**Genome-wide association analysis and Mendelian randomization proteomics identify novel protein biomarkers and drug targets for primary prevention of heart failure**

**Authors:**
Danielle Rasooly, PhD*[1,2], Gina M Peloso, PhD*[2,3], Alexandre C Pereira, MD, PhD[4,5], Hesam Dashti, PhD[1,6], Claudia Giambartolomei, PhD[7,8], Eleanor Wheeler, PhD[9], Nay Aung, MD, PhD[10,11], Brian R Ferolito, MS[2], Maik Pietzner, PhD[9,12], Nicole M Kosik, MPH[2], Liam Gaziano, MPhil[2,13], Daniel C Posner, PhD[2], A Patrícia Bento, PhD[14], Qin Hui, PhD[15,16], Chang Liu, MPH[15], Krishna Aragam, MD[2,6,17], Zeyuan Wang, MPH[15], Brian Charest MS, MPH[2], Jennifer E Huffman, PhD[2], Peter W.F. Wilson, MD[16,18], Lawrence S Phillips, MD[16,19], John Whittaker, PhD[20], Patricia B Munroe, PhD[10,21], Steffen E Petersen, MD, DPhil, MPH[10,11], Kelly Cho, PhD, MPH[1,2], Andrew R Leach, DPhil[14], María Paula Magariños, PhD[14], John Michael Gaziano, MD, MPH[1,2], VA Million Veteran Program[22], Claudia Langenberg, MD, PhD[9,12†], Yan V Sun, PhD, MS,[15,16,23†], Jacob Joseph, MBBS, MD[2,24†], Juan P Casas, MD, PhD[1,2†]

*These first authors contributed equally.
†These senior authors contributed equally.

**Institutions:**
[1] Division of Aging, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02130, USA.
[2] Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, 150. S. Huntington Ave, Boston, MA 02130, USA.
[3] Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Ave Crosstown Centre, Boston, MA 02118 USA.
[4] Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo, Av Dr Eneas de Carvalho Aguiar 54, São Paulo 5403000, Brazil.
[5]Genetics Department, Harvard Medical School, Harvard University, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.
[6] Broad Institute of MIT and Harvard, 415 Main St., Cambridge, MA 02142, USA.
[7] Central RNA Lab, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy.
[8] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA.
[9] MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Addenbrookes Hospital, IMS, Box 285, Cambridge, CB2 0QQ, UK.
[10] William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK.
[11] Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, UK.
[12] Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Kapelle Ufer 2, Berlin 10117, Germany.
[13] BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Worts Causeway, Cambridge, CB1 8RN, UK.

[14] Department of Chemical Biology, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, UK.
[15] Department of Epidemiology, Emory University Rollins School of Public Health, 1518 Clifton Rd NE, Atlanta, GA 30322, USA.
[16] Atlanta VA Health Care System, 1670 Clairmont Rd, Decatur, GA 30033, USA.
[17] Massachusetts General Hospital, Boston, MA 02114, USA.
[18] Division of Cardiology, Department of Medicine, Emory University School of Medicine, 1639 Pierce Dr NE, Atlanta, GA 30322, USA.
[19] Division of Endocrinology and Metabolism, Department of Medicine, Emory University School of Medicine, 1639 Pierce Dr NE, Atlanta, GA 30322, USA.
[20] MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, United Kingdom.
[21] National Institute for Health Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK.
[22] Million Veteran Program membership is provided in the Supplementary Material.
[23] Department of Biomedical Informatics, Emory University School of Medicine, 1639 Pierce Dr NE, Atlanta GA, 30332, USA.
[24] Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02130, USA.

**Authors for correspondence:**
Dr. Juan P Casas: jpcasasromero@bwh.harvard.edu
Dr. Jacob Joseph: jjoseph16@partners.org

**Abstract (150 words):**

We conducted a large-scale meta-analysis of heart failure (HF) genome-wide association studies (GWAS) consisting of over 90,000 HF cases and more than 1 million control individuals of European ancestry to uncover novel genetic determinants for HF. Using the GWAS results and blood protein quantitative loci (pQTLs), we performed Mendelian randomization (MR) and colocalization analyses on human proteins to provide causal evidence for the role of druggable proteins in the genesis of HF. We identified 39 genome-wide significant HF risk variants, of which 18 are previously unreported. Using a combination of MR proteomics and genetic cis-only colocalization analyses, we identified 10 additional putatively causal genes for HF. Findings from GWAS and MR-proteomics identified seven (CAMK2D, PRKD1, PRKD3, MAPK3, TNFSF12, APOC3 and NAE1) proteins as potential targets for interventions to be used in primary prevention of HF.

## INTRODUCTION

Heart failure (HF) is one of the most important threats to the sustainability of health systems for high-income countries such as the United States, and is responsible for a disproportionate and increasing morbidity, mortality, and economic burden in low and middle-income countries[1]. Despite major improvements in the understanding of risk factors for incident HF[2], this knowledge has not yet been fully translated into effective interventions for primary prevention of HF, except for blood pressure (BP) lowering medications[3] and statins[4]. In addition, traditional risk factors for HF, such as high BP, type-2 diabetes mellitus (T2DM), smoking, coronary artery disease (CAD) and obesity, only explain half of all incident HF cases[5], suggesting the need to uncover novel pathways involved in the development of HF.

Due to the inherent attributes of human genetics that minimize the risk of residual confounding and reverse causation[6], large-scale genomic analyses provides an opportunity to uncover causal mechanisms for complex phenotypes such as HF. Recent genome-wide association studies (GWAS) of HF by the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) and the Million Veteran Program (MVP)[7] have identified 26 genomic loci associated with HF[8]. This emerging knowledge has served to identify novel biological mechanisms associated with incident HF and may inform the development of novel interventions for the primary prevention of HF.

Novel technological developments can simultaneously measure thousands of human proteins in a single blood sample. The SOMAscan, one of such technologies, utilizes aptamers, which are short oligonucleotides designed to bind to their protein targets with high specificity[9]. The SOMAscan V4 assay includes 5207 aptamers capable of measuring 4988 unique human proteins, of which 514 are the target of drugs licensed or in clinical phase, 1153 are the target of compounds in pre-clinical phase, and 1377 are proteins predicted to be druggable [9,10]. This offers a unique opportunity for translating the genomic findings of HF into novel interventions for the primary prevention of HF.

Given that human proteins account for the majority of targets for approved drugs to date and that expression or activity is central to the development of human disease[11], leveraging GWAS data of HF and protein quantitative trait loci (pQTL) offers an opportunity to provide mechanistic insight into the causal pathways involved in the emergence of HF as well as to inform novel therapeutic targets.

Here, we conduct a meta-analysis of GWAS on HF from the MVP and the HERMES consortium and leverage our GWAS of HF with pQTLs from the Fenland study to conduct Mendelian randomization (MR) and genetic colocalization analyses on human proteins covered by SOMAscan V4[12]. We then perform extensive downstream analyses covering HF risk factors, cardiac MRI traits, -omics, and downstream transcriptomics analyses to investigate the biological credibility of our genetic findings.

## RESULTS

First, we meta-analyzed GWAS on HF from the HERMES consortium and MVP and identified variants at GW-significance ($p < 5 \times 10^{-8}$) (**Figure 1**). We performed follow-up analysis of the newly discovered HF variants to identify the likely causal gene for each signal and to investigate associations with 12 HF risk factors and 9 left ventricular (LV) cardiac MRI traits. We also conducted pathway enrichment analysis and determined *in silico* druggability. Second, using the GWAS data on SOMAscan V4 proteomics, we selected conditionally independent *cis*-variants, defined as any variant within a $\pm 1$ Mb region of the protein-encoding gene, that associated with plasma levels of SOMAscan proteins ($p < 5 \times 10^{-8}$). We propose that these variants are instrumental variables for measured SOMAscan proteins and conducted two-sample MR analyses using our European-descent GWAS meta-analysis of HF from the MVP and HERMES consortium. We conducted several analyses to minimize confounding and biases. For the MR results that passed our significance threshold (FDR < 5%), we performed genetic colocalization analysis to ensure the MR results were unlikely to be confounded by linkage disequilibrium (LD). For the MR results with evidence of colocalization, we conducted MR and colocalization analyses against HF risk factors and cardiac MRI traits and *cis*-eQTL searches. Then, we conducted a novel multi-step analytical approach to reduce the risk of horizontal pleiotropy. Finally, we annotated our MR hits for druggability.

### Genome-wide meta-analysis identifies 18 novel loci for HF

We performed meta-analyses of genome-wide association results for HF from two studies: MVP ($n_{cases}$=43,344; $n_{controls}$=258,943) and HERMES ($n_{cases}$=47,309; $n_{controls}$=930,014). After quality control, we obtained association results for 10,227,138 genetic variants with HF. We observed 39 variants with genome-wide significant signals with HF, of which 18 variants were > 500KB from a previously reported indexed variant (**Figure 2** and **Table S1**). We performed fine-mapping using GWAS summary statistics (**Figure S1**). We determined the gene closest to the indexed SNP, as well as the gene with the highest score from Polygenic Priority Score (PoPs)[13] within a 500KB region of the indexed SNP (**Table 1**). PoPs takes genome-wide features into account while the nearest gene is based on local information, providing complementary information for annotation of indexed variants (see Methods). For all the genes suggested by the nearest gene and PoPS, we also conducted a gene burden test, using putative Loss-of-Function (pLoF) variants, from the Genebass-UK Biobank resource. *RFX4* and *UBC*, both suggested by PoPs, showed the most significant gene-based p-values with HF (p-values of $9.12 \times 10^{-4}$ and $4.6 \times 10^{-3}$ , respectively). The gene burden tests obtained from Genebass were derived using both a gene-based (mean) and SKAT-O (hybrid variance/mean) approach[14]. From herein, we used genes suggested by PoPs as default to describe the distinct variants.

