Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure


To cite this version:
Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure

Sonia Shah et al.**

Heart failure (HF) is a leading cause of morbidity and mortality worldwide. A small proportion of HF cases are attributable to monogenic cardiomyopathies and existing genome-wide association studies (GWAS) have yielded only limited insights, leaving the observed heritability of HF largely unexplained. We report results from a GWAS meta-analysis of HF comprising 47,309 cases and 930,014 controls. Twelve independent variants at 11 genomic loci are associated with HF, all of which demonstrate one or more associations with coronary artery disease (CAD), atrial fibrillation, or reduced left ventricular function, suggesting shared genetic aetiology. Functional analysis of non-CAD-associated loci implicate genes involved in cardiac development (MYOZ1, SYNPO2L), protein homoeostasis (BAG3), and cellular senescence (CDKN1A). Mendelian randomisation analysis supports causal roles for several HF risk factors, and demonstrates CAD-independent effects for atrial fibrillation, body mass index, and hypertension. These findings extend our knowledge of the pathways underlying HF and may inform new therapeutic strategies.

**A full list of authors and their affiliations appears at the end of the paper.
Heart failure (HF) affects >30 million individuals worldwide and its prevalence is rising. HF-associated morbidity and mortality remain high despite therapeutic advances, with 5-year survival averaging ~50%. HF is a clinical syndrome defined by fluid congestion and exercise intolerance due to cardiac dysfunction. HF results typically from myocardial disease with impairment of left ventricular (LV) function manifesting with either reduced or preserved ejection fraction. Several cardiovascular and systemic disorders are implicated as aetiological factors, most notably coronary artery disease (CAD), obesity and hypertension; multiple risk factors frequently co-occur and the contribution to aetiology has been challenging based on observational data alone. Monogenic hypertrophic and dilated cardiomyopathy (DCM) syndromes are known causes of HF, although they account for a small proportion of disease burden. HF is a complex disorder with an estimated heritability of ~26%. Previous modest-sized genome-wide association studies (GWAS) of HF reported two loci, while studies of DCM have identified a few replicated loci. We hypothesised that a GWAS of HF with greater power would provide an opportunity for: (i) discovery of genetic variants modifying disease susceptibility in a range of comorbid contexts, both through subtype-specific and shared pathophysiological mechanisms, such as fluid congestion; and (ii) provide insights into aetiology by estimating the unconfounded causal contribution of observationally associated risk factors by Mendelian randomisation (MR) analysis.

Herein, we perform a large meta-analysis of GWAS of HF to identify disease associated genomic loci. We seek to relate HF-associated loci to putative effector genes through integrated analysis of expression data from disease-relevant tissues, including statistical colocalisation analysis. We evaluate the genetic evidence supporting a causal role for HF risk factors identified through observational studies using Mendelian randomisation and explore mediation of risk through conditional analysis. In summary, our study identifies additional HF risk variants, prioritises putative effector genes and provides a genetic appraisal of the putative causal role of observationally associated risk factors, contributing to our understanding of the pathophysiological basis of HF.

Results
Meta-analysis identifies 11 genomic loci associated with HF.
We conducted a GWAS comprising 47,309 cases and 930,014 controls of European ancestry across 26 studies from the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) Consortium. The study sample comprised both population cohorts (17 studies, 38,780 HF cases, 893,657 controls) and case-control samples (9 studies, 8,529 cases, 36,357 controls; see Supplementary Notes 2 and 3 for a detailed description of the included studies). Genotype data were imputed to either the 1000 Genomes Project (60%), Haploype Reference Consortium (35%) or study-specific reference panels (5%). We performed a fixed-effect inverse variance-weighted (IVW) meta-analysis relating 8,281,262 common and low-frequency variants (minor allele frequency (MAF) > 1%) to HF risk (Fig. 1). We identified 12 independent genetic variants, at 11 loci associated with HF at genome-wide significance (P < 5 × 10⁻⁸), including 10 loci not previously reported for HF (Fig. 2, Table 1). The quantile–quantile, regional association plots and study-specific effects for each independent variant are shown in Supplementary Figs. 1–3. We replicated two previously reported associations for HF and three of four loci for DCM (Bonferroni-corrected P < 0.05; Supplementary Data 1). Using linkage disequilibrium score regression (LDSC), we estimated the heritability of HF in UK population cohorts (35%) or study-specific reference panels (60%), including 10 loci not previously reported for HF (Fig. 2, Table 1). Using linkage disequilibrium score regression (LDSC), we estimated the heritability of HF in UK European ancestry (930,014 controls).

Fig. 1 Study design and analysis workflow. Overview of study design to identify and characterise heart failure-associated risk loci and for secondary cross-trait genome-wide analyses. GWAS, genome-wide association study; QTL, quantitative trait locus; MAGMA, Multi-marker Analysis of Genomic Annotation; SNP, single-nucleotide polymorphism; miCOJO, multi-trait-based conditional and joint analysis.

