Title

Genetics of myocardial interstitial fibrosis in the human heart and association with disease.

Journal

Nature Genetics, 55(5)

Authors

Nauffal, Victor
Di Achille, Paolo
Klarqvist, Marcus
et al.

Publication Date

2023-05-01

DOI

10.1038/s41588-023-01371-5

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2024 May 21.

Published in final edited form as:

Nat Genet. 2023 May ; 55(5): 777–786. doi:10.1038/s41588-023-01371-5.

Genetics of myocardial interstitial fibrosis in the human heart and association with disease

Victor Nauffal[1,2,11], **Paolo Di Achille**[3,11], **Marcus D. R. Klarqvist**[3,11], **Jonathan W. Cunningham**[1,2,11], **Matthew C. Hill**[2,7,11], **James P. Pirruccello**[2,4,6], **Lu-Chen Weng**[2,7], **Valerie N. Morrill**[2], **Seung Hoan Choi**[2], **Shaan Khurshid**[2,5], **Samuel F. Friedman**[3], **Mahan Nekoui**[2], **Carolina Roselli**[2,8], **Kenney Ng**[9], **Anthony A. Philippakis**[2,3,10], **Puneet Batra**[3], **Patrick T. Ellinor**[2,5,12,*], **Steven A. Lubitz**[2,5,12,*]

[1]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA.

[2]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[3]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[4]Cardiology Division, Massachusetts General Hospital, Boston, MA, USA.

[5]Demoulas Center for Cardiac Arrhythmias and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.

[6]Division of Cardiology, University of California San Francisco, San Francisco, CA, USA.

[7]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.

[8]Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[9]Center for Computational Health, IBM Research, Cambridge, MA, USA.

[10]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[11]These authors contributed equally.

[12]These authors jointly supervised this work.

Abstract

Myocardial interstitial fibrosis is associated with cardiovascular disease and adverse prognosis. To investigate the biological pathways that underlie fibrosis in the human heart, we developed a machine learning model to measure native myocardial T1 time, a marker of myocardial fibrosis, in 41,505 UK Biobank participants who underwent cardiac magnetic resonance imaging. Greater T1 time was associated with diabetes mellitus, renal disease, aortic stenosis, cardiomyopathy, heart

* pellinor@broadinstitute.org; slubitz@mgh.harvard.edu.

Author Contributions

V.N., J.W.C., P.T.E. and S.A.L. conceived the study. M.D.R.K. and P.D.A. ingested and prepared cMRI data. V.N., M.D.R.K., P.D.A., and J.W.C. performed quality control. M.D.R.K. trained machine learning models. V.N., M.D.R.K., and P.D.A. performed the main analyses. M.C.H. performed in vitro experiments. V.N., M.D.R.K., P.D.A., J.W.C., M.C.H., P.T.E. and S.A.L. wrote the paper. J.P.P., L.-C.W., V.N.M., S.H.C., S.K., S.F.F., M.N., C.R., K.N., A.A.P. and P.B. contributed to the analysis plan or provided critical revisions.

Code Availability

Code used to ingest, quality control, and train machine learning models are available at https://github.com/broadinstitute/ml4h under an open-source BSD license.

failure, atrial fibrillation, conduction disease and rheumatoid arthritis. Genome-wide association analysis identified 11 independent loci associated with T1 time. The identified loci implicated genes involved in glucose transport (*SLC2A12*), iron homeostasis (*HFE*, *TMPRSS6*), tissue repair (*ADAMTSL1*, *VEGFC*), oxidative stress (*SOD2*), cardiac hypertrophy (*MYH7B*) and calcium signaling (*CAMK2D*). Using a TGFβ1-mediated cardiac fibroblast activation assay, we found that 9 out of the 11 loci comprised genes that exhibited temporal changes in expression and/or open chromatin conformation supporting their biological relevance to myofibroblast cell state acquisition. By harnessing machine learning to perform large-scale quantification of myocardial interstitial fibrosis using cardiac imaging, we validate associations between cardiac fibrosis and disease, and identify novel biologically relevant pathways underlying fibrosis.

The cardiac extracellular matrix (ECM) is a dynamic compartment that plays key structural and regulatory roles in establishing myocardial tissue architecture and function. Pathologic perturbations to homeostatic turnover of ECM components leads to the progressive development of interstitial fibrosis[1], which is the histological hallmark of several cardiac diseases[2–8]. A myriad of hemodynamic, metabolic and inflammatory stressors contribute to the accelerated development of interstitial fibrosis and associated cardiovascular diseases[9]. There is a critical need to understand the cellular and molecular mechanisms that contribute to pathological cardiac fibrosis in humans, given that their identification would enable the development of targeted anti-fibrotic therapies applicable across a wide-range of cardiovascular diseases. However, progress has been hindered by challenges in reliable non-invasive measurement of interstitial fibrosis at scale in the human heart, and by the lack of adequately powered validation studies of findings from animal or in vitro tissue culture models. The advent of machine learning tools capable of generating imaging-based phenotypes at scale and large biorepositories containing deep phenotyping and genomic data offers a unique opportunity to overcome these challenges.

Native myocardial T1 time measured using cardiac magnetic resonance imaging (cMRI) is a histopathologically validated metric for quantifying interstitial fibrosis in the human heart[10,11]. The UK Biobank (UKB) is a large-scale prospective cohort with rich cMRI[12], genomic, and clinical outcomes data[13]. We sought to use machine learning to quantify fibrosis in over 40,000 study participants who underwent cMRI T1 mapping, assess associations between fibrosis and clinical outcomes, and identify pathways responsible for cardiac fibrosis in humans using genetic association analyses.

RESULTS

Machine learning enables T1 time measurement at scale

We acquired mid-ventricular short-axis cMRI T1 maps for 42,654 participants in the UKB (Fig. 1). In accordance with contemporary guidelines for assessment of cardiac T1 mapping[14–16], we developed a machine learning model to auto-segment myocardial regions of interest within the interventricular septum (IVS) and quantify T1 time (Online Methods). T1 times measured using our machine learning model were highly correlated with manually-derived T1 times in an independent validation set (n = 100; Pearson correlation coefficient r = 0.97, 95% CI 0.95–0.98) (Supplementary Fig. 1). Following T1 map image quality

control, we retained measured T1 time at the IVS for 41,505 participants who constituted our study sample (Supplementary Fig. 2 and Online Methods).

The mean age of participants was 64.0 ± 7.7 years and 48.1% were men. Mean native myocardial T1 time of the study sample was 918.1 ± 41.5 ms (Table 1). Our results were consistent with known sex-specific patterns of higher T1 time in women compared to men (Supplementary Fig. 3)[17,18]. Increasing age was associated with decreased T1 time in females ($\Delta T1(ms)_{/10\ years} = -4.15 \pm 0.64$, $P = 1.9 \times 10^{-36}$) and increased T1 time in males ($\Delta T1\ (ms)_{/10\ years} = 2.52 \pm 0.34$, $P = 2.3 \times 10^{-13}$) after adjusting for comorbidities and medications (Supplementary Fig. 4a). Sub-setting to "healthy" participants free of cardiovascular or metabolic diseases at time of MRI did not alter the sex-specific age-related trends in T1 time (Supplementary Fig. 4b).

Age, sex and body mass index (BMI) contributed to the majority of variation in T1 time in the study sample (Supplementary Table 1). In the sections below, we present findings from multivariable analyses adjusted for demographics, comorbidities and medications relevant to myocardial fibrosis. Additionally, we also provide in the supplementary tables results from models adjusted only for age, sex and BMI for comparison.

T1 time is associated with other cMRI-derived measures

We examined the association of myocardial interstitial fibrosis with cMRI derived-measures of left ventricular and atrial structure and function. Prior to performing association testing, we rank-based inverse normal transformed T1 times. As such, changes in T1 time reported approximate multiples of one standard deviation (SD) of T1 time. In multivariable analysis, lower LV ejection fraction, higher LV mass, and larger LV end systolic volume were associated with increased T1 time. In the left atrium, lower ejection fraction and larger end systolic volume were associated with increased T1 time (Supplementary Table 2). Two-sided *P*-values < 7.1×10^{-3} were considered statistically significant.

T1 time is associated with prevalent diseases

We then investigated whether T1 time is associated with *a priori* selected cardiovascular, metabolic, and systemic inflammatory diseases relevant to myocardial fibrosis by comparing T1 times from participants with prevalent disease to "healthy" participants free of cardiovascular or metabolic disorders at time of cMRI (Online Methods, Fig. 2 and Supplementary Fig. 5). Two-sided *P*-values < 3.1×10^{-3} were considered statistically significant.

Among cardiovascular diseases, higher T1 time was associated with hypertrophic cardiomyopathy ($\Delta T1_{(SD)} = 0.74 \pm 0.17$; $P = 1.3 \times 10^{-5}$), dilated cardiomyopathy ($\Delta T1_{(SD)} = 0.47 \pm 0.12$; $P = 1.3 \times 10^{-4}$), heart failure ($\Delta T1_{(SD)} = 0.41 \pm 0.06$; $P = 1.1 \times 10^{-10}$), atrial fibrillation ($\Delta T1_{(SD)} = 0.21 \pm 0.03$; $P = 7.1 \times 10^{-12}$), atrioventricular node/distal conduction disease ($\Delta T1_{(SD)} = 0.37 \pm 0.06$; $P = 3.7 \times 10^{-11}$), and aortic stenosis ($\Delta T1_{(SD)} = 0.35 \pm 0.10$; $P = 7.6 \times 10^{-4}$). Coronary artery disease ($\Delta T1_{(SD)} = 0.06 \pm 0.03$; $P = 2.1 \times 10^{-2}$) and myocardial infarction ($\Delta T1_{(SD)} = 0.10 \pm 0.04$; $P = 1.8 \times 10^{-2}$) were associated with a trend towards higher T1 time that did not reach the threshold for statistical significance.

T1 time was not associated with hypertension ($\Delta T1_{(SD)}$ = 0.02 ± 0.01; P = 0.18), which is consistent with prior studies that showed a limited ability of T1 mapping to differentiate between individuals with hypertension and controls except among individuals with concomitant left ventricular hypertrophy (LVH) (Fig. 2 and Supplementary Fig. 5)[19]. Accordingly, we stratified participants with hypertension based on presence of concomitant LVH. We found that hypertension was associated with higher T1 times only among participants with concomitant LVH ($\Delta T1_{(SD)}$ = 0.15 ± 0.03; $P = 4.0 \times 10^{-8}$) but not among those without LVH ($\Delta T1_{(SD)} = 4.0 \times 10^{-3}$ ± 0.01; P = 0.75) (Supplementary Fig. 6).

Metabolic disorders, including diabetes mellitus type 1 ($\Delta T1_{(SD)}$ = 0.40 ± 0.07; $P = 4.0 \times 10^{-8}$), diabetes mellitus type 2 ($\Delta T1_{(SD)}$ = 0.27 ± 0.03; $P = 7.7 \times 10^{-20}$), hyperlipidemia ($\Delta T1_{(SD)}$ = 0.18 ± 0.05; $P = 6.4 \times 10^{-5}$), and chronic kidney disease ($\Delta T1_{(SD)}$ = 0.20 ± 0.05; $P = 8.6 \times 10^{-5}$) were associated with higher T1 time (Fig. 2 and Supplementary Fig. 5). Of the systemic inflammatory diseases examined, rheumatoid arthritis was associated with increased T1 time ($\Delta T1_{(SD)}$ = 0.13 ± 0.04; $P = 6.1 \times 10^{-4}$) (Fig. 2 and Supplementary Fig. 5).

We performed a sensitivity analysis comparing disease cases to non-cases (i.e. “healthy” controls and non-cases with other cardiovascular and metabolic diseases) further adjusting for comorbidities and medications and found overall similar findings (Supplementary Table 3 and Supplementary Fig. 7). Stratification by sex yielded consistent disease associations with T1 time among males and females (Supplementary Fig. 8).

