peptide ELISA described on the manufacturer's specification sheets.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HUES8 was obtained from Memorial Sloan Kettering Cancer Center.
Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were utilized.

Human research participants

Policy information about studies involving human research participants

Population characteristics	This project included a total of 181,446 AF cases and 1,468,899 controls, meta-analyzed 68 summary level results from more than 40 primary cohorts. Of these samples 60% were non-Icelandic non-Finnish European, 23% were Icelandic, 9% were East Asian, 7% were Finnish, 0.67% were Admixed African and African-American, 0.47% were Hispanic, 0.1% were Brazilian and 0.04% were South Asian as shown in Supplementary Table 2. The countries of origin for each primary study are listed in Supplementary Table 1. Supplementary Table 24 shows the baseline characteristics of all updated or novel AF GWAS studies. Baseline information for previously published, unchanged and non-overlapping studies have previously been published, as reported in Supplementary Table 23.
Recruitment	No new participants were recruited for this study. This meta-analysis includes summary level results from more than 40 individual studies with varying designs: population-based prospective studies, case-control studies, cohort studies or clinical trials. Details on participant recruitment for each participating study can be found in the supplementary materials.
Ethics oversight	Ethical regulations were followed for this study. Written informed consent was obtained from all participants in this study.The UK Biobank resource was approved by the UK Biobank Research Ethics Committee and all participants provided written informed consent to participate. Use of UK Biobank data was performed under application number 17488 and was approved by the local Massachusetts General Brigham Institutional Review Board. The Institutional Review Board (IRB) at Massachusetts General Hospital reviewed and approved the overall study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.