

1 **Proteome-wide Mendelian Randomization Analysis Identified Potential Drug**
2 **Targets for Myocardial Infarction**

3

4 Xufeng Ye¹, Xinpei Wang¹, Jinzhu Jia^{1,2*†}

5 1Department of Biostatistics, School of Public Health, Peking University, 38 Xueyuan Road,
6 Haidian District, 100191, China.

7 2Center for Statistical Science, Peking University, 100871, China.

8

9 *Corresponding author(s). E-mail(s): jzjia@math.pku.edu.cn;

10 Contributing authors: yexufeng2017@pku.edu.cn; w-xp@pku.edu.cn;

11 †These authors contributed equally to this work.

12

13 **Subtitle: Potential Drug Targets for Myocardial Infarction**

14 **Word count:**

15 **#Correspondence and requests for reprint:**

16 Dr. Jinzhu Jia, PhD, Department of Biostatistics, School of Public Health, Peking University,
17 China. 38 Xueyuan 14 Road, Beijing, 100191 China. Phone: (86) 010-82801528. Email:
18 jzjia@pku.edu.cn

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

Abstract

Background: Myocardial infarction (MI) is a leading cause of global mortality. Finding effective drugs to treat MI is an urgent concern for clinicians. Proteome-wide Mendelian randomization (MR) analysis provides a new way to investigate invaluable therapeutic drug targets more efficiently.

Methods: Using a proteome-wide MR approach, we assessed the genetic predictive causality between thousands of plasma proteins and MI risk. First, by adopting several principles to judge genetic variants associated with plasma proteins and MI risk, we selected a series of suitable variants utilized as instrumental variables (IVs) for the latter Mendelian Randomization (MR) analysis. Second, we performed a proteome-wide MR analysis to select candidate proteins. Third, sensitivity tests including heterogeneity test, reverse causality test, and colocalization analyses were conducted to ensure the robustness of our selected protein. Last, we assessed the drugability of the identified potential drug targets for MI using databases including DrugBank, PharmGKB, and TTD.

Results: Of the identified IVs, 3,156 associated with 1,487 plasma proteins were validated. 15 proteins exhibited significant genetically predicted causal associations (P – value $< 3.362 \times 10^{-5}$) with MI risk, including Plasmin, MSP, Apo B, TAGLN2, LRP4, C1s, Angiostatin, Apo C-III, PCSK9, ANGL4, FN1.4, Apo B, IL-6 sRa, SWAP70, FN, FN1.3. Sensitivity analyses pinpointed Plasmin and Angiostatin for heterogeneity and proteins MSP, Apo B, and Angiostatin for reverse causality effects. Colocalization analysis found several proteins sharing genetic variants with MI, notably Apo B, TAGLN2, LRP4, C1s, Apo C-III, PCSK9 and ANGL4. When the threshold was lowered to 0.7, additional variants SWAP70 could be contained. 7 potential drug targets for MI were identified: SWP70, TAGLN2, LRP4, C1s, Apo C-III, PCSK9, and ANGL4. Drugability assessment categorized these proteins into varying therapeutic potential categories, from successfully drugged targets to those only reported in the literature.

Conclusion: Our comprehensive study elucidated 7 promising drug targets offering profound insights into its molecular dynamics and presenting potential pathways for therapeutic interventions against MI.

Keywords: Myocardial infarction, plasma protein, drug target, Mendelian randomization

1
2
3

4 **Clinical Perspective**

5 1) What Is New?

6 ★ The analysis of thousands of proteins has identified 7 proteins that have a potential causal role
7 in myocardial infarction risk.

8 ★ Four of these ten proteins have drugs approved or in development that target them, and three 5
9 have not been previously reported to be associated with atrial fibrillation risk.

10

11 2) What Are the Clinical Implications?

12 ★ The results of the present study demonstrate new potential drug/therapeutic targets for
13 myocardial infarction.