We additionally examined the association of serum biomarkers and electrocardiogram (ECG) intervals relevant to the examined prevalent diseases with T1 time. Concordant with the association of diabetes mellitus with increased T1 time, individuals with HgbA1c levels in the prediabetes ($\Delta T1_{(SD)}$ = 0.06 ± 0.02; $P = 4.6 \times 10^{-4}$) and diabetes ($\Delta T1_{(SD)}$ = 0.27 ± 0.04; $P = 5.0 \times 10^{-13}$) ranges[20] had higher T1 time as compared to those with normal HgbA1c levels. Increase in T1 time associated with impaired renal function only became apparent with moderate to severe reductions in estimated glomerular filtration rate[21] (<45 $ml.min^{-1}/1.73\ m^2$; ($\Delta T1_{(SD)}$ = 0.22 ± 0.07; $P = 2.6 \times 10^{-3}$). Among the examined ECG intervals, prolonged QRS interval (>120 ms) was associated with increased T1 time ($\Delta T1_{(SD)}$ = 0.28 ± 0.03; $P = 8.9 \times 10^{-19}$) (Supplementary Table 4 and Supplementary Fig. 9). A two-sided P-value threshold of $< 4.6 \times 10^{-3}$ was used for the biomarker and ECG interval analysis.

T1 time is associated with incident cardiovascular disease

We compared incidence of cardiovascular disease among individuals in the top 20th percentile of T1 time distribution to that in the bottom 80th percentile over a median follow-up of 2.54 years, interquartile range 1.63–3.88 years. We analyzed individual diseases and defined two composite endpoints including major arrhythmia (comprising incident sustained ventricular arrhythmia, cardiac arrest or implantable cardioverter defibrillator implantation), and major adverse cardiovascular events (comprising incident myocardial infarction, heart failure, all-cause mortality, atrial fibrillation or major arrhythmia). Participants in the top 20th percentile of T1 time had a higher risk of incident heart failure (HR 1.66, 99% CI 1.04–2.67; $P = 5.4 \times 10^{-3}$), atrial fibrillation (HR 1.62, 99% CI 1.18–2.23; $P = 9.5 \times 10^{-5}$),

and atrioventricular node/distal conduction disease (HR 1.93, 99% CI 1.32–2.83; $P = 8.8 \times 10^{-6}$) compared to those in the lower 80th percentile. Additionally, incident major adverse cardiovascular events (HR 1.74, 99% CI 1.38–2.20; $P = 1.23 \times 10^{-9}$) were more frequent among participants in the top 20th percentile of T1 time compared to those in the lower 80th percentile (Fig. 3, Supplementary Table 5 and Supplementary Fig. 10). There was a trend towards a higher incidence of the composite major arrhythmia endpoint (HR 1.75, 99% CI 0.87–3.55; $P = 4.0 \times 10^{-2}$) in the top 20th percentile of T1 time that did not reach the threshold for statistical significance (Supplementary Table 5 and Supplementary Fig. 10). Two-sided *P*-values < 0.01 were considered statistically significant. A sex-stratified sensitivity analysis showed consistent findings for both males and females (Supplementary Fig. 11).

Lifestyle factors are associated with T1 time

We investigated the impact of lifestyle factors on myocardial interstitial fibrosis stratified by sex and adjusted for diseases associated with T1 time. In both males and females, light-to-moderate alcohol use ($\Delta T1_{(SD)} = -0.21 \pm 0.04$; $P_{males} = 1.0 \times 10^{-9}$; $\Delta T1_{(SD)} = -0.15 \pm 0.03$; $P_{females} = 6.9 \times 10^{-8}$) and heavy alcohol use ($\Delta T1_{(SD)} = -0.22 \pm 0.04$; $P_{males} = 1.7 \times 10^{-10}$; $\Delta T1_{(SD)} = -0.24 \pm 0.03$; $P_{females} = 5.4 \times 10^{-18}$) were associated with lower T1 time as compared to non-drinkers. Current cigarette smoking across both sexes ($\Delta T1_{(SD)} = 0.26 \pm 0.03$; $P_{males} = 1.3 \times 10^{-14}$; $\Delta T1_{(SD)} = 0.24 \pm 0.04$; $P_{females} = 7.5 \times 10^{-11}$) was associated with increased T1 time compared to non-smokers. We found a salutary effect of adequate self-reported physical activity defined as meeting guideline recommendations of ≥ 150 minutes of moderate or ≥ 75 minutes of vigorous activity per week or the equivalent combination on myocardial interstitial fibrosis[22]. Physical activity meeting guideline recommendations was associated with decreased T1 time in both males and females ($\Delta T1_{(SD)} = -0.07 \pm 0.02$; $P_{males} = 1.6 \times 10^{-6}$; $\Delta T1_{(SD)} = -0.08 \pm 0.02$; $P_{females} = 3.1 \times 10^{-8}$). Among participants with BMI < 30 kg/m^2, increasing BMI was associated with lower T1 time in both males and females ($\Delta T1_{(SD)\ /kg/m2} = -0.08 \pm 0.01$; $P_{males} = 1.4 \times 10^{-162}$; $\Delta T1_{(SD)\ /kg/m2} = -0.06 \pm 0.01$; $P_{females} = 1.0 \times 10^{-162}$). The trend reversed among obese male participants with BMI > 30 kg/m^2. Increasing BMI was associated with increased myocardial fibrosis and T1 time ($\Delta T1_{(SD)\ /kg/m2} = 0.03 \pm 0.01$; $P_{males} = 2.4 \times 10^{-7}$) in obese males but was not associated with T1 time in obese females ($\Delta T1_{(SD)\ /kg/m2} = 4.0 \times 10^{-3} \pm 4.6 \times 10^{-3}$; $P_{females} = 0.35$) (Supplementary Table 6 and Supplementary Fig. 12). Two-sided *P*-values < 1.3×10^{-2} within each stratum were considered statistically significant.

Causal effect of diabetes mellitus type 1 on cardiac fibrosis

We explored the causal contribution of cardiovascular risk factors and diseases to myocardial fibrosis using two-sample Mendelian randomization (MR). The inverse variance weighted (IVW) method constituted our primary analysis. We used the MR-Egger method as a sensitivity analysis to account for potential horizontal pleiotropy[23]. We found consistent evidence from both the IVW and MR-Egger method supporting a potential causal effect of diabetes mellitus type 1 ($\Delta T1_{IVW\ (SD)\ /log\ odds\ of\ genetic\ predisposition\ to\ DM1} = 0.007 \pm 0.002$; $P = 1.0 \times 10^{-3}$; $\Delta T1_{MR\text{-}Egger\ (SD)\ /log\ odds\ of\ genetic\ predisposition\ to\ DM1} = 0.01 \pm 0.003$; $P = 1.0 \times 10^{-3}$) on myocardial interstitial fibrosis. This was also supported by inspecting the

scatter plot of the association of individual genetic variants comprising the diabetes mellitus type 1 genetic instrument with both T1 time and diabetes mellitus type 1 (Supplementary Fig. 13). A causal effect of diastolic blood pressure on myocardial T1 time was suggested by the inverse variance weighted method, but this association was not supported by the MR-Egger method or the scatter plot of variant effect sizes. We did not find evidence for a causal effect of BMI, systolic blood pressure, chronic kidney disease, estimated glomerular filtration rate, diabetes mellitus type 2, coronary artery disease, atrial fibrillation or serum lipoproteins on myocardial fibrosis (Supplementary Table 7). Two-sided *P*-values < 3.9×10^{-3} were considered statistically significant.

Genetic association analysis highlights loci relevant to fibrosis

Next, we sought to determine the genetic basis of interstitial fibrosis by performing genetic association analysis of T1 time. The single nucleotide polymorphism (SNP) heritability (h^2_g) of T1 time was 0.13, which is lower than that of other cMRI phenotypes in the UKB such as LV mass (0.26), LV end diastolic volume (0.40), and LV end systolic volume (0.31)[24]. Interestingly, in contrast to the strong association of measured T1 time with other cMRI measures of left ventricular and atrial structure and function, we observed limited genetic correlation of T1 time with these cMRI measures, suggesting that while shared exposures drive the association between these MRI measures and T1 time, distinct biologic pathways contribute to development of myocardial interstitial fibrosis (Supplementary Fig. 14).

We performed a genome-wide association study (GWAS) and discovered 11 genome-wide significant loci (Fig. 4a and Supplementary Table 8). There was no evidence of inflation in our GWAS results ($\lambda_{GC} = 1.053$, LD score regression intercept = 1.0002) (Supplementary Fig. 15). Regional association plots for genome-wide significant SNPs are shown in Supplementary Figure 16.

A regulatory region variant in a promoter-flanking region upstream of the solute carrier *SLC2A12* gene was the most significant lead SNP (rs2627230_T; $P = 8.1 \times 10^{-14}$) and associated with increased T1 time. *SLC2A12* encodes solute facilitated glucose transporter member 12 (GLUT12), a basal and insulin-independent glucose transporter in the heart[25] with previously reported associations with idiopathic dilated cardiomyopathy in humans[26], as well as insulin resistance[27] and kidney disease[28] in animal models. Lead SNPs near *SOD2* (rs9457699_G; $P = 2.0 \times 10^{-11}$) and *VEGFC* (rs365843_T; $P = 3.2 \times 10^{-9}$), two genes with established roles in cardiac hypertrophy and fibrosis in animal models[29,30], were associated with decreased T1 times. Another lead SNP associated with T1 time is an intronic variant in *ADAMTSL1* (rs1576900_A; $P = 3.6 \times 10^{-11}$), a gene encoding an ADAMTS-like protein which is thought to modulate the function of ADAMTS metalloproteinases with integral roles in ECM turnover[31,32]. Additionally, rs6120777_A, an intronic variant in *MYH7B*, was among the lead genome-wide significant variants ($P = 9.2 \times 10^{-11}$). *MYH7B* has been associated with familial hypertrophic cardiomyopathy[33], but no associations between rs6120777_A and hypertrophic cardiomyopathy have been yet reported. Notably, the association of rs6120777_A with T1 time persisted even after exclusion of hypertrophic cardiomyopathy cases in a sensitivity analysis (Supplementary Table 9).

Interestingly, we also identified rs115740542_C ($P = 2.7 \times 10^{-10}$), a variant near the *HFE* gene in perfect linkage disequilibrium (LD) with rs1800562_A, which leads to the missense change p.Cys282Tyr and is the most common cause of hereditary hemochromatosis, an iron overload disorder associated with cardiomyopathy[34]. Additionally, rs855791_A, a missense variant in *TMPRSS6* with previously reported associations with iron homeostasis[35], was associated with T1 time ($P = 5.7 \times 10^{-10}$).

We identified a variant in *CAMK2D* (rs55754224_T; $P = 1.4 \times 10^{-9}$) that has been previously associated with atrial fibrillation[36]. The remaining genome-wide significant variants were located near genes associated with cardiac arrhythmias[37–39], cardiac remodeling[40] and myocyte cytoskeletal proteins[41], including *PPP2R3A*, *PIM1* and *KANK1*, respectively.

In conditional analysis, no additional independent genome-wide significant SNPs were identified. We performed a sensitivity analysis after excluding participants with diseases known to be associated with focal replacement fibrosis (myocardial infarction, heart failure, and dilated/hypertrophic cardiomyopathy) and identified 10 genome-wide significant loci, 9 of which overlapped with the main GWAS loci. The subthreshold variant rs2271426_T, an intronic variant in *PIK3C2B*, reached genome-wide significance ($P = 4.1 \times 10^{-8}$) in the GWAS sensitivity analysis (Supplementary Figs. 17 and 18 and Supplementary Table 9).

Excluding participants with prevalent hereditary hemochromatosis ($n = 62$) did not alter the association with T1 time of the two loci, *H2BC4/HFE* and *TMPRSS6*, associated with iron homeostasis (Supplementary Figs. 19 and 20 and Supplementary Table 10). We then performed a sex-stratified GWAS and identified no novel or sex-specific loci (Supplementary Figs. 21 and 22 and Supplementary Table 11).

ADAMTSL1 and *SLC2A12* expression is associated with fibrosis

Of the 11 lead SNPs identified in the GWAS of T1 time, 8 (or their proxies, $r^2 > 0.8$) were expression quantitative trait loci (eQTL) in the LV or right atrial appendage (Supplementary Table 12). Notably, rs1576900_A, the lead GWAS intronic variant in *ADAMTSL1*, was also the top eQTL variant for *ADAMTSL1* in both the LV and right atrial appendage and associated with decreased *ADAMTSL1* expression and lower T1 time (Supplementary Figs. 23 and 24). In addition, rs2627230_T, the lead GWAS variant in a regulatory region upstream of *SLC2A12*, was the top eQTL variant for *SLC2A12* in the LV and associated with decreased *SLC2A12* expression and higher T1 time (Supplementary Fig. 23).

