Druggable proteins influencing cardiac structure and function: implications for heart failure therapies and cancer related cardiotoxicity

Amand Schmidt (amand.schmidt@ucl.ac.uk)
University College London https://orcid.org/0000-0003-1327-0424
Mimount Bourfiss
UMCU
Abdulrahman Alasiri
UMCU
Sandesh Chopade
University College London, UK; UCL British Heart Foundation Research Accelerator, UK
Esther Puyol-Anton
King's
Marion van Vugt
UMCU https://orcid.org/0000-0002-6634-1989
Sander van der Laan
University Medical Centre Utrecht https://orcid.org/0000-0001-6888-1404
Anneline ter Riele
UMCU
Birgitta Velthuis
UMCU
Nora Franceschini
University of North Carolina at Chapel Hill
Bram Ruijsink
King's College London
Aroon Hingorani
University College London https://orcid.org/0000-0001-8365-0081
Pim van der Harst
UMC Utrecht
Christian Gross
UMCU
Chris Clarkson
UCL
Albert Henry
University College London https://orcid.org/0000-0001-7422-2288
R Lumbers
University College London https://orcid.org/0000-0002-9077-4741
Joshua Bis
Cardiovascular Health Research Unit https://orcid.org/0000-0002-3409-1110
Folkert Asselbergs
University Medical Center Utrecht https://orcid.org/0000-0002-1692-8669
Jessica van Setten
UMCU
Chris Finan
University College London https://orcid.org/0000-0002-3319-1937

Article

Keywords:

Posted Date: January 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1305500/v1

Abstract

Dysfunction of either the right or left ventricle can lead to heart failure (HF) and subsequent morbidity and increased mortality. We performed a genome-wide association study (GWAS) of 16 measurements of biventricular function and structure obtained from cardiac magnetic resonance (CMR). We then used aggregated data from three independent plasma proteome GWAS, and performed *cis*-Mendelian randomization (MR) to identify proteins with a likely causal effect on biventricular traits. The subset of proteins with a robust CMR effect were prioritized through linkage with mRNA expression from the Human Protein Atlas, protein interaction data from IntAct, drug compound information from the British national formulary and ChEMBL, and by identifying plasma proteins with robust effects on HF, atrial fibrillation, non-ischemic cardiomyopathy, dilated cardiomyopathy (DCM), or coronary heart disease. In total, 33 plasma proteins were prioritised, including 25 proteins that were druggable by either an approved or developmental compound. Fifteen proteins could be mapped to compounds with a known cardiovascular indication or side-effect, including repurposing candidates with a causal effect on DCM and/or HF: IL18, IL18R, I17RA, GPC5, LAMC2, PA2GA, CD33, and SLAF7. We additionally found that 13 of the 25 druggable proteins (52%, 95%CI 0.31; 0.72) could be mapped to a compound with an oncological indication or side-effect. To further inform drug development, we performed a drug-target MR phenome-wide scan of these 33 prioritized proteins on 56 clinically relevant traits. We have identified a prioritized set of plasma proteins influencing biventricular traits, providing indispensable leads to facilitate drug development and drug repurposing for cardiac diseases, and explaining observed cardiotoxicities of several targets exploited for the treatment of cancer.

Introduction

Dysfunction of the right or left ventricle, arising due to intrinsic heart muscle or coronary artery disease, or pulmonary or systemic hypertension, leads to the clinical syndrome of heart failure (HF).[1] Heart failure can be accompanied by ventricular hypertrophy or dilation (depending on the cause) and with impairment of either cardiac contraction or relaxation, leading to heart failure syndromes defined according to impaired or preserved ejection fraction.

Despite recent advances offered by drugs such SGLT2 inhibitors for treatment of HF, drug development for cardiac disease has been met with high failure rates, often occurring during costly late-stage clinical testing[2–4]. These late-stage failures are indicative of the poor predictive potential of pre-clinical experiments for cardiac target identification. This is complicated further by the considerable phenotypic heterogeneity that underlies diagnoses such as HF[5], resulting in compounds failing for futility that may genuinely benefit a subset of patients. Conversely, several drugs – predominantly for oncological indications, have been found to cause cardiotoxicity, which may confront patients with treatment induced heart problems[6].

Cardiac magnetic resonance (CMR) imaging is the gold standard for quantification of biventricular function and morphology, and has become an integral diagnostic modality for cardiac diseases; see Supplementary Table 1. Here we utilized CMR images, available from the UK biobank (UKB), to extract measures from both left and right ventricle (LV, RV) using a purpose built highly accurate, fast, deep-learning algorithm[7].

Proteins constitute the majority of drug targets[8], which are increasingly analysed through high throughput assays measuring the levels of hundreds to thousands of (plasma) proteins[9]. To leverage proteins and CMR measurements for drug target validation, we have developed an analytical framework[10] to perform drug target analyses using human genetic data. Specifically, through two-sample drug target Mendelian randomization (MR), we can anticipate the on-target effect a drug target protein will have on disease relevant traits such as CMR measurements. Previously, this approach has been extensively validated for cardiovascular drug targets[11–19].

To prioritize circulating plasma proteins for their involvement with LV and RV traits, relevant for cardiac disease, we first performed a genome-wide association study (GWAS) on sixteen CMR traits measured in up to 36,548 UKB subjects. Subsequently, we format-normalized protein quantitative trait loci (pQTLs) data, sourced from three independent GWAS involving cross-platform measurement of plasma protein concentrations using Somalogic[9], Olink[20] and Luminex[21] assays spanning 3000+ plasma proteins. Drug target MR was used to prioritize proteins on their likely causal contribution to CMR traits. Repurposing opportunities were identified by extracting cardiovascular indications and side-effects from ChEMBL[22] and the British National Formulary (BNF) for drug targets with licensed compounds. Results were further annotated with tissue specific mRNA expression data from the Human Protein Atlas (HPA) database[23], and with information from IntAct[24] and Reactome[25] to identify druggable protein-protein interaction.

Methods

Quantification of LV and RV CMR traits

The current study sourced information from 36,548 UKB subjects who had data on both CMR images and genotyping. To minimize influence of pre-existing conditions, we excluded subjects with prevalent diseases (e.g. myocardial infarction, HF, and congenital heart diseases) known to affect the LV or RV traits; see Supplementary Table 2.

The deep-learning methodology (AI-CMR^{QC}) to extract LV and RV CMR measurements has been previously described, and extensively validated[7]. Briefly, the fully automated and quality-controlled cardiac analysis tool calculates LV and RV traits from cine short axis and 2- and 4-chamber acquisitions. This resulted in structural measures on end-diastolic, end-systolic, or stroke volumes (EDV, ESV, SV), end-diastolic mass (EDM), and, LV mass to EDV ratio (LV-MVR), or functional measures such as ejection fraction (EF), peak ejection rate (PER), peak (atrial) filling rates (PAFR, PFR); Supplementary Figure 1 and Supplementary Table 1. Automatic quality control steps consisted of pre-analysis checks on image-quality (e.g., motion artefacts, erroneous image plane planning), and post-analysis checks on accuracy of the image-analysis (e.g., coverage of the segmentations, detected abnormalities in volume and discrepancies between LV and RV parameters); with automatic detection and removal of outlying observations.

GWAS of CMR traits

We used genotyped and imputed data as provided by UKB[26] (GRCh37 assembly). In brief, samples were genotyped on the Affymetrix BiLEVE and Axiom arrays, with untyped variants imputed using the Haplotype Reference Consortium, 1000 Genomes, and UK10K as reference panels. We excluded samples as recommended by UKB[26], and in addition used the following sample exclusion criteria: discordant self-reported and genetically inferred sex, and genotypical missingness rate above 0.01. Variant quality control included removing variants with minor allele frequency (MAF) below 0.1%, imputation quality below 0.3, and deviation from the Hardy–Weinberg equilibrium (HWE p-value<1×10^{-6}).

Genetic associations with the 16 CMR traits were estimated using BOLT-LMM[27], utilizing a mixed-effects model to account for possible cryptic relatedness and population stratification. The BOLT-LMM models were run using default setting and conditional on age at CMR, sex, body surface area, systolic blood pressure (SBP), genotype measurement batch, 40 principal components (PCs), and assessment centre.

Genetic heritability of CMR variability

BOLT-REML[27] (with default settings, excluding variants with a MAF below 0.1%, HWE P<1×10^{-6} and over 1% missingness) was used to estimate narrow-sense genetic heritability (i.e., the proportion of phenotypic variance explained by common variants), as well as the pairwise genetic correlation between the CMR traits.

Functional and phenotypic annotations, and identification of likely causal loci

Lead variants were identified through LD-clumping (LD: linkage disequilibrium) within a one megabase flanking region, applying a pairwise r-squared threshold of 0.001.

Putative causal genes were identified through manual curation of purpose built locus-view plots, where AFS, MB, JvS, and CF independently determined the most likely causal genes (Supplementary File 2). Locus-view plots combined variant specific CMR associations around each respective lead variant (+/- 250 kbp flanking region) with information on regional genes and their exon structure. These plots were enhanced with an incidence (i.e., boolean) matrix annotating genes on 23 criteria, including whether the gene was coding, encoded a target for drug compounds with known cardio-metabolic (side-)effects, had a *cis*-MR CMR association, previous associations with cardio-metabolic traits sourced from GWAS catalog[28], the presence of mRNA expression or splice-sites in cardiac or vascular tissues from GTEx, *trans* protein associations with other CMR loci, protein-protein interactions between CMR associated proteins; see Supplementary File 2 for a detailed exposition.