Phenotypic effects of HF-associated variants. Next, we investigated associations between the identified loci and other traits that may provide insights into aetiology. First, we queried the NHGRI-EBI GWAS Catalog and a large database of genetic associations in UK Biobank (http://www.nealelab.is/uk-biobank), and identified several biomarker and disease associations at each locus (Supplementary Data 2 and 3). Second, we tested for associations of identified loci with ten known HF risk factors, including cardiac structure and function measures, using GWAS summary data (Supplementary Data 4). Six sentinel variants were associated with CAD, including established loci, such as 9p21/CDKN2B-AS1 and LPA. Four variants were associated with atrial fibrillation (AF), a common antecedent and sequela of HF. To estimate whether the HF risk effects were mediated wholly or in part by risk factors upstream of HF (e.g., CAD), we conditioned HF GWAS summary statistics on nine HF risk factors using Multi-trait Conditional and Joint Analysis (mtCOJO). Conditioning on AF attenuated the HF risk effect by >50% for the PTTX2/FAM241A locus but not other AF-associated loci (KLHL3, SYNPO2/LAPG5), conditioning on CAD fully attenuated effects for two of the six CAD loci (LPA, 9p21/CDKN2B-AS1) and conditioning on body mass index (BMI) ablated the effect of the FTO locus (Supplementary Fig. 4, Supplementary Data 5). Next, we performed hierarchical agglomerative clustering of loci based on cross-trait associations to identify groups related to HF subtypes (Fig. 3). Among HF loci not associated with CAD, a group of four clustered together, of which two (KLHL3 and SYNPO2/LAPG5) were associated with AF and two (BAG3 and CDKN1A) with reduced LV systolic function (fractional shortening (FS); Bonferroni-corrected P < 0.05); we highlight the results for these loci in our reporting of subsequent analyses to identify candidate genes. Notably, genetic loci and other traits that may provide insights into aetiology. First, we queried the NHGRI-EBI GWAS Catalog and a large database of genetic associations in UK Biobank (http://www.nealelab.is/uk-biobank), and identified several biomarker and disease associations at each locus (Supplementary Data 2 and 3). Second, we tested for associations of identified loci with ten known HF risk factors, including cardiac structure and function measures, using GWAS summary data (Supplementary Data 4). Six sentinel variants were associated with CAD, including established loci, such as 9p21/CDKN2B-AS1 and LPA. Four variants were associated with atrial fibrillation (AF), a common antecedent and sequela of HF. To estimate whether the HF risk effects were mediated wholly or in part by risk factors upstream of HF (e.g., CAD), we conditioned HF GWAS summary statistics on nine HF risk factors using Multi-trait Conditional and Joint Analysis (mtCOJO). Conditioning on AF attenuated the HF risk effect by >50% for the PTTX2/FAM241A locus but not other AF-associated loci (KLHL3, SYNPO2/LAPG5), conditioning on CAD fully attenuated effects for two of the six CAD loci (LPA, 9p21/CDKN2B-AS1) and conditioning on body mass index (BMI) ablated the effect of the FTO locus (Supplementary Fig. 4, Supplementary Data 5). Next, we performed hierarchical agglomerative clustering of loci based on cross-trait associations to identify groups related to HF subtypes (Fig. 3). Among HF loci not associated with CAD, a group of four clustered together, of which two (KLHL3 and SYNPO2/LAPG5) were associated with AF and two (BAG3 and CDKN1A) with reduced LV systolic function (fractional shortening (FS); Bonferroni-corrected P < 0.05); we highlight the results for these loci in our reporting of subsequent analyses to identify candidate genes. Notably, genetic
associations with DCM at the BAG3 locus have been reported previously\(^{10,11}\).

**Tissue-enrichment analysis.** We performed gene-based association analyses using MAGMA\(^{26}\) to identify tissues and aetiological pathways relevant to HF. Thirteen genes were associated with HF at genome-wide significance, of which four were located within 1 Mb of a sentinel HF variant and expressed in heart tissue (Supplementary Data 6). Tissue specificity analysis across 53 tissue types from the Genotype-Tissue Expression (GTEx) project identified the atrial appendage as the highest ranked tissue for gene expression enrichment, excluding reproductive organs (Supplementary Fig. 5). We sought to map candidate genes to the HF loci by assessing the functional consequences of sentinel variants (or their proxies) on gene expression, and protein structure/abundance using quantitative trait locus (QTL) analyses.

**Variant effects on protein coding sequence.** Since the identified HF variants were located in non-coding regions, we investigated if sentinel variants were in linkage disequilibrium (LD, \(r^2 > 0.8\)) with non-synonymous variants with predicted deleterious effects. We identified a missense variant in BAG3 (rs2234962; \(r^2 = 0.99\) with sentinel variant rs17617337) associated previously with DCM and progression to HF, and three missense variants in SYNPO2L (rs34163229, rs3812629 and rs60632610; all \(r^2 > 0.9\) with sentinel variant rs4746140)\(^{16,11,27}\). All four missense variants had Combined Annotation Dependent Depletion scores >20, suggesting deleterious effects (Supplementary Data 7).