In transcriptome-wide association analysis, increased expression of *ADAMTSL1* in left ventricular tissue ($P = 5.3 \times 10^{-9}$) and right atrial appendage ($P = 1.3 \times 10^{-7}$) was associated with increase in myocardial interstitial fibrosis as measured by T1 time. On the other hand, increased expression of *SLC2A12* in the LV ($P = 1.9 \times 10^{-10}$) was associated with decrease in myocardial interstitial fibrosis (Fig. 4b and Supplementary Tables 13 and 14).

Multi-omic assessment of T1 time-associated loci

To further evaluate T1 time genome-wide significant loci, we implemented a cellular assay for fibroblast activation whereby primary human cardiac fibroblasts are stimulated

with TGFβ1 (Fig. 5a). We profiled the transcriptional and epigenomic characteristics of these TGFβ1 stimulated fibroblasts temporally using RNA-seq and ATAC-seq. Principal component analysis of the transcriptomic data revealed that the TGFβ1 stimulated cardiac fibroblasts followed a distinct trajectory (Fig. 5b). Differential expression analysis (false discovery rate (FDR) < 0.01 and $|\log_2$ fold change$| > 1$) comparing unstimulated controls to cells treated with TGFβ1 for 72 hours (Supplementary Table 15) uncovered several regulators of cardiac fibrosis, including *MEOX1* (Fig. 5c)[42]. To interrogate our GWAS loci, we examined genes nearest to the lead SNPs within a locus in addition to non-overlapping genes within these loci that were identified in our eQTL lookup. Notably, *MYH7B*, *TMPRSS6*, *NCK1-AS1*, *U91328.19* and *PNLDC1* are not expressed in cardiac fibroblasts and could not be assessed. Of the 14 genes expressed in cardiac fibroblasts, 5 genes within 5 distinct loci showed evidence of differential gene expression with TGFβ1 treatment at FDR < 0.01 (Fig. 5d).

Next, we interrogated the open chromatin landscapes in cardiac myofibroblasts. Consistent with the efficacy of our assay, we observed differential chromatin accessibility at canonical downstream TGFβ targets, including *IGFBP3* (Fig. 5e). Overall, the epigenetic and transcriptional trajectories were similar (Fig. 5f). Differential chromatin accessibility analysis (FDR $< 1 \times 10^{-10}$, Supplementary Table 16) across stimulated fibroblasts uncovered 15,428 peaks that separated into 6 clusters (Fig. 5g). Among these peaks, we identified an enrichment for SMAD and TEAD binding sites, which have recently been implicated as regulators of fibroblast-to-myofibroblast cell state transitions, via *de novo* motif enrichment analysis (Fig. 5g)[43]. Of the 19 prioritized genes, 7 genes within 7 distinct loci had annotated differentially accessible peaks with TGFβ1 treatment (FDR $< 1 \times 10^{-10}$, $|\log_2$ fold change$| > 1$) (Fig. 5h). Intersecting our gene set with the differentially expressed genes and ATAC peaks, we identified three overlapping genes including *VEGFC*, *KANK1* and *PIM1* (Fig. 5h). Consistent with a decrease in mRNA expression, we observed a decrease in promoter accessibility in *KANK1*, in addition to additional changes in chromatin accessibility (Fig. 5i).

Using our prioritized gene set, 9 of the 11 genome-wide significant loci comprised genes that showed evidence of differential transcriptional or epigenetic signatures with TGFβ1 stimulation. Expanding our gene set to include all genes within 250 kb from lead variants in our loci ($n_{genes} = 139$), we identified additional 7, 13 and 2 genes that with TGFβ1 treatment were differentially expressed only, had annotated differentially accessible peaks only or both, respectively (Supplementary Fig. 25). Using the expanded gene set, all 11 loci included genes that showed responsiveness to TGFβ1 treatment.

DISCUSSION

We developed an automated machine-learning model to measure myocardial interstitial fibrosis in over 40,000 participants in the UKB. We identified associations between myocardial fibrosis and diabetes mellitus, renal disease, aortic stenosis, cardiomyopathy, atrial fibrillation, conduction disease, and rheumatoid arthritis. MR analysis provided evidence for a causal effect of diabetes mellitus type 1 on myocardial interstitial fibrosis. Furthermore, greater myocardial fibrosis at the time of cMRI was associated with incident

cardiovascular disease over a median follow-up of 2.5 years. In the first large-scale GWAS of T1 time in the human heart, we identified 11 independent loci implicating genes involved in biological pathways relevant to fibrosis, including glucose transport (*SLC2A12*), iron homeostasis (*HFE*, *TMPRSS6*), tissue repair (*ADAMTSL1*, *VEGFC*), oxidative stress (*SOD2*), cardiac hypertrophy (*MYH7B*), and calcium signaling (*CAMK2D*). Using a cellular assay for TGFβ1-mediated cardiac fibroblast activation, we demonstrate that the GWAS loci are enriched for genes that exhibit transcriptional and epigenetic changes following treatment with TGFβ1, further supporting their functional relevance to myocardial fibrosis. Overall, the heritability of myocardial interstitial fibrosis as measured by native myocardial T1 time was relatively low, emphasizing the important contribution of non-genetic environmental and lifestyle factors to cardiac fibrosis.

Our findings have several major implications. First, our results highlight the role of glucose homeostasis and diabetes in myocardial fibrosis and pinpoint potential additional pathways for further interrogation. Our strongest GWAS variant fell within a regulatory region upstream of *SLC2A12*, which encodes an insulin-independent glucose transporter (GLUT12) highly expressed in the heart[25]. Additionally, in MR analysis, we found evidence supporting a potential causal association between diabetes mellitus types 1 and increased myocardial T1 time. GLUT12 knock-out in zebrafish leads to development of heart failure and a diabetic phenotype[27], consistent with our TWAS results, suggesting that decreased expression of *SLC2A12* in cardiac tissue was associated with increased interstitial fibrosis.

Second, pathways involved in tissue repair were associated with myocardial fibrosis. Increased human cardiac expression of *ADAMTSL1*, which encodes an ADAMTS-like protein that lacks catalytic activity and is thought to modulate the function of ADAMTS metalloproteinases with integral roles in ECM turnover[31,32], was associated with higher myocardial interstitial fibrosis in this study. The exact effect of ADAMTSL-1 on ADAMTS metalloproteinases remains unknown; however, homology between mammalian ADAMTSL-1 and invertebrate papilin, a known inhibitor of ADAMTS-2, has been reported[44,45]. Studies in mice with cardiac-specific overexpression of *Adamts2* have shown an abrogated pressure overload-induced hypertrophic response[46]. Thus, potential inhibition of ADAMTS-2 activity may explain the increased myocardial fibrosis associated with increased expression of *ADAMTSL1* in the human myocardium. *VEGFC*- and *VEGFD*-mediated lymphangiogenesis has been associated with cardiac repair, and knock-out zebrafish models go on to develop severe cardiac hypertrophy and myocardial interstitial fibrosis[30]. In this study, the lead SNP rs365843_T tagging *VEGFC* was associated with increased expression of *VEGFC* in human right atrial appendage tissue and with lower T1 time reflecting decreased myocardial interstitial fibrosis. In line with the above findings, TGFβ1 treatment of human cardiac fibroblasts was associated with decreased expression of *VEGFC* in our cellular assay. Thus, our current findings extend those from animal models and suggest a role for reparative pathways involving *ADAMTS* and *VEGFC* in reducing myocardial fibrosis in the human heart.

Third, our results shed light on the role of myocardial oxidative stress in the development of myocardial interstitial fibrosis. Lead SNP rs9457699_G is an eQTL for *SOD2* in the human LV and was associated with increased expression of *SOD2* and lower T1 time.

This is congruent with findings from *Sod2* knockout mice, which exhibit increased levels of oxygen reactive species with associated myocardial fibrosis and development of dilated cardiomyopathy[29].

Fourth, several established pathways involved in myocardial fibrosis were implicated in our results. We provide further evidence for the role of *CAMK2D* in pathologic cardiac remodeling and fibrosis[47,48] via association of rs55754224_T, an intronic variant in *CAMK2D*, with increased myocardial T1 time. rs55754224_T has been previously reported to be associated with increased risk of atrial fibrillation[36], an atrial arrhythmia in which atrial remodeling and myocardial fibrosis are central pathologic features[7]. *PIM1* has been shown to play a role in antagonizing cell senescence[49] and reducing myocardial infarct size[50] and was implicated in our results.

Fifth, we identified two genes, *HFE* and *TMPRSS6*, associated with iron homeostasis in our GWAS even following exclusion of participants with prevalent hemochromatosis. Iron deposition in the heart alters myocardial tissue magnetic properties and is associated with lower T1 time and with development of iron overload cardiomyopathy[51].

Sixth, we found an association between modifiable lifestyle behaviors and burden of myocardial interstitial fibrosis. Our findings emphasize the importance of abstaining from cigarette smoking and regular physical activity for cardiovascular health. Extremes of weight gain or loss were associated with greater myocardial fibrosis, particularly in males. The J-shaped association between BMI and cardiovascular outcomes has been increasingly recognized in epidemiologic studies[52]. Alcohol use was associated with decreased T1 time in both males and females which has been reported in prior studies[53]. Prolonged heavy alcohol consumption (>50 drinks per week for > 10 years) is known to be associated with alcohol-induced cardiomyopathy[54]; however, this population is not well-represented in the UKB.

Our study has several limitations. First, contrast agents were not used in the UKB cMRI protocol, which prohibited the assessment of late gadolinium enhancement and extracellular volume fraction. Second, the T1 maps in the UKB are obtained at a single mid-ventricular short-axis slice, and we cannot be certain that a single slice is representative of myocardial fibrosis throughout the LV. Third, the presence of tissue edema and paramagnetic ions may alter T1 time independent of interstitial fibrosis. Fourth, UKB participants are predominantly of European ancestry, and findings from our genetic analysis may not apply to other ancestries. As new multi-ancestry biorepositories with deep phenotyping emerge, further validation of our study findings in an independent multi-ancestry sample will be possible. Fifth, longer follow-up and continued imaging of UKB participants will allow for better powered analyses examining the prognostic role of myocardial interstitial fibrosis for cardiovascular disease in the future. Notably, our findings of increased risk of incident major arrhythmia associated with increased interstitial fibrosis should be interpreted with caution given the relatively limited number of events ($n = 75$) examined. Sixth, phenotypic characterization of the UKB was performed using disease/procedure codes and self-report using surveys, which may be subject to misclassification bias. Seventh, our cellular assay for cardiac fibrosis was based on TGFβ1 stimulation of cardiac fibroblasts; however, multiple

other cell types and non-TGFβ1 dependent pro-fibrotic pathways contribute to myocardial interstitial fibrosis, which could not be assessed by this assay.

In conclusion, machine learning enables quantification of myocardial interstitial fibrosis at scale. Our study yields insights into novel biological pathways underlying cardiac fibrosis and prioritizes several pathways relevant to myocardial fibrosis for further investigation.

ONLINE METHODS

In the sections below, we provide a detailed description of the methods used in this manuscript. Briefly, we trained a machine learning model to segment cardiac T1 maps from the UKB and measure native myocardial T1 time at the IVS. We examined the associations between T1 time and cardiometabolic risk factors, cardiovascular diseases, serum biomarkers, ECG intervals and lifestyle factors. We performed a two-sample MR analysis to examine the causal effect of cardiovascular risk factors and disease on myocardial interstitial fibrosis. We then performed genome- and transcriptome-wide association analyses of native myocardial T1 time. Lastly, we explored the relevance of the identified T1 time-associated loci to myocardial fibrosis by profiling the transcriptional and open chromatin characteristics of genes within these loci using a cellular assay for TGFβ1-mediated cardiac fibroblast activation.

Study design and sample

The UKB is a prospective cohort of 502,629 individuals from the UK enrolled between 2006–2010 with deep phenotyping, imaging and multiple genomic data types. The cohort design has been previously described[13,55]. Briefly, around 9.2 million individuals 40–69 years old living in England, Scotland, and Wales were invited to participate in the study, and 5.4% agreed to participate. Extensive questionnaire data, physical measures, and biological samples were collected at baseline, with ongoing data collection in large subsets of the cohort, including repeated assessments and multimodal imaging. Starting in 2014, 42,654 participants have returned for the first multi-modal imaging visit, including cMRI with T1 mapping, allowing for the assessment of myocardial interstitial fibrosis[12]. All study participants are followed longitudinally for health-related outcomes through linkage to national health-related datasets.