Format normalization of cross-platform protein quantitative trait loci

Genetic association with plasma protein concentration were available from the following sources: Somalogic measurements on 3,301 participants of the INTERVAL cohort[9], Luminex assays on 6,861 Framingham participants[21], and OLINK assays on 30,931 individuals from the SCALLOP consortium[20]. Framingham provided pQTLs from GWAS of common variants, as well as from an exome GWAS, which were concatenated here. For the less than 1% variant overlap between the two Framingham arrays we selected results with the smallest standard error; representing the highest degree of precision.

The GWAS files were normalized using a purpose built normalization pipeline (https://bit.ly/3pxBKXU), standardizing file structures, mapping variants against the same genome assembly, assigning UniProt identifiers, and providing annotations with Variant Effect Predictor (VEP), Polymorphism Phenotyping v2 (PolyPhen), and Combined Annotation Dependent Depletion (CADD).

Mendelian randomization of plasma protein effects on CMR traits

MR was subsequently employed to ascertain the likely causal consequences of protein concentration on the 16 CMR traits. To prevent potential influence of study specific factors, all drug target MR were conducted per contributing study/consortium, performing separate analyses for SCALLOP, Framingham and INTERVAL. Specifically, drug target MR was conducted by selecting variants from a 100 kbp windows around the *cis* gene known to encode the protein, clumping variants to an LD of 0.40, where residual LD was modelled using a generalized least square (GLS) model[29] and a 5,000 random sample of UKB participants. To reduce the risk of "weak-instrument bias"[30], we selected genetic variants with an F-statistic of 15 or higher, furthermore due to the absence of sample overlap between protein concentration GWAS and CMR GWAS any *potential* weak-instrument bias would act towards a null effect, reducing power rather than increasing type 1 errors.

MR analyses were conducted using the GLS implementation of the inverse-variance weighted (IVW) estimator, as well as with an Egger correction protecting against horizontal pleiotropy[31]. To minimize the potential influence of horizontal pleiotropy, we excluded variants with large leverage or outlier statistics, and used the Q-statistic to identify possible remaining violations[32]. Finally, a model selection framework was applied to select the most appropriate estimator (IVW or MR-Egger) for each specific protein effect estimate[32,33].

Protein prioritization

After accounting for multiplicity (see below), we identified druggable proteins with a CMR association, and through linkage with ChEMBL and BNF extracted cardiovascular related indications and side effects (see Supplemental note 1). The BNF draws information from drug medication inserts, scientific literature, regulatory authorities, and professional bodies, and is jointly authored by the British Medical Association and the Royal Pharmaceutical Society. ChEMBL[22] was extracted for information in clinically used drug targets (from FDA approved drugs) and information on drug targets that are in early phase consideration. ChEMBL was also used to identify proteins that are *potentially* druggable as described by Finan *et al*[8].

In addition to prioritizing drugged and druggable proteins with a CMR association, we identified a subset of proteins with a concordant risk increasing or risk decreasing effect. Specifically, results were coded towards the cardiac function or structure improving direction by multiplying estimates for EDV, ESV, EDM, and MVR by -1, and retaining the original effect direction for the remaining traits. Inferentially, this would mean that we can evaluate protein level increases according to potential beneficial cardiac effect rather than simply increase in CMR trait values. A concordant set of prioritized proteins was identified by selecting proteins with at least three CMR associations passing multiple testing correction, that were either all in the beneficial positive direction or the detrimental negative direction (i.e., without directionally discordant results). The above classification of beneficial vs harmful CMR effect direction is of course imperfect and simplifies the more complex relationship observed in observational studies. For example, LV-EF has been shown to have a u-shaped association with mortality[34]. As such these heuristic orientation are used here as a first filtering step, followed by more direct ascertainment on clinical cardiac outcomes through a phenom-wide scan (see below).

These prioritized proteins included drugged and druggable targets, as well as proteins that may not yet be druggable. To determine the potential pharmacological relevance of these proteins we subsequently identified the distance to the nearest (additional) druggable protein based on the in IntAct[24] protein-protein interaction database as modelled in Reactome[25] (accessed April 2021). Here distance reflected the number of protein-protein interactions between the “index” protein and the next druggable protein, where a distance of 1 represents a direct link.

Drug target phenome-wide scan to anticipate effects of prioritized targets

The CMR prioritized set of drugged, druggable, concordant CMR effect, and nearest druggable proteins were further pruned on an association with HF, non-ischemic cardiomyopathy (CM), dilated cardiomyopathy (DCM), atrial fibrillation (AF), and/or coronary heart disease (CHD), using the drug target MR pipeline described before. Next, for CMR prioritized proteins with a cardiac trait association we evaluated their effects on 56 clinically relevant traits, combining drug target MR with a phenome-wide scan to further inform potential on-target protein effects in future drug development programs.

Assessing tissue specificity of prioritized targets

The set of prioritized CMR-associated plasma proteins with cardiac effects was annotated by exploring their tissue specific mRNA expression from the Human Protein Atlas (HPA)[23]. Sourcing the consensus expression obtained by normalizing TPM (transcripts per million) values from three independent transcriptomics datasets: GTEx[35], Fantom5[36], and HPA’s own RNAseq experiments[23].

The normalized human expression data were used to determine a proteins tissue specificity[37], ranging from 0 (ubiquitous expression across all tissues) to 1 (tissue-specific expression). Differentially overexpressed tissues were identified by comparing tissue specific expression against average expression, testing against a standard normal quantile of 1.96.

Quality control and multiple testing

LD score regression[38] was used to explore the possibility of any remaining bias due to population stratification or cryptic relatedness – finding no cause for concern (Supplementary Table 3). Genetic loci were identified using the traditional genome-wide threshold of 5.00×10^{-8}, and a conservative threshold of 7.14×10^{-9}. The latter simply accounts for multiplicity by performing a Bonferroni correction based on the seven PCs necessary to explain over 90% of the CMR trait variance (Supplementary Figure 2).

Based on the described instrument selection criteria we had sufficient genetic variants to robustly assess 892 unique proteins. Accounting for the same seven PCs described above and the number of proteins, the MR effect estimates with the CMR traits were evaluated using an alpha of 7.81×10^{-6}. The phenome-wide scan drug target analysis of CMR prioritized plasma proteins were evaluated using a multiplicity corrected alpha of 1.24×10^{-5}. Under the null-hypothesis the p-values of a group of tests follow a uniform distribution between zero and one[41]. Hence, to additionally explore the potential impact of multiple testing, we performed CMR-trait specific “overall” null-hypothesis tests, comparing the empirical p-value distribution (using Kolmogorov-Smirnov “KS”-tests) against the uniform distribution expected under the null-hypothesis[41].

Unless otherwise specified, any remaining hypothesis tests were evaluated using an alpha of 0.05, and all point estimates (odds ratios [OR] or mean differences) refer to a unit change of the independent variable; typically, one standard deviation in plasma protein level (MR results) or an increase in risk allele (GWAS results). To better illustrate concordance, and only where specified, MR results were orientated towards the cardiac beneficial effect direction by multiplying EDV, ESV, EDM, and MVR MR estimates by -1.

Results

UK biobank participants with LV and RV CMR measurements

CMR measurements were obtained from a sample of 36,548 UKB subjects utilizing an extensively validated deep-learning approach[7]. On average, subjects were 63.9 (standard deviation, SD: 7.6) years old, 18,879 (51.8%) were women. Participants had a mean SBP of 138.2 mmHg (SD: 18.4), a mean diastolic blood pressure (DBP) of 78.6 mmHg (SD: 10.0), and a mean heart rate of 62.5 bpm (SD: 10.2); see Supplementary Table 4.

Genomic loci associated with measures of RV and LV structure and function

We performed GWAS on 16 CMR traits, leveraging genotyped and imputed variants from the Affymetrix BiLEVE and Axiom arrays, and applying BOLT-LMM conditional on age, sex, body surface area, SBP, genotype measurement batch, 40 PCs and assessment centre.

The 91 unique lead variants (Figure 1-2, Supplementary Tables 5) were mapped to 53 genes based on independent review of annotated local-view plots (Supplementary file 2). This resulted in 16 genes for RV-ESV, 15 for LV-EF, 14 for LV-MVR, 14 for LV-ESV, 12 for RV-EF, 10 for RV-EDV, 6 for LV-EDV end LV-EDM, 5 for RV-SV, 4 for LV-SV, 2 for RV-PER, 1 for RV-PAFR, and none for RV-PFR, LV-PFR, LV-PER, and LV-PAFR. We identified five novel locus-trait associations for RV-PAFR (*SCN10A*), RV-PER (*ALDH2* and *HLA-B*), and LV-SV and LV-MVR (both *HLA-B)*. Twenty-six genes were associated with multiple measures. For example variants mapped to *TTN* were associated with 11, *BAG3* with 6, and both *TMEM43* and *ATXN2* with 5 measures. Of these multi-trait genes, 12 were associated with both LV and RV measures: *TTN, BAG3, TMEM43, ATXN2, PROB1, DMPK, ZNF572, PLEC, HSPB7, HLA-B, SPON1, and OBSCN* (Figure 1 and Supplementary Table 4).

Genetic heritability of CMR traits and pairwise genetic correlation

BOLT-REML was used to estimate the amount of phenotypic variation that could be explained by narrow-sense genetic heritability (Figure 2). Heritability estimates ranged between 36% and 31% for both RV and LV measurements of EDV and ESV, as well as LV-EDM. For LV-MVR, EF and SV of both ventricles heritability ranged between 20% and 29%. Despite an absence of GWAS hits for PFR, LV-PER and LV-PAFR, heritability of these traits was between 6% and 12%.

The pairwise genetic correlation (Figure 3) indicated that genetic variants for SV and PER measurements (both LV and RV) were highly correlated (correlation coefficient close to 1.0), as were genetic variants associated with EDV and ESV traits from both ventricles, and variants for LV-PFR and RV-PFR. LV-EDM had a moderately strong correlation (around 0.70) with SV, PER, ESV, EDV of both ventricles. Finally, variants for LV-MVR, RV-EF, and LV-EF showed a positive correlation among themselves (maximum 0.68), and negative correlation with EDV, ESV, EDM, and SV traits (maximum -0.86).