14

15

16 **Non-standard Abbreviations and Acronyms**

MI	myocardial infarction
FDA	Food and Drug Administration
GWAS	genome-wide association study
IV	instrumental variable
IVW	inverse variance weighted method
LD	linkage disequilibrium
MAF	minor allele frequency
MHC	major histocompatibility complex
MR	mendelian randomization
pQTL	quantitative trait locus
SNP	single nucleotide polymorphism
TTD	therapeutic target database

17

18

19

20

21

22

1

2

3 **Introduction**

4 Myocardial Infarction (MI) ranks among the primary cardiovascular diseases world-wide, which
5 mainly originates from insufficient or interrupted blood supply via the coronary arteries, leading to
6 cardiac muscle damage or necrosis [1]. Despite advancements in risk factor management and
7 pharmaceutical interventions aiming to alleviate the burden of MI, its incidence and prevalence
8 have steadily surged in the past two decades and will be rising in the forthcoming decades [2].
9 Contemporary therapeutic strategies, encompassing both extant and drugs under development,
10 often come under scrutiny due to concerns regarding their safety and efficiency [3]. Consequently,
11 the pursuit for more efficient and safe drug targets is at the forefront of clinicians and researchers'
12 priorities.

13

14 *Proteomic Considerations in MI Pathophysiology*

15 Proteins play a vital role in various biological processes at the cellular level, establishing
16 themselves as crucial components for potential drug targets [4–6]. Numerous studies have
17 underscored a significant correlation between levels of various plasma proteins and the risk of MI
18 [7–9]. However, the inherent constraints of observational research, such as residual confounding
19 and potential reverse causality, render these significant observations non-representative of clear
20 causal relationships [10, 11].

21

22 *Advancements in Cardiovascular Risk Factor Targeting*

23 Cardiovascular risk factor identification and targeted causal inference have made remarkable
24 strides in the prevention and treatment of coronary artery diseases, including MI. Propelled by
25 modern medical advancements, epidemiological studies have investigated hundreds of biomarkers
26 (circulating molecules, physiological traits, and other pathologies) and targets linked with
27 coronary artery disease. [12, 13] The potency and replicability of associations, coupled with the
28 inherent mechanisms delineated by foundational medicine, have led researchers to causally link
29 certain novel biomarkers and targets with MI [14, 15]. Subsequent interventions aiming to prevent
30 these identified “risk factors” have been initiated. However, differences often emerge between
31 these findings and those observed in epidemiological settings [16]. The ambiguity surrounding the
32 causal roles of numerous biomarkers during disease progression together with the steep costs of
33 drug development programs reveals the necessity to distinguish between “causality” and mere
34 “association” [17].

35

36 *Mendelian Randomization (MR) as a Research Tool*

37 MR offers a robust statistical apparatus, furnishing a unique approach to investigating causal
38 relationships between exposures and outcomes [18, 19]. By using genetic variations as
39 instrumental variables to estimate the causal effects of plasma proteins on MI, we can obtain more

1 precise causal estimates than conventional observational studies [14–17]. Notably, MR studies on
2 plasma proteins and MI predominantly feature a few proteins grounded in extant knowledge,
3 sidelining a variety of other proteins potentially correlated with MI risk.

4

5 *Research Aims and Methodology*

6 This study endeavors to leverage the MR framework to holistically analyze potential causal
7 relationships between thousands of plasma proteins and MI risk, streamlining the identification of
8 novel MI drug targets. Our methodology incorporates the selection of independent genetic loci,
9 significantly correlated with protein levels and proximate to protein-encoding gene locations, as
10 instrumental variables, substantially mitigating the possibility for violating instrumental variable
11 assumptions. We further fortify our results through a series of sensitivity analyses. We anticipated
12 that this investigation will shed comprehensive insights, serving as a cornerstone for MI drug
13 target research and development.

14

15 **METHODS**

16 **Data Source And Disclosure Statement**

17 The database for genetic variations associated with plasma proteins has been sourced from
18 deCODE genetics/Amgen, Inc., Reykjavik, Iceland. The database for genetic variations associated
19 with myocardial infarction (MI) has been obtained from two major repositories: the UK BioBank
20 and the CARDIoGRAM+C4D Consortium. Both of these databases are publicly accessible, in
21 accordance with data sharing policies.

22

23 **Discovery analysis**

24 **Mendelian randomization**

25 Figure 1 illustrates the working principle of Mendelian randomization. Firstly, we define certain
26 mathematical symbols as depicted in the figure. We denote the independent instrumental variables
27 (IVs) as g_i . Here, X represents exposure, Y stands for the outcome, and U signifies the
28 confounders. The aggregate effects of g_i on X and Y are represented as β_{X_i} and β_{Y_i} , respectively.