Use of UKB data was performed under application number 7089 and was approved by the local Massachusetts General Hospital institutional review board. All participants signed informed consent prior to participation in the UKB.

cMRI T1 mapping protocol and image quality control

A standardized non-contrast enhanced cMRI protocol using a clinical wide bore 1.5 Tesla scanner (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany) is performed on all cMRI substudy participants. The scanner is equipped with 48 receiver channels, a 45 mT/m and 200 T/m/s gradient system, an 18-channel anterior body surface coil used in combination with 12 elements of an integrated 32 element spine coil and ECG gating for cardiac synchronization. The imaging protocol includes: 3 long-axis cines, 1 short-axis cine, phase contrast sequence at the left ventricular outflow tract, 3

segment short-axis tagging and midventricular short-axis T1 mapping. Native T1 mapping within a single breath hold was performed using the Shortened Modified Look-Locker Inversion recovery (ShMOLLI, WIP780B) technique. The following imaging parameters for T1 mapping were implemented: field of view 360 × 236 mm, voxel size 0.9 × 0.9 × 8.0, flip angle 35 degrees and TR/TE 2.6/1.07 ms[12]. T1 maps were generated online and stored in the UKB imaging database.

To date, the UKB MRI core lab has only released raw T1 maps to UKB researchers. As such, we developed our own automated pipeline to measure native myocardial T1 time from raw T1 maps (Fig. 1). First, we set up an automatic procedure to identify raw T1 map series among the several files provided by the UKB under the category “Experimental shMOLLI sequence images” (UKB Field ID 20214). Preliminary explorations indicated that the T1 map series names contained the “t1map” keyword and explicitly mentioned the MRI “sax (short-axis)” view. Therefore, we discarded all series with names not containing either of the two keywords. Then, we set up a standardized quality control process using a custom online tool to streamline the review of the 42,654 selected T1 maps by four experienced MRI reviewers (V.N., M.D.R.K., P.D.A., and J.W.C.; Supplementary Fig. 26). Native myocardial T1 time was measured at the interventricular septum (IVS)[14]. All images were reviewed and assessed for overall image quality as well as artifacts involving myocardial segments (IVS and LV free wall). A cardiologist (V.N.) reviewed all images that were flagged by any of the four reviewers and made a final ascertainment on image quality and extent of artifact involving myocardial segments. Off-axis images and those with severe distortion of overall image pixel intensity were excluded (Supplementary Fig. 27). Artifact within a myocardial segment was deemed major if it affected at least one-third of the segment of interest (Supplementary Fig. 27). T1 maps with major artifacts involving the IVS (n = 1,149) were excluded from the final T1 time study sample (Supplementary Fig. 2). T1 maps with major artifacts restricted to the LV free wall were included and contributed to the final T1 time study sample (Supplementary Fig. 27). Artifacts were 3-fold higher within the LV free wall segment as compared to the IVS. Current consensus guidelines recommend using the IVS to avoid susceptibility artifacts from lung, liver or veins and for accurate and reproducible measurement of native myocardial T1 time[14–16].

Semantic segmentation and T1 time measurement

Six hundred (500 training, 100 validation) T1 maps were randomly selected and used to develop our machine learning model (Supplementary Fig. 2). To train a machine learning model to segment the IVS, two cardiologists (V.N. and J.W.C.) labeled all cardiac structures within the short-axis T1 maps (350 V.N.; 250 J.W.C.). Fifty T1 maps were labeled by both readers to allow for assessment of inter-reader reliability. Cardiac structures that were labeled included: IVS, LV free wall, papillary muscles/trabeculae, LV blood pool, RV free wall and RV blood pool. Additionally, a region of interest encompassing the mid-myocardium within the IVS and excluding the blood pool on either side was delimited (Supplementary Fig. 28). Pixel intensity values were transformed to T1 times using the accompanying T1 map legend. Native myocardial T1 time was measured as the median T1 time for all pixels within the corresponding IVS region of interest.

The manual tracing procedure, called semantic segmentation, displayed high inter-reader concordance at the IVS between the cardiologist-labeled segmentations, as measured in 50 overlapping cMRI acquisitions (Sørensen-Dice coefficients: 0.84, 95% CI 0.76–0.92). Additionally, IVS-derived T1 times were highly correlated between the two readers (Pearson correlation coefficient $r = 0.95$, 95% CI 0.92–0.97) (Supplementary Fig. 29).

We then trained a machine learning model to identify the IVS using the cardiologist-segmented data as truth labels. The machine learning model had high accuracy for segmenting the IVS when tested in the validation set (Sørensen-Dice coefficients: 0.82, 95% CI 0.70–0.94). Model predictions were then post-processed. First, we subtracted the blood pool, papillary muscles/trabeculae, LV and RV free walls, and any non-cardiac structures and retained the auto-segmented IVS. Second, to minimize the potential of contamination by residual blood pool, trabeculae or non-cardiac structures, we applied a sequence of morphological operations (i.e., skeletonization followed by dilation with a 3-pixel kernel) to the IVS segment and generated representative mid-myocardial regions of interest within the IVS. In the validation set, automatically generated T1 times were highly correlated with T1 times derived from manually traced regions of interest (Pearson correlation coefficient $r = 0.97$, 95% CI 0.95–0.98) (Supplementary Fig. 1).

Next, the machine learning model was used to segment the IVS in the remaining 42,054 cMRIs not used for model training or validation, followed by automated selection of regions of interest and measurement of T1 time. To maximize the quality of the generated T1 times, we manually reviewed all 42,654 T1 maps to exclude low-quality acquisitions and major artifacts affecting the IVS (see above section: cMRI T1 mapping protocol and image quality control). Following quality control, we retained measured T1 time at the IVS for 41,505 participants who constituted our study sample (Supplementary Fig. 1). Mean T1 time of the study sample was 918.1 ± 41.5 ms (Table 1). These values are consistent with previously reported T1 times in a smaller study from the UKB[56] including 11,882 cMRIs.

Machine learning model development

For segmenting cardiac structures in cMRI T1 maps, we employed the DenseNet-121 architecture[57] as the base encoder model in a U-Net model[58] that was pre-trained on ImageNet[59]. DenseNets are constructed with two principal building blocks: (1) dense blocks comprising of batch normalization, the non-linear ReLU activation function, and 3×3 convolutions of increasing number of channels that are propagated from previous layers to enable efficient gradient flow; and (2) transition blocks that compress the number of channels by half using channel-wise convolutions (1 × 1), and perform a spatial reduction by a factor of 2 by using an average pooling layer of stride 2 and pool size 2. The U-Net architecture contains long-range skip connections that allow for pixel-accurate segmentation by sharing feature information along a contracting-expansive path. This is achieved by concatenating features at each downsampling in the encoder with the corresponding features at each upsampling step. These ‘skip connections’ preserve contextual and spatial information.

The inputs for this model were the cMRI T1 maps with size 288 × 384 × 3. The models were trained with the Adam optimizer[60] with a learning rate set to a cosine decay policy decaying

from 0.0001 to 0 over 100 epochs, weight decay of 0.0001, categorical cross-entropy as the loss function, and a batch size of 16. No additional hyperparameter search or ablation studies were performed.

For all training data, the following augmentations (random permutations of the training images) were applied: random shifts in the XY-plane by up to ± 16 pixels and rotations by up to ± 5 degrees around its center axis.

Phenotype derivation and association with T1 time

Prevalent cardiometabolic, cardiovascular and systemic inflammatory diseases at time of first visit for cMRI as well as incident cardiovascular events were ascertained using International Classification of Diseases, 9th and 10th editions, codes and OPCS Classification of Interventions and Procedures version 4 codes as well as self-report using surveys (Supplementary Table 17). Derived myocardial T1 times were rank-based inverse normal transformed and thus, reported changes in T1 times are dimensionless and reflect approximately multiples of 1 SD of the underlying quantitative trait. Multiple linear regression was used to assess the association of a priori selected prevalent cardiometabolic, systemic inflammatory and cardiovascular diseases relevant to myocardial fibrosis at time of MRI with T1 time by comparing disease cases to “healthy” controls free of cardiovascular and metabolic disorders. Healthy participants were selected to be free of prevalent dilated cardiomyopathy, hypertrophic cardiomyopathy, heart failure, atrial fibrillation, atrioventricular node/distal conduction disease, hypertension, diabetes mellitus, aortic stenosis, chronic kidney disease, hemochromatosis, and rheumatoid arthritis. Multivariable models were adjusted for age at MRI visit, sex, BMI, MRI scanner, beta blocker use, angiotensin-converting enzyme inhibitor/angiotensin receptor blocker use, statin use and mineralocorticoid receptor antagonist use. We observed a J-shaped relationship between BMI and T1 time with higher BMI being associated with lower T1 time until the obesity threshold (BMI > 30) was exceeded, at which time BMI was associated with increased T1 time. As such, we modeled BMI as a linear spline with a knot at 30 kg/m^2. We performed three sensitivity analyses. First, we performed an analysis comparing cases to non-cases and further adjusting for comorbidities including myocardial infarction, diabetes mellitus type1, diabetes mellitus type 2 and chronic kidney disease. Second, we performed a sex-stratified analysis to examine trends of association across males and females. Third, to examine the impact of the presence of left ventricular hypertrophy (LVH) on the association of hypertension with T1 time, we performed a stratified analysis by LVH status. LVH was defined as left ventricular mass indexed to body surface area > 72 g/m^2 in males and > 55 g/m^2 in females using recently defined reference values in the UKB[61]. A *P*-value threshold of $< 3.1 \times 10^{-3}$ (0.05/16) was used to determine statistically significant associations after adjusting for multiple testing.

We leveraged biomarker and ECG data in the UKB to examine the association of T1 time with serum biomarkers and ECG intervals associated with the examined prevalent diseases. Examined serum biomarkers were measured at time of enrollment in the UKB and included hemoglobin A1c, lipoproteins, C-reactive protein (CRP), creatinine, cystatin c and insulin-like growth factor 1 (IGF-1). Of 41,505 participants with T1 time measured, 33,241

had serum biomarker data available. The examined ECG intervals included automated measurements of the QRS interval, P-wave duration and PR interval (GE Marquette 12SL ECG analysis program) from supine resting 12-lead ECGs performed at time of MRI. We excluded ECGs from participants with a diagnosis of Wolff-Parkinson-White syndrome, a paced rhythm, atrial fibrillation or atrial flutter at time of ECG, 2nd or 3rd degree AV block, digoxin use or class I or III antiarrhythmic drug use. Of 41,505 participants with T1 time measured, 28,602 had ECG data that met our inclusion criteria. Estimated glomerular filtration was calculated using the CKD-Epi equation incorporating both serum creatinine and cystatin c[62]. We categorized biomarkers based on clinically established thresholds where possible[20,21,63,64]. CRP and IGF-1 were categorized into top decile vs. lower 90th percentile. Multiple linear regression was used to assess the association of biomarkers and ECG intervals with T1 time further adjusting for age at MRI, sex, BMI, MRI scanner, myocardial infarction, beta blocker use, angiotensin converting enzyme inhibitor/angiotensin receptor blocker use, statin use and mineralocorticoid receptor antagonist use. Two-sided *P*-values < 4.6×10^{-3} (0.05/11) were considered statistically significant.

A time-to-event analysis was performed to assess the association of T1 time with incident cardiovascular events. Follow-up time was defined as time from MRI visit to first occurrence of the outcome of interest, death or last follow-up (April 30th, 2020). For each incident disease analysis, study participants with prevalent disease at time of MRI were excluded as they were not at risk for the outcome of interest. We then stratified the cohort into the upper 20th and lower 80th percentile of T1 time. Using a multivariable Cox proportional hazards model adjusted for age at MRI, sex, BMI, MRI scanner, myocardial infarction, diabetes mellitus type 1, diabetes mellitus type 2, chronic kidney disease, beta blocker use, angiotensin converting enzyme inhibitor/angiotensin receptor blocker use, statin use, and mineralocorticoid receptor antagonist use, we examined the association of T1 time with incident cardiovascular events. Adjusted Kaplan-Meier curves were constructed to compare incidence rate of cardiovascular events between the two groups. The validity of the proportional hazards assumption was verified by examining the Schoenfeld residuals. Two-sided *P*-values < 0.01 (0.05/5) were considered statistically significant. All statistical tests were performed using R version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) (R, Core Team 2020).