**Prioritisation of putative effector genes by expression analysis.** We then sought to identify candidate genes for HF risk loci by assessing their effects on gene expression. Given that cardiac dysfunction defines HF and that HF-associated genes by MAGMA analysis were enriched in heart tissues, we first looked for expression quantitative trait loci (eQTL) in heart tissues (LV, left atrium, and RAA, right atrium auricular region) from the Myocardial Applied Genomics Network (MAGNet) and GTEx projects. Three of 12 variants were significantly associated with the expression of one or more genes located in cis in at least one heart tissue (Bonferroni-corrected \(P < 0.05\); Supplementary Data 8). For several of the identified HF loci, extra-cardiac tissues are likely to be relevant; for example, liver is reported to mediate effects of the LPA locus\(^{28}\). To further explore these effects, we then analysed results from a large whole-blood eQTL dataset (\(n = 31,684\)) and found associations with cis-gene expression (\(P < 5 \times 10^{-8}\)) for 8 of 12 sentinel variants (Supplementary Table 1)\(^{29}\). For most HF variants, heart eQTL associations were consistent with those for blood traits; however, for intronic HF sentinel variants in BAG3, CDKN1A and KLHL3 we detected expression of the corresponding gene transcripts in blood only.

Next, to prioritise among candidate genes identified through eQTL associations, we estimated the posterior probability for a common causal variant underlying associations with gene expression and HF at each locus, by conducting pairwise Bayesian colocalisation analysis\(^{30}\). We found evidence for colocalisation (posterior probability >0.7) for MYOZ1 and SYNPO2L in heart, PSRC1 and ABO in heart and blood; and CDKN1A in blood (Supplementary Data 8, Supplementary Table 1). PSRC1 and MYOZ1 were also implicated in a transcriptome-wide association analysis performed using predicted gene expression based on GTEx human atrial and ventricular expression reference data (Supplementary Table 2). Using serum pQTL data from the INTERVAL study (\(N = 3,301\)), we also identified significant concordant cis associations for BAG3 and ABO (Supplementary Data 9)\(^{31}\).

The evidence linking candidate genes with HF risk loci is summarised in Supplementary Table 3, and candidate genes are described in Supplementary Note 1. At HF risk loci associated with reduced systolic function or AF, but not with CAD, the annotated functions of candidate genes related to myocardial disease processes, and traits that may influence clinical expressivity, such as renal sodium handling. For example, the sentinel variant at the SYNPO2L/AGAP5 locus was associated with expression of MYOZ1 and SYNPO2L, encoding two \(\alpha\)-actinin binding Z-disc cardiac proteins. MYOZ1 is a negative regulator of calcineurin signalling, a pathway linked to pathological hypertrophy\(^{32,33}\) and SYNPO2L is implicated in cardiac development and sarcomere maintenance\(^{34}\). The HF sentinel variant at the BAG3 locus was in high LD with a non-synonymous variant associated previously with DCM\(^{11}\), and was associated with decreased cis-gene expression in blood. BAG3 encodes a Z-disc-associated protein that mediates selective macroautophagy and promotes cell survival through interaction with apoptosis regulator BCL2\(^{35}\). CDKN1A encodes p21, a potent cell cycle inhibitor that mediates post-natal cardiomyocyte cell cycle arrest\(^{36}\) and is implicated in LMNA-mediated cellular stress

**Fig. 2 Manhattan plot of genome-wide heart failure associations.** The x-axis represents the genome in physical order; the y-axis shows \(-\log_{10}(P)\) values for individual variant association with heart failure risk from the meta-analysis (\(n = 977,323\)). Suggestive associations at a significance level of \(P < 1 \times 10^{-5}\) are indicated by the blue line, while genome-wide significance at \(P < 5 \times 10^{-8}\) is indicated by the red line. Meta-analysis was performed using a fixed-effect inverse variance-weighted model. Independent genome-wide significant variants are annotated with the nearest gene(s).
### Table 1 Variants associated with heart failure at genome-wide significance.

<table>
<thead>
<tr>
<th>Gene</th>
<th>rsID</th>
<th>Chr</th>
<th>Position (bp)</th>
<th>Function</th>
<th>Nearest gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CELSR2</td>
<td>rs660240</td>
<td>2</td>
<td>109817838</td>
<td>UTR3 C/T</td>
<td>0.79, 1.06 (1.04–1.08)</td>
</tr>
<tr>
<td>PITX2, FAM241A</td>
<td>rs17042102</td>
<td>4</td>
<td>111668626</td>
<td>Intergenic A/G</td>
<td>0.12, 1.12 (1.09–1.14)</td>
</tr>
<tr>
<td>KLHL3</td>
<td></td>
<td></td>
<td></td>
<td>Intronic G/A</td>
<td>0.77, 1.05 (1.03–1.14)</td>
</tr>
<tr>
<td>CDKN1A, LPA</td>
<td>rs140570886</td>
<td>6</td>
<td>161013013</td>
<td>Intronic C/T</td>
<td>0.02, 1.24 (1.16–1.3)</td>
</tr>
<tr>
<td>9p21/CDKN2B-AS1</td>
<td>rs1556516</td>
<td>9</td>
<td>22100176</td>
<td>ncRNA C/G</td>
<td>0.48, 1.06 (1.05–1.08)</td>
</tr>
<tr>
<td>SYNPO2L, AGAP5</td>
<td>BAG3</td>
<td></td>
<td></td>
<td>Intronic T/A</td>
<td>0.47, 1.04 (1.03–1.08)</td>
</tr>
<tr>
<td>ATXN2</td>
<td>rs56094641</td>
<td>16</td>
<td>53806453</td>
<td>Intronic G/A</td>
<td>0.42, 1.06 (1.04–1.08)</td>
</tr>
</tbody>
</table>

