# Proteome-wide Mendelian randomization identifies candidate causal proteins for cardiovascular diseases

Chen Li<sup>1</sup>, Nicolas De Jay<sup>1</sup>, Supriya Sharma<sup>1</sup>, Katrina A. Catalano<sup>1</sup>, Emily R. Holzinger<sup>1</sup>, Venkatesh Sridharan<sup>2</sup>, Zhaoqing Wang<sup>3</sup>, Lei Zhao<sup>2</sup>, Joseph D. Szustakowski<sup>1</sup>, Ching-Pin Chang<sup>2</sup>, Joseph C. Maranville<sup>1</sup>, Erika M. Kvikstad<sup>1</sup>

- 1. Informatics and Predictive Sciences, Bristol-Myers Squibb, USA
- 2. Immunology and Cardiovascular Thematic Research Center, Bristol-Myers Squibb, USA
- 3. Translational Medicine, Bristol-Myers Squibb, USA

## Abstract

### Background:

Cardiovascular diseases (CVD) remain the leading cause of death worldwide, while a lack of clarity on underlying mechanisms has hindered development of novel therapies. Integration of human genetics and proteomics across different ancestries can provide novel, affordable, and systematic approach for target identification and prioritization.

### Methods:

Mendelian randomization approach was applied to unravel causal associations between 2,940 circulating proteins and 21 CVD. Genome-wide summary statistics from the largest genetic mapping of human plasma proteome and meta-analyses on CVD across FinnGen, UK Biobank and Biobank Japan were used. Forward and reverse causation were studied to distinguish respective targets and biomarkers. Genetic instruments for Europeans and East Asians were derived separately and applied to the cardiovascular outcomes in cohorts from corresponding ancestries. We further prioritized drug targets by integrating biological, clinical and population study evidence from cross-database annotations and literature review. Single-cell enrichment analysis and phenome-wide causality scan were performed to further understand target mechanism of action.

#### **Results:**

We found 221 novel candidate causal proteins that impacted risk of one or more CVD through forward MR, and 16 biomarkers whose expression levels were affected by CVD through reverse MR (Bonferroni-adjusted *P*-value <= 0.05). Forward and reverse MR found largely non-overlapping proteins among CVD (only 2 overlapped: LGALS4 and MMP12), suggesting distinct proteomic causes and consequences of CVD. Many of the candidate causal proteins (73.4%) identified are supported by strong literature evidence for a role in immune response and atherosclerotic lesion formation, angiogenesis and vascular remodeling, myogenesis and cardiac progenitor cell differentiation, and energy metabolism. Single cell integration further prioritized ADAM23 for cardiomegaly, PAM for stable angina pectoris and ventricular arrythmia and LPL for peripheral artery disease, whose transcript expression were enriched in cardiomyocytes. Three protein functional groups were highlighted in the phenome-wide scan for their specific

enrichment for CVD, including blood coagulation and fibrinolysis, angiogenesis and vascular remodeling, and cell proliferation and myogenesis.

#### Conclusions:

Our study identified potential therapeutic targets for CVD and distinguished them from biomarkers due to reverse causation. This study provides human genetics-based evidence of novel candidate genes, a foundational step towards full-scale causal human biology-based drug discovery for CVD.

# 1 Introduction

2

3 Recent development in novel therapeutic treatments for cardiovascular disease (CVD) has helped to improve survival and reduce hospitalization of patients.<sup>1,2</sup> Despite this progress, CVD 4 remains the leading cause of death across ancestries and geographies.<sup>3</sup> Innovation and 5 6 investments in developing new therapies for CVD have successfully driven discovery of many 7 novel drugs, including some first- or best-in-class therapies, such as mavacamten, a targeted 8 inhibitor of cardiac myosin, for treatment of patients with obstructive hypertrophic 9 cardiomyopathy;<sup>4</sup> inhibitors of SGLT2 (dapagliflozin, empagliflozin) to treat heart failure regardless of left ventricle ejection fraction and diabetes comorbidity.<sup>5-7</sup> vericiguat, a stimulator 10 of sGC, to treat heart failure patients with reduced ejection fraction;<sup>8</sup> inhibitors of PCSK9 11 (alirocumab, evolocumab) for hypercholesterolemia and atherosclerotic CVD.<sup>9,10</sup> However, in 12 13 recent years, the number of drugs entering all phases of clinical trials and drugs approved by regulatory bodies (including the US Food and Drug Administration) for CVD has declined, 14 especially when comparing to drugs in other therapeutic areas, such as oncology.<sup>11</sup> This is 15 possibly due to a relatively higher cost of trials, as cardiovascular trials are often larger in size 16 17 and longer in duration to manifest primary endpoints that satisfy regulatory requirements.<sup>12</sup> The challenges confronted in cardiovascular drug development call for new approaches to 18 19 increase accuracy and efficiency of trials at lower costs, and studies have shown drug targets 20 with robust human genetic evidence are more likely to be successful in clinical trials.<sup>13</sup>

21

22 Recent advances in omics technologies provide opportunities to discover novel therapeutic strategies in an unbiased, rapid, and cost-effective manner. High throughput immunoaffinity-23 24 based proteomics platform is emergingly applied to systematically quantify protein abundance in large cohorts.<sup>14</sup> Deep profiling of circulating proteins with integration of genomics in a large 25 26 population cohort allows us to understand proteins associated with disease at an unprecedented scale with the possibility of unraveling novel pathways involved in CVD 27 28 pathogenesis. Observational studies have shown cross-sectional associations between proteins and certain CVD,<sup>15</sup> but causal relationships are unclear due to potential confounding effects and 29

reverse causations. Genome-wide association studies (GWAS) of circulating proteins identify
 genetic variants that regulate protein levels in plasma, which can be used as instrumental
 variables (IVs) to infer causality under the framework of Mendelian randomization (MR).

33

34 Studies have demonstrated the feasibility of MR methodology in identifying protein targets for 35 CVD but are often restricted to a small number of targeted proteins, one or a few CVD and one ancestry population.<sup>16-18</sup> In this study, we leveraged the largest genetic mapping of circulating 36 proteome that has recently been published as a public data resource, which measures 2,940 37 proteins in up to 50,000 individuals in UK Biobank,<sup>19</sup> to systematically identify disease 38 39 associations across 21 CVD ranging from rare conditions of myocarditis and cardiomyopathy to 40 common conditions of heart failure and ischemic stroke. To distinguish targets of protein 41 expression levels affecting disease risks and biomarkers that vary protein expressions as consequences of diseases, we conducted forward and reverse MR to delineate bi-directional 42 causality. We prioritized newly identified targets and biomarkers with further evidence from 43 44 cross-database evaluation of pathophysiological function and clinical implication and analyzed 45 potential therapeutic and adverse effects on a wide spectrum of diseases to respectively show 46 opportunities for drug repurposing and caveats for safety issues.