Association analysis of lifestyle factors with T1 time

Lifestyle factors including alcohol use, cigarette smoking, physical activity and obesity were examined for association with T1 time. Alcohol use data was based on self-report in the UKB. All participants completed a touchscreen questionnaire at their initial assessment containing a series of questions pertaining to alcohol use. Participants reporting weekly drinking were shown a chart with common alcoholic beverages and corresponding units of alcohol and asked to report the average number of drinks consumed per week by category. Participants reporting drinking one to three times a month or on special occasions were asked to report the average number of drinks consumed per month. We aggregated this data and converted the number of drinks to grams of alcohol based on units defined by the National Health Service (https://www.nhs.uk/live-well/alcohol-advice/calculating-alcohol-units/) (e.g., 1 standard glass of wine = 2.1 units = 16.8 grams of

alcohol). Participants who responded with “prefer not to answer,” “I do not know,” or with incomplete responses were excluded. According to the National Institute on Alcohol Abuse and Alcoholism, light-to-moderate alcohol use, termed “drinking in moderation”, was defined as ≤ 98 grams/week for women and ≤ 196 grams/week for men, with 1 standard US drink containing 14 grams of alcohol. Heavy alcohol use was defined as > 98 grams/week for women and > 196 grams/week for men (https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking). Cigarette smoking status was categorized into non-smokers, prior smokers and current smokers. Physical activity was assessed using self-reported activity questionnaires. Adequate physical activity was defined as self-reported activity that met standard American Heart Association guidelines[22] (≥ 150 minutes of moderate or ≥7 5 minutes of vigorous intensity activity, or equivalent combination per week). Finally, for BMI and given the observed J-shaped relationship with T1 time, we report the association of BMI with T1 time separately for those with BMI < 30 and ≥ 30 kg/m^2. We examined the association of lifestyle factors with T1 time stratified by sex and adjusted for age at MRI, myocardial infarction, diabetes mellitus type 1, diabetes mellitus type 2, heart failure, atrial fibrillation, chronic kidney disease, angiotensin converting enzyme inhibitor use, angiotensin receptor blocker use, beta blocker use, mineralocorticoid receptor antagonist use and statin use. We additionally adjusted for BMI modeled as a linear spline with knot at 30 kg/m^2 when BMI was not the examined outcome. Within each sex stratum, two-sided *P*-values < 1.3×10^{-2} (0.05/4) were considered statistically significant.

Genomic data, imputation, sample and variant quality control

In total, 488,377 UKB participants were genotyped using either one of two overlapping arrays, the UK BiLEVE Axiom Array or the UKB Axiom Array. Prior to imputation, a number of quality control filters were applied to the genotype data. Variants with > 5% missing rate, minor allele frequency < 0.0001 and that violated Hardy-Weinberg Equilibrium (*P*-value < 1×10^{-12}) were excluded. Additionally, samples that were identified as outliers for genotype missingness rate (> 5%) and heterozygosity were also excluded. These filters resulted in a genotype dataset that included 670,730 autosomal variants in 487,442 samples. Imputation into the Haplotype Reference Consortium (HRC) and UK10K+ 1000G phase 3 reference panels was carried out using IMPUTE4. The imputation process resulted in a dataset with 93,095,623 autosomal SNPs and short indels in 487,442 individuals[13].

Of 42,654 study participants who underwent cMRI with T1 mapping, 41,635 had imputed genetic data available. Sample and variant quality control filters were applied prior to conducting genetic association analyses. Samples with sex chromosome aneuploidy and those with discordant genetically inferred and self-reported sex were excluded. One of each pair of third-degree relatives or closer was excluded. Variants with imputation quality score (INFO) < 0.3 and those with minor allele frequency < 0.01 were excluded. Following quality control, our dataset included 40,399 individuals with 9,853,972 SNPs and short indels. Among the 40,399 study participants with adequate quality genetic data, 39,339 had T1 time data that passed quality control available and constituted the study sample for the genome-wide association analysis (Supplementary Fig. 2).

Two-sample MR analysis methods

We performed a two-sample MR analysis to examine the causal effect of cardiovascular risk factors and disease on myocardial interstitial fibrosis as measured using T1 time. Exposures examined included BMI, systolic blood pressure, diastolic blood pressure, chronic kidney disease, estimated glomerular filtration rate, diabetes mellitus type 1, diabetes mellitus type 2, atrial fibrillation, coronary artery disease, low density lipoprotein, high density lipoprotein, total cholesterol and triglycerides. We searched the public domain for summary statistics of the examined exposures in large study samples of predominantly European ancestry and prioritized the inclusion of study samples that did not overlap with our UKB study sample where possible. If ancestry-specific results were reported, we used the European ancestry results. Details of included summary statistics are summarized in Supplementary Table 18.

We derived genetic instruments for each cardiovascular risk factor or disease using the publicly available summary statistics after removing strand-ambiguous variants. We first subset to genome-wide significant variants ($P < 5 \times 10^{-8}$) and pruned the variants (LD threshold of $r^2 < 0.01$ within a 1,000-kb window) using 1000G multi-ancestry or European phase 3 LD data, based on ancestral composition of available summary statistics, to identify a subset of independent variants associated with each cardiovascular risk factor or disease. Next, we harmonized the external summary statistics with our T1 time summary statistics by transforming all genomic coordinates to Genome Reference Consortium Human build 37, harmonizing strand orientation and aligning effect alleles and effect estimates. We excluded from the genetic instrument variants with minor allele frequency < 1% in our study sample or with poor imputation quality (INFO score < 0.3) (Supplementary Table 18). We performed two-sample MR using the IVW method in the Mendelian Randomization package in R[65]. We additionally used the MR-Egger method as a sensitivity analysis. The MR-Egger directional pleiotropy test assesses for the presence of horizontal pleiotropy, and the MR-Egger regression test yields pleiotropy-robust causal estimates[23]. Two-sided *P*-values < 3.85×10^{-3} (0.05/13) were considered statistically significant after accounting for multiple testing. Finally, scatter plots of the association of the individual genetic variants included in each genetic instrument with both T1 time and the exposure under investigation were examined to assess for plausible causal associations.

Genome-wide common variant association analysis methods

We performed a common variant genome-wide association analysis of T1 time using a fixed effect linear regression model in PLINK 2.0[66]. The models were adjusted for age at MRI, sex, MRI scanner, genotyping array, and first ten principal components of genetic ancestry. Rank-based inverse normal transformation was applied to the measured myocardial T1 times. As such, effect size estimates in the GWAS are dimensionless and reflect approximately multiples of 1 SD of the underlying quantitative trait. A two-sided *P*-value < 5×10^{-8} was used to define genome-wide significant common variants, and $5 \times 10^{-8} <$ *P*-value < 1×10^{-6} denoted suggestive loci. Distinct genomic loci were defined by starting with the SNP with the lowest *P*-value, excluding other SNPs within 500 kb, and iterating until no SNPs remained. The independently significant SNPs with the lowest *P*-value at each genomic locus are termed lead SNPs. We then performed a conditional

analysis adjusting for the imputed allele dosage of each lead SNP to examine for additional independent genome-wide significant SNPs within a locus. We performed three sensitivity analyses. First, we repeated the above GWAS of T1 time after exclusion of individuals with prevalent diseases at time of MRI known to be associated with focal replacement fibrosis including heart failure, dilated cardiomyopathy, hypertrophic cardiomyopathy and myocardial infarction (n_{GWAS} = 38,339). Second, we repeated the GWAS after excluding 62 participants with prevalent hereditary hemochromatosis (n_{GWAS} = 39,277) to examine whether the identified genome-wide significant loci associated with iron homeostasis were driven by hereditary hemochromatosis cases. Third, we performed a sex-stratified GWAS for males (n_{GWAS} = 19,025) and females (n_{GWAS} = 20,314) separately.

LD score regression analysis was performed using ldsc version 1.0.0[67]. With ldsc, the genomic control factor (lambda GC) was partitioned into components reflecting polygenicity and inflation, using the software's defaults.

Regional association plots were generated with LocusZoom (version 1.4)[68] using LD data from the 1000G phase 3 European reference panel. In instances where lead SNPs were not part of the 1000G phase 3 reference panel, in-sample LD was calculated using PLINK 1.9.

Heritability and genetic correlation analysis

SNP-heritability of T1 time was assessed using BOLT-REML v2.3.4[69]. We also computed genetic correlation between T1 time and other cMRI measures including LV end diastolic and systolic volumes, LV mass, LV ejection fraction, left atrial end diastolic and systolic volumes and left atrial ejection fraction using ldsc version 1.0.0[70]. These cMRI-based phenotypes from the UKB have been described previously[71,72].

EQTL and TWAS methods

We performed an expression quantitative trait locus look-up using version 8 of the Genotype-Tissue Expression (GTEx) database[73]. In-sample LD was calculated for all variants within 1 Mb of genome-wide significant lead SNPs using PLINK 1.9. List of proxy SNPs for each lead SNP were generated using an LD r^2 threshold of > 0.8. We searched the GTEx v8 database for statistically significant differential gene expression in right atrial appendage and left ventricular tissues associated with the lead SNPs and their proxies. When no significant differential gene expression associated with the lead SNP was identified, significant findings from the closest proxy were reported.

We then performed a transcriptome-wide association analysis to test the mediating effects of gene expression levels in right atrial appendage and left ventricular tissue on T1 time. We used pre-computed transcript expression reference weights derived using elastic net models from S-PrediXcan on GTEx v8 eQTL data for the right atrial appendage and LV[74]. S-Predixcan was then run with its default settings. A Bonferroni-corrected two-sided P-value threshold < 7.5×10^{-6} (0.05/6,637 genes tested) was used to define significant gene expression-phenotype associations.

Cardiac fibrosis assay and profiling

We used primary human cardiac fibroblasts (ACBRI5118, Cell Systems) for all *in vitro* experiments. Cells were cultured in Lonza FGM-3 Cardiac Fibroblast Growth Medium (CC-4525). The cardiac fibroblast activation assay was initiated by performing a media exchange for starvation media (CC-4525 without FBS and supplements) containing TGFβ1 (Sigma T7039) at 10 ng/ml. Controls were given starvation media only. Total RNA was extracted using the Direct-zol RNA Miniprep kit (Zymo Research, R2051). Standard RNA-seq libraries were generated and sequenced by Genewiz on an Illumina NovaSeq. Sequenced reads were aligned to the human genome (GRCh38) using Salmon (version 1.8.0)[75]. Differential expression analysis was performed with DESeq2 (1.30.1)[76].

Fifty thousand cardiac fibroblasts were used as input for ATAC-seq, following the OMNI-ATAC-seq protocol[77]. Transposed DNA was purified with a Qiagen PCR MinElute kit (Qiagen 28004), and final ATAC-seq libraries were purified with a 1.8X 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 with default paired-end settings (version 2.3.4.3)[78]. Next, all non-nuclear and unmapped paired reads were discarded. Duplicated reads were removed with the Picard MarkDuplicates function, default settings. Visualization of ATAC-seq signals was done with Homer (version 4.10)[79], and all reads were normalized by read count where scores represent read count per bp per 1×10^7 reads. Motif enrichment analysis was performed with Homer using the findMotifsGenome.pl function with default parameters. Peak calling was carried out with MACS2 (callpeak–nomodel–broad) (version 2.2.6) using all ATAC-seq libraries as input. Reads were counted for each condition from the comprehensive peak file by using bedtools (multicov module) (version v2.26.0)[80]. Quantile normalization of ATAC-seq data sets was performed with CQN (version 1.36.0)[81], and offsets were fed into DESeq2 to quantify differential accessibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to acknowledge the contributions of the UKB participants without whom this work would not have been possible. National Institutes of Health (NIH) T32HL007604 grant to V.N. NIH 1R01HL092577, NIH K24HL105780, American Heart Association (AHA) 18SFRN34110082, Foundation Leducq 14CVD01, and MAESTRIA 965286 grants to P.T.E. Scholar award from the Sarnoff Cardiovascular Research Foundation and NIH grant K08HL159346 to J.P.P. NIH 1R01HL139731, NIH R01HL157635, and AHA 18SFRN34250007 grants to S.A.L. AHA Postdoctoral fellowship 18SFRN34110082 to L.-C.W. NIH 5T32HL007208-42 grant to M.C.H. Some artwork incorporated into Figure 5a was created with BioRender.com.