With the exception of rs6945340/*HIP1* and rs79682748/*SGIP1*, all other distinct variants for HF had an association (defined as 0.01/number of secondary traits, $p < 1 \times 10^{-4}$) with at least one HF risk factor (**Figure 3A**). Five variants had the largest number of associations with HF risk factors: rs9352691/*PHIP* (blood pressure, body mass index (BMI), high-density lipoprotein cholesterol (HDL-C), alcohol consumption, and atrial

fibrillation (AF)), rs12992672/*TMEM18* (BMI, HDL-C, type-2 diabetes mellitus (T2DM), AF and smoking), rs4755720/ HSD17B12 (BMI, HDL-C, T2DM and CAD), rs233806/*BANK1* (blood pressure, HDL-C and BMI) and rs959388/*PRKD1* (BMI, smoking, and blood pressure), details in **Table S2**. We observed that the directionality of the associations with HF risk factors was concordant with the findings on HF risk in 32 out of the 42 (76%) associations. HDL-C and diastolic BP accounted for nine of the 10 discordant associations (**Figure S2**).

Only three variants (rs3820888/*SPATS2L,* rs4755720/*HSD17B12* and rs72688573/*FAF1)* showed at least one association ($p < 1 \times 10^{-4}$) with LV cardiac-MRI traits (**Figure S3** and **Table S2**). The rs3820888/*SPATS2L* variant was associated with six LV cardiac-MRI traits and AF; all these associations were directionally concordant with the HF findings. The rs4755720/*HSD17B12* variant was associated with LV end-diastolic volume indexed to body surface area and four HF risk factors, and rs72688573/*FAF1* was associated with LV mass to end-diastolic volume ratio and two HF risk factors, see details in **Table S2**.

**MR Proteomics and colocalization identifies 10 genes for HF**

We used 2,900 *cis*-pQTLs across 1,557 genes from the Fenland study as proposed instruments for two-sample MR of proteomics with HF. We found 16 genes passed our MR threshold (FDR < 5%), of which 10 genes also showed suggestive evidence of colocalization between HF and pQTL signals (posterior probably of Hypothesis 4 (PP.H4): one common causal variant > 0.5) for at least one of the instruments, see details in **Table 2 and Table S3**. Except for *ENPEP*, no other gene that colocalized was within 500KB of a known HF GWAS loci. For those genes with more than one instrument, we did not observe any evidence of heterogeneity based on Cochran's Q statistic according to the IVW model, **Table 2**.

Except for *ENPEP*, *TNXB*, and *SIRPA*, all the other genes that passed MR and colocalization thresholds with HF also showed an association (defined as MR $p < 1 \times 10^{-4}$ and colocalization: PP.H4 > 0.5) with at least one of the 12 HF risk factors (**Figure 3B, Table S4**). We observed that the directionality of the MR associations with HF risk factors was concordant with the MR findings on HF in 10 out of the 14 (71%) associations. HDL-C, LDL-C and systolic BP accounted for all discordant associations. Only the *TNFSF12* gene showed an association with a LV cardiac MRI trait that passed statistical thresholds for MR and colocalization, see details on **Table S4 to S5.** In order to provide some further mechanistic insight, we investigated if the *cis*-pQTL instruments for the 10 MR genes were also *cis*-eQTLs ($p < 5 \times 10^{-8}$). Twelve of the 18 proposed instruments were also *cis*-eQTLs in at least one tissue. None of the *cis*-pQTLs used as proposed instruments for *TNXB*, *APOC3*, and *APOH* genes showed a *cis*-eQTL association (**Table S6**).

In our assessment of horizontal pleiotropy (see Methods and **Figure S4**), the 18 proposed instruments for the 10 MR genes were associated ($p < 5 \times 10^{-8}$) with 251 proteins or gene-expression using SOMAscan V4, Fenland study and eQTLGen, respectively (**Table S7)**. For 217 of the 251 proteins/gene-expression, we identified at least one *cis*-pQTL or *cis*-eQTL at $p < 5 \times 10^{-8}$ associated with protein levels based on the SOMAscan V4 Fenland

study, or gene expression based on eQTLGen. We then conducted two-sample MR of these secondary proteins/genes expression against HF and identified four genes (*TP53*, *ZNF259, ACVR2A, and MYRF*) that passed multiple testing thresholds (0.05/217, $p < 2 \times 10^{-4}$, **Table S7**). These four secondary genes correspond to the following genes identified by MR proteomics as hits for HF: *TNXB* (*ACVR2A* and *MYRF), APOH (TP53) and APOC3 (ZNF259). TP53* and *ACVR2A* were in a different biological pathway than *APOH* and *TNXB*, respectively, suggesting potential horizontal pleiotropy. *ZNF259* and *MYRF* did not retrieve any biological pathways; hence, it is unknown if these are due to horizontal pleiotropy. We then determined protein-protein interaction (PPI) networks for APOH and TNXB proteins using Enrichr and GPS-Prot databases. The Enrichr's PPI Hub Protein pathways reported interactions between APOH and CDC42, AKT1, TP53, and GRB2 (adjusted p-values < 0.04), while the GPS-Prot showed that the APOH protein is directly connected to TP53 with a confidence > 0.6 **(Figure S5).** No significant interaction was identified for the TNXB and ACVR2A proteins.

**Pathway enrichment analysis recovers pathways relevant to HF**

We used previously published and our newly identified HF GWAS variants (n=40) together with the 18 proposed instruments for the 10 MR-proteomics genes associated with HF and conducted gene pathway enrichment analysis using GTEx V8. These 58 variants associated with 1,605 GTEx V8 *cis*-eQTLs ($p < 1 \times 10^{-4}$), corresponding to a total of 165 unique genes. For the most represented tissues, see **Table S8**. After restricting the analysis to pathways described in Gene Ontology, KEGG, and Reactome, we observed 56 enriched pathways (FDR < 5%). Biological pathways to highlight include muscle adaptation (adjusted p-value = 0.03), ventricular system development (p = 0.03), sarcomere organization (p = 0.04), regulation of vasculature development (p = 0.04), and aldosterone-regulated sodium reabsorption (p = 0.04), details on **Figure S6** and **Table S9**.

For the 18 GWAS distinct variants on HF, we determined the *differential* gene expression associated with the novel HF variants ($p < 1 \times 10^{-4}$) in each GTEx V8 tested tissue (heart atrial, heart ventricle, artery aorta, adipose, liver, kidney, and whole-blood tissues, and transformed cultured fibroblasts). We then used the set of differentially expressed genes to conduct over-representation analysis on a per-tissue basis (**Figure S7**). A total of 605 enriched pathways had at least two differentially expressed genes, with heart-left ventricle being the tissue with the most significantly enriched pathways (n=393). The rs6945340/*HIP1* variant showed the largest number of enriched pathways (n=391, all tissues) with heart left-ventricle being the primary tissue. Pathways to highlight for this variant include Krebs cycle, respiratory electron transport chain (both with $p= 4.8 \times 10^{-30}$) and oxidative phosphorylation ($p = 3.2\ 10^{-5}$). Further details on enriched pathways for the 18 variants in all tested tissues are available in **Table S10**.

**Mouse knock-out models for novel genes identified by GWAS or MR-proteomics**

We queried for knock-out (KO) mouse models, using the Mouse Genomics (MGI) resource, for evidence that modification of the target (i.e. novel genes identified by our GWAS or MR-proteomics) produces a phenotype relevant to HF. In thirteen genes (8 GWAS and 5 MR-proteomics genes), we retrieved evidence of a KO associated with

cardiovascular abnormalities. KO models on *CAMK2D*, *PRKD1*, *MAPK3*, *NAE1*, *SLC39A8*, *PHIP*, *RFX4*, *SCARB1* and *TNXB* showed phenotypes such as myocardial abnormalities, dilated cardiomyopathy, abnormal response to cardiac infarction, and cardiac hypertrophy, suggesting an intrinsic role in heart function regulation **(Table S11)**.

**Druggability**

A total of seven novel genes from the GWAS (*CAMK2D*, *PRKD1* and *PRKD3)* and MR-proteomics (*MAPK3*, *TNFSF12*, *APOC3* and *NAE1*) were identified to encode proteins that are predicted to be druggable (*CAMK2D*) or targets for 14 unique drugs that are either licensed or in the clinical phase (*PRKD1*, *PRKD3*, *MAPK3*, *TNFSF12*, *APOC3* and *NAE1*). Except for drugs targeting Apolipoprotein C-III mRNA, Volanesorsen and AKCEA-APO-CIII-LRx evaluated for familial chylomicronemia syndrome, all the other 12 drugs are either licensed or under clinical investigation for cancer (n=10 (*MAPK3*, *PRKD1*, *PRKD3* and *NAE1*)) or autoimmune disorders (n=2, (*TNFSF12*)). In four of the seven druggable genes, we were able to use our MR findings to infer the type of pharmacological action (agonist versus antagonist) needed to prevent HF and compared this against the pharmacological action of the existing drugs with a single target (which are most likely to reproduce genetic findings). Through this process, we observed a match in one gene (*APOC3*); and for the other druggable genes (*MAPK3*, *NAE1*, and *TNFSF12*), the existing drugs were an inhibitor/antagonist, while MR suggested an agonist, details on **Table S12**.