47

# 48 Methods

49

#### 50 Study Overview

51

Figure 1 summarizes the overview of our study design – data processing and analysis workflow.
The study consists of two parts, the primary analysis, a bi-directional MR that disentangle causal
relationships between proteins and CVD, and secondary analyses to strengthen evidence for
target prioritization, including single cell integration to assess cell-type enrichment of targets
prioritized and differential gene expression between cardiomyopathy and healthy cardiac cell
types; manual curation of biological functions and cross-database annotation of protein

- 58 druggability and disease association; a phenome-wide MR scan to assess pleiotropy and
- 59 specificity of targets and biomarkers across disease categories.
- 60

#### 61 MR analyses

62

Various methods have been developed to conduct MR with different capabilities in type I and type II error control, and degrees of tolerance to invalid instruments.<sup>20</sup> We used the inverse variance weighted MR as the primary method because of its conceptually simple design and relatively greater power to detect less significant but true positive results, and other robust methods as sensitivity analyses, including MR Egger, median weighted and mode weighted MR.

69 Forward-MR

70

71 Forward MR investigated causal roles of proteins on CVD risk, where proteins are exposures and 72 CVD are outcomes. IVs for proteins are derived from summary statistics from the largest GWAS 73 of plasma proteins, where genetic associations were assessed with 2,940 plasma proteins in 74  $\sim$ 54,000 UK Biobank participants, of which 34,557 participants are of European ancestry and 262 participants are East Asians.<sup>19</sup> Independent lead variants (also known as protein 75 76 guantitative trait loci, or pQTLs) with p-values less than 5 X  $10^{-8}$  were selected as IVs for 77 European and East Asian (Linkage disequilibrium (LD) clumped variants in LD  $R^2 < 0.1$  with each other and located within 500kb around lead variants) cohorts separately. LD was calculated 78 based on a randomly selected subset of Caucasian (n = 10,000) or all East Asian (n = 2,783) 79 participants in the UK Biobank for Europeans and East Asians, respectively.<sup>21</sup> For IVs absent from 80 the CVD GWAS datasets, we found the best proxies available (p-values < 5 X  $10^{-8}$ , in LD R<sup>2</sup> > 0.7 81 82 with index variants, and within 500kb distance from the index variants). We assessed *cis*- and 83 trans-pQTLs as IVs separately. cis-pQTLs are defined as genetic variants within 1Mb on either 84 side of the start and the stop codon of a protein-coding gene. Proteins that significantly affected 85 CVD risk with Bonferroni-adjusted P value < 0.05 (Bonferroni adjust for the total number of 86 proteins and diseases tested) were identified as target candidates and followed-up with

downstream analyses of multi-trait colocalization, cross-database protein annotation and
phenome-wide MR scan.

89

90 Reverse-MR

91

92 Reverse MR studies protein level changes caused by CVD incidence, where diseases are 93 exposures and proteins are outcomes. IVs for CVD are independent lead variants (LD clumped 94 variants with p-value < 5 X  $10^{-8}$ , in LD R<sup>2</sup> < 0.1 between each other and located within 500kb 95 around lead variants) derived from summary statistics of GWAS on 21 CVD where CVD are 96 identified by ICD-10 and phecode in Europeans (UK Biobank and FinnGen meta-analysis, n = 449,000) and East Asians (Biobank Japan, n = 179,000). IV definitions for all CVD are based on 97 the GWAS meta-analyses in respective cohorts,<sup>22</sup> except for hypertrophic and dilated 98 cardiomyopathy, a more cardiomyopathy-focused GWAS with greater statistical power was used 99 for Europeans for these 2 relatively rare cardiovascular disorders.<sup>23</sup> Ancestry-specific LD was 100 101 estimated in corresponding cohorts as illustrated in the Forward-MR analysis. For IVs absent 102 from the proteomics GWAS datasets, we found the best proxies available using the same criteria 103 as described in the forward-MR analysis. Proteins that are significantly affected by CVD 104 (Bonferroni-adjusted P value < 0.05, Bonferroni adjust for the total number of proteins and 105 diseases tested) are identified as potential biomarkers and compared against significant results 106 from the forward-MR to rule out reverse causation.

107

#### 108 Cross-database protein annotation

109

110 Protein biological functions are annotated with cross-referencing multiple databases, including

111 UniProt and RefSeq. We also annotated protein druggability by interrogating Drug-Gene-

112 Interaction (DGI) database,<sup>24</sup> ChEMBL v33,<sup>25</sup> PharmGKB,<sup>26</sup> and DrugCentral Postgres v14.5<sup>27</sup>

113 databases to find therapeutic targets of approved or under clinical development. To evaluate

114 protein associations with diseases, we queried GWAS Catalog<sup>28</sup> for population-level evidence, as

115 well as ClinVar<sup>29</sup> and ClinGen<sup>30</sup> for clinically suggested implications.

116

#### 117 Phenome-wide MR scan

118

119 We ran MR to disentangle bi-directional causal relationships between 2,940 proteins and 159 120 disease outcomes in European and East Asian cohorts separately. For East Asian cohort, an 121 additional 38 quantitative biomarkers and 23 medication usage phenotypes were tested. 122 Diseases were defined by ICD-10 and phecode by a previous GWAS meta-analysis in UK Biobank, 123 FinnGen and Biobank Japan, where detailed description of the 3 cohorts and disease definitions 124 are shown.<sup>22</sup> The diseases cover a wide range of categories, with an average of 8 diseases 125 representing each category. Significant results from forward- and reverse-MR are followed up with the phenome-wide MR scan to detect off-targets effects and opportunities for drug 126 127 repurposing. Diseases that have  $\geq$  50 cases were included in the forward- and reverse-MR 128 analyses and at least one genome-wide significant locus in the reverse-MR analysis. Enrichment 129 was assessed via the over representation analysis to determine whether a protein functional 130 group (G) was over-represented in a disease category (D), and enrichment P value was calculated under hypergeometric distribution using the formula:  $P = 1 - \sum_{i=0}^{k-1} \frac{\binom{N}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$ , where N 131 132 is the product of the total number of diseases across disease categories and the total number of 133 proteins in G, M is the product of the total number of diseases in D and the total number of 134 proteins in G, n is the number of significant protein signals in G across disease categories, and k 135 is the number of significant protein signals in G and D. Enrichment significance threshold is 136 Bonferroni-corrected for the total number of protein functional groups and the total number of 137 disease categories.

138

#### 139 Single-cell analysis

140

To link circulating proteins to proteins enriched in cell types originating in the heart tissues or demonstrate specificity for cardiomyopathy *vs* healthy hearts, we analyzed cell type enrichment for significant genetic signals from MR analyses using single-cell RNA sequencing data from left

ventricular samples and assessed differential gene expression between cardiomyopathy andhealthy hearts.

- 146
- 147 Single-cell RNA sequencing data curation

148

We curated publicly available single-cell RNA sequencing data from a recently published cohort of 42 human left ventricular samples (11 DCM, 15 HCM, 16 non-failing hearts).<sup>31</sup> Cell and celltype metadata, and gene counts adjusted for ambient RNA were retrieved from the Broad Institute's Single Cell Portal (accession SCP1303). Cell-type enrichment and differential gene expression data were retrieved from figure source data or supplementary tables of the corresponding publication.