29

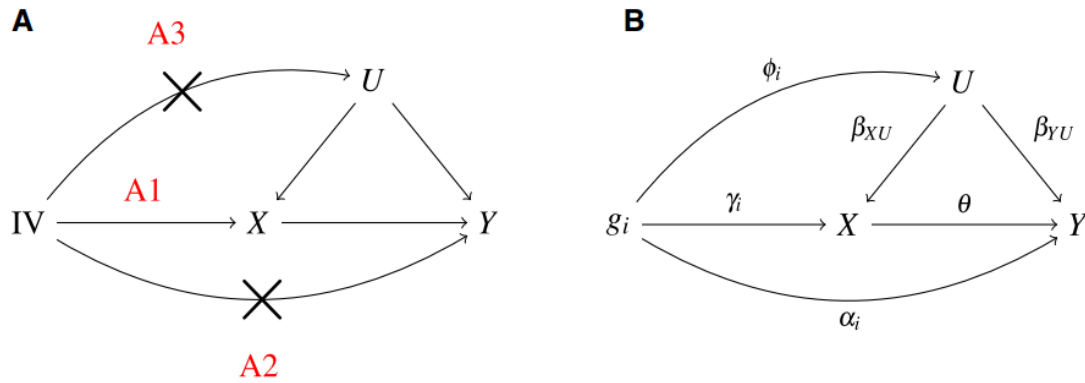


Fig. 1 The working principle of Mendelian randomization

To avoid the potential for biased inference from horizontal pleiotropy (where IVs affect Y not just through X), Mendelian randomization requires three core instrumental variable (IV) assumptions [20]:

(A1) Relevance assumption: The genetic variants used as instruments are associated with the exposure. Mathematically, the IV is correlated with exposure X; that is, $\gamma_i \neq 0$.

(A2) Independence assumption: The genetic variation and the outcome of interest share no common causes (i.e., no confounders). Mathematically speaking, the IV is uncorrelated with the outcome Y given the exposure X; that is, $\alpha_i = 0$.

(A3) Exclusion restriction assumption: There are no independent pathways between the genetic variation and the outcome apart from through the exposure. Mathematically, the IV is uncorrelated with unmeasured confounders U; namely, $\phi_i = 0$.

Given the above assumptions, we have equations (1) and (2):

$$\beta_{Xi} = \gamma_i + \beta_{XU} \cdot \phi_i \quad (1)$$

$$\begin{aligned} \beta_{Yi} &= \theta \cdot \beta_{Xi} + \alpha_i + \beta_{YU} \cdot \phi_i \\ &= \theta \cdot \gamma_i + \beta_{XU} \cdot \phi_i + \alpha_i + \beta_{YU} \cdot \phi_i \end{aligned}$$

$$= \theta \cdot \beta_{Xi} + r_i \quad (\text{where } r_i = \alpha_i + \beta_{YU} \cdot \phi_i) \quad (2)$$

Under the circumstance of assumption (A2) and (A3), $\alpha_i = 0$ $\phi_i = 0$. Then, we collect causal effect $\theta = \beta_{Yi} / \beta_{Xi}$, which is called wald ratio.

IV Selection and Validation

Genetic variations utilized as IVs need to adhere to the three assumptions previously stated. The first of these assumptions postulates a correlation between genetic variations and the exposure, (various types of plasma proteins). To substantiate this assumption, we drew upon the findings of a genome-wide association study (GWAS) conducted by Feringstad, E. et al[21]. This study delved

1 into the associations of 27.2 million genetic variants with 4,907 aptamers, quantifying 4,719
2 plasma proteins in a sample of 35,559 Icelanders (courtesy of the Icelandic Cancer Project and
3 deCODE genetics, Reykjavík, Iceland). Subsequent to this, we applied rigorous filters to this pool
4 of pQTLs, eliminating: 1. Non-biallelic single nucleotide polymorphism (SNPs); 2. SNPs with
5 minor allele frequency (MAF) < 0.01; 3. Variants located either on the sex chromosomes or within
6 the human major histocompatibility complex (MHC) region.