**In-silico trials**

We searched for genetic associations (for the GWAS hits) and conducted two-sample MR (for the MR proteomics hits) to evaluate safety and efficacy outcomes relevant for the primary prevention trials on HF. Seven of the 18 GWAS distinct variants and two of the 10 MR-proteomics genes were additionally associated ($p < 1 \times 10^{-4}$) with efficacy outcomes (CAD, T2DM) in the same direction as HF (**Table S13).** None of the 18 distinct GWAS variants or the 10 MR-proteomics genes showed an association with a safety binary (cancers (lung, prostate, colorectal, breast), chronic kidney disease and Alzheimer's disease) or continuous (liver enzymes, creatinine) outcome at $p < 1 \times 10^{-4}$.

**Comparison with Global Biobank Meta-analysis Initiative (GBMI) on HF**

An unpublished study from the GBMI reporting a multi-ancestry HF GWAS (68,408 HF cases and 1,286,331 controls) identified 11 potentially novel loci for HF[15]. We compared these associations with our HERMES-MVP GWAS and determined that seven of the 11 GBMI variants were associated ($p < 5 \times 10^{-8}$) in our HF meta-analysis. Two GBMI loci correspond to the same variants (rs10455872/*PLG* and rs600038/ *SURF1*) previously reported by HERMES or MVP, and an additional five loci were in LD ($r^2$ range: 0.39 to 1) with our findings **(Table S14)**. Finally, two GBMI GWAS variants (rs17035646 and rs61208973) showed suggestive evidence of association in our HF GWAS ($p < 0.003$).

## DISCUSSION

Our genetic analysis on HF consisting of 90,653 cases identified 18 distinct HF-associated variants through GWAS and an additional 10 putatively causal genes for HF through MR and colocalization using proteomic instruments. Our study expands the knowledge on the biological pathways associated with all HF risk loci discovered to date and identifies seven druggable genes as potential targets for intervention for the primary prevention of HF.

We conducted several strategies to provide biological credibility to our 18 distinct GWAS variants. First, 16 of the 18 variants showed genetic associations with HF risk factors that were directionally concordant with the HF findings, and several LV cardiac MRI traits. Second, over-representation analysis using differentially expressed genes by each GWAS variant identified the heart left ventricle as the most significantly enriched tissue and recovered several pathways of HF relevance. Third, systematic querying on KO mouse models identified *CAMK2D, PRKD1, PHIP, RFX4, SLC39A8* and *SCARB1,* novel genes found by our GWAS, of which the KO models presented phenotypes relevant to HF. Novel variants to highlight include rs3820888/*SPATS2L* and rs4755720/*HSD17B12* that showed associations with HF risk factors and LV cardiac MRI traits. The rs3820888/*SPATS2L* variant showed evidence of colocalization with six cardiac MRI traits, including LVEF, LV mass to end-diastolic volume ratio, and AF, all of which were directionally concordant with the HF findings. Previous GWAS have also indicated that the same variant was also associated with QT interval[16]. The rs4755720/*HSD17B12* variant colocalized with LV end-diastolic volume indexed to BSA and HF risk factors that were directionally concordant with the HF findings, all showing a protective effect. Previous GWAS indicated that this variant, as well as others in strong LD, associated with a reduction in adiposity measures and an increase in lung function metrics, suggesting that cardiometabolic fitness may explain the association with HF[17–19].

We conducted MR-proteomic analyses to uncover the putative causal role of human proteins in HF. Ten genes passed our genetic colocalization test, of which nine were also not in LD with a previously reported HF variant, minimizing the probability of confounding by LD. Seven of the 10 genes showed associations with at least one HF risk factor, and in the majority (71%) of these associations, the point estimate was directionally concordant with the MR findings on HF.

Four (*MAPK3*, *PRKD1*, *CAMK2D* and *PRKD3*) of the seven druggable genes identified by our analyses encode proteins with serine/threonine kinase activity. These four genes associated with HF risk factors in a manner that is concordant with the findings on HF. *CAMK2D* also showed a suggestive association ($p = 9 \times 10^{-4}$) with LV mass. In support of our findings, a mouse model with deletion of *MAPK3*/*MAPK1* genes developed cardiac hypertrophy and ventricular dilation followed by reduced ventricular performance[20]. *CAMK2D*, *PRKD1*, *PRKD3* are calcium/calmodulin dependent protein kinases known to be associated with cardiac pathophysiology. Protein Kinase-D, encoded by *PRKD1* gene, appears to be a regulator of myocardial structure and function. Mice with a deletion of *PRKD1* in cardiomyocytes were reported to be resistant to stress induced hypertrophy in response to pressure overload, angiotensin-II and adrenergic activation[21]. Calcium/Calmodulin-Dependent Protein Kinase II (CamKII) is composed of four chains,

one of which, delta (δ), is encoded by the *CAMK2D* gene. CamKII-δ is largely expressed in cardiac tissue (confirmed by our pathway enrichment analysis), where it regulates proteins involved in calcium handling, excitation-contraction coupling, activation of hypertrophy, cell death and inflammation[22]. Several case-control studies have shown an upregulation of cardiac CamKII-δ expression and activity in patients with HF, dilated cardiomyopathy and diabetic cardiomyopathy. In support of this, several experimental studies in animal models of dilated cardiomyopathy and HF have shown that chemical inhibition of CamKII led to protection from cardiac dysfunction, adverse cardiac remodeling, and cardiac arrhythmias[22]. More recently, administration of a novel ATP-competitive CaMKII-δ oral inhibitor (RA306) in a dilated cardiomyopathy mouse model led to an improvement of ejection fraction[23]. This oral inhibitor offers the opportunity to test the causal role of CamKII-δ through clinical trials for the prevention of HF. Interestingly, *CAMK2D* gene was also associated with AF, confirming an association demonstrated by in-vitro and animal models of AF[22].

Additional druggable genes identified were *APOC3*, *TNFSF12*, and *NAE1*. The *APOC3* gene is known for its associations with lipids, and CAD, which were confirmed in our MR and colocalization analysis and were directionally concordant with HF risk. Apolipoprotein C-III mRNA is targeted by two different antisense oligonucleotides (ASO), Volanesorsen and AKCEA-APO-CIII-$L_{Rx}$, evaluated for familial chylomicronemia syndrome. Phase 3 trials on Volanesorsen have shown an increase in LDL-C levels and thrombocytopenia, which make it an unlikely candidate for long-term use for the prevention of HF.[24] AKCEA-APO-CIII-$L_{Rx}$ is an ASO liver specific that appears to have a better safety profile, and may be more suitable for long-term use[25]. *TNFSF12* gene encodes for the TNF superfamily member 12 protein; increased levels of this protein were associated with a risk reduction in HF according to our MR and colocalization findings. Similar, directionally concordant, findings were reported by recent MR proteomics (using various proteomics platforms) against ischemic stroke[26]. These results are consistent with the finding that *TNFSF12* is MR associated and colocalized with AF, a risk factor for both ischemic stroke and HF. In addition, we observed a clear reduction in LV mass to end-diastolic volume ratio and a suggestive ($p = 2 \times 10^{-3}$) increase in LVEF, both directionally concordant with a risk reduction in HF. Transgenic mice and adenoviral-mediated gene expression models have also pointed to a role of *TNFSF12* in the development of dilated cardiomyopathy and severe cardiac disfunction[27]. *NAE1* gene encodes NEDD8 activating enzyme E1 subunit 1 protein, and our MR and colocalization findings showed this gene was associated with lower values of blood pressure, which coincides with the reduced risk on HF.

Strengths of the current analysis are multiple. First, the large number of HF cases included in our analysis led us to identify new variants and putatively causal genes for HF through GWAS and MR proteomics. Second, we used three complementary strategies— nearest gene (local method), PoPs (global method) and pLoF— to assign the most likely gene responsible for the GWAS signal with HF. Through this process, we observed agreements in 11 of 18 GWAS variants, which provided some degree of confidence in the gene-prioritization. For the remaining nine variants there is some uncertainty in the gene responsible for the association, highlighting the challenge in assigning the gene responsible for GWAS loci[28–30]. Third, we provide biological credibility for most of our

genetic findings through extensive and complementary analysis covering HF risk factors, LV cardiac MRI, and -omics. Fourth, in seven MR hits for HF, we showed that our proposed instruments, in addition to associations with HF risk factors or LV cardiac MRI traits, were also associated with gene expression, and protein levels all acting in *cis*. Fifth, KO models of several genes identified through GWAS and MR developed highly relevant phenotypes to HF and in some cases (CAMK2D), specific pharmacological inhibition showed reversibility of the HF-phenotypes. Six, the lack of associations between the distinct GWAS loci and MR genes with safety outcomes used in the primary prevention trials of HF provides some reassurance on target safety profiles.

The degree of credibility on the causality of proteins identified by MR depends on whether the MR assumptions are valid. First, our colocalization analysis on HF, risk factors for HF and LV cardiac MRI traits makes confounding by LD unlikely. The selection of *cis*-variants as proposed instruments minimizes the chances of horizontal pleiotropy. To even further minimize chances of horizontal pleiotropy, we developed a novel analysis that attempted to empirically test the relevant conditions needed for horizontal pleiotropy to invalidate MR. First, we looked for secondary proteins or gene-expression associated with our MR protein hits, and then evaluated if those secondary proteins/gene-expression were associated with HF and fall in a biological or PPI pathway outside our protein hits. After doing this, only *TNXB* showed some evidence of horizontal pleiotropy. Interestingly, *cis*-pQTLs used as instruments for TNXB were not associated with *cis*-eQTLs, HF risk factors or LV cardiac MRI traits.