Competing Interests

M.D.R.K., P.D.A, S.F.F. and P.B. are supported by grants from Bayer AG and IBM applying machine learning in cardiovascular disease. P.B. serves as a consultant for Novartis and Prometheus Biosciences. P.B. is employed by Flagship Pioneering as of January 4, 2023. C.R. is supported by a grant from Bayer AG to the Broad Institute focused on the development of therapeutics for cardiovascular disease. S.A.L. is employed at Novartis Institutes for Biomedical Research as of July 18, 2022. S.A.L. received sponsored research support from Bristol Myers Squibb / Pfizer, Bayer AG, Boehringer Ingelheim, Fitbit, and IBM, and has consulted for Bristol Myers Squibb / Pfizer, Bayer AG, Blackstone Life Sciences, and Invitae previously. P.T.E. receives sponsored research support from Bayer

AG, Novartis, Myokardia and Quest. L.-C.W. receives sponsored research support from IBM to the Broad Institute. The remaining authors have no disclosures.

Data Availability

UK Biobank data are made available to researchers from research institutions with genuine research inquiries, following IRB and UK Biobank approval. Genome-wide association analysis summary statistics are available from the Downloads page of the Cardiovascular Disease Knowledge Portal (broadcvdi.org). Raw and processed next-generation sequencing data have been deposited at the NCBI Gene Expression Omnibus under accession number GSE225336. Genome Reference Consortium Human Build 37 (GRCh37) data is publicly available at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/. Genome Reference Consortium Human Build 38 (GRCh38) data is publicly available at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/. Genotype-Tissue Expression version 8 datasets are publicly available at https://www.gtexportal.org/home/datasets. All other data are contained within the article and its supplementary information, or are available upon reasonable request to the corresponding author.

REFERENCES

1. del Monte-Nieto G, Fischer JW, Gorski DJ, Harvey RP & Kovacic JC Basic Biology of Extracellular Matrix in the Cardiovascular System, Part 1/4. Journal of the American College of Cardiology 75, 2169–2188 (2020). [PubMed: 32354384]
2. aus dem Siepen F et al. T1 mapping in dilated cardiomyopathy with cardiac magnetic resonance: quantification of diffuse myocardial fibrosis and comparison with endomyocardial biopsy. European Heart Journal - Cardiovascular Imaging 16, 210–216 (2015). [PubMed: 25246502]
3. Frangogiannis NG & Kovacic JC Extracellular Matrix in Ischemic Heart Disease, Part 4/4. Journal of the American College of Cardiology 75, 2219–2235 (2020). [PubMed: 32354387]
4. Díez J, González A & Kovacic JC Myocardial Interstitial Fibrosis in Nonischemic Heart Disease, Part 3/4. Journal of the American College of Cardiology 75, 2204–2218 (2020). [PubMed: 32354386]
5. Bing R et al. Imaging and Impact of Myocardial Fibrosis in Aortic Stenosis. JACC Cardiovasc Imaging 12, 283–296 (2019). [PubMed: 30732723]
6. Nguyen TP, Qu Z & Weiss JN Cardiac fibrosis and arrhythmogenesis: the road to repair is paved with perils. J Mol Cell Cardiol 70, 83–91 (2014). [PubMed: 24184999]
7. Ling L et al. Diffuse Ventricular Fibrosis in Atrial Fibrillation. Journal of the American College of Cardiology 60, 2402–2408 (2012). [PubMed: 23141493]
8. Chen Z et al. Myocardial tissue characterization by cardiac magnetic resonance imaging using T1 mapping predicts ventricular arrhythmia in ischemic and non–ischemic cardiomyopathy patients with implantable cardioverter-defibrillators. Heart Rhythm 12, 792–801 (2015). [PubMed: 25533585]
9. Kong P, Christia P & Frangogiannis NG The pathogenesis of cardiac fibrosis. Cell Mol Life Sci 71, 549–574 (2014). [PubMed: 23649149]
10. Mewton N, Liu CY, Croisille P, Bluemke D & Lima JAC Assessment of Myocardial Fibrosis With Cardiovascular Magnetic Resonance. Journal of the American College of Cardiology 57, 891–903 (2011). [PubMed: 21329834]
11. Diao K et al. Histologic validation of myocardial fibrosis measured by T1 mapping: a systematic review and meta-analysis. J Cardiovasc Magn Reson 18, 92 (2017).
12. Petersen SE et al. UK Biobank's cardiovascular magnetic resonance protocol. J Cardiovasc Magn Reson 18, 8 (2015).
13. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). [PubMed: 30305743]

14. Messroghli DR et al. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). J Cardiovasc Magn Reson 19, 75 (2017). [PubMed: 28992817]
15. Rogers T et al. Standardization of T1 measurements with MOLLI in differentiation between health and disease--the ConSept study. J Cardiovasc Magn Reson 15, 78 (2013). [PubMed: 24025486]
16. Puntmann VO, Peker E, Chandrashekhar Y & Nagel E T1 Mapping in Characterizing Myocardial Disease: A Comprehensive Review. Circ Res 119, 277–299 (2016). [PubMed: 27390332]
17. Liu C-Y et al. Evaluation of age-related interstitial myocardial fibrosis with cardiac magnetic resonance contrast-enhanced T1 mapping: MESA (Multi-Ethnic Study of Atherosclerosis). J Am Coll Cardiol 62, 1280–1287 (2013). [PubMed: 23871886]
18. Roy C et al. Age and sex corrected normal reference values of T1, T2 T2* and ECV in healthy subjects at 3T CMR. J Cardiovasc Magn Reson 19, 72 (2017). [PubMed: 28934962]
19. Treibel TA et al. Extracellular volume quantification in isolated hypertension - changes at the detectable limits? J Cardiovasc Magn Reson 17, 74 (2015). [PubMed: 26264919]
20. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Diabetes Care 43, S14–S31 (2020). [PubMed: 31862745]
21. Stevens PE, Levin A, & Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. Ann Intern Med 158, 825–830 (2013). [PubMed: 23732715]
22. Arnett DK et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 140, (2019).
23. Bowden J, Davey Smith G & Burgess S Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 44, 512–525 (2015). [PubMed: 26050253]
24. Pirruccello JP et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. Nat Commun 11, 2254 (2020). [PubMed: 32382064]
25. Waller AP et al. GLUT12 functions as a basal and insulin-independent glucose transporter in the heart. Biochim Biophys Acta 1832, 121–127 (2013). [PubMed: 23041416]
26. Heidecker B et al. The gene expression profile of patients with new-onset heart failure reveals important gender-specific differences. Eur Heart J 31, 1188–1196 (2010). [PubMed: 20031959]
27. Jiménez-Amilburu V, Jong-Raadsen S, Bakkers J, Spaink HP & Marín-Juez R GLUT12 deficiency during early development results in heart failure and a diabetic phenotype in zebrafish. J Endocrinol 224, 1–15 (2015). [PubMed: 25326603]
28. Linden KC et al. Renal expression and localization of the facilitative glucose transporters GLUT1 and GLUT12 in animal models of hypertension and diabetic nephropathy. Am J Physiol Renal Physiol 290, F205–213 (2006). [PubMed: 16091581]
29. Sharma S et al. SOD2 deficiency in cardiomyocytes defines defective mitochondrial bioenergetics as a cause of lethal dilated cardiomyopathy. Redox Biology 37, 101740 (2020). [PubMed: 33049519]
30. Vivien CJ et al. Vegfc/d-dependent regulation of the lymphatic vasculature during cardiac regeneration is influenced by injury context. NPJ Regen Med 4, 18 (2019). [PubMed: 31452940]
31. Perrucci GL, Rurali E & Pompilio G Cardiac fibrosis in regenerative medicine: destroy to rebuild. J. Thorac. Dis 10, S2376–S2389 (2018). [PubMed: 30123577]
32. Kelwick R, Desanlis I, Wheeler GN & Edwards DR The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. Genome Biol 16, 113 (2015). [PubMed: 26025392]
33. Chen P et al. MYH7B variants cause hypertrophic cardiomyopathy by activating the CaMK-signaling pathway. Sci China Life Sci 63, 1347–1362 (2020). [PubMed: 32207065]
34. Alexander J & Kowdley KV HFE-associated hereditary hemochromatosis. Genet Med 11, 307–313 (2009). [PubMed: 19444013]

35. Ramsay AJ, Hooper JD, Folgueras AR, Velasco G & López-Otín C Matriptase-2 (TMPRSS6): a proteolytic regulator of iron homeostasis. Haematologica 94, 840–849 (2009). [PubMed: 19377077]

36. Roselli C et al. Multi-ethnic genome-wide association study for atrial fibrillation. Nat Genet 50, 1225–1233 (2018). [PubMed: 29892015]

37. Heijman J, Ghezelbash S, Wehrens XHT & Dobrev D Serine/Threonine Phosphatases in Atrial Fibrillation. J Mol Cell Cardiol 103, 110–120 (2017). [PubMed: 28077320]

38. Lubbers ER & Mohler PJ Roles and regulation of protein phosphatase 2A (PP2A) in the heart. Journal of Molecular and Cellular Cardiology 101, 127–133 (2016). [PubMed: 27832939]

39. Ramirez AH et al. Novel rare variants in congenital cardiac arrhythmia genes are frequent in drug-induced torsades de pointes. Pharmacogenomics J 13, 325–329 (2013). [PubMed: 22584458]

40. Zhu N et al. Pim-1 Kinase Phosphorylates Cardiac Troponin I and Regulates Cardiac Myofilament Function. Cell Physiol Biochem 45, 2174–2186 (2018). [PubMed: 29544221]

41. Pan W et al. Structural insights into ankyrin repeat-mediated recognition of the kinesin motor protein KIF21A by KANK1, a scaffold protein in focal adhesion. J Biol Chem 293, 1944–1956 (2018). [PubMed: 29217769]

42. Alexanian M et al. A transcriptional switch governs fibroblast activation in heart disease. Nature 595, 438–443 (2021). [PubMed: 34163071]

43. Xiao Y et al. Hippo pathway deletion in adult resting cardiac fibroblasts initiates a cell state transition with spontaneous and self-sustaining fibrosis. Genes Dev 33, 1491–1505 (2019). [PubMed: 31558567]

44. Wight TN The ADAMTS Proteases, Extracellular Matrix, and Vascular Disease: Waking the Sleeping Giant(s)! ATVB 25, 12–14 (2005).

45. Hirohata S et al. Punctin, a Novel ADAMTS-like Molecule, ADAMTSL-1, in Extracellular Matrix. Journal of Biological Chemistry 277, 12182–12189 (2002). [PubMed: 11805097]

46. Wang X et al. Critical Role of ADAMTS2 (A Disintegrin and Metalloproteinase With Thrombospondin Motifs 2) in Cardiac Hypertrophy Induced by Pressure Overload. Hypertension 69, 1060–1069 (2017). [PubMed: 28373586]

47. Willeford A et al. CaMKIIδ-mediated inflammatory gene expression and inflammasome activation in cardiomyocytes initiate inflammation and induce fibrosis. JCI Insight 3, 97054 (2018). [PubMed: 29925681]

48. Ling H et al. Requirement for Ca2+/calmodulin–dependent kinase II in the transition from pressure overload–induced cardiac hypertrophy to heart failure in mice. J. Clin. Invest 119, 1230–1240 (2009). [PubMed: 19381018]

49. Ebeid DE et al. PIM1 Promotes Survival of Cardiomyocytes by Upregulating c-Kit Protein Expression. Cells 9, 2001 (2020). [PubMed: 32878131]

50. Muraski JA et al. Pim-1 regulates cardiomyocyte survival downstream of Akt. Nat Med 13, 1467–1475 (2007). [PubMed: 18037896]

51. Torlasco C et al. Role of T1 mapping as a complementary tool to T2* for non-invasive cardiac iron overload assessment. PLoS One 13, e0192890 (2018). [PubMed: 29466447]

52. Song X et al. Cardiovascular and all-cause mortality in relation to various anthropometric measures of obesity in Europeans. Nutrition, Metabolism and Cardiovascular Diseases 25, 295–304 (2015).