Genetic appraisal of HF risk factors. Although many risk factors are associated with HF, only myocardial infarction and hypertension have an established causal role based on evidence from randomised controlled trials (RCTs). Important questions remain about causality for other risk factors. For instance, type 2 diabetes (T2D) is a risk factor for HF, yet it is unclear if the association is mediated via CAD risk or by direct myocardial effects, which may have important preventative implications. Accordingly, we investigated potential causal roles for modifiable HF risk factors, using GWAS summary data. First, we estimated the genetic correlation ($r_g$) between HF and 11 related traits, using bivariate LDSC. For eight of the eleven traits tested, we found evidence of shared additive genetic effects with estimates of $r_g$ ranging from $-0.25$ to $0.67$ (Supplementary Table 4). The estimated CAD-HF $r_g$ was $0.67$, suggesting $45\% (r_g^2)$ of variation in genetic risk of HF is accounted for by common genetic variation shared with CAD, and that the remaining genetic variation is independent of CAD.

Next, we estimated the causal effects of the 11 HF risk factors using Generalised Summary-data-based Mendelian Randomisation, which accounts for pleiotropy by excluding heterogenous variants based on the heterogeneity in dependent instrument (HEIDI) test (Methods, Supplementary Fig. 6, Supplementary Data 10). Consistent with evidence from RCTs and genetic studies, we found evidence for causal effects of higher diastolic blood pressure (DBP; OR = 1.30 per 10 mmHg, $P = 9.13 \times 10^{-21}$) and systolic blood pressure (SBP; OR = 1.18 per 10 mmHg, $P = 4.8 \times 10^{-23}$), and higher risk of CAD (OR = 1.36, $P = 1.67 \times 10^{-70}$) on HF. We note that the effect estimates for variant associations with blood pressure, included as instrumental variables, were adjusted for BMI, which may attenuate the estimated causal effect on HF. We found a s.d. increment of BMI (equivalent to $4.4 \text{ kg m}^{-2}$ (men) $- 5.4 \text{ kg m}^{-2}$ (women) accounted for a $74\%$ higher HF risk ($P = 2.67 \times 10^{-50}$), consistent with previous reports. We identified evidence supporting causal effects of genetic liability to AF (OR of HF per 1 log odds higher AF = 1.19, $P = 1.40 \times 10^{-75}$) and T2D (OR of HF per 1 log odds higher T2D = 1.05, $P = 6.35 \times 10^{-05}$) and risk of HF. We did not find supportive evidence for a causal role for higher heart rate (HR) or lower glomerular filtration rate (GFR) despite reported observational associations. We then performed a sensitivity analysis to explore potential bias arising from the inclusion of case-control samples by repeating the Mendelian randomisation analysis, using HF GWAS estimates generated from population-based cohort studies only. The results of this analysis were consistent with those generated from the overall sample (Supplementary Table 5).

To investigate whether risk factor effects on HF were mediated by CAD and AF, we performed analyses conditioning for CAD and AF using mCJO. We observed attenuation of the effect of T2D after conditioning for CAD (OR = 1.02, $P = 0.19$), suggesting at least partial mediation by CAD risk rather than through direct myocardial effects of hyperglycaemia. Similarly, the effects of low-density lipoprotein cholesterol (LDL-C) were fully explained by effects of CAD on HF risk (OR = 1.00, $P = 0.80$).

Responses of $KLHL3$ is a negative regulator of the thiazide-sensitive Na⁺Cl⁻ cotransporter (SLC12A3) in the distal nephron; loss of function variants cause familial hyperkalaemic hypertension (FHHt) by increasing constitutive sodium and chloride resorption. The sentinel variant at this locus was associated with decreased gene expression and could predispose to sodium and fluid retention. Notably, thiazide diuretics inhibit SLC12A3 to restore sodium and potassium homeostasis in FHHt and are effective treatments for preventing hypertensive HF.
Conversely, the effects of blood pressure, BMI, and triglycerides (TGs) were only partially attenuated, suggesting causal mechanisms independent of those associated with AF and CAD (Fig. 4, Supplementary Data 10).

### Discussion

We identify 12 independent variant associations for HF risk at 11 genomic loci by leveraging genome-wide data on 47,309 cases and 930,014 controls, including 10 loci not previously associated with HF. The identified loci were associated with modifiable risk factors and traits related to LV structure and function, and include the strongest associations signals from GWAS of CAD (9p21, LPA)\(^1\), AF (PITX2)\(^1\) and BMI (FTO)\(^2\). Conditioning for CAD, AF and blood pressure traits demonstrated that the effects of some loci (e.g., 9p21/CDKN2B-AS1) were mediated wholly via risk factor trait associations (e.g., CAD); however, for 8 of 12 variants the attenuation of effects was <50%, suggesting alternative mechanisms may be important. Those loci associated with reduced LV systolic function or AF mapped to candidate genes implicated in processes of cardiac development, protein homeostasis and cellular senescence. We use genetic causal inference methods to explore the syndromic heterogeneity and causal biology of HF, and to provide insights into aetiology. Mendelian randomisation analysis confirms previously reported causal effects for BMI and provides evidence supporting the causal role of several observationally linked risk factors, including AF, elevated blood pressure (DBP and SBP), LDL-C, CAD, TGs and T2D. Using conditional analysis, we demonstrate CAD-independent effects for AF, BMI, blood pressure and estimate that the effects of T2D are mostly mediated by an increased risk of CAD.