155

#### 156 Cell-type enrichment

157

158 To determine whether a gene exhibited significantly enriched expression in one or more cell 159 types, pseudobulk profiles were generated from healthy and patient donors.<sup>31</sup> A gene was 160 considered to be enriched for a cell type if it satisfies the following criteria: (1) the gene expression was > 4X higher for the cell type compared to the other cell types (one versus rest) with  $FDR_{BH}$ -161 162 adjusted P-value < 0.01, (2) at least 25% of nuclei in the cell type expressed the gene, and (3) a 163 classifier trained on the gene expression predicted whether a nucleus belongs to the cell type with AUC > 0.6. For visualization, we included genes if their expression were significantly enriched 164 165 in at least one cell type and sorted them according to (1) the cell type in which nuclei displayed 166 the highest average expression, (2) the average Gini coefficient calculated from the average nuclei 167 expression and the percentage of nuclei expressing that gene.

168

169 Differential gene expression between cardiomyopathy and healthy hearts

170

171 Chaffin et al. analyzed pseudobulk profiles using a limma-voom model regressing gene expression

in each cell type on a disease group (HCM vs control and DCM vs control) adjusted for age and

- sex. A gene was differentially expressed if the expression was 50% higher in the cardiomyopathy
- 174 than non-failing cardiac cell types with a FDR<sub>BH</sub>-adjusted *P*-value < 0.01.<sup>31</sup>
- 175

#### 176 Analysis software

- 177
- 178 All analyses were performed using the statistical software R (Version 4.1.3). We used the PLINK
- 179 1.9 beta and 2.0 alpha (https://www.cog-genomics.org/plink/) to generate LD matrix for each
- ancestry and perform LD clumping, the TwoSampleMR R package
- 181 (https://github.com/MRCIEU/TwoSampleMR) to perform MR analyses,<sup>32</sup> and the Seurat (
- 182 https://satijalab.org/seurat/) and the edgeR R packages to perform single cell analyses.<sup>33,34</sup>
- 183

#### 184 Ethical Statement

- 185
- 186 The MR analyses were based on publicly available GWAS summary statistics and ancestry-
- 187 specific LD matrices were estimated using UK Biobank individual-level data (Applications 26041
- and 65851). The included GWAS all received informed consent from study participants and have
- 189 been approved by pertinent local institutional review board.
- 190

# 191 **Results**

192

## 193 Target discovery with forward-MR – causal effects of proteins on each CVD risk

194

195 In the forward-MR analysis in European ancestry, 221 proteins significantly affect at least one

196 CVD risk (*P*-value < 1.27 X 10<sup>-7</sup> or 2.09 X 10<sup>-7</sup> using *trans*- or *cis*-pQTL as instrumental variables,

- 197 respectively), which spanned multiple Olink panels, including cardiometabolic, inflammation,
- neurology, and oncology panels (Figure 2). Up to 60% significant proteins affected only one CVD,
- 199 while several proteins were causally associated with a highly pleiotropic set of diseases across
- 200 CVD (Figure 2), such as LPA with angina pectoris, atrial fibrillation, ventricular arrhythmia,
- 201 myocardial infarction, peripheral arterial disease, ischemic stroke, cardiac valvular disease, and

202 cardiomegaly, possibly through dual pathological attributes of lipoprotein(a) in procoagulant 203 effects of apo(a) and atherogenic and proinflammatory effects of oxidized apoB-related phospholipids.<sup>35</sup> Studies suggested LPA as an emerging therapeutic target for reducing 204 lipoprotein(a) levels, thereby lowering risks of atherosclerotic CVD independent of lowering 205 low-density lipoprotein cholesterol levels,<sup>36</sup> with positive results reported by several ongoing 206 207 clinical trials.<sup>37</sup> Most of the targets identified were novel, while several established targets for 208 cardiovascular complications were also been found, including PCSK9 for myocardial infarction, 209 stroke, and coronary revascularization, ANGPTL3 for reducing low-density lipoprotein 210 cholesterol, and ECE1 for congestive heart failure and hypertension (Table S1). Many of the 211 targets (62%) identified are supported by strong literature evidence for a role in immune 212 response and atherosclerotic lesion formation, angiogenesis, and vascular remodeling, 213 myogenesis and cardiac progenitor cell differentiation, and energy metabolism (Table S2). Full 214 forward MR results are provided in Table S4.

215

# Biomarker discovery with reverse-MR – causal effects of CVD on each protein level 217

218 We identified 16 biomarkers in total, among which 5 were from the East Asian analyses (ERBB3, 219 SIRT5, CXCL13, SUSD5, TTR) shown in Table S5 and Figure 3. Of note, ERBB3, a member of 220 membrane-bound tyrosine kinase receptor family that activates cell proliferation and 221 differentiation, found only in the East Asian population, has recently been shown to have cardioprotective effects on cardiomyocyte survival and angiogenesis under stress.<sup>38</sup> Genetic 222 223 polymorphisms of ERBB3 were reported to be associated with coronary artery disease in Han 224 and Uygur ancestries of China,<sup>39</sup> and circulating protein levels with overweight-related hypertension in the same population.<sup>40</sup> Only 2 biomarkers (LGALS4, MMP12) were overlapped 225 226 with the targets derived from the forward-MR analyses (Table S5), suggesting distinct roles of 227 proteins on CVD pathogenesis. We also found NPPB, the gene encoding preprohormone 228 (propro-B-type natriuretic peptide, preproBNP) in cardiomyocytes, as a biomarker specifically to 229 dilated cardiomyopathy. The preproBNP is cleaved into proBNP, and then further processed into 230 2 circulating fragments -- the biologically active BNP and the inactive N-terminal proBNP, both of which are routinely used in clinical diagnosis and treatment management in heart failure.
Studies have suggested that BNP may play different roles in heart failure with preserved and
reduced ejection fraction (HFpEF and HFrEF, respectively),<sup>41</sup> and BNP or NT-proBNP were less
effective biomarkers for HFpEF due to their average lower values in HFpEF than in HFrEF, which
dropped into a normal range in some patients.<sup>42</sup> Our finding is consistent with these studies
that BNP-related biomarkers are mainly indicative for heart diseases with reduced left ventricle
ejection fraction.