53. Voskoboinik A et al. Relation of Alcohol Consumption to Left Ventricular Fibrosis Using Cardiac Magnetic Resonance Imaging. Am J Cardiol 123, 460–465 (2019). [PubMed: 30473327]

54. Fernández-Solà J Cardiovascular risks and benefits of moderate and heavy alcohol consumption. Nat Rev Cardiol 12, 576–587 (2015). [PubMed: 26099843]

METHODS-ONLY REFERENCES

55. Sudlow C et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12, e1001779 (2015). [PubMed: 25826379]

56. Puyol-Antón E et al. Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with uncertainty-based quality-control. J Cardiovasc Magn Reson 22, 60 (2020). [PubMed: 32814579]
57. Huang G, Liu Z, van der Maaten L & Weinberger KQ Densely Connected Convolutional Networks. arXiv:1608.06993 [cs] (2018).
58. Ronneberger O, Fischer P & Brox T U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs] (2015).
59. Deng J et al. ImageNet: A large-scale hierarchical image database. in 2009 IEEE Conference on Computer Vision and Pattern Recognition 248–255 (IEEE, 2009). doi:10.1109/CVPR.2009.5206848.
60. Kingma DP & Ba J Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs] (2017).
61. Petersen SE et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. J Cardiovasc Magn Reson 19, 18 (2017). [PubMed: 28178995]
62. Inker LA et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med 367, 20–29 (2012). [PubMed: 22762315]
63. Stone NJ et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J. Am. Coll. Cardiol 63, 2889–2934 (2014). [PubMed: 24239923]
64. Grundy SM et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J. Am. Coll. Cardiol (2018) doi:10.1016/j.jacc.2018.11.002.
65. Yavorska OO & Burgess S MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. International Journal of Epidemiology 46, 1734–1739 (2017). [PubMed: 28398548]
66. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 4, 7 (2015).
67. Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291–295 (2015). [PubMed: 25642630]
68. Boughton AP et al. LocusZoom.js: Interactive and embeddable visualization of genetic association study results. Bioinformatics 37, 3017–3018 (2021). [PubMed: 33734315]
69. Loh P-R et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 47, 1385–1392 (2015). [PubMed: 26523775]
70. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236–1241 (2015). [PubMed: 26414676]
71. Pirruccello JP et al. Deep Learning of Left Atrial Structure and Function Provides Link to Atrial Fibrillation Risk. medRxiv (2021) doi:10.1101/2021.08.02.21261481.
72. Khurshid S et al. Deep Learning to Predict Cardiac Magnetic Resonance-Derived Left Ventricular Mass and Hypertrophy From 12-Lead ECGs. Circ Cardiovasc Imaging 14, e012281 (2021). [PubMed: 34126762]
73. GTEx Consortium et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017). [PubMed: 29022597]
74. GTEx Consortium et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 9, 1825 (2018). [PubMed: 29739930]
75. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419 (2017). [PubMed: 28263959]
76. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
77. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14, 959–962 (2017). [PubMed: 28846090]

78. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). [PubMed: 22388286]

79. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010). [PubMed: 20513432]

80. Quinlan AR BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47, 11.12.1–34 (2014).

81. Hansen KD, Irizarry RA & Wu Z Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13, 204–216 (2012). [PubMed: 22285995]

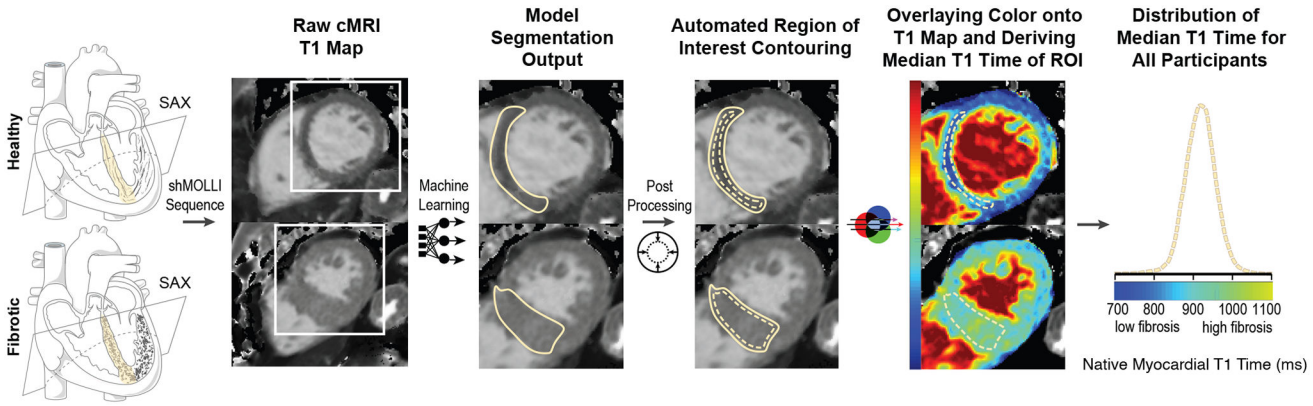

Figure 1 |. Overview of the automated pipeline for native myocardial T1 time measurement at the interventricular septum using machine learning.

A representative healthy heart and one with increased interstitial fibrosis are shown for illustration. Cardiac T1 mapping using the Shortened Modified Look-Locker Inversion (shMOLLI) recovery sequence was performed at the mid-ventricular short-axis. A machine-learning model trained on the raw MRI T1 maps generated automated segmentation of the interventricular segment (solid yellow contour) followed by selection of representative myocardial regions of interest (dashed yellow contour) using morphological operations. T1 map color legends were then used to transform pixel intensities within the region of interest into T1 times. For each participant, the median T1 time by ROI was calculated and used as the representative T1 time. MRI, magnetic resonance imaging; ROI, region of interest; SAX, short-axis; shMOLLI, Shortened Modified Look-Locker Inversion. The heart schematics were drawn by using pictures from Servier Medical Art, which were further modified. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/). T1 maps shown are reproduced by kind permission of UK Biobank ©.

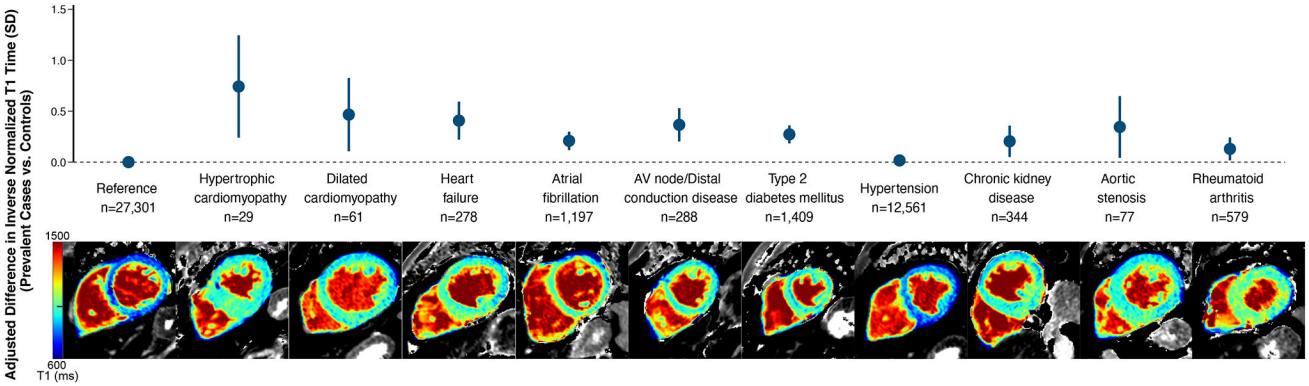

Figure 2 |. Change in native myocardial T1 time associated with prevalent cardiovascular, metabolic and systemic inflammatory diseases as compared to healthy controls.

Healthy controls free of prevalent dilated cardiomyopathy, hypertrophic cardiomyopathy, heart failure, atrial fibrillation, atrioventricular node/distal conduction disease, hypertension, diabetes mellitus, aortic stenosis, chronic kidney disease, hemochromatosis and rheumatoid arthritis constituted the reference group. Numbers of controls or cases with available native myocardial T1 time are shown below each category. For each disease, a representative T1 map of a case is provided from the study sample. Multiple linear regression was implemented and a two-sided *P*-value threshold adjusted for multiple testing of $< 3.1 \times 10^{-3}$ was used to define statistically significant associations. Data are presented as mean adjusted change in T1 time along with $(1 - \alpha)*100$ (%) confidence intervals. Confidence intervals were constructed using the adjusted two-sided α (3.1×10^{-3}) for multiple testing. AV, atrioventricular. T1 maps shown are reproduced by kind permission of UK Biobank ©.

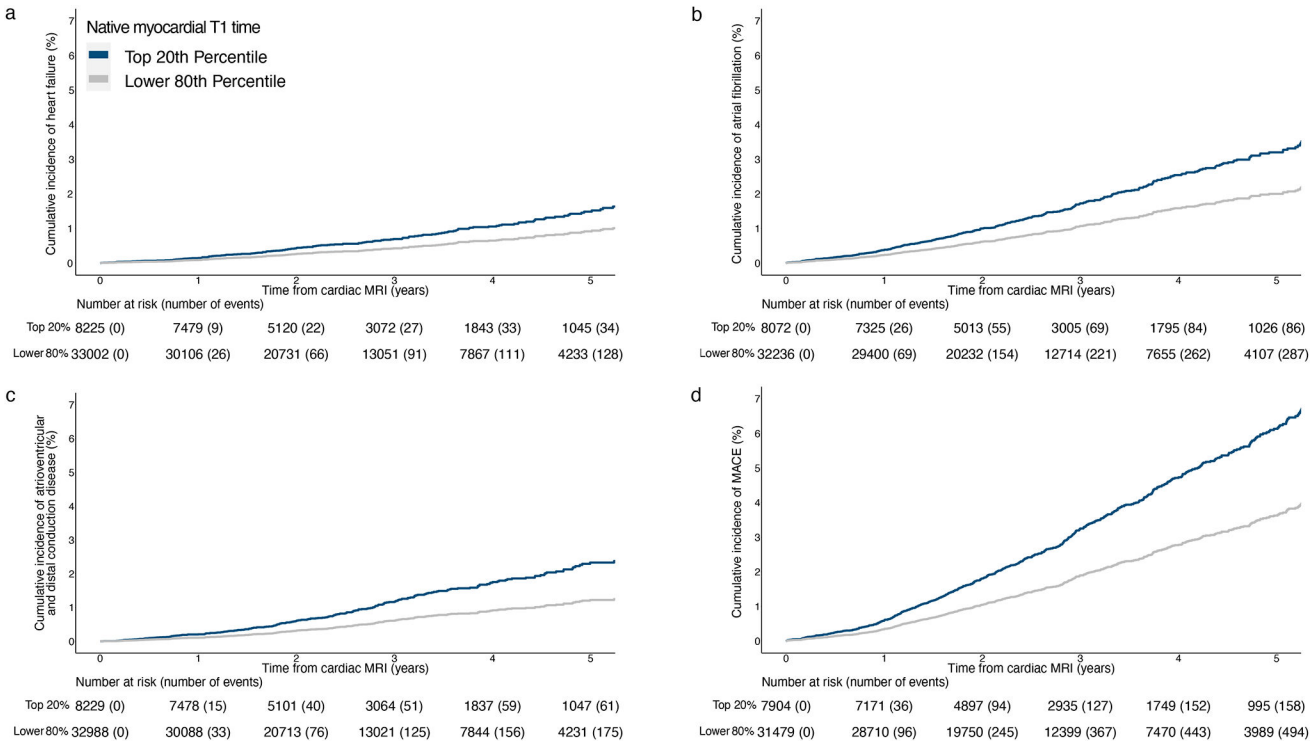

Figure 3 |. Adjusted cumulative incidence of heart failure, atrial fibrillation, atrioventricular node/distal conduction disease and MACE stratified by top 20th percentile vs. lower 80th percentile of native myocardial T1 time.

MACE, major adverse cardiovascular events.