Broader phenotypic effects of the CMR genes

Extracting data from GWAS catalogue (Figure 2, Supplementary Figure 3), we found genes identified for their association with one or more RV and LV measures were frequently associated with CMR traits from previous studies (e.g., LV dilatation, LV mass, and fractal dimension), with electrocardiographic traits (e.g., PR segmentation, QRS duration, QT interval), blood pressure and heart rate; as well as with plasma concentration of various apolipoproteins and cholesterol-containing lipoproteins. The following CMR genes were previously associated with a number of cardiac diseases including: AF (*SYNPO2L, TBX5, IGF1R, GOSR2, TTN, SCN10A, CDKN1A, MYO18B, KCNH2)* hypertrophic cardiomyopathy (*HSPB7, SYNPO2L, BAG3, NSF, FHOD3, CDKN1A, SMARCB1*), DCM (*BAG3, FHOD3, TTN, SMARCB1)*, HF (*SYNPO2L, BAG3*); and CHD (*ATXN2, ALDH2, PTPN11, GOSR2*); Supplementary Figure 3.

Identifying plasma protein levels with an effect on CMR traits

We initially linked our putative CMR genes to BNF and ChEMBL and identified 18 genes which encoded a druggable protein (Supplementary Table 4-6). These include genes encoding drug targets for compounds with indications and/or side-effects for AF, HF, CHD, chronic obstructive pulmonary disease (COPD) and diabetes (*IGFR1, KCNH2, KCNK3, PDE5A, SCN10A*, Supplementary Tables 5-6).

We next expanded this analysis to use drug target MR to directly identify casual plasma proteins for CMR traits, which additionally provides effect directions that can inform the type of drug compound effect (i.e, inhibiting or activating compound effects). Specifically, we identified circulating proteins with a causal effect on RV and/or LV measures by performing a two-sample *cis*-MR, combining aggregated genetic data on protein levels from three sources (SCALLOP, Framingham, and INTERVAL) with the GWAS discovery analysis undertaken here (see methods). We found that 304 proteins were associated with at least one CMR trait (Supplementary Figure 4), with the number of associated proteins ranging from 62 for LV-ESV to 33 for LV-PAFR (Figure 4). The Kolmogorov-Smirnov test provided strong evidence that these results were not driven by multiple testing (Supplementary Figure 5).

Next, we identified I) 18 “drugged” CMR associated proteins that were targeted by a licensed drug compound (Supplementary Figure 6), II) 21 “druggable” proteins which were amenable to small-molecule perturbation or monoclonal antibody inhibition (Supplementary Figure 7), and III) 30 proteins with directionally “concordant” effects on three or more CMR traits which either all had beneficial or detrimental effects (Supplementary Figure 8). The set of concordant proteins contained 25 entries which were not part of the drugged or druggable sets, hence through linkage to IntAct data we next identify the “nearest” set of druggable proteins which contained 8 proteins for which we had plasma pQTL data. The nearest druggable were identified by counting the number of protein-protein interactions between the indexing protein and the nearest druggable protein. This resulted in a set of 72 proteins associated with LV and/or RV measurements (18 drugged, 21 druggable, 25 concordant and 8 nearest druggable), which we prioritized further through *cis*-MR to identify 33 proteins which were involved with the following cardiac outcomes: HF, DCM, non-ischemic CM, AF, and/or CHD.

Drugged CMR proteins: repurposing opportunity

Eight of the 18 druggable proteins could be associated with cardiac traits through *cis*-MR (Figure 5): IL6RA, CO6A1, CD33, CAH6, COFA1, TIE2, LAMC2, I17RA, and SLAF7.

CD33, I17RA, SLAF7 affected HF; CO6A1, I17RA affected non-ischemic CM; CAH6, LAMC2 were associated with DCM; IL6RA, CD33, COFA1 with AF, and finally IL6RA and TIE2 affected CHD. Focussing on the proteins affecting multiple cardiac traits, we found that increased levels of I17RA (Interleukin-17 receptor A) predominantly improved LV function (Supplementary Figure 6) and decreased the risk of HF OR 0.97 (95%CI 0.96; 0.98) and non-ischemic CM OR 0.94 (95%CI 0.92; 0.96). I17RA is targeted by the anti-inflammatory monoclonal antibody (mAb) brodalumab (Figure 6). IL6RA (Interleukin-6 receptor subunit alpha) had a directionally discordant effect on 9 LV and RV traits (Supplementary Figure 6), with increased levels being associated with decreased risk of AF OR 0.95 (0.94; 0.96) and CHD OR 0.94 (0.93; 0.94). Noting that genetic instruments for IL6R are associated with reduced membrane bound IL6[11], we find directionally concordant effects by IL6R inhibiting compounds such as tocilizumab decreasing cardiovascular risk (Figure 6). CD33 (Myeloid cell surface antigen CD33) was found to reduce LV-EDM and is targeted by mAb such as gemtuzumab which are indicated in oncology and have documented cardiovascular side effects

(Figure 6). Increased levels of CD33 decreased the risk of HF OR 0.96 (95%CI 0.95; 0.98) and AF OR 0.96 (95%CI 0.89; 1.03). Similarly, SLAF7 (SLAM family member 7) and TIE2 (Angiopoietin-1 receptor) are both inhibited by compounds with an oncological indication with known cardio-metabolic side-effects, and non-oncological indications such as amyloidosis (SLAF7) and CHD (TIE2); Figure 6. Through MR we found that SLAF7 improved RV-EF and RV-PAFR function, but nevertheless increased the risk of HF OR 1.07 (95%CI 1.05; 1.08), TIE2 beneficially affected CMR traits with an LV-EF effect of 0.43% (95%CI 0.32; 0.55), RV-ESV -0.68 ml (95%CI -0.89; -0.48), and an RV-PAFR effect of 5.47 ml/s (95%CI 4.15; 6.79), while increasing the risk of CHD OR 1.10 (95%CI 1.06; 1.15).

Druggable CMR proteins: de novo developmental targets

We identified 11 *druggable* proteins (out of 21 in total) with an effect on a cardiac outcome (Figure 5, Supplementary Figure 7, Supplementary Tables 11-12): TNF12, ICOSL, IL8, TDGF1, LYAM1, PA2GA, TNR5, MK03, MFGM, ERAP2, ERAP1.

PA2GA, MK03 affected HF; TNF12, TDGF1, TNR5, MFGM affected non-ischemic CM; TNF12, ICOSL, TNR5, MFGM, were associated with DCM; TNF12, IL8, TDGF1, LYAM1, MK03, ERAP1 associated with AF, and finally IL8, TDGF1, ERAP2 and ERAP1 with CHD. Focussing on proteins with an effect on multiple cardiac traits, we found that TNF12 (Tumor necrosis factor ligand superfamily member 12) decreased the risk of non-ischemic CM OR 0.82 (95%CI 0.77; 0.88), AF OR 0.90 (95%CI 0.89; 0.91), and DCM OR 0.80 (95%CI 0.75; 0.85). Higher concentration of TNF12 improved LV dimensions but increased LV-EDM. TNF12 is inhibited by two phase 1 compounds indicated for neoplasm and rheumatoid arthritis. Higher levels of IL8 (Interleukin-8) IL8 increased LV-EDM, while improving RV-PER, and decreased the risk of HF OR 0.74 (95%CI 0.69; 0.81) and AF OR 0.83 (95%CI 0.77; 0.89), while increasing the risk of CHD OR 1.18 (95%CI 1.11; 1.25); Figure 5, Supplementary Figure 7 & Tables 11-12). IL8 is the target of mAb in development for treatment of neoplasms and chronic lung disease (Figure 6, Supplementary Table 8). TDGF1 (teratocarcinoma-derived growth factor 1), targeted by a developmental immunoconjugate BIIB015 for treatment of tumours, improved LV and RV cardiac traits (EF, SV, PER, RV-PFR), and decreased the risk of CHD, non-ischemic CM OR 0.93 (95%CI 0.92; 0.94), and increased the risk of AF OR 1.01 (95%CI 1.01; 1.01); Figure 5, Supplementary Figure 7, Tables 11-12. MK03 (Mitogen-activated protein kinase 3) is inhibited by multiple ERK1/2 kinase compounds for treatment of neoplasms and associated with improved RV-ESV and RV-EF, and decreased the risk of HF OR 0.85 (95%CI 0.80; 0.91) and AF OR 0.86 (95%CI 0.82; 0.91). ERAP1 and ERAP2 (Endoplasmic reticulum aminopeptidase 1 and 2, forming a protein complex[39]), both improved LV and RV CMR measurements (Supplementary Figure 6), and are both inhibited by the same compound tosedostat (currently in development for oncology). Higher ERAP1 was associated with an increased risk of non-ischemic CM OR 1.10 (95%CI 1.07; 1.13), and decreased risk of AF OR 0.99 (95%CI 0.98; 0.99) and HF OR 0.98 (95%CI 0.97; 0.98), while higher levels of ERAP2 in turn increased the risk of CHD OR 1.03 (95%CI 1.02; 1.03); Figure 5.