238

# Phenome-wide MR screening for specificity and pleiotropy of targets and biomarkers

241 We expanded our analysis to evaluate all proteins and all diseases in the forward and reverse 242 MR across the two ancestries to assess pleiotropy and specificity of targets and biomarkers 243 that trigger or respond to one or multiple diseases (Table S6). For targets, 38 out of 190 targets 244 (trans-forward-MR) were specific to one disease and 51 to one disease category; for biomarkers, 245 4 out of 16 biomarkers were specific to one disease and the diseases were all in the 246 cardiovascular category, such as NPPB specific to dilated cardiomyopathy. Several functional 247 groups showed significant enrichment for their causality on CVD and not the other disease categories, including blood coagulation and fibrinolysis (enrichment P-value =  $3.54 \times 10^{-7}$ ), 248 249 angiogenesis and vascular remodeling  $(2.29 \times 10^{-6})$ , and cell proliferation and myogenesis (2.38)250 X 10<sup>-5</sup>). Full enrichment results of protein functional group across disease categories are shown 251 in Table S7. To assess pleiotropy, 37.8% (14 out of 37) proteins in the functional group of 252 immune and inflammatory response exerted significant causal effects on at least 20 different 253 diseases across disease categories, whereas in the other top 10 largest functional groups, there 254 was  $\leq 1$  protein implicated with  $\geq 20$  diseases. The pleiotropic signature of the immune and 255 inflammatory response group was consistent when altering the pleiotropy threshold to 15, 10 256 or 5 proteins (Table S8), suggesting involvement of this functional group in the pathogenesis of a 257 variety of diseases. Apart from this group, several other proteins of different functional groups 258 showed extensive pleiotropy, such as CTSB and MST1 in regulating autophagy and apoptosis, 259 and TNXB in mediating cell-cell and cell-matrix interaction. Biomarkers were associated with a

- smaller number of diseases on average than targets (mean [SD]: 4.56 [3.29] vs 8.71 [12.28]),
- and they were not enriched for any disease categories (Figure 4).
- 262

# 263 Single-cell integration elucidates cell context-dependent mechanisms of targets

264

265 To gain insights into possible cell context-dependent mechanisms of these targets and 266 biomarkers, we examined their gene expression patterns along major cardiac cell types of the left ventricle identified in a recently published single-cell late-stage DCM and HCM heart atlas.<sup>31</sup> 267 268 Fifty-four out of 235 proteins (22%) had significantly enriched gene expression in at least one 269 major cardiac cell population (Figure 5A). The enriched cell types of each protein were in line 270 with its biological functions (Figure 5B). For example, ADAM23, a protein involved in cell-cell 271 and cell-matrix interactions and cardiac remodeling had significantly elevated expression in 272 cardiomyocytes relative to other cell types. Cardiac-specific conditional knockout of Adam23 in 273 mice exhibited cardiac hypertrophy and fibrosis, whereas transgenic mice overexpressing 274 Adam23 in the heart exhibited reduced cardiac hypertrophy in response to pressure overload.<sup>43</sup> 275 ADAM23 was a significant causal signal in both cis- and trans-MR, specifically for cardiomegaly, 276 while other paralogs in the ADAM protein family, such as ADAM15 and ADAM9, were more 277 pleiotropic, broadly affecting a highly heterogeneous spectrum of diseases. PAM, a known 278 marker of cardiomyocytes encoding the major atrial membrane protein involved in proANP 279 containing secretory granule biosynthesis, was enriched in cardiomyocytes. Similarly, LPL, which 280 encodes a key enzyme in hydrolysis of triglyceride and catabolism of triglyceride-rich 281 lipoprotein, was significantly elevated in cardiomyocytes and adipocytes. Mislocalization of LPL 282 to the cell surface of cardiomyocytes has been associated with cardiomyopathy pathogenesis in 283 mice.<sup>44</sup> Our findings suggest that cell-type-dependent mechanisms of targets emerge as 284 compelling focal points for potential therapeutic hypotheses development. 285

286 Next, we investigated whether our prioritized targets were differentially expressed in

cardiomyopathy and healthy donor heart biopsies in a cell-type specific manner. Indeed, 102

proteins (43.4%) were differentially expressed at the transcriptional level in at least one cardiac

- cell type (Figure 5D). NPPB showed significantly higher expression in cardiomyocytes from DCM,
- but not HCM, than non-failing hearts (Figure 5C). FURIN, mediating proteolytic cleavage of non-
- 291 functional proBNP to its active hormone BNP, expressed at a significantly lower level in
- fibroblasts, endothelial cells and macrophages from DCM than non-failing hearts, suggesting
- agonists of FURIN may help restore cardioprotective activity of BNP in DCM.



### 294

**Figure 1. Study Overview.** GWAS: genome-wide association study, pQTL: protein quantitative trait loci, EUR:

European, EAS: East Asian, LD: linkage disequilibrium, ORF: open reading frame, CVD: cardiovascular disease, IVs:
 instrumental variables, MR: Mendelian randomization.



ups: groups 1,2 and 3 corresponding to *P*-value  $\geq$  FDR<sub>BH</sub> threshold of 0.05, *P*-value between Bonferroni-corrected a

311 nd FDR<sub>BH</sub> thresholds of 0.05 and P -value < Bonferroni-corrected threshold of 0.05. Significant causal associations fo 312 r each CVD were shown in the panels B (*P* -value > 1 X 10<sup>-50</sup>) and C (*P* -value  $\leq$  1 X 10<sup>-50</sup>). Effect coefficient  $\beta$  was plot 313 ted as X-axis and  $-\log_{10}(P - value)$  as Y-axis. Dashed vertical line indicates  $\beta = 0$ . Each dot was labeled with protein na 314 me abbreviation colored by the corresponding protein Olink Panel. Ang: Angina pectoris, AF: Atrial flutter/fibrillatio 315 n, CVD: Cardiac valvular disease, Car: Cardiomegaly, CeAn: Cerebral aneurysm, CHF: Chronic heart failure, DC: Dilat 316 ed cardiomyopathy, EV: Esophageal varix, HC: Hypertrophic cardiomyopathy, IH: Intracerebral hemorrhage, IS: Ische 317 mic stroke, MI: Myocardial infarction, Myo: Myocarditis, Per: Pericarditis, PAD: Peripheral arterial disease, SAP: Stab 318 le angina pectoris, SH: Subarachnoid hemorrhage, CA Un: Unruptured cerebral aneurysm, UAP: Unstable angina pe 319 ctoris, Var: Varicose, VA: Ventricular arrhythmia.

- 320
- 321





#### 323 Figure 3. Discovery of biomarkers – circulating proteins that respond to cardiovascular disease (CVD)

324 development. Reverse-MR results using inverse variance weighted method across ancestries are shown. Proteome-325 wide Manhattan plot - causal effect of each CVD on each protein in the panel A (European) and B (East Asian) 326 ancestries. CVD that had at least one genome-wide significant variant association were shown. Significant causal 327 associations for each CVD were shown in the panels C (European) and D (East Asian). CVD that had causal effects 328 on at least one protein were shown. Figures were represented in the same manner as in the Figure 2. AF: Atrial 329 flutter/fibrillation, Ang: Angina pectoris, CHF: Chronic heart failure, CVD: Cardiac valvular disease, CeAn: Cerebral 330 aneurysm, DC: Dilated cardiomyopathy, EV: Esophageal varix, HC: Hypertrophic cardiomyopathy, IS: Ischemic 331 stroke, MI: Myocardial infarction, PAD: Peripheral arterial disease, SAP: Stable angina pectoris, UAP: Unstable 332 angina pectoris, Var: Varicose.

333

334



336

В

335



337 338

Figure 4. Phenome-wide MR scan to evaluate target and biomarker specificity and pleiotropy. Heatmap

339 representation of Z-scores from A. forward-MR results using trans-pQTLs as instrumental variables and B. reverse-340 MR results, with inverse variance weighted method in the European ancestry. For Z-scores of absolute values > 20, 341 they are truncated to a maximum absolute value of 20. The color gradient, from green (negative) to red (positive), 342 illustrates direction and magnitude of causal associations between proteins (Y-axis) and diseases (X-axis). Diseases 343 and proteins are grouped by their different categories and labeled with different colors on the top and the left-side 344 of the heatmap, respectively.