The heterogeneity of aetiology and clinical manifestation of HF are likely to have reduced statistical power. We identify a modest number of genetic associations for HF compared to other cardiovascular disease GWAS of comparable sample size, such as for AF, suggesting that an important component of HF heritability may be more attributable to specific disease subtypes than components of a final common pathway\(^17\). Subsequent studies will explore emerging opportunities to define HF subtypes and longitudinal phenotypes in large biobanks and patient registries at scale using standardised definitions based on diagnostic codes, imaging and electronic health records. We speculate that future analysis of HF subtypes may yield additional insights into the genetic architecture of HF to inform new approaches to prevention and treatment.

### Methods

#### Samples

Participants of European ancestry from 26 cohorts (with a total of 29 distinct datasets) with either a case-control or population-based study design were included in the meta-analysis, as part of the HERMES Consortium. Cases included participants with a clinical diagnosis of HF of any aetiology with no inclusion criteria based on LV ejection fraction; controls were participants without HF. Definitions used to adjudicate HF status within each study are detailed in the definitions used to adjudicate HF status within each study are detailed in the indications used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the nitions used to adjudicate HF status within each study are detailed in the

#### Genotyping and imputation

All studies used high-density genotyping arrays and performed genotype calling and pre-imputation quality control (QC), as reported in Supplementary Data 13. Studies performed imputation using one or more of the following reference panels: 1000 Genomes (Phase 1 or Phase 3)\(^4\), Hapmap 2 NCBI build 36\(^4\). Haplotypes Reference Consortium (HRC)\(^5\), the Estonian Whole-
Conditional Mendelian randomisation analyses of HF risk factors. Forest plot of HF risk factors with significant causal effect HF risk estimated using Mendelian randomisation, implemented with GSMR. Diamonds represent the odds ratio and the error bars indicate the 95% confidence interval. The unadjusted estimates represent the risk of HF as estimated from the HF GWAS data, while the adjusted estimates represent risk of HF conditioned, using GWAS summary statistics for atrial fibrillation (adjusted for AF) or coronary artery disease (adjusted for CAD) estimated using the mcCOJO method. For binary traits (coronary artery disease, atrial fibrillation and type 2 diabetes), the MR estimates represent average causal effect per natural-log odds increase in the trait risk. For continuous traits, the MR estimates represent average causal effect per standard deviation increase in the reported unit of the trait. LDL, low-density lipoprotein; HDL, high-density lipoprotein; CAD, coronary artery disease; AF, atrial fibrillation.

Meta-analysis. Meta-analysis of summary data was conducted using the fixed-effect IVW approach implemented in METAL (released March 2011)68. Variants were included if they were present in at least half of all studies. We tested for inflation of the meta-analysis test statistic due to cryptic population structure by estimating the LDSC intercept, implemented using LDSC v1.0.85. As the LDSC intercept indicated no inflation (LD score intercept of 1.0069), no further correction was applied to the meta-analysis summary estimates. To identify variants independently associated with HF, we analysed the genome-wide results using FUMA v1.3.269, selecting a random sample of 10,000 UK Biobank participants of European ancestry as an LD reference dataset70. Variants were filtered using a P < 5 × 10−8 and independent genomic loci were LD-pruned based on an r2 < 0.1. We calculated Cochran’s Q and I2 statistics to assess whether the effect estimates for HF sentinel variants were consistent across studies71.

Heritability estimation. To estimate the proportion of HF risk explained by common variants we estimated heritability h2 on the liability scale, using LDSC on the UK Biobank summary data (6,504 HF cases, 387,652 controls), assuming a population prevalence of 2.5%14. This approach assumes that a binary trait has an underlying continuous liability, and above a certain liability threshold an individual becomes affected. We can then estimate the genetic contribution to the continuous liability. Sample ascertainment can change the distribution of liability in the sampled individuals and needs to be adjusted for, which requires making assumptions about the population prevalence of the trait.

LD reference dataset. A LD reference was created, including 10,000 UK Biobank participants of European ancestry, based on HRC-imputed genotypes (referred henceforth as UKB10K). European individuals were identified by projecting the UK Biobank samples onto the 1000 G Phase 3 samples. A genomic relationship matrix was constructed using HapMap3 variants, filtered for MAF > 0.01, P_HWE < 10−6 and missingness < 0.05 in the European subset, and one member of each pair of samples with observed genomic relatedness >0.05 was excluded to obtain a set of unrelated European individuals. Random sampling without replacement was used to extract a subset of 10,000 unrelated individuals of European ancestry. Variants with a minor allele count > 5, a genotype probability > 0.9 and imputation quality > 0.3 were converted to hard calls. This LD reference dataset was used for downstream summary-based analysis and for identifying SNP proxies.