Nearest druggable proteins with directionally concordant CMR effects

Next, we identified 30 proteins with directional concordant effects on three or more CMR traits (i.e., with all beneficial or detrimental effects), and mapped these indexing proteins to their (next) nearest druggable protein (Figure 7, Supplementary Figure 8). This resulted in drugged and druggable proteins that either directly interacted with an indexing protein or were separate by at most one protein-protein interaction (Figure 7). Some of these indirectly drugged and druggable proteins had known cardio-metabolic indications and/or side-effects (Figure 7, Supplementary Figure 9). For example, PPAC (low molecular weight phosphotyrosine protein phosphatase) beneficially affected LV-PFR, RV-EDV, and RV-ESV, and while not druggable itself, interacted with eight druggable proteins (Figure 7, Supplementary Figure 7). Six of these PPAC related proteins (PDE4D, GBRG1, PRS7, VWF, RARA, 5HT1E) were targeted by inhibiting compounds with a recorded cardio-metabolic indication or side effect (Figure 7, Supplementary Figure 9, Supplementary Tables 13-14).

The set of concordant proteins included 25 that were not included in the drugged or druggable set (Supplementary Figure 8), and through identification of the nearest druggable protein we identified an additionally 8 druggable proteins with plasma pQTL data (EGFR, FA10, PAI1, MET, LYAM2, SYUA, IL6RB, IL18) that were not included previously. Pruning this combined set of proteins (i.e., the set of directionally concordant proteins and the indirectly drugged or druggable proteins they interacted with) on the presence of cardiac outcome effects (Figure 5) resulted in the following set of prioritized proteins: BAG3, C1QC, PGLT1 affected HF; BAG3, PATE4, affected non-ischemic CM; MANBA, NCAM2, BAG3, C1QC, GPC5, IL18R, were associated with DCM; LYAM2, PPAC, BGH3 associated with AF, and finally MANBA, UD16, SPA12 affected CHD.

Focussing on proteins with an effect on multiple cardiac traits, we found that higher concentrations of MANBA improved 5 CMR traits (ESV, EF, LV-PFR), and decreased CHD and DCM risk: OR 0.93 (95%CI 0.91; 0.96) and OR 0.76 (95%CI 0.72; 0.81) respectively. BAG3 (BAG family molecular chaperone regulator 3) improved 6 CMR traits and decreased the risk of HF OR 0.75 (95%CI 0.72; 0.79), non-ischemic CM OR 0.30 (95%CI 0.25; 0.36), and DCM OR 0.14 (95%CI 0.11; 0.17). Higher levels of C1QC (Complement C1q subcomponent subunit C) detrimentally affected 3 CMR traits but nevertheless decreased the risk of HF OR 0.97 (95%CI 0.96; 0.98) and DCM OR 0.86 (95%CI 0.82; 0.90). Higher UD16 (UDP-glucuronosyltransferase 1-6) worsened 4 LV CMR traits, and increased the risk of DCM OR 1.62 (95%CI 1.46; 1.80) and CHD OR 1.06 (95%CI 1.04; 1.08)

Tissue expression and phenome-wide scan of likely on-target clinical effects.

We next explored mRNA expression and performed a phenome-wide scan of the anticipated on-target effects of increased protein concentration of the 33 prioritized proteins who affected LV and RV measurements, as well had an effect on cardiac outcomes (Figures 5 & 8, Supplementary Figures 10-11).

Tissue specificity did not differ between CMR prioritized proteins and non-prioritized proteins (p-value = 0.20). We did observe a significant difference in tissue-specific expression (p-value 9.01×10^{-3}), with prioritized plasma proteins more frequently higher expressed in spleen, lymph node, liver, granulocytes, kidney, pancreas, and lung tissues (Supplementary Figure 11).

In addition to the cardiac outcomes these proteins were prioritized on, the *cis*-MR phenome-wide scan showed that these proteins were frequently associated with DBP, SBP, ECG measurement during exercise, lipid fraction such as (HDL-C, Apo-A1, triglycerides, LDL-C, and Apo-B), estimated glomerular filtration rate (eGFR), body mass index (BMI), glycosylated haemoglobin (HbA1c), c-reactive protein, lung function (FEV1, FVC, PEF), and carotid intima-media thickness (cIMT) (Figure 8); protein specific results are presented in Figure 5, Supplementary Figure 10 and Table 16.

Discussion

In the current study we derived 16 traits of left and right ventricular structure and function from CMR imaging, and utilized GWAS to identify 87 genetic variants associated with one or more measurement. We prioritised 51 genes that likely drive the discovered genetic association, 25 of which affected multiple CMR traits, and with 12 affecting both LV and RV traits. Independently, we leveraged drug target MR to identify 33 plasma proteins associating with RV or LV measurements and robust effects on cardiac outcomes, including HF and DCM. To further inform drug development, we conducted a phenome-wide scan assessing the potential on-target effects perturbation of these 33 proteins may have on 56 clinically relevant traits. We found that 15 (60% 95%CI 39; 79) of the 25 drugged or druggable proteins were targeted by compounds with a cardiovascular indication or side effect (Table 1).

While the number of discovered genes and genetic heritability differed considerably across CMR traits (16 genes for RV-ESV, compared to zero for RV-PFR, LV-PER, and LV-PAFR), the genetic contribution was balanced across both ventricles, and variants for LV and RV measurements were often highly correlated, suggesting similar genetic burden between LV and RV traits. The relatively high number of multi-CMR loci that affected both LV and RV (12/51 or 24%) additionally supported a shared genetic background. Principal component analysis of the CMR measurement further found that 7 PCs explain more than 90% of the phenotypic variation, where typically LV and RV of a specific trait contributed to the same PC (Supplementary Figure 2). Similar observations were seen in the *cis*-MR analyses assessing the causal effects of plasma proteins on CMR traits: 19 proteins out of 33 (58%) affected both ventricles; p-value 5.74×10^{-5}, see Table 1).

In the current paper, we used enhanced locus-view plots including extensive annotations (see methods and Supplemental File 2) relevant to identify the putative casual gene. Four raters (AFS, MB, JvS and CF) independently reviewed these data, with discordance resolved through consensus. Compared to previous CMR GWAS[40,41], which reported on the nearest gene instead of the likely causal gene, we found 33 novel genes of which 19 (58%) were linked with cardiac-metabolic traits sourced from GWAS catalog (Supplementary Figure 3). Additionally, our list of genes included 8 drugged proteins with known cardio-metabolic indications or side effects, providing further support for these likely causal genes. Rediscovered CMR genes included variants previously associated with DCM, HCM, AF, and CHD providing orthogonal support for LV and RV involvement; Supplementary Figure 3. We uniquely determined LV and RV PER, PFR and PAFR, where PFR is especially relevant for HF with preserved EF. Through *cis*-MR of plasma pQTL we identified seven proteins that affected PFR as well as HF or DCM risk: UD16, MANBA, TNR5, TNF12, MFGM, CPC5, and BAG3, where the last five proteins were (indirectly) drugged or druggable, providing important leads for drug development; Table 1.

The *cis*-MR analysis leveraged three distinct plasma pQTL resources, distilling a prioritized set of 33 plasma proteins that affect both CMR traits and cardiac outcomes; Table 1. While these proteins were prioritized on robust association with CMR traits and cardiac outcome, the association between a proteins' CMR effect direction and cardiac outcome effect direction (both categorized as 'beneficial', 'harmful', or 'mixed', the latter for multiple directionally discordant protein effects) did not reach significance (p-value 0.85). This likely reflects imperfect understanding of the relation between CMR traits and disease, as mentioned before some of the considered CMR traits show a u-shaped relationship with death[34]. Furthermore, given the strong (observational and genetic) correlation between CMR traits, inference might be further improved by considering CMR traits jointly. Despite these caveats, we found that that the identified subset of drugged and druggable proteins is enriched for compounds with known cardiac indication or side effect: 60% (95%CI 39; 79) in our analysis, compared to 15% (172/1151) from a look-up of targets with a level 1 ATC code for cardiovascular system. Additionally, it is worth considering that all of the considered RV or LV related proteins showed on-target effects on known cardio-metabolic risk factors which may offer alternative and directionally discordant pathways to cardiac disease, offering a further explanation the observed difference between CMR effect direction and cardiac outcome (Figure 5, 8).

Our analyses have highlighted multiple drug targets that affected the joint risk of multiple cardiac outcomes (Figure 5 and Table 1). Some of these proteins are closely linked. For example, ERAP1 (Endoplasmic reticulum aminopeptidase 1) affects CHD, AF, and non-ischemic CM, and is closely related to ERAP2 which showed a directionally opposing effect on CHD. This discordance of effects might be explained by similarly opposing effect on known CVD risk factors such as Lp(a), DBP, and carotid plaque. Both ERAP1 and ERAP2 play a major role in peptide trimming for presentation on major histocompatibility complex (MHC) Class I molecules[42], which is involved with cardiomyocyte pathogenesis[43]. Similarly, TNF12 decreases the risk of non-ischemic CM, DCM, and AF, and promotes IL8 concentration which we linked to a lower risk of AF and HF (and higher risk of CHD). IL8 concentration has been previously associated with HF and AF outcomes, supporting these observations[44,45]. We found that higher plasma concentrations of BAG3 affected multiple CMR traits as well as HF, DCM, and non-ischemic CM risk. BAG3 is indirectly drugged through a direct protein interaction with HSP7C (heat shock cognate 71 kDa protein), where compounds for HSP7C have been documented to affect cardiac function and calcium handling[46].

Through BNF and ChEMBL linkage we found that 13 (52%, 95%CI 0.31; 0.72) of the 25 druggable proteins were targeted by a compound with an oncological indication (Table 1). For example, CD33 and SLAF7, together with CD38 (which did not have plasma pQTL data) are targeted by mAbs for multiple myeloma[47]. The high degree of oncological targets suggests that some of the reported cardiotoxicity[48] (e.g., by tyrosine kinase inhibitors such as TIE2) may likely be due to on-target effects, which are resistant to potential compound improvements. Because inhibiting oncological compounds are used to prevent cancer progression, compounds activating these proteins may not necessarily cause novel neoplasms. Activator compounds may nevertheless influence the growth of any existing undiagnosed neoplasms, and hence a change in action type should very carefully explored. Aside from the oncological targets, we found many additional repurposing opportunities, for example the PA2GA (Phospholipase A2) inhibitor varespladib previously failed to show a beneficial CHD effect (which we confirmed)[49], whereas we found convincing effect of PA2GA on CMR traits and HF.