Although most of our variants and genes showed associations with HF risk factors that were biologically concordant with HF risk, some discordant associations were observed. HDL-C and diastolic BP accounted for most of these discordant associations. It has been reported that higher levels of diastolic BP may be protective on HF[31,32], instead of deleterious as we assumed, while the HDL-C association with HF seems to be non-linear[32], which was not accounted for in our MR analysis that included HDL-C as a co-variable. We validated seven of the 11 variants reported in an unpublished HF analysis by GBMI[15]. Future GWAS meta-analysis including larger releases of MVP, All of US and GBMI will provide chances for replication as well as to include non-white populations[15,33,34]. Although the absence of HF sub-types in this analysis most certainly decreased our ability to detect signals specific to HF sub-types, it does not invalidate the ones identified. Evidence from primary prevention trials using HF as an outcome (as our genetic study) that uncovered the benefits of BP lowering therapies and statins indicates the plausibility for translation of our genetic findings. Although our design attempted to emulate a primary prevention trial on HF, further studies with access to individual participant data that reliably recreate eligibility criteria and outcome ascertainment that cover efficacy (including HF sub-types) and safety outcomes are needed.

In conclusion, we discovered a total of 18 distinct novel HF-associated variants and 10 putatively causal genes for HF through GWAS and MR-proteomics with evidence of biological plausibility. The new mechanisms and pathways together with the seven druggable genes discovered provide a tractable path for the translation of our genomic findings for the primary prevention of HF.

## METHODS

### Genome-wide association study for HF

We performed a fixed effects inverse-variance weighted meta-analysis HF from the published MVP (n=302,258) and HERMES (n=964,057)[8] GWAS using METAL[35] (version release 2020-05-05) in a total of 1,266,315 individuals. We removed variants with a MAF < 0.5%, resulting in 10,227,138 associations.

We used FUMA[36] to annotate our results using the default settings. We defined distinct variants to have an R2 < 0.6 and determined the associations that were > 500KB from a previously reported indexed variant in MVP and HERMES. We used the closest gene to the indexed variant and the top gene per locus identified by PoPs to prioritize genes for our GWA-significant ($p < 5 \times 10^{-8}$) loci.

The PoPS method[13] is a new gene prioritization method that identifies the causal genes by integrating GWAS summary statistics with gene expression, biological pathway, and predicted protein-protein interaction data. First, as part of the PoPS analysis, we used MAGMA to compute gene association statistics (z-scores) and gene-gene correlations from GWAS summary statistics and LD information from the 1000 Genomes. Next, PoPS performs marginal feature selection by using MAGMA to perform enrichment analysis for each gene feature separately. The model is fit by generalized least squares (GLS), and MAGMA results are used to perform marginal feature selection, retaining only features that pass a nominal significance threshold ($p < 0.05$). Then, PoPS computes a joint enrichment of all selected features simultaneously in a leave one chromosome out (LOCO) framework. The gene features employed by PoPS are listed here: https://github.com/FinucaneLab/gene_features. Finally, PoPS computes polygenic priority scores for each gene by fitting a joint model for the enrichment of all selected features. The PoP score for a gene is independent of the GWAS data on the chromosome where the gene is located. The PoPS analysis returned scores for a total of 18,383 genes per set of GWAS datasets. We then annotated our GWAS loci with the Ensembl genes in a 500kb window and selected the highest PoP score gene in the locus as the prioritized gene.

### Associations of HF GWAS variants with HF risk factors and LV cardiac MRI traits

For genetic variants that passed the GWAS threshold for HF ($p < 5 \times 10^{-8}$), we determined genetic associations for 12 HF risk factors and 9 LV cardiac MRI traits derived from available GWAS. Data on HF risk factors was obtained from European-descent GWAS studies: BMI[37], smoking[38], alcohol intake frequency[39], AF[40], diastolic and systolic BP [41], T2DM[42], CAD[43], LDL-C[44], HDL-C[44], estimated glomerular filtration rate (eGFR)[29], and chronic obstructive airways disease (COPD)[38].

For LV cardiac MRI traits, we determined genetic associations from two separate publications. Seven LV cardiac MRI measurements in 36,041 participants of the UK Biobank from Pirruccello *et al*[45] and LV mass and LV mass to end-diastolic volume ratio from cardiac MRI in 42,157 UK-Biobank participants from Aung *et al (unpublished)* using automated CMR analysis techniques and LV GWAS techniques[46,47].

We used $p < 1 \times 10^{-4}$ (0.01/number of secondary to HF traits tested in the manuscript) to account for multiple testing. For associations that passed our p-value threshold, we evaluated whether the directionality of HF risk factors associations was concordant with findings on HF; for example, for a variant that showed an increased risk of HF, we expect a positive association with a deleterious risk factor.

**Mendelian randomization on 1557 proteins and HF**

***Selection of proposed pQTL instruments***

We obtained pQTLs from a genome-proteome-wide association study in the Fenland study of 10,708 participants of European-descent[12] (retrieved from www.omiscience.org). The genome-proteome-wide association study was conducted using 10.2 million genetic variants and plasma abundances of 4,775 distinct protein targets (proteins targeted by a least one aptamer) measured using the SOMAscan V4 assay[12]. Significant genetic variant pQTLs were defined as passing a Bonferroni p-value threshold of $p < 1.004 \times 10^{-11}$. Approximate conditional analysis was performed to detect secondary signals for each genomic region identified by distance-based clumping of association statistics[12]. To diminish the likelihood of horizontal pleiotropy, we restricted proposed instrumental variants to (lead and secondary signals) *cis*-pQTLs using a p-value threshold of $p < 5 \times 10^{-8}$ in marginal statistics, where *cis* is defined as any variant within a $\pm 1$ Mb region of the protein-encoding gene. A total of 2,900 *cis*-pQTLs across 1557 genes (mean=1.9, min=1, max=14) covering an equal number of proteins from the Fenland study were used as proposed instruments for 2-sample MR of proteomics against HF.

**Mendelian Randomization and colocalization**

We performed two-sample MR using the TwoSampleMR package in R (https://mrcieu.github.io/TwoSampleMR/) [48,49]. The Wald Ratio was used for instruments with one variant and the inverse-variance weighted MR method was used for instruments with multiple variants. We tested the heterogeneity across variant-level MR estimates, using the Cochrane Q method (mr_heterogeneity option in TwoSampleMR package) and plotted the effects of the variants on the proteins against the effects of the variants on HF to validate our instruments when more than one variant was included. We defined significant MR results using a False Discovery Rate (FDR) of 0.05 calculated by the Benjamini-Hochberg method (corresponding p-value = $5 \times 10^{-4}$).

For results that passed our MR threshold (FDR<0.05), we performed colocalization between the HF GWAS and *cis*-pQTL instruments for the protein of interest using the coloc package[50–52] in R (https://chr1swallace.github.io/coloc, using default priors). We performed colocalization using marginal (unadjusted) pQTL results as well as results conditional on each of the instruments used in the MR using windows of +/-200KB from the instruments for the conditional results and +/-200KB from the TSS for the marginal results. Statistically significant MR hits with a posterior probability for hypothesis 4 (PP.H4) > 0.5 (the probability of a shared causal variant) for at least one instrumental variant were then investigated further with default priors (probability of shared causal variant for trait 1 and trait 2 is $P1 = P2 = 1 \times 10^{-4}$, probability of shared causal variant across two traits is $P12 = 1 \times 10^{-5}$). We also investigated if the *cis*-pQTL instruments for genes that passed both MR and colocalization thresholds were also *cis*-eQTLs ($p < 5 \times$

$10^{-8}$). Tissues used were whole blood from eQTLGen and heart atrial, heart ventricle, artery aorta, adipose, liver, kidney tissues, and transformed cultured fibroblasts from GTEx V8.

**MR and colocalization for HF risk factors and cardiac MRI traits**

For proteins that passed both MR and colocalization thresholds, we conducted 2-sample MR analyses of these proteins, using *cis*-pQTLs from the Fenland study as proposed instruments, against 12 HF risk factors and 9 cardiac MRI traits described in the previous section, see **Supplementary Material** for details on traits and datasets. For the MR results that passed a p-value threshold of $p < 1 \times 10^{-4}$, we conducted colocalization analyses as previously described. We defined significant findings as those that passed thresholds for MR ($p < 1 \times 10^{-4}$) and colocalization (PP.H4 > 0.5).

**Assessment of horizontal pleiotropy**

For statistical findings that passed the MR and colocalization thresholds, we evaluated the possibility that horizontal pleiotropy may invalidate our findings. The pipeline of analysis is depicted in **Figure S4**. *Step-1*: We determined if our *cis*-pQTLs were associated ($p < 5 \times 10^{-8}$) with other proteins levels included in SOMAscan V4 or with gene expression using data from eQTLGen. *Step-2*: We queried if the genes (including genes that encode SOMAscan proteins) identified in *Step-1* were within 1MB of the risk loci for HF identified by GWAS conducted to date. *Step-3*: We conducted two-sample MR to identify if the secondary genes/proteins (identified in *Step-1*) were associated with HF, using a Bonferroni-corrected p-value (0.05/number of unique genes/proteins identified in *Step-1*). We leveraged as proposed instruments the lead *cis*-pQTL ($p < 5 \times 10^{-8}$) from the Fenland study, and if it was not available, we used the lead *cis*-eQTL ($p < 5 \times 10^{-8}$) identified from eQTLGen. *Step-4*: We then mapped all secondary genes/proteins identified in *Step-3* to Reactome/KEGG pathways; and compared if these pathways are on the same (vertical pleiotropy) or different (horizontal pleiotropy) pathway as that associated with the primary genes identified through MR proteomics for HF. To further investigate the physiological functionalities of our findings retrieved in *Step-4*, we queried two databases: the Enrichr[53–55] , an interactive gene knowledge discovery database, and the GPS-Prot server[56], a platform with aggregated information about protein-protein interactions.