7
8 To satisfy Assumptions 2 and 3, we employed more stringent criteria in selecting pQTLs: 1. Only
9 pQTLs located within 200kb upstream of the transcription start site or 200kb downstream of the
10 termination site of their corresponding protein-coding genes were included and labeled as
11 cis-pQTLs; 2. For these cis-pQTLs, we counted the number of proteins associated with each
12 pQTL, as well as those linked to this pQTL in linkage disequilibrium ($LD R^2 > 0.5$). We then
13 excluded pQTLs that had associations with more than five proteins. It's worth noting that our
14 criteria for selecting cis-pQTLs are stricter than those in many standard studies. Our primary aim
15 is to ensure the instrumental variables (IVs or pQTLs) we choose influence the outcome (MI)
16 solely through their corresponding proteins, thus minimizing the impact of horizontal pleiotropy
17 on our MR estimates.

18
19 Finally, to mitigate the impact of correlation between IVs on the MR estimates, we clumped the
20 above-selected IVs (pQTLs). In alignment with Albert's research, we controlled LD R^2 between
21 IVs below 0.4[22]. This threshold is backed by evidence suggesting that the issue of unstable
22 coefficient estimation due to collinearity begins to manifest when the LD R^2 between independent
23 variables exceeds 0.36[23]. Subsequently, we assessed the strength of each IV for every protein.
24 Proteins with insufficient IV strength (with an F – statistic < 10) were excluded, ensuring that the
25 IVs we employed effectively represented genetic proxies of protein levels. Through these rigorous
26 quality control measures, we ensured our chosen IVs for MR analysis faithfully adhered to the
27 foundational three assumptions of Mendelian Randomization.

28 29 **Statistical analysis**

30 For our primary MR analysis, we utilized the IVs shortlisted from the aforementioned
31 procedures. The data for our outcome variable, MI, was extracted from a GWAS encompassing
32 approximately 831,000 individuals of European ancestry, an amalgamation of several contributory
33 cohorts. Within this, the UK Biobank provided data on 17,505 MI cases and 454,212 controls. It
34 pinpointed 1,966 single-nucleotide polymorphisms (SNPs) spread across 31 loci, all exhibiting
35 associations with MI and surpassing the genome-wide significance threshold ($P=5*10^{-8}$). In
36 contrast, the CARDIoGRAM+C4D Consortium contributed information on 61,000 MI cases,
37 578,000 controls, and a sweeping range of 8,126,035 SNPs. Of these, 4,419 variants, dispersed
38 over 80 loci, were deemed significant.

39
40 In scenarios where only one IV corresponded to a protein, we employed the wald ratio method to
41 estimate the causal effect. However, for proteins linked to two or more IVs, we use MR-Egger
42 [24], the inverse variance weighted method (IVW) [25], weighted mode method [26], simple
43 mode method [24] and weighted median method [27] for our MR estimation. To ensure precision,

1 we use Bonferroni correction to adjust the MR results. Associations reflecting a P –value < 0.05/
2 (the number of proteins assessed in MR) were classified as significant. We use R package
3 “TwoSampleMR” (<https://github.com/MRCIEU/TwoSampleMR>) to execute MR analysis.
4

5 Sensitivity analysis

6 Horizontal Pleiotropy Test

7 Horizontal pleiotropy arises when a genetic variation affects the outcome via path- ways other
8 than the primary exposure, which can introduce biases into previously calculated MR estimates.
9 To address this, we employ Egger regression [28] to assess the presence of horizontal pleiotropy.
10 Theoretically, in the absence of horizontal pleiotropy, IVs impact the outcome solely through the
11 exposure. Within this regression model, the slope quantifies the effect of exposure on the outcome,
12 while the intercept serves as an indicator for pleiotropy. A significant deviation of the intercept
13 from zero signals the existence of horizontal pleiotropy.
14

15 Heterogeneity Test

16 A fundamental assumption of Mendelian randomization (MR) is that genetic variations correlate
17 with the exposure but are independent of any confounding factors that could bias the causal
18 association between the exposure and the outcome. However, the genetic variations utilized as
19 instrumental variables (IVs) might sometimes violate this assumption. This can occur if they are
20 related to other factors influencing the outcome or different genetic variations impact the exposure
21 through diverse pathways. In MR analysis, the heterogeneity test serves as an essential tool to
22 assess if the genetic variations used as IVs align with the MR assumptions.