In the current study, we uniquely combined GWAS of UKB derived CMR measurements of RV and LV structure and functions, with cross-platform GWAS of the plasma proteome, to perform *cis*-MR of protein effect on these CMR measurement, as well as 56 additionally clinically relevant traits. By leveraging orthogonal lines of evidence on genetic expression, mRNA expression, protein interactions, and drug compound indications and side effects, we were able to identify a robust set of proteins related with CMR and cardiac outcomes. The synthesis of these independent evidence sources complimented the already robust analytic techniques used. Genetic analyses were conducted using methods such as BOLT-LMM and BOLT-REML, which appropriately account for any potential

population admixture or relatedness[27]. Drug target MR analyses were guarded against horizontal pleiotropy by removing variants with either high leverage or heterogeneity statistics (as potential outliers), and a model selection framework was used to apply the MR-Egger correction (which is unbiased in the presence of 100% pleiotropic variants)[33]. Furthermore, results were strenuously corrected for multiplicity accounting for the correlation between CMR traits through PCA, where prioritization based on multiple lines of evidence results in an indirect replication that will further reduce the false positive rate. This plausibility of false positives driving these results was further explored through Kolmogorov-Smirnov test, comparing the observed p-value distributions against the p-value distribution expected when all results are false positive, finding considerable difference between both (Supplementary Figure 5).

Nevertheless, the following potential limitations deserve consideration. While we did not exclude individuals from non-European ancestry and corrected for any potential population stratification bias through efficient linear-mixed-models[27], the majority of participants were of European decent and hence generalizability of our results should be confirmed. There are some caveats that suggest that drug target MR analysis may be more useful as a reliable test of effect direction. This is because drugs that inhibit a target usually do so by modifying its function not its concentration, whereas genetic variants used in MR analysis usually affect protein expression and therefore concentration. Furthermore, while trials of drug compounds are closely monitored, and followed for a fixed period, allowing for exploration of induction-times[50]. MR estimates are considered to reflect a life-long exposure, but in the absence of serial assessment, possible changes across age are difficult to explore, as are disease induction-times. For these reasons, we suggest that drug target MR offers a robust indication of effect direction but may not directly anticipate the effect magnitude of pharmacologically interfering with a protein and position our findings as a resource to inform ongoing and future drug trials[51].

In conclusion, through large scale analyses of the genome and plasma proteome, and linkage to mRNA expression, protein interactions and drug compounds, we have identified a prioritized set of 33 proteins with a robust CMR and cardiac outcome fingerprint and determined anticipated effects of protein perturbation through a highly powered phenome-wide scan. Our analyses provide a detailed overview of potential targets for repurposing or *de novo drug* development for cardiac therapies.

Data Availability

Source GWAS CMR data have been deposited XX. We additionally leveraged pQTL data from there sources, which can be accessed from: https://ega-archive.org/studies/EGAS00001002555 (interval), https://zenodo.org/record/2615265#.YbNUwr3MLOg (scallop), ftp://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_pQTLs/ (Framingham). Genomcis data for the *cis*-MR phewas was sourced for CHD cases from CardiogramplusC4D[52] (http://www.cardiogramplusc4d.org/); non-ischemic CM from (https://www.ebi.ac.uk/gwas/publications/30586722); DCM from (https://www.ebi.ac.uk/gwas/publications/33677556) HF from HERMES[53] (https://www.ebi.ac.uk/gwas/publications/31919418); AF from AFgen[54] (http://csg.sph.umich.edu/willer/public/afib2018/); stroke (subtypes) from MEGASTROKE[55] (http://www.megastroke.org/index.html) and intracranial aneurysms and subarachnoid haemorrhage data from (https://cd.hugeamp.org/) ; the presence of carotid plaque and carotid artery intima media thickness were available from a meta-analysis of the Cohorts for Aging Research in Genomic Epidemiology (CHARGE)[56] and University College London Edinburgh Bristol (UCLEB)[57] (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v6.p1); glycemic traits, blood lipid, exercise ECG, and lung function measurement were sourced from (http://www.nealelab.is/uk-biobank); type 2 diabetes[58] from DIAGRAM (http://diagram-consortium.org/index.html); BMI from GAINT[59] (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files); Asthma from (https://www.ebi.ac.uk/gwas/publications/32296059); CRP from (https://www.ebi.ac.uk/gwas/publications/30388399); the CKDGen consortium provided GWAS associations on blood urea nitrogen (BUN), estimated glomerular filtration rate (eGFR), and chronic kidney disease[60] (http://ckdgen.imbi.uni-freiburg.de/); Alzheimer's disease data was sourced from (https://ctg.cncr.nl/software/summary_statistics); Parkinson's disease data from (https://www.ebi.ac.uk/gwas/publications/31701892); Lewy body dementia from (https://www.ebi.ac.uk/gwas/publications/33589841); oncological data were sourced from (https://github.com/Wittelab/pancancer_pleiotropy). We additionally leveraged data from the Human Protein Atlas (https://www.proteinatlas.org/about/download), IntAct[24] and Reactome[25] (https://reactome.org/download-data), ChEMBL (https://chembl.gitbook.io/chembl-interface-documentation/downloads) and from Finan et al. (https://www.science.org/doi/10.1126/scitranslmed.aag1166, Table S1).

Code Availability

Analyses were conducted using Python v3.7.4 (for GNU Linux), Pandas v0.25, Numpy v1.15, Matploblib, Seaborn (for GNU Linux), and ggplot2 (for GNU Linux). Scripts and data necessary to generate the illustrations have been deposited: XX.

Declarations

Author's contributions

AFS, MB, FWA, JvS, CF, contributed to the idea and design of the study. AFS, MB, AA and CF performed the analyses. AFS, JvS, MB, CF drafted the manuscript. All authors provided critical input on the analyses and the drafted manuscript.

Funding and role of funding sources

AFS is supported by BHF grant PG/18/5033837 and the UCL BHF Research Accelerator AA/18/6/34223. MB is supported by the Alexandre Suerman Stipend of the UMC Utrecht (2017). CF and AFS received additional support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre. ADH is an NIHR Senior Investigator. JvS and MvV are supported by Dutch Heart Foundation grant 2019T045. ATR is supported

by the CardioVascular Research Initiative of the Netherlands Heart Foundation (CVON 2015-12 eDETECT and 2012-10 PREDICT) and the Dutch Heart Foundation grant number 2015T058. SWvdL is funded through grants from the Netherlands CardioVascular Research Initiative of the Netherlands Heart Foundation (CVON 2011/B019 and CVON 2017-20) and is additionally supported by the ERA-CVD program (grant number: 01KL1802), the EU H2020 TO_AITION (grant number: 848146), and the Leducq Fondation 'PlaqOmics'. AA is supported by King Abdullah International Medical Research Center (KAIMRC). This work was supported by grant [R01 LM010098] from the National Institutes of Health (USA) and by EU/EFPIA Innovative Medicines Initiative 2 Joint Undertaking BigData@Heart grant n° 116074, as well as by the UKRI/NIHR Multimorbidity fund Mechanism and Therapeutics Research Collaborative MR/V033867/1 and the Rosetrees Trust.

Conflict of interest statements

AFS and FWA have received Servier funding for unrelated work. SWvdL has received Roche funding for unrelated work.

Guarantor

AFS, CF, AA, performed the presented analyses. AFS had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgement

This research has been conducted using the UK Biobank Resource under Application Number 24711. The authors are grateful to UK Biobank participants. UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government, and the Northwest Regional Development Agency. It has also had funding from the Welsh Assembly Government and the British Heart Foundation.

Prior postings and presentations

A preprint version of this manuscript has been deposited on biorxiv: https://doi.org/.

References

1. Ponikowski, P. *et al.* 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal* **37**, 2129–2200 (2016).
2. Redfield, M. M. *et al.* Isosorbide Mononitrate in Heart Failure with Preserved Ejection Fraction. N Engl J Med **373**, 2314–2324 (2015).
3. Gheorghiade, M. *et al.* Effect of Vericiguat, a Soluble Guanylate Cyclase Stimulator, on Natriuretic Peptide Levels in Patients With Worsening Chronic Heart Failure and Reduced Ejection Fraction: The SOCRATES-REDUCED Randomized Trial. JAMA **314**, 2251–2262 (2015).
4. Fordyce, C. B. *et al.* Cardiovascular drug development: is it dead or just hibernating? J Am Coll Cardiol **65**, 1567–1582 (2015).
5. Andersson, C. *et al.* Risk factor-based subphenotyping of heart failure in the community. PLOS ONE **14**, e0222886 (2019).
6. Albini, A. *et al.* Cardiotoxicity of Anticancer Drugs: The Need for Cardio-Oncology and Cardio-Oncological Prevention. J Natl Cancer Inst **102**, 14–25 (2010).
7. Ruijsink, B. *et al.* Fully Automated, Quality-Controlled Cardiac Analysis From CMR: Validation and Large-Scale Application to Characterize Cardiac Function. JACC: Cardiovascular Imaging **13**, 684–695 (2020).
8. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. Science Translational Medicine **9**, (2017).
9. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. Nature (2018) doi:10.1038/s41586-018-0175-2.
10. Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. Nature Communications **11**, 3255 (2020).
11. Swerdlow, D. I. *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. The Lancet **379**, 1214–1224 (2012).
12. Swerdlow, D. I. *et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials*. vol. 385 (2015).
13. Schmidt, A. F. *et al.* Phenome-wide association analysis of LDL-cholesterol lowering genetic variants in PCSK9. BMC Cardiovascular Disorders **240**, (2019).
14. Schmidt, A. F. *et al.* PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. The Lancet Diabetes & Endocrinology **5**, 97–105 (2017).
15. Schmidt, A. F. *et al.* Cholesteryl ester transfer protein (CETP) as a drug target for cardiovascular disease. Nat Commun **12**, 5640 (2021).