345

346

А



- 352
- 353





357 Figure 5. Enrichment of gene expression across different cell types in left ventricle biopsies. A. Overall enrichment 358 across 42 left ventricle samples from 11 dilated cardiomyopathy (DCM), 15 hypertrophic cardiomyopathy (HCM) 359 and 16 non-failing hearts combined. Proteins with significantly enriched gene expression in at least one cell type 360 are shown. Average expression is scaled to center at 0 and capped to a range of -2.5 to 2.5 with a color gradient 361 from yellow to purple. The percentage of nuclei expressing each gene is proportional to the size of the dots. The 362 significant enrichment of gene expression (FDR<sub>BH</sub>-adjusted P-value < 0.01) in cell types are highlighted with black 363 circles around the dots. Dots are omitted for genes with less than 0.5% of nuclei within a cell type. B. Box plots of 364 single-cell gene expression in log<sub>2</sub> counts per million (CPM) of ADAM23, PAM and LPL in each cell type. FDR<sub>BH</sub>-365 adjusted P-values are reported only for genes in significantly enriched cell types. C. Box plots of single-cell gene 366 expression (log<sub>2</sub>CPM) of NPPB and FURIN in non-failing, DCM and HCM heart samples stratified by cell types. Red 367 asterisk highlights genes that are significantly differentially expressed in failing vs non-failing heart tissues. D. 368 Differential gene expression in DCM or HCM vs non-failing heart samples stratified by cell types. The log<sub>2</sub> fold 369 change (LFC) of gene expression is scaled to center at 0 and capped to a range of -2 to 2, corresponding to a range 370 of 0.25 to 4 times difference in gene expression. The significant difference of gene expression is labeled with 371 asterisk in the square box.

372

# 373 **Discussion**

374

- 375 Main Findings
- 376

Leveraging genetic IVs to infer causalities between circulating proteins and incident diseases in a proteome-wide and phenome-wide manner help to identify candidate drivers of disease and biomarkers at an unprecedently large scale. In this study, we integrated proteomics MR with single cell analysis to prioritize novel targets for CVD. We found 221 causal signals, the majority of which conform to known disease biology through pathways such as immune response and atherosclerotic lesion formation, angiogenesis and vascular remodeling, myogenesis and cardiac

- progenitor cell differentiation, and energy metabolism. Reverse-MR identified 16 biomarkers
- 384 whose protein expressions were affected by disease status, 5 of which were exclusively found in

385 the East Asian population, demonstrating the value of multi-ancestry populations in identifying 386 drug targets via genetic studies of molecular traits (omics). Only two of the targets were also 387 biomarkers, the LGALS4 and the MMP12, suggesting distinct causes and consequences of 388 circulating proteins on CVD. Some established targets for cardiovascular drugs approved or in 389 clinical development have been found, such as PCSK9 inhibitors (bococizumab, alirocumab and 390 evolocumab), ECE1 inhibitor (daglutril) and ANGPTL3 inhibitor (evinacumab), supporting the 391 validity of our approach. About half of the candidate causal genes (112/235) have been linked 392 to CVD or CVD-related risk factors in previous GWAS, genetic linkage or clinical studies, the rest 393 of genes are implicated with CVD for the first time. Among the novel candidate causal genes to 394 highlight, AGER, a member of the immunoglobulin superfamily of cell surface molecules, was 395 linked to pathogenesis of atherothrombotic diseases, of which upregulated expression was 396 found in human atherosclerotic plaques and polymorphisms associated with myocardial infarction and ischemic stroke;<sup>45</sup> SELE, the E-selectin, and SELPLG, a glycoprotein ligand that 397 398 binds to the E-, P- and L-selectin, mediating recruitment of leukocytes on vascular surfaces 399 during initial steps of inflammation, were shown to be involved in development of 400 atherosclerosis and thrombosis with increased plasma protein levels observed in various CVD;<sup>46</sup> 401 MDK, a secreted heparin-binding growth factor that regulates many biological processes, 402 including cell proliferation, differentiation, migration and survival, was shown to have 403 cardioprotective effects against ischemia and reperfusion injury via attenuating cardiomyocyte apoptosis.<sup>47</sup> Our findings provide evidence supporting therapeutic hypotheses underpinning 3 404 405 established or under investigated cardiovascular drug targets and more broadly showcase the 406 potential of large-scale integration of multi-omics in understanding causal human biology of 407 complex disease for novel drug target discovery.

408

409 The values and limitations of bidirectionality in MR analyses

410

411 Observational studies found associations without a capability to distinguish causal and

412 confounding and are often lack of ascertainment of directionality of associations. We applied a

413 bi-directional MR approach to disentangle causal roles of circulating protein level changes on

414 disease risk – to discover targets, and changes on proteomic expression profiles under disease 415 circumstances – to identify biomarkers. Distinct proteomic signatures were described as causes 416 or consequences of different CVD, however, we were unable to distinguish disease incidence 417 and progression, and thereby to assess proteomic changes underpinning disease progression 418 due to limitations on disease outcome definitions from genome-wide meta-analyses. To achieve 419 the full potential of bi-directional MR and characterize dynamic influence of proteomic changes 420 along disease trajectories, a longitudinal GWAS on progression endpoints will be needed. 421 MMP12 is a notable example, of which the mechanism of action can be further elucidated with 422 the longitudinal data. It was reported that MMP12 was upregulated after myocardial infarction, 423 and then attenuated by endogenous inhibitors, such as tissue inhibitor of metalloproteinases 424 (TIMPs) that provides a negative feedback loop to regulate a temporal succession of events that 425 promote myocardial wound healing while limiting tissue damage.<sup>48</sup> MMP12 levels are 426 temporally fine-tuned after myocardial infarction, orchestrating a series of events including 427 inflammation, fibrosis, angiogenesis and collagen degradation in order to achieve an optimal 428 scar formation and prevent left ventricular dysfunction and heart failure prognosis.<sup>49</sup> In our 429 findings, we confirmed that MMP12 was a biomarker that increased expression after myocardial 430 infarction, and activation of MMP12 reduced risk of peripheral arterial disease and ischemic 431 stroke, but because of limitations on data availability of the exact time point when MMP12 432 protein level was quantified after ischemic injury, we were not able to assess the temporal 433 regulation of MMP12 during cardiac remodeling upon the injury. Further studies that include 434 disease progression endpoints can help corroborate and extend our findings to characterize 435 dynamic protein level changes of cardiovascular targets and biomarkers.