**Pathway enrichment analysis**

We conducted enrichment analysis to identify biological pathways associated with HF risk loci (established and novel) that passed the GWAS p-value thresholds. For each locus, we selected the top variant and then identified *cis*-eQTLs (within a 1Mb region) from GTEx V8 in any tissue associated with the top variants and extracted all genes with a $p < 1 \times 10^{-4}$. We merged all retrieved genes to a gene set that was then used for inquiry for the enriched pathways. This set of genes was set forth to an over-representation analysis using the pathways described in Gene Ontology, KEGG and Reactome. Selected pathways were those significantly enriched at an FDR < 0.05.

Additionally, we explored the downstream transcriptional consequences associated with the distinct variants identified by our GWAS on HF and those not previously reported. We

used the distinct variants and conducted a differential gene-expression analysis (using a dominant model) for all transcripts available in GTEx V8 for heart atrial, heart ventricle, artery aorta, adipose, liver, kidney, transformed cultured fibroblasts and whole-blood tissues. After fitting models for our variants, we retrieved all genes differentially expressed at a $p < 1 \times 10^{-4}$ and conducted an enrichment pathway analysis (through an overrepresentation analysis, as described above). Enrichment analyses were performed using the R packages clusterProfiler and enrichplot[57].

**Querying the MGI database:**

We queried the Mouse Genome Informatics (MGI, http://www.informatics.jax.org/) resource for all candidate genes from our novel GWAS hits list or those suggested as causal from our MR/colocalization approach. MGI uses a standardized nomenclature, and controlled vocabularies such as the Mouse Developmental Anatomy Ontology, the Mammalian Phenotype Ontology and the Gene Ontologies. As MGI extracts and organizes data from primary literature, we have parsed all system abnormalities associated with models on all of the queried genes[58]. For models that displayed cardiovascular abnormalities, we have hand-curated the abnormalities and organized them into 3 distinct groups associated with (1) congenital heart malformations, (2) myocardial abnormalities, and (3) vascular abnormalities.

**Druggability annotations**

Proteins encoded by genes identified in the GWAS and MR analyses for HF were annotated with drug tractability information based on information provided by OpenTargets[10,59,60] (release 2021-03-08). OpenTargets tractability system stratified drug targets into nine mutually exclusive groups (termed "buckets") based on the drug type and the stage of the drug discovery pipeline. For easier interpretation, we regrouped the original buckets into four mutually exclusive groups, as follows: *Licensed drugs*: bucket-1 for antibodies, small molecules and other modalities. *Drugs in clinical development:* buckets 2 and 3 for antibodies, small molecules, and other modalities. *Compounds in preclinical phase*: buckets 4 and 5 for small molecules. *Predicted druggable*: buckets 6 to 8 for small molecules plus buckets 4 and 5 for antibodies. The remaining proteins were considered non-druggable. For genes that were the target of licensed drugs, we checked whether the disease indication was also a risk factor for HF, as this may introduce a bias analogous to confounding by indication in MR.

**Table 1. Loci reported for HF in meta-analysis of HERMES and MVP HF GWAS datasets.**
Genes that are druggable or predicted to be druggable are highlighted in bold.

| rsID | Chr | Pos | Nearest Gene | PoPs gene | CADD Phred Score | NEA | EA | MAF | Beta | SE | P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Novel variants*** | | | | | | | | | | | |
| rs4755720 | 11 | 43628749 | *HSD17B12* | *HSD17B12* | 7.526 | C | T | 0.400 | -0.037 | 0.006 | 8.14E-11 |
| rs7766436 | 6 | 22598259 | *HDGFL1* | *HDGFL1* | 1.902 | C | T | 0.286 | 0.037 | 0.006 | 5.18E-10 |
| rs3820888 | 2 | 201180023 | *SPATS2L* | *SPATS2L* | 6.261 | C | T | 0.378 | -0.034 | 0.006 | 1.20E-09 |
| rs12992672 | 2 | 632592 | *TMEM18* | *TMEM18* | 1.061 | G | A | 0.172 | 0.045 | 0.007 | 1.70E-09 |
| rs10846742 | 12 | 125308682 | *SCARB1* | *UBC* | 0.277 | G | A | 0.173 | -0.046 | 0.008 | 1.73E-09 |
| rs17620390 | 4 | 114384328 | ***CAMK2D*** | ***CAMK2D*** | 1.907 | C | A | 0.284 | -0.037 | 0.006 | 1.90E-09 |
| rs72688573 | 1 | 50746997 | *FAF1* | *FAF1* | 6.834 | C | T | 0.022 | -0.122 | 0.021 | 4.31E-09 |
| rs10938398 | 4 | 45186139 | *GNPDA2* | *N/A* | 1.094 | G | A | 0.421 | 0.033 | 0.006 | 4.50E-09 |
| rs6945340 | 7 | 75100124 | *POM121C* | *HIP1* | 3.58 | C | T | 0.208 | -0.040 | 0.007 | 5.89E-09 |
| rs7564469 | 2 | 145258445 | *ZEB2* | *GTDC1* | 19.56 | C | T | 0.165 | -0.043 | 0.007 | 6.66E-09 |
| rs7977247 | 12 | 107259470 | *RIC8B* | *RFX4* | 1.868 | C | T | 0.434 | 0.032 | 0.006 | 1.07E-08 |
| rs1016287 | 2 | 59305625 | *FANCL* | *N/A* | 19.22 | C | T | 0.280 | 0.037 | 0.006 | 1.11E-08 |
| rs959388 | 14 | 30169987 | ***PRKD1*** | ***PRKD1*** | 0.596 | G | T | 0.417 | -0.031 | 0.006 | 1.30E-08 |
| rs233806 | 4 | 103212846 | *SLC39A8* | *BANK1* | 9.713 | C | T | 0.207 | -0.037 | 0.007 | 1.57E-08 |
| rs17038861 | 2 | 37233265 | *HEATR5B* | ***PRKD3*** | 0.403 | G | T | 0.195 | 0.039 | 0.007 | 2.35E-08 |
| rs9352691 | 6 | 79785607 | *PHIP* | *PHIP* | 6.373 | C | T | 0.365 | 0.032 | 0.006 | 2.65E-08 |
| rs10520390 | 19 | 46327831 | *SYMPK* | *DMWD* | 2.849 | G | C | 0.059 | 0.074 | 0.013 | 2.87E-08 |
| rs79682748 | 1 | 66989719 | *SGIP1* | *SGIP1* | 4.719 | G | A | 0.018 | -0.155 | 0.028 | 3.00E-08 |
| ***Previously reported variants*** | | | | | | | | | | | |
| rs7859727 | 9 | 22102165 | *CDKN2B* | *CDKN2A* | 1.448 | C | T | 0.488 | 0.061 | 0.006 | 3.11E-29 |
| rs2634071 | 4 | 111669220 | *PITX2* | *PITX2* | 1.622 | C | T | 0.219 | 0.079 | 0.007 | 3.64E-29 |
| rs11642015 | 16 | 53802494 | *FTO* | *RPGRIP1L* | 4.826 | C | T | 0.432 | 0.058 | 0.006 | 2.69E-25 |
| rs10455872 | 6 | 161010118 | *LPA* | ***PLG*** | 0.146 | G | A | 0.074 | -0.104 | 0.011 | 8.20E-23 |
| rs3176326 | 6 | 36647289 | *CDKN1A* | *CDKN1A* | 10.88 | G | A | 0.173 | -0.068 | 0.007 | 2.51E-22 |
| rs602633 | 1 | 109821511 | *PSRC1* | *CELSR2* | 8.63 | G | T | 0.207 | -0.054 | 0.007 | 5.19E-16 |
| rs1739833 | 1 | 16331108 | *C1orf64* | *ZBTB17* | 4.205 | C | T | 0.331 | -0.048 | 0.006 | 7.89E-15 |
| rs17617337 | 10 | 121426884 | *BAG3* | *BAG3* | 0.079 | C | T | 0.218 | -0.050 | 0.007 | 8.88E-14 |
| rs600038 | 9 | 136151806 | *ABO* | *SURF1* | 7.596 | C | T | 0.215 | -0.049 | 0.007 | 9.44E-14 |
| rs34163229 | 10 | 75406912 | *SYNPO2L* | *SEC24C* | 24 | G | T | 0.133 | -0.056 | 0.008 | 7.40E-13 |
| rs113437066 | 17 | 65836220 | *BPTF* | *BPTF* | 2.588 | ATTT | A | 0.197 | 0.061 | 0.010 | 1.81E-10 |
| rs11746435 | 5 | 137006762 | *KLHL3* | *HNRNPA0* | 7.216 | T | A | 0.229 | 0.042 | 0.007 | 2.04E-10 |
| rs2832275 | 21 | 30602994 | *BACH1* | *LTN1* | 1.08 | T | A | 0.139 | -0.047 | 0.008 | 2.90E-10 |
| rs7795282 | 7 | 74122857 | *GTF2I* | *GTF2IRD1* | 0.262 | G | A | 0.221 | -0.042 | 0.007 | 7.69E-10 |
| rs12933292 | 16 | 69566309 | *NFAT5* | *NFAT5* | 0.403 | G | C | 0.425 | 0.034 | 0.006 | 8.96E-10 |
| rs216199 | 17 | 2200871 | *SMG6* | *SMG6* | 3.442 | C | T | 0.388 | -0.037 | 0.006 | 1.11E-09 |
| rs2013002 | 12 | 112200150 | ***ALDH2*** | *ATXN2* | 4.336 | C | T | 0.415 | 0.033 | 0.006 | 5.68E-09 |
| rs17163345 | 1 | 222806218 | *MIA3* | *MIA3* | 7.541 | G | A | 0.270 | -0.034 | 0.006 | 2.15E-08 |
| rs3764351 | 17 | 37824339 | *PNMT* | *MED1* | 5.182 | G | A | 0.340 | -0.033 | 0.006 | 2.27E-08 |

Abbreviations: CADD, Combined Annotation Dependent Depletion; NEA, non-effect allele; EA, effect allele; MAF, minor allele frequency; SE, standard error.