16. Holmes, M. V. *et al.* Secretory phospholipase A2-IIA and cardiovascular disease: A mendelian randomization study. Journal of the American College of Cardiology **62**, 1966–1976 (2013).
17. Casas, J. P. *et al.* PLA2G7 Genotype, lipoprotein-associated phospholipase A2 activity, and coronary heart disease risk in 10 494 cases and 15 624 controls of european ancestry. Circulation **121**, 2284–2293 (2010).
18. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. bioRxiv 627398 (2019) doi:10.1101/627398.
19. Gordillo-Marañón, M. *et al.* Validation of lipid-related therapeutic targets for coronary heart disease prevention using human genetics. Nat Commun **12**, 6120 (2021).
20. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab **2**, 1135–1148 (2020).
21. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun **9**, 3268 (2018).
22. Gaulton, A. *et al.* The ChEMBL database in 2017. Nucleic Acids Res **45**, D945–D954 (2017).
23. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. Nat Biotechnol **28**, 1248–1250 (2010).
24. Orchard, S. *et al.* The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res **42**, D358-363 (2014).
25. Jassal, B. *et al.* The reactome pathway knowledgebase. Nucleic Acids Res **48**, D498–D503 (2020).
26. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. Nature **562**, 203–209 (2018).
27. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet **47**, 284–290 (2015).
28. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res **47**, D1005–D1012 (2019).
29. Burgess, S., Zuber, V., Valdes-Marquez, E., Sun, B. B. & Hopewell, J. C. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. Genetic Epidemiology **41**, 714–725 (2017).
30. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in mendelian randomization studies. International Journal of Epidemiology **40**, 755–764 (2011).
31. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. International Journal of Epidemiology **44**, 512–525 (2015).
32. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in medicine **36**, 1783–1802 (2017).
33. Bowden, J. *et al.* Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. Int J Epidemiol **47**, 1264–1278 (2018).
34. Wehner, G. J. *et al.* Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie? Eur Heart J **41**, 1249–1257 (2020).
35. Carithers, L. J. & Moore, H. M. The Genotype-Tissue Expression (GTEx) Project. Biopreserv Biobank **13**, 307–308 (2015).
36. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. Nature **507**, 455–461 (2014).
37. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform **18**, 205–214 (2017).
38. Bulik-Sullivan, B. K. *et al.* LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. Nat Genet **47**, 291–295 (2015).
39. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat Immunol **6**, 689–697 (2005).
40. 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).
41. Pirruccello, J. P. *et al.* Genetic Analysis of Right Heart Structure and Function in 40,000 People. *bioRxiv* 2021.02.05.429046 (2021) doi:10.1101/2021.02.05.429046.
42. Agrawal, N. & Brown, M. A. Genetic associations and functional characterization of M1 aminopeptidases and immune-mediated diseases. Genes Immun **15**, 521–527 (2014).
43. Gröschel, C. *et al.* T helper cells with specificity for an antigen in cardiomyocytes promote pressure overload-induced progression from hypertrophy to heart failure. Sci Rep **7**, 15998 (2017).
44. Nymo, S. H. *et al.* Inflammatory cytokines in chronic heart failure: interleukin-8 is associated with adverse outcome. Results from CORONA. Eur J Heart Fail **16**, 68–75 (2014).
45. Guo, Y., Lip, G. Y. H. & Apostolakis, S. Inflammation in atrial fibrillation. J Am Coll Cardiol **60**, 2263–2270 (2012).
46. Kim, Y.-K. *et al.* Deletion of the Inducible 70-kDa Heat Shock Protein Genes in Mice Impairs Cardiac Contractile Function and Calcium Handling Associated With Hypertrophy. Circulation **113**, 2589–2597 (2006).
47. Petrucci, M. T. & Vozella, F. The Anti-CD38 Antibody Therapy in Multiple Myeloma. Cells **8**, E1629 (2019).
48. Han, X., Zhou, Y. & Liu, W. Precision cardio-oncology: understanding the cardiotoxicity of cancer therapy. NPJ Precis Oncol **1**, 31 (2017).

49. Nicholls, S. J. *et al.* Varespladib and cardiovascular events in patients with an acute coronary syndrome: the VISTA-16 randomized clinical trial. JAMA **311**, 252–262 (2014).
50. The HPS3/TIMI55–REVEAL Collaborative Group. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. N Engl J Med **377**, 1217–1227 (2017).
51. Schmidt, A. F., Hingorani, A. D. & Finan, C. Human Genomics and Drug Development. Cold Spring Harb Perspect Med a039230 (2021) doi:10.1101/cshperspect.a039230.
52. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. **47**, 1121–1130 (2015).
53. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nature Communications **11**, 1–12 (2020).
54. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nature Genetics **50**, 1234–1239 (2018).
55. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes | Nature Genetics. https://www.nature.com/articles/s41588-018-0058-3.
56. Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. Circ Cardiovasc Genet **2**, 73–80 (2009).
57. Shah, T. *et al.* Population genomics of cardiometabolic traits: design of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium. PloS one **8**, e71345 (2013).
58. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nature Genetics **50**, 559–571 (2018).
59. 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).
60. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat. Genet. **51**, 957–972 (2019).

Table 1

Table 1 Summarizing the prioritized plasma protein information

Protein (uniprot id)	Nearest druggable protein (uniport id)	Druggability	Ventricle	CMR effect	Cardiac effect	Clinical development phase	No. compounds	Compound Action types	CVD indication or side effects	Oncolog indicatic or side effects
MANBA (O00462)		Currently not druggable	Both	Beneficial	Beneficial	0	0			
NCAM2 (O15394)		Currently not druggable	Both	Beneficial	Beneficial	0	0			
TNF12 (O43508)		Directly druggable	Left	Mixed	Beneficial	1	2	Inhibitor		✓
ICOSL (O75144)		Directly druggable	Left	Beneficial	Harmful	1	1	Inhibitor		
BAG3 (O95817)	HSP7C (P11142)	Indirectly druggable	Both	Beneficial	Beneficial	3	1	Inhibitor		
C1QC (P02747)		Currently not druggable	Both	Harmful	Beneficial	0	0			
IL6RA (P08887)		Directly drugged	Both	Mixed	Beneficial	4	4	Antagonist, Inhibitor	✓	✓
PATE4 (P0C8F1)		Currently not druggable	Right	Harmful	Harmful	0	0			
IL8 (P10145)	IL18 (Q14116)	Directly druggable	Both	Mixed	Mixed	2	2	Inhibitor	✓	✓
CO6A1 (P12109)		Directly drugged	Both	Mixed	Harmful	4	2	Hydrolytic Enzyme	✓	
TDGF1 (P13385)		Directly druggable	Both	Beneficial	Mixed	1	1	Binding Agent		✓
LYAM1 (P14151)		Directly druggable	Both	Mixed	Beneficial	3	3	Antagonist, Inhibitor	✓	
PA2GA (P14555)		Directly druggable	Left	Mixed	Harmful	3	2	Inhibitor	✓	
ISK2 (P20155)	LYAM2 (P16581)	Indirectly druggable	None	None	Harmful	3	3	Antagonist, Inhibitor	✓	
UD16 (P19224)		Currently not druggable	Left	Harmful	Harmful	0	0			
CD33 (P20138)		Directly drugged	Left	Harmful	Beneficial	4	6	Binding Agent, Other	✓	✓
CAH6 (P23280)		Directly drugged	Right	Harmful	Harmful	4	1	Inhibitor		
PPAC (P24666)	5HT1E (P28566),HDA10 (Q969S8),PDE4D (Q08499),RARA (P10276),VWF (P04275),P2RY4 (P51582),PRS7 (P35998),GBRG1 (Q8N1C3)	Indirectly drugged	Both	Beneficial	Harmful	4	94	Agonist, Allosteric Antagonist, Antagonist, Inhibitor, Inverse Agonist, Modulator, Partial Agonist, Positive Allosteric Modulator, Positive Modulator	✓	✓
TNR5 (P25942)		Directly druggable	Both	Mixed	Harmful	2	5	Agonist, Antagonist, Inhibitor, Partial Agonist		✓
MK03 (P27361)		Directly druggable	Right	Beneficial	Beneficial	2	3	Inhibitor		✓
COFA1		Directly	Both	Mixed	Harmful	4	2	Hydrolytic	✓	

(P39059)		drugged						Enzyme		
GPC5 (P78333)	SYUA (P37840),TPH1 (P17752)	Indirectly drugged	Right	Harmful	Harmful	4	4	Inhibitor	✓	✓
TIE2 (Q02763)		Directly drugged	Both	Beneficial	Harmful	4	8	Inhibitor	✓	✓
MFGM (Q08431)		Directly druggable	Right	Beneficial	Harmful	1	1	Binding Agent		
IL18R (Q13478)		Indirectly druggable	Left	Beneficial	Beneficial	2	2	Cross-Linking Agent, Inhibitor	✓	
LAMC2 (Q13753)		Directly drugged	Right	Beneficial	Harmful	4	1	Hydrolytic Enzyme	✓	
BGH3 (Q15582)		Currently not druggable	Both	Harmful	Beneficial	0	0			
ERAP2 (Q6P179)		Directly druggable	Both	Beneficial	Harmful	2	1	Inhibitor		✓
SPA12 (Q8IW75)		Currently not druggable	Both	Beneficial	Harmful	0	0			
PGLT1 (Q8NBL1)		Currently not druggable	Both	Beneficial	Harmful	0	0			
I17RA (Q96F46)		Directly drugged	Both	Mixed	Beneficial	4	1	Antagonist	✓	
SLAF7 (Q9NQ25)		Directly drugged	Right	Beneficial	Harmful	4	1	Inhibitor	✓	✓
ERAP1 (Q9NZ08)		Directly druggable	Both	Beneficial	Mixed	2	1	Inhibitor		✓

N.b. indirectly drugged or indirectly druggable proteins were mapped to the nearest druggable protein (Figure 7) potentially resulting in tied proteins at the same distance. Tied proteins were both included in further analyses, for example resulting in 8 nearest druggable proteins for PPAC explaining the large number of mapped compounds. For indirectly drugged/druggable proteins the CMR and cardiac effects represents the indexing protein for which we had plasma pQTL (the first column) and the drug compound data are from the nearest druggable protein(s). The nearest druggable protein of ISK2, LYAM2 was available as pQTL data and here we instead list the LYAM2 effects on CMR and cardiac traits.