436

437 Concordance and differences of using *cis*- and *trans*-pQTLs as IVs in MR

438

Using *cis*- and *trans*-pQTLs as IVs represent fundamentally distinct principles of IV selection,
 thereby affecting interpretation of the forward-MR results. *cis*-pQTLs are physically proximal to
 sentinel variants, and therefore less likely to violate the MR assumption of pleiotropy, while
 *trans*-pQTLs include more IVs that provide greater statistical power and the IVs representing

443 genes that may have biological roles in regulating homeostasis of the associated proteins 444 through signaling transduction pathways or protein-protein interaction.<sup>50</sup> Forward-MR identified 445 221 proteins that showed significant causal effects on at least one CVD, among which 59 are 446 common targets identified using either cis- or trans-pQTLs. One hundred twenty-nine (68.6%) 447 targets were uniquely identified in the *trans*-pQTLs-based MR analysis, accounting for 81.7% of 448 targets with biological annotations that implied a pathophysiological role in CVD, suggesting the 449 value of using *trans*-pQTLs in the MR analysis for exploring novel therapeutic hypotheses. 450 Functional follow-up analyses are needed for validating *trans*-pQTLs derived targets. On the 451 contrary, some targets are found only in the *cis*-pQTLs-based MR analysis, for example, NOTCH1 452 for cardiac valvular disease. NOTCH1 is a Notch receptor that releases its intracellular domain as 453 transcription factor upon activation, which plays an essential role in cell fate determination, and cell proliferation, differentiation and apoptosis during organogenesis throughout the embryo.<sup>51</sup> 454 455 Loss-of-function mutations in NOTCH1 was found to cause aortic valve disease in autosomal-456 dominant human pedigrees that showed a wide spectrum of developmental aortic valve 457 anomalies and severe valve calcification.<sup>52</sup> Agonists of NOTCH1 may be able to treat calcification 458 of the aortic valve, the third leading cause of CVD in adults. Taken together, proteomic MR with 459 cis- and trans-pQTLs provides complementary approaches for therapeutic target identification 460 with overlapped but unparalleled insights.

461

462 Impact of applying different methods for multiple testing correction

463

464 The forward-MR generated more targets that reached the Bonferroni-corrected significance 465 threshold than biomarkers from the reverse-MR. We noted that Bonferroni-correction that 466 corrected for the total number of proteins and phenotypes analyzed in the phenome-wide MR 467 scan can be overly stringent due to intercorrelation within proteins and phenotypes. Using the 468 Bonferroni approach controlled type 1 error with a slight loss in power, given the relatively low 469 success rates for investigational targets passing through clinical trial development,<sup>53</sup> the strict 470 control on false positives at the first step of target discovery can help prioritize targets for 471 downstream development. If using a more lenient approach to adjust for multiple testing, such

472 as the Benjamini–Hochberg method of false discovery rate, we found the number of significant 473 proteins for forward-MR increased from 221 (7.52%) to 1,349 (69.54%), for reverse-MR from 16 474 (0.54%) to 1,045 (35.54%), while resulting in a larger number of significant findings, more of 475 which can be false positives that require extensive functional appraisal with labor-intensive and 476 time-consuming experiments developed on animal models. External validation using 477 independent cohorts can validate our results on top of a stringent significance threshold 478 applied, however, at the time when this manuscript is written up, the pQTLs we used are from 479 the largest proteomic GWAS recently performed under the UK Biobank Pharma Proteomics 480 Project.<sup>19</sup> Future validation is required when pQTLs from equally-powered, independent studies 481 are available. 482 483 Benefits and limitations in incorporating multi-ancestry element in MR study 484

485 Leveraging multi-ancestry data for drug discovery can help discover novel candidate causal 486 genes. For example, two nonsense mutations of PCSK9 first reported in humans were found in 487 an African-American population and associated with a substantial reduction of low-density lipoprotein cholesterol.<sup>54</sup> This finding demonstrated that a lifelong inhibition of PCSK9 protected 488 489 against coronary heart disease without noticeable safety issues. Multi-ancestry element has 490 recently been increasingly incorporated in population genetic studies, such as GWAS, finemapping and polygenic risk score studies, but rarely applied in MR studies.<sup>55,56</sup> In our study, we 491 492 identified 5 biomarkers exclusively detectable in the East Asian population, highlighting the 493 values of leveraging multi-ancestorial populations in identifying candidate causal signals. We 494 benefited from the relatively large sample size of the Biobank Japan, one of the largest non-495 European population cohorts with genome-wide genetic and medical records data available, 496 which provided adequate statistical power in constructing IVs of disease endpoints and thereby 497 detecting significant signals in the reverse MR, while hindered from the relatively small sample 498 size of the pQTL study in this population, limiting us from identifying more potential target 499 signals in the forward MR.

500

#### 501 Conclusions

502

- 503 Our study evaluated causal relationships between 3,000 circulating proteins and 21 CVD using
- 504 biobank-scale genetic association summary statistics for proteins and phenotypes in a bi-
- 505 directional MR framework. We prioritized novel candidate causal signals for CVD and
- 506 interrogated their pleiotropy and specificity in a phenome-wide causality survey. This study
- 507 provides human genetics-based evidence of novel candidate genes, a foundational step towards
- 508 full-scale causal human biology-based drug discovery for CVD.
- 509

# 510 **References**

511

- McClellan M, Brown N, Califf RM, Warner JJ. Call to Action: Urgent Challenges in
   Cardiovascular Disease: A Presidential Advisory From the American Heart Association.
   *Circulation*. 2019;139:e44-e54. doi: 10.1161/CIR.0000000000652
- Mensah GA, Wei GS, Sorlie PD, Fine LJ, Rosenberg Y, Kaufmann PG, Mussolino ME, Hsu
   LL, Addou E, Engelgau MM, et al. Decline in Cardiovascular Mortality: Possible Causes
   and Implications. *Circ Res*. 2017;120:366-380. doi: 10.1161/CIRCRESAHA.116.309115
- Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM,
   Beaton AZ, Boehme AK, Buxton AE, et al. Heart Disease and Stroke Statistics-2023
   Update: A Report From the American Heart Association. *Circulation*. 2023;147:e93-e621.
   doi: 10.1161/CIR.00000000001123
- Olivotto I, Oreziak A, Barriales-Villa R, Abraham TP, Masri A, Garcia-Pavia P, Saberi S,
   Lakdawala NK, Wheeler MT, Owens A, et al. Mavacamten for treatment of symptomatic
   obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double blind, placebo-controlled, phase 3 trial. *Lancet*. 2020;396:759-769. doi: 10.1016/S0140 6736(20)31792-X
- Peikert A, Goyal P, Vaduganathan M, Claggett BL, Kulac IJ, Miao ZM, Vardeny O,
   Kosiborod MN, Desai AS, Jhund PS, et al. Dapagliflozin in Heart Failure With Mildly
   Reduced or Preserved Ejection Fraction According to Polypharmacy Status. *JACC Heart Fail.* 2023. doi: 10.1016/j.jchf.2023.05.014
- Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, Januzzi J, Verma S, Tsutsui
   H, Brueckmann M, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart
   Failure. N Engl J Med. 2020;383:1413-1424. doi: 10.1056/NEJMoa2022190