**Table 2. Protein-hits for heart failure identified through Mendelian randomization that passed an FDR threshold of 5%.** Genes that passed a colocalization threshold of *PP.H4>0.5* are highlighted in bold.

| Protein Gene Name | Number of SNPs | Odds Ratio | 95% CI | P-value | pHet† | FDR | Coloc PP.H4* | Druggability classification (Chemical Modality) | Near known HF gene** |
|---|---|---|---|---|---|---|---|---|---|
| ***ITIH4*** | 2 | 1.13 | (1.07, 1.17) | 2.51E-07 | 0.66 | 3.90E-04 | 0.97 | Non-druggable | No |
| ***APOC3*** | 1 | 1.19 | (1.11, 1.28) | 1.74E-06 | NA | 1.35E-03 | 0.99 | Advanced Clinical Phase (Oligonucleotide) | No |
| ***MAPK3*** | 3 | 0.95 | (0.93, 0.97) | 6.70E-06 | 0.41 | 3.48E-03 | 0.52 | Phase (Small molecule) | No |
| ***TNFSF12*** | 2 | 0.96 | (0.94, 0.98) | 1.78E-05 | 0.05 | 6.94E-03 | 0.79 | Clinical Phase 1 (Antibody) | No |
| *TNFSF12* | 2 | 1.02 | (1.01, 1.03) | 2.89E-05 | 0.11 | 8.99E-03 | 0.01 | Non-druggable | *ABO* |
| ***APOH*** | 2 | 0.96 | (0.94, 0.98) | 5.24E-05 | 0.83 | 1.36E-02 | 0.89 | Non-druggable | No |
| *B3GNT8* | 2 | 0.97 | (0.96, 0.99) | 9.35E-05 | 0.96 | 2.08E-02 | 0.48 | Non-druggable | No |
| *NTN4* | 2 | 1.08 | (1.04, 1.13) | 1.10E-04 | 0.68 | 2.14E-02 | 0.04 | Non-druggable | No |
| ***DLL1*** | 1 | 0.87 | (0.8, 0.93) | 1.53E-04 | NA | 2.65E-02 | 0.75 | Non-druggable | No |
| *MST1* | 3 | 1.02 | (1.01, 1.03) | 1.99E-04 | 0.11 | 3.10E-02 | 0.37 | Non-druggable | No |
| ***ENPEP*** | 4 | 0.96 | (0.94, 0.98) | 3.12E-04 | 0.18 | 4.27E-02 | 0.74 | Non-druggable | *PITX2, FAM241A* |
| ***NAE1*** | 1 | 0.82 | (0.74, 0.91) | 3.55E-04 | NA | 4.27E-02 | 0.6 | Phase (Small molecule) | No |
| ***TNXB*** | 1 | 1.03 | (1.02, 1.05) | 3.56E-04 | NA | 4.27E-02 | 0.61 | Non-druggable | No |
| ***SIRPA*** | 1 | 0.98 | (0.97, 0.99) | 3.94E-04 | NA | 4.39E-02 | 0.56 | Non-druggable | No |
| *EBI3* | 1 | 0.75 | (0.64, 0.89) | 4.44E-04 | NA | 4.61E-02 | 0.01 | Non-druggable | No |
| *IL27* | 1 | 0.75 | (0.64, 0.89) | 4.44E-04 | NA | 4.61E-02 | 0.4 | Non-druggable | No |

*Posterior probability of H4 (one common causal variant) from colocalization of pQTL and GWAS results.
**Previously reported HF GWAS gene for instruments in GWAS loci (within 500KB up or down from each loci).
†MR pHet were measured by Cochran's Q test for heterogeneity.
Abbreviations: MR, Mendelian randomization; FDR, false discovery rate, PP.H4, posterior probability of H4.
Note that an OR>1 indicates an increase in protein corresponding with an increase in HF risk or vice versa, suggesting that the therapeutic solution may be an inhibitor; an OR<1 indicates either a decrease in protein levels corresponding with an increase in HF risk or an increase in protein levels corresponding with a decrease in HF risk, suggesting the therapeutic solution may be an agonist.

**Figure 1. Schematic diagram of the datasets and analyses.**

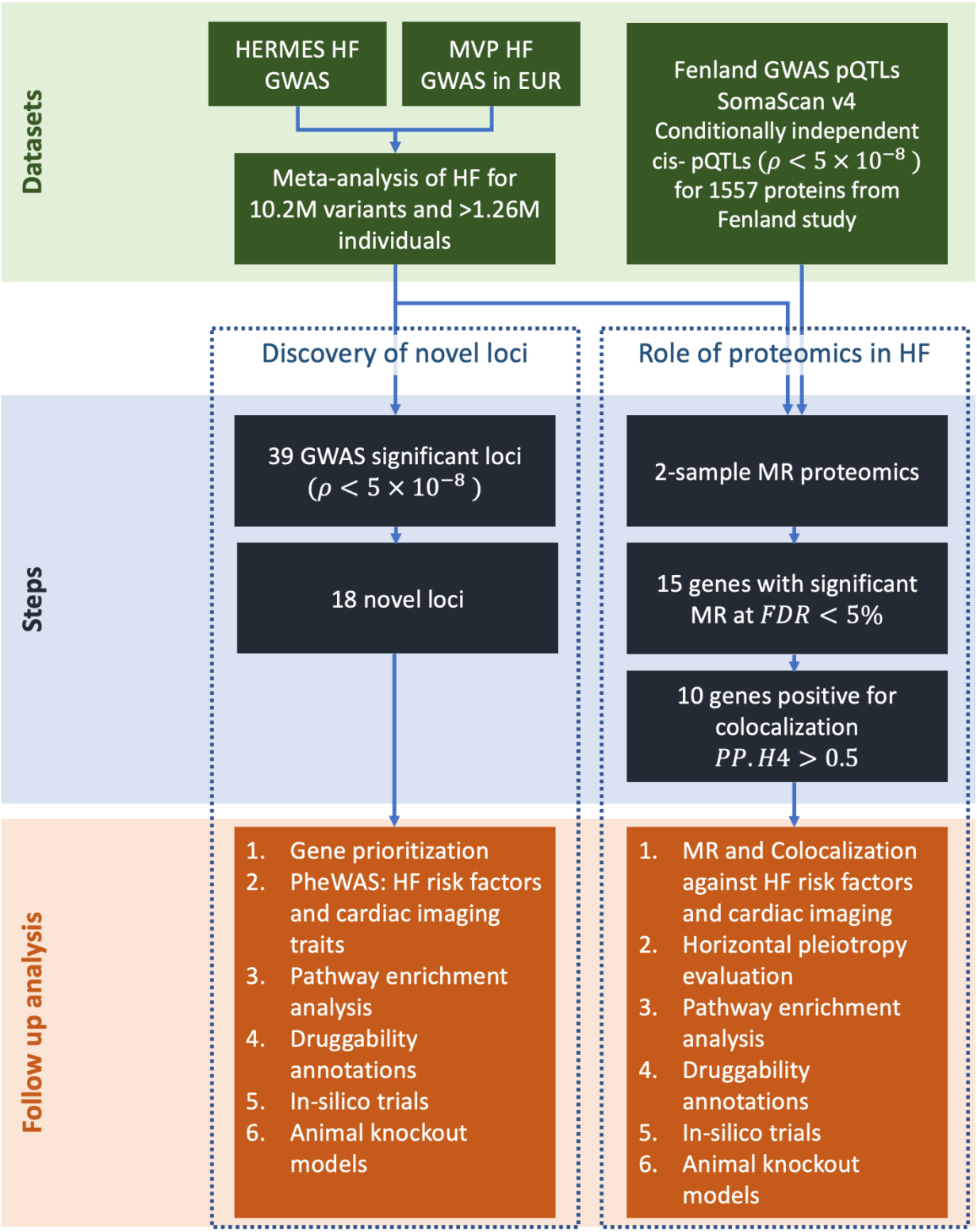
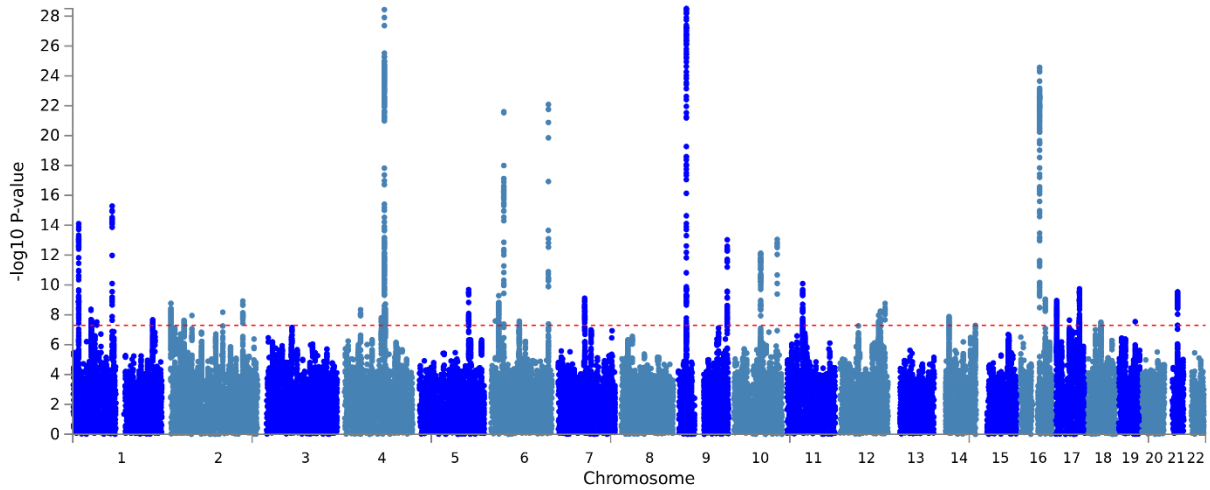