Figures

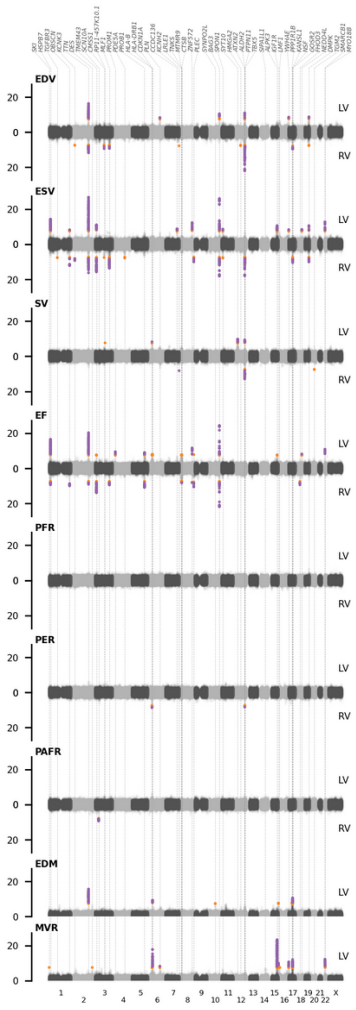

Figure 1

Manhattan plots of genome-wide CMR associations with genomic annotations

N.b. Purple dots indicate associations that pass the conservative significant threshold of 7.14×10^{-9}, with orange dots association between 5.00×10^{-8} and 7.14×10^{-9}, labels indicate the lead gene in the region. LV, left-ventricle; RV, right-ventricle; EDV, end-diastolic volume; ESV, end-systolic volume; SV, stroke volume; EF, ejection fraction; PER, peak ejection rate; PARF/PFR, peak (atrial) filling rate; EDM, end-diastolic mass; MVR, ratio between end diastolic mass and volume. Results are based on an analysis of 36,548 subjects.

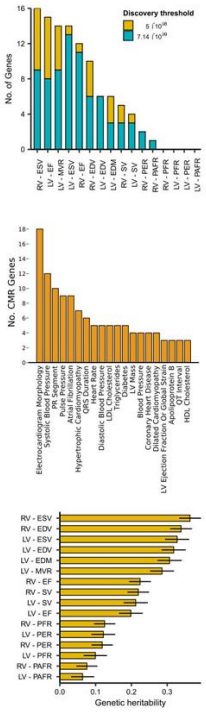

Figure 2

Aggregated GWAS results and genetic heritability estimates

N.b. Top panel depicts the number of significant putative causal genes per CMR trait and significance threshold. The middle panel provides the top 20 most frequent trait associations of the discovered CMR genes, sourced from GWAS catalog. The bottom panel provides the genetic heritability estimates with 95% confidence intervals. Results are based on an analysis of 36,548 subjects, a star indicates significant p-value at an alpha of 0.05.

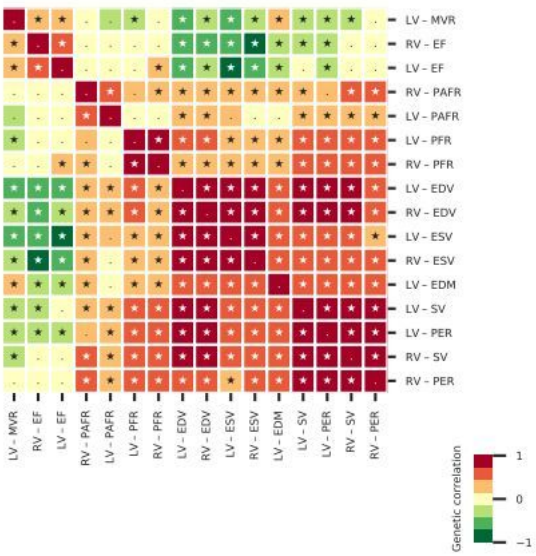

Figure 3

The pairwise genetic correlation between 16 CMR traits

N.b. LV, left-ventricle; RV, right-ventricle; EDV, end-diastolic volume; ESV, end-systolic volume; SV, stroke volume; EF, ejection fraction; PER, peak ejection rate; PARF/PFR, peak (atrial) filling rate; EDM, end-diastolic mass; MVR, ratio between end diastolic mass and volume. Results are based on an analysis of 36,548 subjects. Star-annotated cells indicate significant associations at an alpha of 0.05.

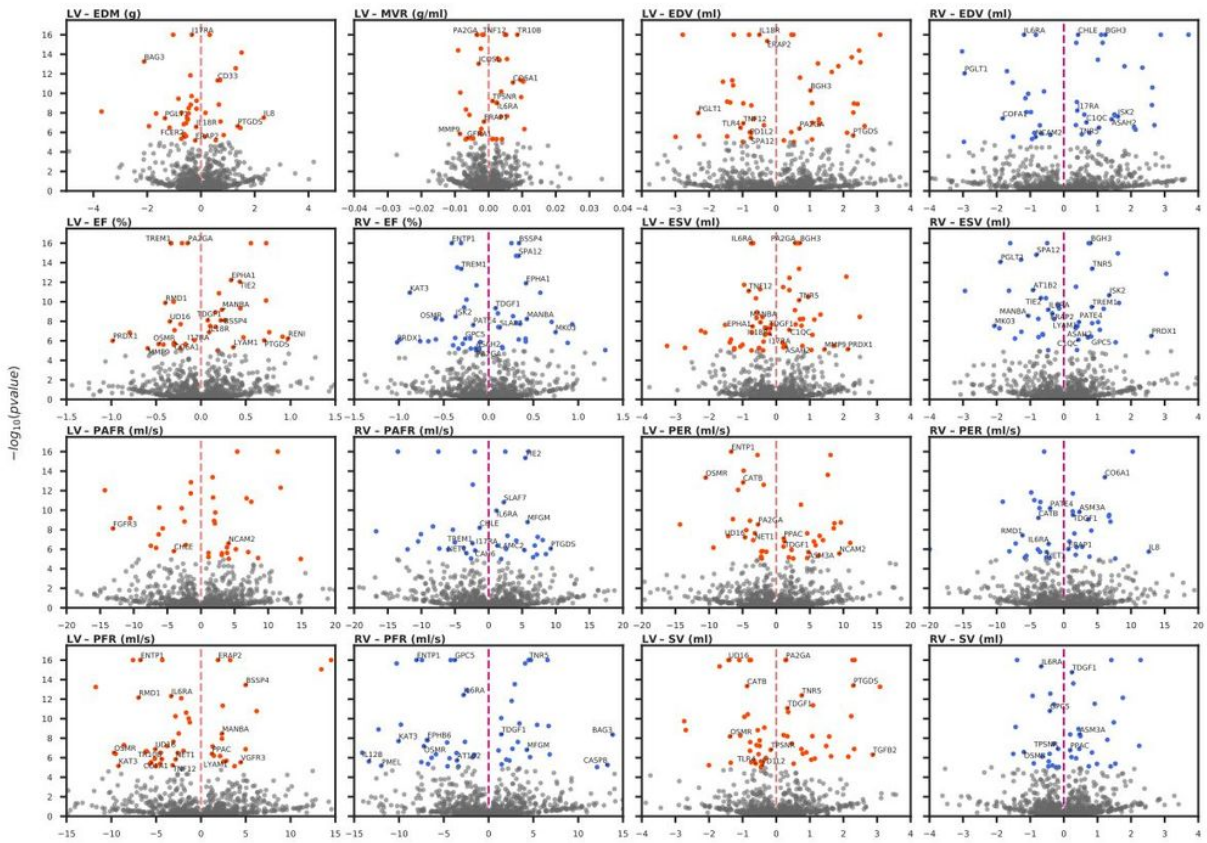

Figure 4

Volcano plots of the plasma protein effect on CMR traits

N.b. Proteins were annotated if they were part of the drugged, druggable, directionally concordant or nearest druggable protein sets. Results were coloured by left and right ventricle if they pass a p-value threshold of 7.81×10^{-6}. The Mendelian randomization effects per unit (in standard deviation) change in the protein are plotted on the x-axis, against the $-\log_{10}$(p-value) on the y-axis.