Gene set enrichment analysis. A gene-based and gene set enrichment analysis of variant associations was performed using MAGMA60, implemented by FUMA v1.3.269. This analysis was performed using summary-level meta-analysis results. First, a gene-based association analysis to identify candidate genes associated with HF was conducted. Second, a tissue enrichment analysis of HF-associated genes was performed using gene expression data for 30 tissues from GTEx. Finally, a gene set enrichment analysis was performed based on pathway annotations from the Gene Ontology database72. For all MAGMA analyses, multiple testing was accounted for by Bonferroni correction.

Missense consequences of sentinel variants and proxies. We queried the protein coding consequence of the sentinel variants and proxies (r2 > 0.8) using the Combined Annotation Dependent Depletion (CADD) score73, implemented using FUMA v1.3.269. The CADD score integrates information from 63 distinct functional annotations into a single quantitative score, ranging from 1 to 99, based on variant rank relative to all 8.6 billion possible single nucleotide variants of the human reference genome (GRCh37). Sentinel SNPs or proxies with CADD score > 20 were identified. A CADD score of 20 indicates that the variant is ranked in the top 0.1% of highest scoring variants, while a CADD score of 30 indicates the variant is ranked in the top 0.01% of human protein impacting variants.

Expression quantitative trait analysis. To determine if HF sentinel variants had cis effects on gene expression, we queried two cQTL datasets based on RNA sequencing of human heart tissue—the GTEx v7 resource74 and the MAGNeT repository (http://www.med.upenn.edu/magnet/). The GTEx v7 sample included 227,912 V and 264,920 non-diseased tissue samples from Europeans (82.9%) and African Americans (15.1%) individuals. The MAGNeT repository included 89 LV and 101 LA tissue samples obtained from rejected donor tissue from hearts with no

Study-level GWA analysis. GWA analysis for each study was performed locally according to a common analysis plan, and summary-level estimates were provided for meta-analysis. Autosomal single-nucleotide polymorphisms (SNPs) were tested for association with HF using logistic regression, assuming additive genetic effects. For the Cardiovascular Health Study, HF association estimates were generated by analysis of incident cases using a Cox proportional hazards model. All studies included age and sex (except for single-sex studies) as covariates in the regression models. Principal components (PCs) were included as covariates for individual studies as appropriate. The following tools were used for study-level GWA analysis: ProbABEL18, mach2dat (http://www.wnc.edu/~yunmii/software.html), QuickT-est69, PLINK260, SNPTTEST37 or R62, as detailed in Supplementary Data 13.

QC on study summary-level data. QC of summary-level results for each study was performed according to the protocol described in Winkler et al.63. In brief, we used the EasyQC tool to harmonise variant IDs and alleles across studies and to compare reported allele frequencies with allele frequencies in individuals of European ancestry from the 1000 Genomes imputation reference panel64. We inspected P–Z plots (reported P value against P value derived from the Z-score), beta and s.e. distributions, and Manhattan plots to check for consistency and to identify spurious associations. For each study, variants were removed if they satisfied any one of the following criteria: imputation quality < 0.5, MAF < 0.01, absolute betas and s.e. > 10. As recommended in Sinnott et al.45 and Johnson et al.66, more stringent QC measures were applied to studies where genotyping of cases and controls was performed on different platforms. This included more stringent thresholds for removing SNPs with low-quality imputation, and where available, individuals genotyped on both platforms were used to remove SNPs with low concordance rates between the two platforms. To check two independent genomic inflation factor (λGEC). For three studies, where some degree of genomic inflation was observed (λGEC > 1.1), genomic control correction was applied (Supplementary Data 13)77.
evidence of structural disease; and 89 LV samples from individuals with DCM, the total number of SNP–gene associations. Baseline characteristics for the MAGnet dataset were consistent with a single common causal variant hypothesis. We tested all genes with significant cis–eQTL association by analysing all variants within a 200 kilobase window around the gene using eQTL summary data for heart tissues and whole blood, and HF summary data from present study. We set the prior probability (or the association) only related with the cis–eQTL association, only with HF exposures with both traits as $10^{-4}$, $10^{-4}$ and $10^{-4}$. For each gene, we report the posterior probability that the association with gene expression and HF risk is driven by a single causal variant. We consider a posterior probability of $\geq 0.7$ as providing evidence, supporting a causal role for the gene as a mediator of HF risk.

Transcriptome-wide association analysis. We employed the S-PrediXcan method implemented in the MetaXcan software (https://github.com/hakkinlab/MetaXcan) to identify genes whose predicted expression levels in heart tissue are associated with HF risk. Prediction models trained on GTExv7 heart tissue datasets were applied to the HERMES meta-analysis results. Only models that significantly predicted gene expression in the GTEx eQTL dataset (false discovery rate < 0.05) were considered. A total of 4859 genes were tested in left ventricle tissue and 4467 genes for right atrial appendage. Genes with an association $P$ = $5.36 \times 10^{-8}$ ($0.05/4859$) that were considered to have gene expression profiles significantly associated with HF.

Protein quantitative trait analysis in blood. We queried both cis- and trans-protein QTL (pQTL) associations based on measures for serum proteins mapping to 3000 genes in 3301 healthy individuals from the INTERVAL study. We accounted for multiple testing by adjusting a significance threshold of $P < 0.05$ for the total number of tests for all variants and proteins tested (36,000 tests).