**Figure 2. Manhattan plots showing associations with HF from (a) GWAS meta-analysis on n= 1,266,315 individuals and (b) MR-wide proteomics. a.** Manhattan plot showing the -log$_{10}$(P-value) of association for each SNP from the GWAS meta-analysis plotted on the y-axis against genomic position on the x-axis. The red dotted line corresponds to genome-wide significance threshold ($P < 5 \times 10^{-8}$). The summary statistics of independent lead SNPs are noted in Supplementary Table 1. **b**. Manhattan plot showing the -log$_{10}$-transformed FDR-adjusted P-value of association for each gene plotted against genomic position on the x-axis. The blue line corresponds to an FDR threshold of 5% and points are color coded by drug tractability information based on data provided by OpenTargets; green for druggable genes.

**(a)**




**(b)**

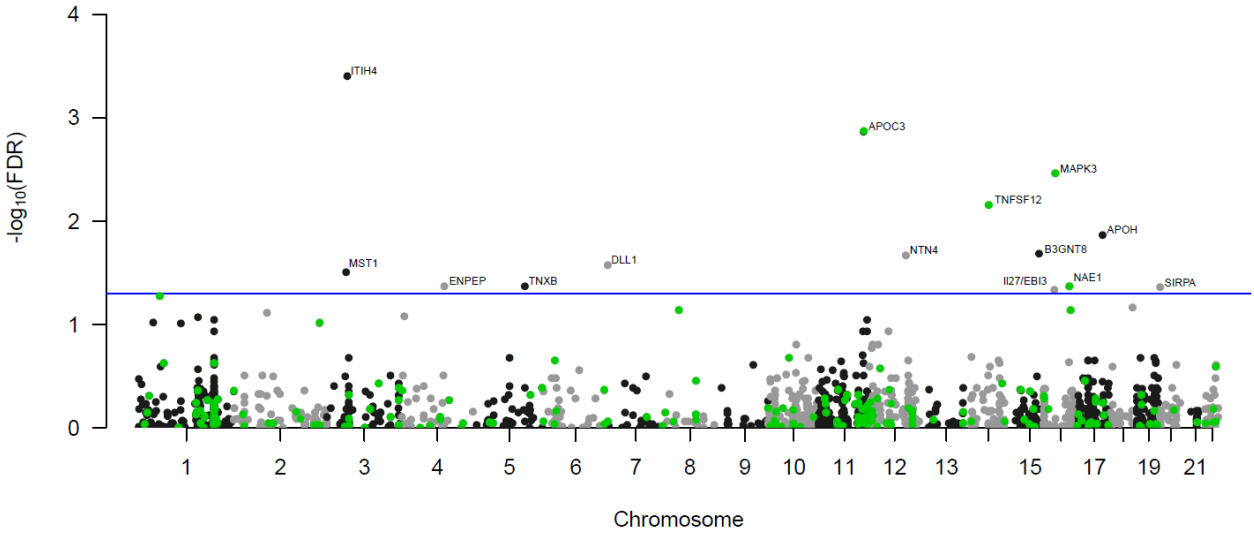


**Figure 3a. Genetic associations of 18 HF loci against risk factors for HF.** The color of the bubble corresponds to the beta coefficient of the genetic association between the loci (x-axis) and trait (y-axis). Blue corresponds to a negative and red corresponds to a positive beta coefficient. The size of each bubble corresponds to the negative logarithm of the association p-value; larger size corresponds to lower p-values. Loci are grouped by druggable and non-druggable genes. Associations with passed the p-value threshold ($p < 1 \times 10^{-4}$) are denoted by a yellow diamond. AC: alcohol consumption, AF: atrial fibrillation, BMI: body mass index, CAD: coronary artery disease, COPD: chronic obstructive pulmonary disease, DBP: diastolic blood pressure, eGFR: estimated glomerular filtration rate, HDL-C: high-density lipoprotein cholesterol, LDL-C: low-density lipoprotein cholesterol, SBP: systolic blood pressure, SMK: smoking, T2D: type 2 diabetes.

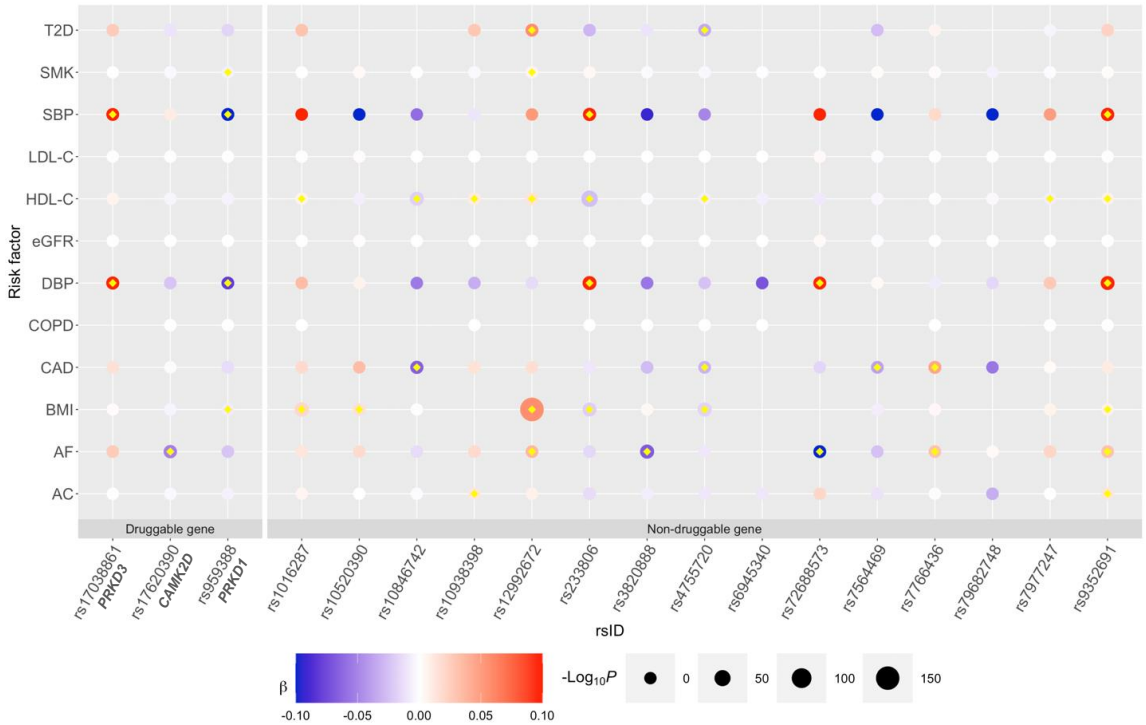
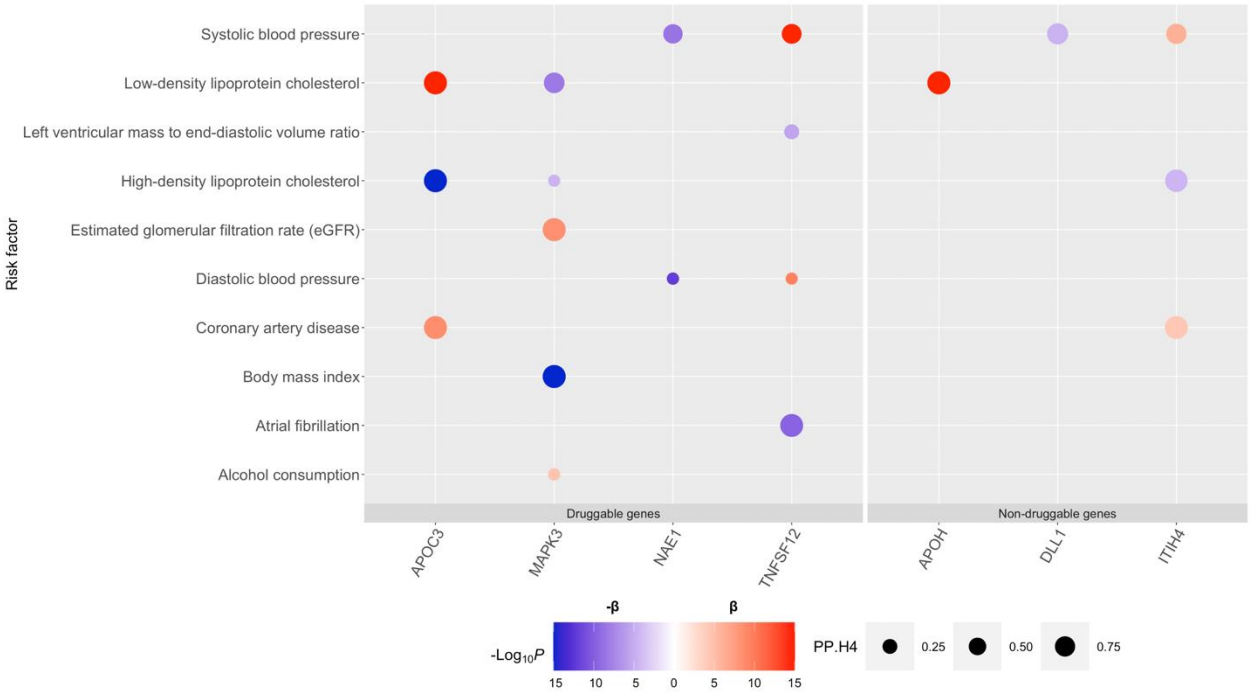