534 7. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Bohm M, Brunner-La Rocca HP, Choi 535 DJ, Chopra V, Chuquiure-Valenzuela E, et al. Empagliflozin in Heart Failure with a 536 Preserved Ejection Fraction. N Engl J Med. 2021;385:1451-1461. doi: 537 10.1056/NEJMoa2107038 538 8. Armstrong PW, Pieske B, Anstrom KJ, Ezekowitz J, Hernandez AF, Butler J, Lam CSP, 539 Ponikowski P, Voors AA, Jia G, et al. Vericiguat in Patients with Heart Failure and Reduced 540 Ejection Fraction. N Engl J Med. 2020;382:1883-1893. doi: 10.1056/NEJMoa1915928 541 9. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, 542 Wang H, Liu T, Wasserman SM, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med. 2017;376:1713-1722. doi: 543 544 10.1056/NEJMoa1615664 545 Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, Stroes ES, Langslet G, 10. 546 Raal FJ, El Shahawy M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. N Engl J Med. 2015;372:1489-1499. doi: 547 10.1056/NEJMoa1501031 548 549 Hwang TJ, Lauffenburger JC, Franklin JM, Kesselheim AS. Temporal Trends and Factors 11. 550 Associated With Cardiovascular Drug Development, 1990 to 2012. JACC Basic Transl Sci. 551 2016;1:301-308. doi: 10.1016/j.jacbts.2016.03.012 Fordyce CB, Roe MT, Ahmad T, Libby P, Borer JS, Hiatt WR, Bristow MR, Packer M, 552 12. 553 Wasserman SM, Braunstein N, et al. Cardiovascular drug development: is it dead or just hibernating? J Am Coll Cardiol. 2015;65:1567-1582. doi: 10.1016/j.jacc.2015.03.016 554 555 13. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, 556 Wang J, et al. The support of human genetic evidence for approved drug indications. Nat 557 Genet. 2015;47:856-860. doi: 10.1038/ng.3314 558 14. Smith JG, Gerszten RE. Emerging Affinity-Based Proteomic Technologies for Large-Scale 559 Plasma Profiling in Cardiovascular Disease. Circulation. 2017;135:1651-1664. doi: 560 10.1161/CIRCULATIONAHA.116.025446 561 15. Ferrannini E, Manca ML, Ferrannini G, Andreotti F, Andreini D, Latini R, Magnoni M, 562 Williams SA, Maseri A, Maggioni AP. Differential Proteomics of Cardiovascular Risk and Coronary Artery Disease in Humans. Front Cardiovasc Med. 2021;8:790289. doi: 563 564 10.3389/fcvm.2021.790289 565 16. Henry A, Gordillo-Maranon M, Finan C, Schmidt AF, Ferreira JP, Karra R, Sundstrom J, 566 Lind L. Arnlov J, Zannad F, et al. Therapeutic Targets for Heart Failure Identified Using 567 Proteomics and Mendelian Randomization. Circulation. 2022;145:1205-1217. doi: 568 10.1161/CIRCULATIONAHA.121.056663

Shah S, Henry A, Roselli C, Lin H, Sveinbjornsson G, Fatemifar G, Hedman AK, Wilk JB,
Morley MP, Chaffin MD, et al. Genome-wide association and Mendelian randomisation
analysis provide insights into the pathogenesis of heart failure. *Nat Commun*.
2020;11:163. doi: 10.1038/s41467-019-13690-5

- 18. Rasooly D, Peloso GM, Pereira AC, Dashti H, Giambartolomei C, Wheeler E, Aung N,
  574 Ferolito BR, Pietzner M, Farber-Eger EH, et al. Genome-wide association analysis and
  575 Mendelian randomization proteomics identify drug targets for heart failure. *Nat*576 *Commun.* 2023;14:3826. doi: 10.1038/s41467-023-39253-3
- 577 19. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A,
  578 Robins C, Vasquez-Grinnell SG, et al. Plasma proteomic associations with genetics and
  579 health in the UK Biobank. *Nature*. 2023. doi: 10.1038/s41586-023-06592-6
- 58020.Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using581summary data. *Genet Epidemiol*. 2020;44:313-329. doi: 10.1002/gepi.22295
- 582 21. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D,
  583 Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and
  584 genomic data. *Nature*. 2018;562:203-209. doi: 10.1038/s41586-018-0579-z
- Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T,
  Yamamoto K, Akiyama M, et al. A cross-population atlas of genetic associations for 220
  human phenotypes. *Nat Genet*. 2021;53:1415-1424. doi: 10.1038/s41588-021-00931-x
- Tadros R, Francis C, Xu X, Vermeer AMC, Harper AR, Huurman R, Kelu Bisabu K, Walsh R,
  Hoorntje ET, Te Rijdt WP, et al. Shared genetic pathways contribute to risk of
  hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat Genet*.
  2021;53:128-134. doi: 10.1038/s41588-020-00762-2
- 592 24. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC,
  593 Griffith OL, Griffith M. DGIdb 3.0: a redesign and expansion of the drug-gene interaction
  594 database. *Nucleic Acids Res.* 2018;46:D1068-D1073. doi: 10.1093/nar/gkx1143
- Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Felix E, Magarinos MP,
  Mosquera JF, Mutowo P, Nowotka M, et al. ChEMBL: towards direct deposition of
  bioassay data. *Nucleic Acids Res.* 2019;47:D930-D940. doi: 10.1093/nar/gky1075
- 598 26. Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, Klein TE. An
  599 Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for
  600 Personalized Medicine. *Clin Pharmacol Ther*. 2021;110:563-572. doi: 10.1002/cpt.2350

601	27.	Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, Mathias SL, Nelson SJ, Oprea TI.
602		DrugCentral: online drug compendium. Nucleic Acids Res. 2017;45:D932-D939. doi:
603		10.1093/nar/gkw993

604 605 606	28.	Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Gunes O, Hall P, Hayhurst J, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. <i>Nucleic Acids Res</i> . 2023;51:D977-D985. doi: 10.1093/nar/gkac1010
607 608 609	29.	Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. ClinVar: improving access to variant interpretations and supporting evidence. <i>Nucleic Acids Res</i> . 2018;46:D1062-D1067. doi: 10.1093/nar/gkx1153
610 611 612	30.	Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. ClinGenthe Clinical Genome Resource. <i>N</i> <i>Engl J Med</i> . 2015;372:2235-2242. doi: 10.1056/NEJMsr1406261
613 614 615 616	31.	Chaffin M, Papangeli I, Simonson B, Akkad AD, Hill MC, Arduini A, Fleming SJ, Melanson M, Hayat S, Kost-Alimova M, et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. <i>Nature</i> . 2022;608:174-180. doi: 10.1038/s41586-022-04817-8
617 618 619	32.	Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. <i>Elife</i> . 2018;7. doi: 10.7554/eLife.34408
620 621 622	33.	Chen Y, Lun AT, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. <i>F1000Res</i> . 2016;5:1438. doi: 10.12688/f1000research.8987.2
623 624 625	34.	Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. <i>Genome Biol</i> . 2019;20:296. doi: 10.1186/s13059-019-1874-1
626 627	35.	Nachtigall M, Nachtigall RH. The Role of Lipoprotein(a) in Cardiovascular Diseases. <i>JAMA</i> . 2021;326:2078. doi: 10.1001/jama.2021.16747
628 629 630 631	36.	O'Donoghue ML, Fazio S, Giugliano RP, Stroes ESG, Kanevsky E, Gouni-Berthold I, Im K, Lira Pineda A, Wasserman SM, Ceska R, et al. Lipoprotein(a), PCSK9 Inhibition, and Cardiovascular Risk. <i>Circulation</i> . 2019;139:1483-1492. doi: 10.1161/CIRCULATIONAHA.118.037184
632 633 634 635	37.	Tsimikas S, Karwatowska-Prokopczuk E, Gouni-Berthold I, Tardif JC, Baum SJ, Steinhagen- Thiessen E, Shapiro MD, Stroes ES, Moriarty PM, Nordestgaard BG, et al. Lipoprotein(a) Reduction in Persons with Cardiovascular Disease. <i>N Engl J Med</i> . 2020;382:244-255. doi: 10.1056/NEJMoa1905239
636 637 638	38.	Hedhli N, Kalinowski A, K SR. Cardiovascular effects of neuregulin-1/ErbB signaling: role in vascular signaling and angiogenesis. <i>Curr Pharm Des</i> . 2014;20:4899-4905. doi: 10.2174/1381612819666131125151058