Association of HF risk loci with other phenotypes. We queried associations (with $P < 1 \times 10^{-5}$) of sentinel variants and proxies ($r^2 > 0.6$) with any trait in the NHGRI-EBI Catalog of published GWAS (accessed 21 January 2019). We report associations (where $P < 1 \times 10^{-5}$) for the sentinel variants with traits in the UK Biobank cohort using the MRBase Phenome database (http://phenomas.mrbase.org/, accessed 17 January 2019). The database contains GWA summary data for 4203 phenotypes measured in 361,194 unrelated individuals of European ancestry from the UK Biobank data. We queried GWAS data for ten traits related to HF risk factors, endophenotypes and related disease traits using summary-level data from the largest available GWAS study (either publicly available or through agreement with study investigators). The following phenotypes were considered: fractional shortening (FS), LV dimension, DCM; AF, CAD, HDL-C, LDL-C, TP53, BMLD, SBP and DBP. In DCM, a GWAS was performed in the UKB discovery sample of European ancestry. The primary genome-wide $P$ value for all cases defined by the presence of ICD10 code I42.0 as a main/secondary diagnosis or primary/secondary cause of death with non-cases as references, using PLINK2. Logistic regression was performed with adjustment for age, sex, genotyping array and the first ten PCs.

Hierarchical agglomerative clustering. We performed hierarchical agglomerative clustering on a locus level using the complete linkage method based on the association with related traits as described above. Where a sentinel variant is not available we use the traits summary statistics from a common proxy in place of the sentinel variant. For the LPA locus, we used associations for a proxy of the more common variant (rs557530499). Dissimilarity structure was calculated using Euclidean distance based on $Z$-score (beta of continuous traits or log odds of disease risk divided by s.e.) of the cross-trait associations. We accounted for multiple testing by adjusting the significance threshold for association.

Genetic correlation analysis. We estimated genetic correlation between HF and 11 risk factors using LDSC on the GWAS summary statistics for each trait: AF, CAD, HDL-C, LDL-C, high-density lipoprotein cholesterol (HDL-C), TGs, T2D, BMI, SBP, DBP and HR and estimated GER.

Mendelian randomisation analysis. We performed two sample Mendelian randomisation analysis using the Generalised summary data-based Mendelian randomisation (GSMR) implemented in GCTA v1.91.7beta3. To identify independent SNP instruments for each exposure, GWAS-significant SNPs ($P < 5 \times 10^{-8}$) for each risk factor were pruned ($r^2 < 0.05$; LD window of 10,000 kb; using the UKB10K LD reference). We then estimated the causal effect of the risk factor on the disease trait according to the MR paradigm. The HEIDI test implemented in GSMR was used to detect and remove (if HEIDI $P < 0.01$) variants showing horizontal pleiotropy i.e., having independent effects on both exposure and outcome, as such variants do not satisfy the underlying assumptions for valid instruments. As sensitivity analyses, we estimated the causal effects of known risk factors on HF risk other statistical methodology and software—the R package TwoSampleMR was used to select independent variant instruments for the exposure using the same parameters as per the GSMR analysis ($P < 5 \times 10^{-8}$; $r^2 < 0.05$; LD window of 10,000 kb), except the TwoSampleMR package uses the 10,000 Genomes as the LD reference. Causal estimates based on the IVW, MR-Egger and median-weighted methods were then calculated using the Mendelian Randomisation R package. To enable comparison of MR estimates between traits, we present effect estimates corresponding to the rank of HF for a 1-s.d. higher risk factor of interest. Where the original GWAS conducted rank-based inverse normal transformation (RINT) of a trait prior to GWAS, we used the per-allele beta coefficients following RINT to approximate the equivalent values on the standardised scale, as has been conducted previously. To determine if the causal effects of the continuous risk factors on HF were mediated via their effects on CAD or AF risk, we repeated the GSMR analysis after conditioning the HF summary statistics on CAD and AF GWAS summary statistics, as described below.

Conditional analysis. To estimate the effects of HF risk variants after adjusting for risk factors which showed a significant causal effect on HF in the MR analyses, we performed the mCJOJO on summary data, as implemented in GCTA v1.91.7beta3. HF summary statistics were adjusted for FATD, CAD, HDL-C, LDL-C, TGs, DBP, SBP and BMI using GWAS summary data. The UKB10K LD reference was used.

Reporting summary. Further information is provided in the Nature Research Reporting Summary.

Data availability. The datasets generated during this study are available from the corresponding author upon reasonable request. The summary GWAS estimates for this analysis are available on the Cardiovascular Disease Knowledge Portal (http://www.broadcvdi.org/).