**Figure 3b. MR and colocalization estimates of MR-proteomic genes-hits against HF risk factors.** This bubble plot shows MR estimates for which $p < 1 \times 10^{-4}$. The size of each bubble corresponds to the posterior probability for hypothesis 4 derived from colocalization. The color of the bubble corresponds to the beta coefficient derived from MR. Blue corresponds to a negative association and red corresponds to a positive association; note that a positive $\beta$ indicates either an increase in protein levels corresponding to an increase in HF risk or a decrease in protein levels corresponding to a decrease in HF risk, while a negative $\beta$ indicates either a decrease in protein levels corresponding to an increase in HF risk or an increase in protein levels corresponding to a decrease in HF risk. The intensity of the color corresponds to $-\log_{10}$(p-value) for the strength of association in the MR. Loci are grouped by druggable and non-druggable genes. *TNXB*, *SIRPA* and *ENPEP* genes are not included as these had no MR estimates on HF risk factors that pass the $p < 1 \times 10^{-4}$ threshold.



## ACKNOWLEDGEMENTS

We are grateful to all the MVP investigators; a list of MVP investigators can be found in Supplementary Materials. This research is supported by funding from the Department of Veterans Affairs Office of Research and Development, Million Veteran Program Grant I01-CX001737 (PI: Phillips) and I01-BX004821 (PI: Wilson/Cho). This publication does not represent the views of the Department of Veterans Affairs or the United States Government. We also acknowledge the VA Merit Grant I01-CX001025 (PI: Wilson/Cho).

The Fenland study was approved by the National Health Service (NHS) Health Research Authority Research Ethics Committee (NRES Committee – East of England Cambridge Central, ref. 04/Q0108/19), and all participants provided written informed consent. We are grateful to all Fenland volunteers and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. The Fenland Study (10.22025/2017.10.101.00001) is funded by the Medical Research Council (MC_UU_12015/1). We further acknowledge support for genomics from the Medical Research Council (MC_PC_13046). Proteomic measurements were supported and governed by a collaboration agreement between the University of Cambridge and SomaLogic.

PBM and SEP acknowledge support from National Institute for Health Research (NIHR) Biomedical Research Centre at Barts. NA acknowledges support from the NIHR Integrated Academic Training programme which supports his Academic Clinical Lectureship post. CG has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754490 – MINDED project.

## REFERENCES

1. Roth, G. A. *et al.* Global Burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 Study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
2. Roger, V. L. Epidemiology of heart failure: A contemporary perspective. *Circ. Res.* **128**, 1421–1434 (2021).
3. Blood Pressure Lowering Treatment Trialists' Collaboration. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. *Lancet* **397**, 1625–1636 (2021).

4. Nissen, S. E. *et al.* Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *N. Engl. J. Med.* **352**, 29–38 (2005).
5. Dunlay, S. M., Weston, S. A., Jacobsen, S. J. & Roger, V. L. Risk factors for heart failure: a population-based case-control study. *Am. J. Med.* **122**, 1023–1028 (2009).
6. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33**, 30–42 (2004).
7. Joseph, J. *et al.* Genetic architecture of heart failure with preserved versus reduced ejection fraction. *bioRxiv* 2021.12.01.21266829 (2021) doi:10.1101/2021.12.01.21266829.
8. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
9. Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
10. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
11. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
12. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
13. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *bioRxiv* (2020) doi:10.1101/2020.09.08.20190561.
14. Karczewski, K. *et al.* Systematic single-variant and gene-based association testing of 3,700 phenotypes in 281,850 UK Biobank exomes. *bioRxiv* (2021) doi:10.1101/2021.06.19.21259117.
15. Wu, K.-H. H. *et al.* Polygenic risk score from a multi-ancestry GWAS uncovers susceptibility of heart failure. *bioRxiv* 2021.12.06.21267389 (2021) doi:10.1101/2021.12.06.21267389.

16. Verweij, N. *et al.* The genetic makeup of the electrocardiogram. *Cell Syst.* **11**, 229-238.e5 (2020).
17. Karlsson, T. *et al.* Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. *Nat. Med.* **25**, 1390–1395 (2019).
18. Hoffmann, T. J. *et al.* A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics* **210**, 499–515 (2018).
19. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
20. Kehat, I. *et al.* Extracellular signal-regulated kinases 1 and 2 regulate the balance between eccentric and concentric cardiac growth. *Circ. Res.* **108**, 176–183 (2011).
21. Fielitz, J. *et al.* Requirement of protein kinase D1 for pathological cardiac remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3059–3063 (2008).
22. Swaminathan, P. D., Purohit, A., Hund, T. J. & Anderson, M. E. Calmodulin-dependent protein kinase II: linking heart failure and arrhythmias. *Circ. Res.* **110**, 1661–1677 (2012).
23. Beauverger, P. *et al.* Reversion of cardiac dysfunction by a novel orally available calcium/calmodulin-dependent protein kinase II inhibitor, RA306, in a genetic model of dilated cardiomyopathy. *Cardiovasc. Res.* **116**, 329–338 (2020).
24. Witztum, J. L. *et al.* Volanesorsen and triglyceride levels in familial chylomicronemia syndrome. *N. Engl. J. Med.* **381**, 531–542 (2019).
25. Esan, O. & Wierzbicki, A. S. Volanesorsen in the treatment of familial chylomicronemia syndrome or hypertriglyceridaemia: Design, development and place in therapy. *Drug Des. Devel. Ther.* **14**, 2623–2636 (2020).
26. Chong, M. *et al.* Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome. *Circulation* **140**, 819–830 (2019).

27. Jain, M. *et al.* A novel role for tumor necrosis factor-like weak inducer of apoptosis (TWEAK) in the development of cardiac dysfunction and failure. *Circulation* **119**, 2058–2068 (2009).
28. Gallagher, M. D. & Chen-Plotkin, A. S. The post-GWAS era: From association to function. *Am. J. Hum. Genet.* **102**, 717–730 (2018).
29. Stanzick, K. J. *et al.* Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat. Commun.* **12**, 4350 (2021).
30. Votava, J. A. & Parks, B. W. Cross-species data integration to prioritize causal genes in lipid metabolism. *Curr. Opin. Lipidol.* **32**, 141–146 (2021).
31. Uijl, A. *et al.* Risk factors for incident heart failure in age- and sex-specific strata: a population-based cohort using linked electronic health records. *Eur. J. Heart Fail.* **21**, 1197–1206 (2019).
32. Emerging Risk Factors Collaboration *et al.* Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993–2000 (2009).
33. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *bioRxiv* 2021.11.19.21266436 (2021) doi:10.1101/2021.11.19.21266436.
34. Levin, M. G. *et al.* Multi-ancestry multivariate genome-wide analysis highlights the role of common genetic variation in cardiac structure, function, and heart failure-related traits. *bioRxiv* 2021.08.03.21261508 (2021) doi:10.1101/2021.08.03.21261508.
35. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
36. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
37. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

38. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
39. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
40. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
41. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
42. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
43. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
44. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
45. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, 2254 (2020).
46. Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **20**, 65 (2018).
47. Aung, N. *et al.* Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation* **140**, 1318–1330 (2019).
48. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

49. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).

50. Giambartolomei, C. *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).

51. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

52. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet.* **16**, e1008720 (2020).

53. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

54. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).

55. Xie, Z. *et al.* Gene set knowledge discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).

56. Fahey, M. E. *et al.* GPS-Prot: a web-based visualization platform for integrating host-pathogen interaction data. *BMC Bioinformatics* **12**, 298 (2011).

57. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (N Y)* **2**, 100141 (2021).

58. Shaw, D. R. Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. *Curr. Protoc. Bioinformatics* **56**, 1.7.1-1.7.16 (2016).

59. Brown, K. K. *et al.* Approaches to target tractability assessment – a practical perspective. *Medchemcomm* **9**, 606–613 (2018).

60. Schneider, M. *et al.* The PROTACtable genome. *Nat. Rev. Drug Discov.* **20**, 789–797 (2021).