Maitusong B, Xie X, Ma YT, Fu ZY, Yang YN, Li XM, Liu F, Chen BD, Gai MT. Association
between ErbB3 genetic polymorphisms and coronary artery disease in the Han and
Uyghur populations of China. *Int J Clin Exp Med*. 2015;8:16520-16527.

- 40. Zhu L, Fang Z, Jin Y, Chang W, Huang M, Chen Y, Yao Y. Circulating ERBB3 levels are
  inversely associated with the risk of overweight-related hypertension: a cross-sectional
  study. *BMC Endocr Disord*. 2021;21:130. doi: 10.1186/s12902-021-00793-8
- Harada E, Mizuno Y, Shono M, Maeda H, Yano N, Tokunaga Y, Yasue H. Abstract 15352:
  Plasma Levels of BNP are Lower in Patients With Heart failure With Preserved EF (HFpEF)
  as Compared With Those With Reduced EF (HFrEF). *Circulation*. 2015;132:A15352A15352. doi: doi:10.1161/circ.132.suppl\_3.15352
- 42. Januzzi JL, Jr., Myhre PL. The Challenges of NT-proBNP Testing in HFpEF: Shooting Arrows
  in the Wind. *JACC Heart Fail*. 2020;8:382-385. doi: 10.1016/j.jchf.2020.03.003
- 43. Xiang M, Luo H, Wu J, Ren L, Ding X, Wu C, Chen J, Chen S, Zhang H, Yu L, et al. ADAM23
  in Cardiomyocyte Inhibits Cardiac Hypertrophy by Targeting FAK AKT Signaling. J Am *Heart Assoc.* 2018;7:e008604. doi: 10.1161/JAHA.118.008604
- 44. Yagyu H, Chen G, Yokoyama M, Hirata K, Augustus A, Kako Y, Seo T, Hu Y, Lutz EP, Merkel
  M, et al. Lipoprotein lipase (LpL) on the surface of cardiomyocytes increases lipid uptake
  and produces a cardiomyopathy. *J Clin Invest*. 2003;111:419-426. doi: 10.1172/JCI16751
- 45. Zee RY, Romero JR, Gould JL, Ricupero DA, Ridker PM. Polymorphisms in the advanced
  glycosylation end product-specific receptor gene and risk of incident myocardial
  infarction or ischemic stroke. *Stroke*. 2006;37:1686-1690. doi:
  10.1161/01.STR.0000226994.93914.6c
- 46. Barbaux SC, Blankenberg S, Rupprecht HJ, Francomme C, Bickel C, Hafner G, Nicaud V,
  Meyer J, Cambien F, Tiret L. Association between P-selectin gene polymorphisms and
  soluble P-selectin levels and their relation to coronary artery disease. *Arterioscler Thromb Vasc Biol.* 2001;21:1668-1673. doi: 10.1161/hq1001.097022
- 47. Horiba M, Kadomatsu K, Yasui K, Lee JK, Takenaka H, Sumida A, Kamiya K, Chen S,
  Sakuma S, Muramatsu T, et al. Midkine plays a protective role against cardiac
  ischemia/reperfusion injury through a reduction of apoptotic reaction. *Circulation*.
  2006;114:1713-1720. doi: 10.1161/CIRCULATIONAHA.106.632273
- Mouton AJ, Rivera Gonzalez OJ, Kaminski AR, Moore ET, Lindsey ML. Matrix
  metalloproteinase-12 as an endogenous resolution promoting factor following
  myocardial infarction. *Pharmacol Res.* 2018;137:252-258. doi:
- 672 10.1016/j.phrs.2018.10.026

673 49. DeLeon-Pennell KY, Meschiari CA, Jung M, Lindsey ML. Matrix Metalloproteinases in
674 Myocardial Infarction and Heart Failure. *Prog Mol Biol Transl Sci.* 2017;147:75-100. doi:
675 10.1016/bs.pmbts.2017.02.001

- 50. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, Gutteridge A, Erola P, Liu
  57 Y, Luo S, et al. Phenome-wide Mendelian randomization mapping the influence of the
  678 plasma proteome on complex diseases. *Nat Genet*. 2020;52:1122-1131. doi:
  679 10.1038/s41588-020-0682-6
- 51. Southgate L, Sukalo M, Karountzos ASV, Taylor EJ, Collinson CS, Ruddy D, Snape KM,
  Dallapiccola B, Tolmie JL, Joss S, et al. Haploinsufficiency of the NOTCH1 Receptor as a
  Cause of Adams-Oliver Syndrome With Variable Cardiac Anomalies. *Circ Cardiovasc Genet.* 2015;8:572-581. doi: 10.1161/CIRCGENETICS.115.001086
- 684 52. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD,
  685 Srivastava D. Mutations in NOTCH1 cause aortic valve disease. *Nature*. 2005;437:270686 274. doi: 10.1038/nature03940
- 68753.Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human688genetics. Nat Rev Drug Discov. 2013;12:581-594. doi: 10.1038/nrd4051
- 689 54. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low
  690 LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354:1264-1272.
  691 doi: 10.1056/NEJMoa054013
- 55. Zhao H, Rasheed H, Nost TH, Cho Y, Liu Y, Bhatta L, Bhattacharya A, Global Biobank Metaanalysis I, Hemani G, Davey Smith G, et al. Proteome-wide Mendelian randomization in
  global biobank meta-analysis reveals multi-ancestry drug targets for common diseases. *Cell Genom*. 2022;2:None. doi: 10.1016/j.xgen.2022.100195
- 56. Zheng J, Zhang Y, Zhao H, Liu Y, Baird D, Karim MA, Ghoussaini M, Schwartzentruber J,
  Dunham I, Elsworth B, et al. Multi-ancestry Mendelian randomization of omics traits
  revealing drug targets of COVID-19 severity. *EBioMedicine*. 2022;81:104112. doi:
  10.1016/j.ebiom.2022.104112
- 700