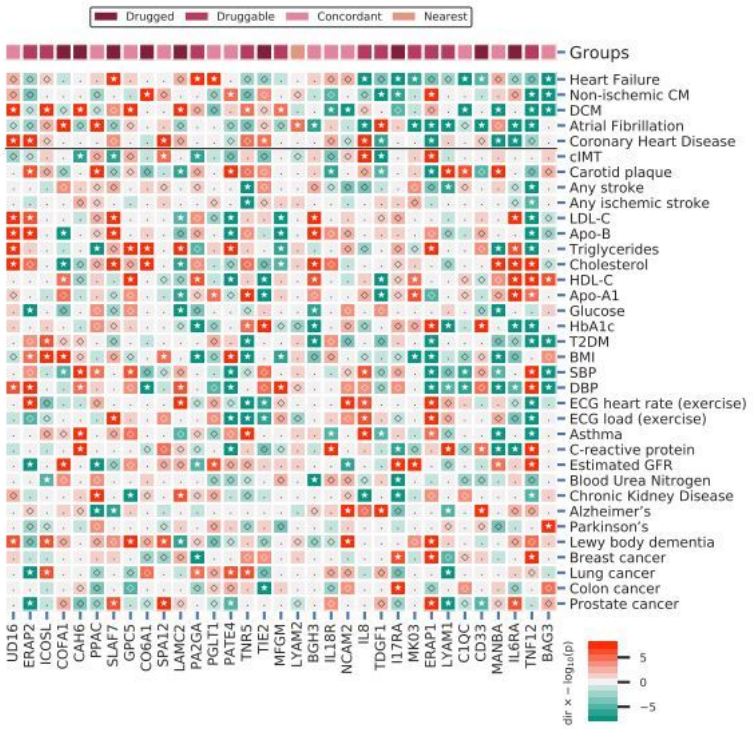

Figure 5

A phenome-wide scan of CMR prioritized proteins associated with one or more cardiac outcome.

N.b. Proteins were curated on having a multiplicity corrected p-value < 1.29×10^{-5} with one or more of the following cardiac traits: heart failure (HF), dilated cardiomyopathy (DCM), non-ischemic CM, atrial fibrillation (AF), or coronary heart disease (CHD). P-values passing the 0.05 threshold are indicated by an open diamond with stars indicating results passing a threshold of 1.29×10^{-5}. Cells were coloured by effect direction times $-\log_{10}$(p-value); where p-values were truncated at 8 for display purposes. The top column indicates whether the CMR associated proteins were identified as drugged, druggable, directionally concordant, or nearest druggable protein. Y-axis abbreviations: DCM: dilated cardiomyopathy; cIMT: carotid artery intima media thickness; T2DM: type 2 diabetes; BMI: body mass index; DBP/SBP diastolic/systolic blood pressure; estimated GFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; LDL-C: low-density lipoprotein cholesterol; HDL-C: high-density lipoprotein cholesterol; Apo-B: apolipoprotein-B; Apo-A1: apolipoprotein-A1; HbA1c: glycated hemoglobin; ECG: electrocardiography; FVC: forced vital capacity; FEV1 forced expiratory volume during the first second; PEF: peak expiratory flow. Note that all 56 phenome-wide traits are presented in Supplemental Figure 10.

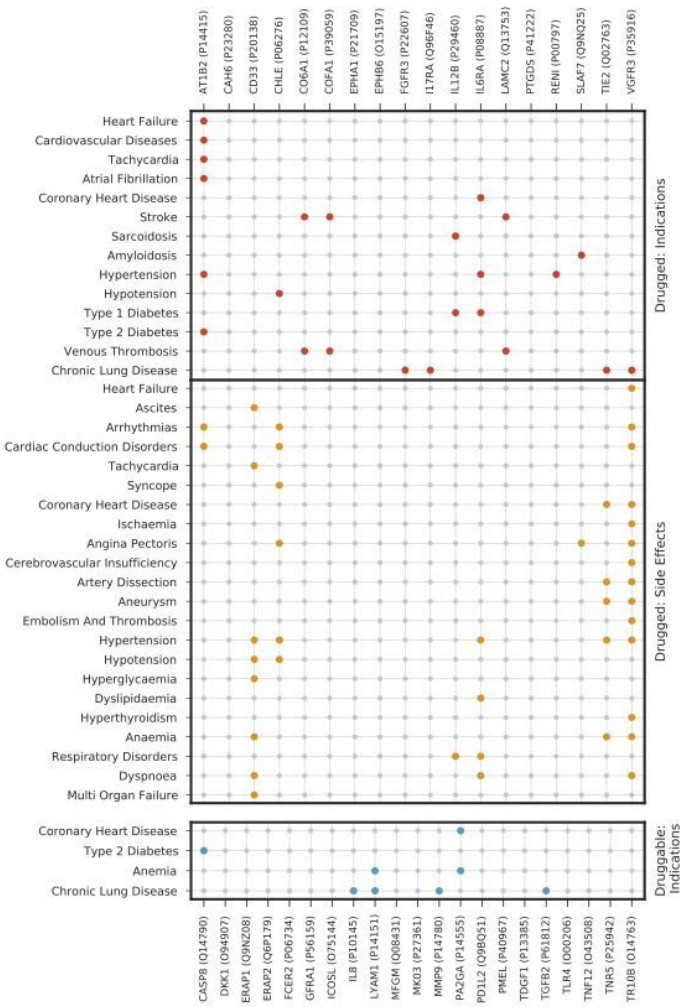

Figure 6

Incidence matrix of cardio-metabolic related indications and side effects of compounds targeting drugged and druggable proteins associated with CMR traits.

N.b. coloured dots represents an established link between the compound and trait; data were extracted from BNF and ChEMBL[22].

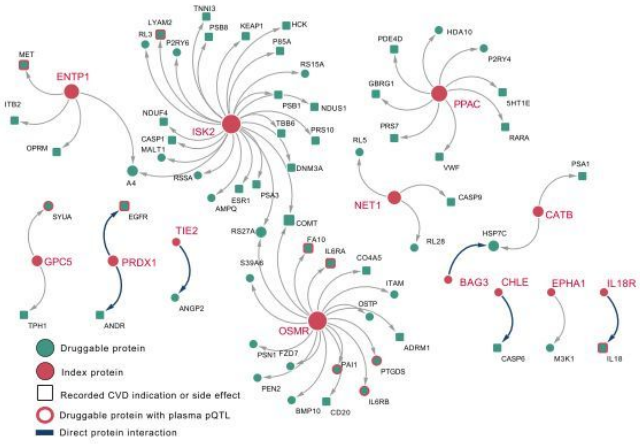

Figure 7

A network of plasma proteins with a directionally concordant CMR effect (index, orange circles) and their nearest druggable protein (green square with record CVD indication or side effect).

N.b. Directly interacting proteins are presented as a thick blue arrow, the remaining druggable proteins were separated by a single intermediate protein. In the presence of ties all druggable proteins with the same distance are presented. Druggable proteins with a CVD indication or side-effect (based on BNF and ChEMBL) are presented as square. Additionally, non-indexing proteins with available plasma pQTL data are represented by an orange outline.

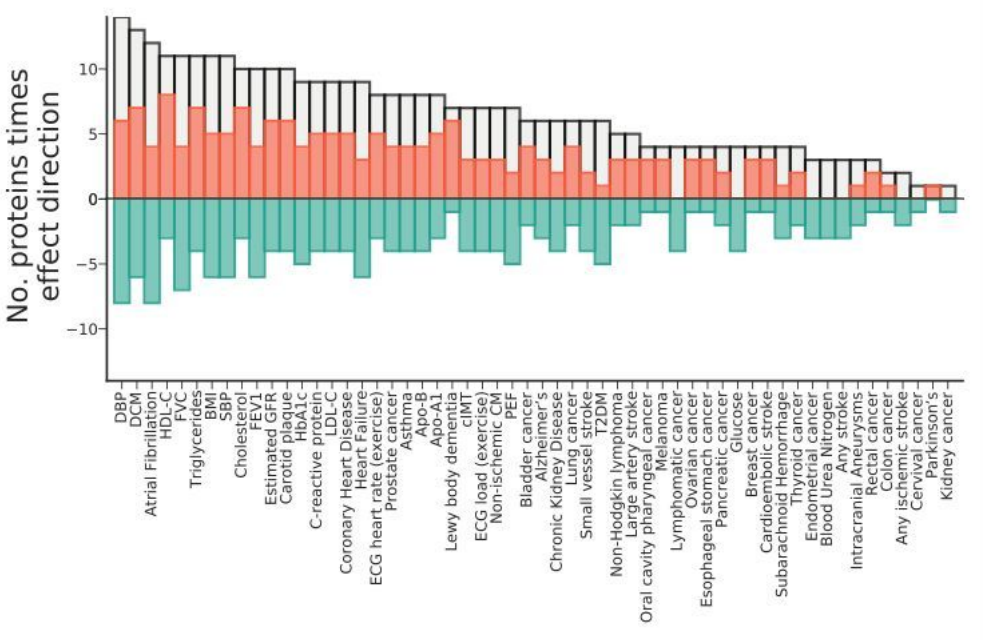

Figure 8

The likely causal consequences of a standard deviation increase of thirty-three CMR and cardiac outcome prioritized plasma proteins.

N.b. Results are presented by the effect direction of the pQTL MR (orange increasing effect, green decreasing effect, and white the total counts irrespective of direction). Counts reflect the number drug target MR estimates that passed a multiplicity corrected p-value threshold of 7.81×10^{-6}. DCM: dilated cardiomyopathy; cIMT: carotid artery intima media thickness; T2DM: type 2 diabetes; BMI: body mass index; DBP/SBP diastolic/systolic blood pressure; estimated GFR: estimated glomerular filtration rate; BUN: blood urea nitrogen; LDL-C: low-density lipoprotein cholesterol; HDL-C: high-density lipoprotein cholesterol; Apo-B: apolipoprotein-B; Apo-A1: apolipoprotein-A1; HbA1c: glycated hemoglobin; ECG: electrocardiography; FVC: forced vital capacity; FEV1 forced expiratory volume during the first second; PEF: peak expiratory flow;

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1.pdf
- SupplementaryFile2.pdf