Received: 8 July 2019; Accepted: 18 November 2019; Published online: 09 January 2020

References
Competing interests

J.B.W., L.B., Xing Chen, C.L.H., M.W.N. and A. Malarstig are current or former employees of Pfizer who may hold Pfizer stock and/or stock options. I.D.B. and I.C. are employees of Regeneron Genetics Center. M.E.D. is an employee of Regeneron Pharmaceuticals. W.M. reports grants and personal fees from Siemens Diagnostics, grants and personal fees from Aegeron Pharmaceuticals, grants and personal fees from AMGEN, grants and personal fees from AstraZeneca, grants and personal fees from Danone Research, personal fees from Hoffmann LaRoche, personal fees from MSD, grants and personal fees from Pfizer, personal fees from Sanofi, personal fees from Synageva, grants and personal fees from BASF, grants from Abbott Diagnostics, grants and personal fees from Numares AG, grants and personal fees from Berlin-Chemie, employment with Synlab Holding Deutschland GmbH, all outside the submitted work. M.L.O. reports grant support from Glaas/SmithKline, Eisai, Janssen, Merck and AstraZeneca. B.M.F. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. V.S. participated in a conference trip sponsored by Novo Nordisk and received a honorarium from the same source for participating in an advisory board meeting. He also has ongoing research collaboration with Bayer Ltd. B.T. is a full-time employee of Servier. S.A.L. receives sponsored research support from Bristol Myers Squibb/Pfizer, Bayer AG and Boehringer Ingelheim, and has consulted for Abbott, Quest Diagnostics and Bristol Myers Squibb/Pfizer. M.V.H. has collaborated with Boehringer Ingelheim in research, and in accordance with the policy of the The Clinical Trial Service Unit and Epidemiological Studies Unit (University of Oxford), did not accept any personal payment. P.T.E. receives sponsored research support from Bayer AG, and has consulted with Bayer AG, Novartis and Quest Diagnostics. D.I.S. is a full-time employee of BenevolentAI. R.T.L. has received research grants from Pfizer. The remaining authors declare no competing interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13690-5.

Correspondence and requests for materials should be addressed to R.T.L.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. Institute of Cardiovascular Science, University College London, London, UK. British Heart Foundation Research Accelerator, University College London, London, UK. Institute of Health Informatics, University College London, London, UK. Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. National Heart, Lung, and Blood Institute’s and Boston University’s Framingham Heart Study, Framingham, MA, USA. deCODE genetics/Amgen Inc., Sturlugata 8, 101, Reykjavik, Iceland. Health Data Research UK London, University College London, London, UK. Cardiovascular Medicine unit, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden. Pfizer Worldwide Research & Development, 1 Portland St, Cambridge, MA, USA. Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. Department of Epidemiology and Biostatistics, Imperial College London, St Mary’s Campus, London W2 1PG, UK. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, St Mary’s Campus, London W2 1PG, UK. Department of Clinical Sciences, Lund University, Malmö, Sweden. Department of Cardiology, Herlev Gentofte Hospital, Herlev Ringvej 57, 2650 Herlev, Denmark. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. Department of Neurobiology, Care Sciences and Society/Section of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden. School of Health and Social Sciences, Dalarna University, Falun, Sweden. Regeneron Genetics Center, 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA. Department of Biostatistics, University of Washington, Seattle, WA, USA. Department of Medicine, University of Washington, Seattle, WA, USA. Division of Cardiology, Department of Medicine, Emory University Medical Center, Atlanta, GA, USA. Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, Texas, USA. Department of Molecular and Functional Genomics, Geisinger, Danville, PA, USA. Division of Preventive Medicine, Brigham and Women’s Hospital, Boston, MA 02215, USA. Harvard Medical School, Boston, MA 02115, USA. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. Novartis Institutes for Biomedical Research, Cambridge, MA, USA. Department of Biostatistics, University of Liverpool, Liverpool, UK. Department of Medicine (Nephrology, Hypertension and Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Heidelberg, Germany. The National Institute for Health Research University College London Hospitals Biomedical Research Centre, London, UK. The Alan Turing Institute, London, United Kingdom. Division of Molecular & Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. Cardiovascular Division, Department of Medicine, University of Minnesota, Minneapolis, MN, USA. Regeneron Pharmaceauticals, Cardiovascular Research, 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia. Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK.
Regeneron Genetics Center

Goncalo Abecasis\textsuperscript{22}, Joshua Backman\textsuperscript{22}, Xiaodong Bai\textsuperscript{22}, Suganthish Balasubramanian\textsuperscript{22}, Nilanjan Banerjee\textsuperscript{22}, Aris Baras\textsuperscript{22}, Leland Barnard\textsuperscript{22}, Christina Beechert\textsuperscript{22}, Andrew Blumenfeld\textsuperscript{22}, Michael Cantor\textsuperscript{22}, Yating Chai\textsuperscript{22}, Jonathan Chung\textsuperscript{22}, Giovanni Coppola\textsuperscript{22}, Amy Damasi\textsuperscript{22}, Frederick Dewey\textsuperscript{22}, Aris Economides\textsuperscript{22}, Gisu Eom\textsuperscript{22}, Caitlin Forsythe\textsuperscript{22}, Erin D. Fuller\textsuperscript{22}, Zhenhua Gu\textsuperscript{22}, Lauren Gurski\textsuperscript{22}, Paloma M. Guzzardo\textsuperscript{22}, Lukas Habegger\textsuperscript{22}, Young Hahn\textsuperscript{22}, Alicia Hawes\textsuperscript{22}, Cristopher van Hout\textsuperscript{22}, Marcus B. Jones\textsuperscript{22}, Shareef Khalid\textsuperscript{22}, Michael Lattari\textsuperscript{22}, Alexander Li\textsuperscript{22}, Nan Lin\textsuperscript{22}, Daren Liu\textsuperscript{22}, Alexander Lopez\textsuperscript{22}, Kia Manooghehr\textsuperscript{22}, Jonathan Marchini\textsuperscript{22},