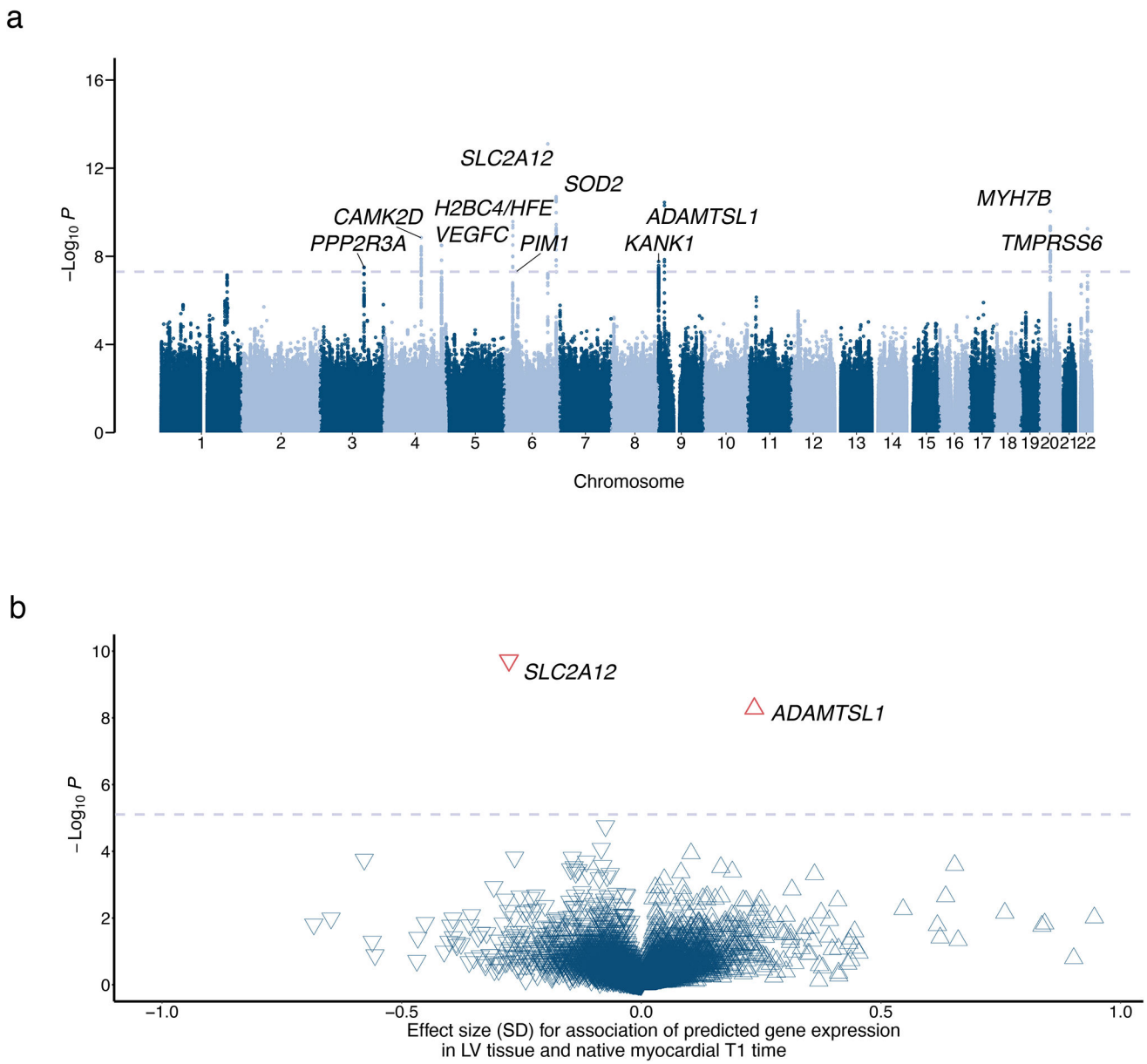

Figure 4 |. Genome-wide and transcriptome-wide association analyses.

a, Native myocardial T1 time genome-wide association results across 22 autosomes. Nearest genes are used for annotation. A fixed effect multiple linear regression model was implemented. The dashed grey line represents the threshold for genome-wide significance (two-sided *P*-value < 5×10^{-8} adjusted for multiple testing). **b**, Volcano plots depicting transcriptome-wide association results for native myocardial T1 time using human left ventricular tissue gene expression from GTEx v8 and S-PrediXcan. Upward facing triangles reflect increased T1 time associated with increased gene expression in left ventricular tissue. Downward facing triangles reflect decreased T1 time associated with increased gene expression in left ventricular tissue. The dashed grey line represents the threshold for transcriptome-wide significance (two-sided *P*-value < 7.5×10^{-6} adjusted for multiple testing).

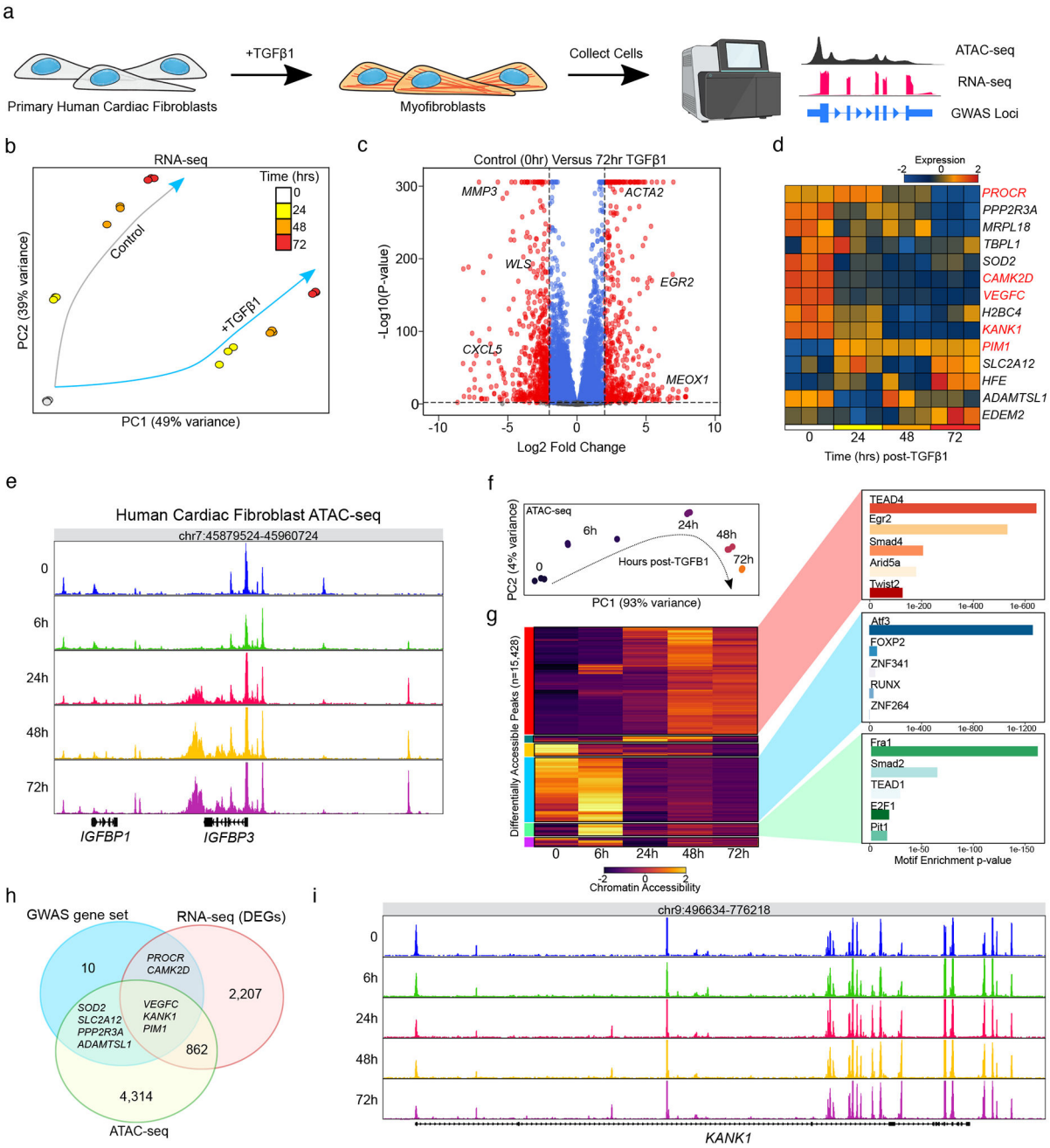

Figure 5 |. Multi-omic examination of human cardiac fibroblast activation.

a, Graphical schematic describing the cardiac fibroblast activation experiments created with BioRender.com. **b**, Principal component analysis of control and TGFβ1-treated cardiac fibroblast RNA-seq. **c**, Volcano plot displaying differentially expressed genes between control (0 h) and stimulated cardiac fibroblasts (72 h post-TGFβ1 treatment) assessed using a generalized linear model implementing a negative binomial distribution. Expected differentially expressed cardiac fibrosis regulator genes are labeled. Red dots indicate significantly differentially expressed genes (FDR < 0.01) with |$\log_2$ fold-change| > 1.0 associated with TGFβ1 treatment. **d**, Heatmap of normalized expression levels for

prioritized genes associated with GWAS loci and expressed in cardiac fibroblasts. Red labels indicate that the gene is significantly differentially expressed (FDR < 0.01). **e**, Genome browser tracks showing expected enhanced chromatin accessibility in cardiac fibroblasts around *IGFBP1* and *IGFBP3*, which are known to mediate TGFβ1-induced cardiac fibroblast activation. Labels indicate hours (h) post-TGFβ1 treatment. **f**, Principal component analysis of ATAC-seq data from activated cardiac fibroblasts. **g**, left, Heatmap displaying differential chromatin accessibility analysis presented as normalized accessibility counts. Columns represent average accessibility for 3 replicates. Right, *de novo* motif enrichment analysis carried out on the clusters of differentially accessible ATAC-seq peaks. **h**, Venn diagram showing the intersection of significantly differentially expressed genes from RNA-seq, differentially accessible peaks enriched in TGFβ- treated fibroblasts (FDR $< 1 \times 10^{-10}$), and prioritized genes associated with GWAS loci. **i**, Representative genome browser track of ATAC-seq data depicting decreased chromatin accessibility in a promoter flanking region upstream of the *KANK1* locus following TGFβ1 treatment concordant with RNA-seq data demonstrating decreased expression with TGFβ1 treatment. Genomic coordinates are based on Genome Reference Consortium Human Build 38.

Table 1 |

Study sample characteristics at time of first visit for cMRI

Baseline characteristics	***n* (%)**
Participants	41,505
Age at MRI (mean (SD))	64.0 (7.7)
Male	19,956 (48.1)
Body mass index, kg/m^2 (mean (SD))	26.5 (4.3)
Cigarette smoking *	
Never	25,663 (62.4)
Prior	14,009 (34.1)
Current	1,422 (3.5)
Alcohol use *	
None	1,948 (5.5)
Light-to-moderate	19,436 (54.6)
Heavy	14,198 (39.9)
Adequate physical activity *	26,612 (73.0)
Hypertension	12,561 (30.3)
Diabetes mellitus type 2	1,409 (3.4)
Diabetes mellitus type 1	162 (0.4)
Hyperlipidemia	433 (1.0)
Chronic kidney disease	344 (0.8)
Coronary artery disease	2,457 (5.9)
Myocardial infarction	840 (2.0)
Dilated cardiomyopathy	61 (0.1)
Hypertrophic cardiomyopathy	29 (0.1)
Heart failure	278 (0.7)
Aortic stenosis	77 (0.2)
Atrial fibrillation	1,197 (2.9)
Ventricular arrhythmia/History of cardiac arrest	118 (0.3)
Atrioventricular node/Distal conduction disease	288 (0.7)
Rheumatoid arthritis	579 (1.4)
Beta blocker	2,235 (5.4)
ACE-inhibitor/ARB	5,709 (13.8)
Mineralocorticoid receptor antagonist	78 (0.2)
Statins	8,024 (19.3)
Native myocardial T1 time, ms (mean (SD))	918.1 (41.5)
LV mass, g (mean (SD))	90.1 (29.6)
LV ejection fraction, % (mean (SD))	60 (6)
LV end systolic volume, ml (mean (SD))	58.1 (18.9)

Baseline characteristics	***n* (%)**
LV end diastolic volume, ml (mean (SD))	143.3 (32.7)
LA end systolic volume, ml (mean (SD))	33.6 (14.1)
LA end diastolic volume, ml (mean (SD))	72.8 (20.2)
LA ejection fraction, % (mean (SD))	54.8 (8.6)

Values are presented as the number (percentage) unless otherwise specified. A subset of the study sample had data available for each cMRI parameter: LA end systolic volume/LA end diastolic volume/LA ejection fraction (n = 37,234); LV end systolic volume/LV end diastolic volume/LV ejection fraction/LV mass (n = 40,869).

*Data for cigarette smoking, alcohol use and self-reported physical activity was available for 41,094, 35,582 and 36,462 out of 41,505 study participants (percentage reported reflects proportion of those with non-missing data). Adequate physical activity is defined as ≥ 150 min of moderate or ≥ 75 min of vigorous activity or equivalent combination per week. ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; cMRI, cardiac magnetic resonance imaging; SD, standard deviation; LA, left atrium; LV, left ventricle.