23 A widely employed method to conduct this heterogeneity test is Cochran’s Q test[29]. The specific
24 steps are:

25 (1) For each genetic variation, define $\hat{\beta}_i^{IV} = \frac{\hat{\beta}_{Yi}}{\hat{\beta}_{Xi}}$ with an approximate standard error given by

26 $SE(\hat{\beta}_i^{IV}) = \frac{SE(\hat{\beta}_{Yi})}{\hat{\beta}_{Xi}}$, where $\hat{\beta}_{Xi}$ and $\hat{\beta}_{Yi}$ represent the effects of the IV on the exposure and
27 outcome, respectively.

28 (2) Compute the statistic $Q = \sum_i \omega_i (\hat{\beta}_i^{IV} - \hat{\beta}^{IV})^2$, where $\omega_i = SE(\hat{\beta}_i^{IV})^{-2}$ is the inverse
29 variance weight and $\hat{\beta}^{IV} = \frac{\sum_i \omega_i \hat{\beta}_i^{IV}}{\sum_i \omega_i}$ is the inverse variance-weighted estimate based on all
30 genetic variations (fixed effects).

31 (3) Under the null hypothesis, the statistic Q follows a chi-squared distribution $\chi^2(I - 1)$. If this
32 is significant, it indicates heterogeneity among the genetic variations, suggesting potential
33 violations of MR assumptions.
34

35 Reverse causation

36 Although the motivation of Mendelian randomization is to avoid bias from unmeasured

1 confounding and reverse causation, the method may invalidate causal conclusions in several
2 circumstances [30]: Genetic association with the risk factor is not primary, Feedback mechanism
3 and Cross-generational effects. To enhance the robustness of our Mendelian randomization, we
4 implemented MR-Steiger method [31] to undertake a directionality test. Additionally, we conduct
5 a reverse Mendelian Randomization to verify whether there exists causation from MI to plasma
6 proteins, which utilize MI as an exposure and plasma proteins as outcome.

8 **Colocalization**

9 Proximal variants often inherit together and therefore are usually correlated, which is termed as
10 linkage disequilibrium law in genetics. Two traits might be causally influenced by distinct variants
11 that happen to be correlated, providing a pathway between genetic variation and an outcome not
12 mediated by the exposure, which could violate the exclusion restriction assumption of Mendelian
13 randomization. We employ colocalization analysis to determine whether a specific genetic locus's
14 variant is associated with both a plasma protein (a particular exposure) and MI (a health outcome).
15 The central premise of colocalization analysis is to determine whether a particular variation at a
16 genetic locus is associated with both an exposure and a health outcome. We use R package "coloc"
17 (<https://github.com/chriswallace/coloc>) to execute colocalization analysis.

18 **Potential Drug target**

19 To investigate the current state of drug development focused on these identified proteins, we
20 consulted three specialized databases: the DrugBank Database, the PharmGKB Database, and the
21 Therapeutic Target Database (TTD), each with its own unique features. The DrugBank Database
22 [32] offers comprehensive data on FDA- approved as well as experimental drugs, including details
23 on drug targets, mechanisms of action, and drug interactions. PharmGKB[33], provides
24 international drug labeling information over various countries, including the United States,
25 Canada, Europe, Switzerland, and Japan. The TTD Database [34] offers valuable literature-based
26 information on known therapeutic protein and nucleic acid targets, targeted disease conditions,
27 pathways information and the respective drugs or ligands for each target.

29 **Ethics statement**

30 Patients or the public were not involved in the design, conduct, reporting or dissemination plans of
31 our research.

32

1 Results

2 IV selection and validation

3 Ferkingstad, E. et al's research delved into the associations of 27.2 million genetic variants with
4 4,907 aptamers, quantifying 4,719 plasma proteins in a sample of 35,559 Icelanders (courtesy of
5 the Icelandic Cancer Project and deCODE genetics, Reykjavik, Iceland). From this extensive
6 dataset, we earmarked 4,631 proteins as potential exposures and identified 28,191 pQTLs
7 associated with these proteins (P -value $< 1.8 \times 10^{-9}$, (0.05/27.2 million)) as candidate IVs for our
8 Mendelian Randomization (MR) analysis. After excluding IVs to satisfy the assumption of MR
9 analysis, we select 3674 IVs associated with 1502 plasma proteins.

10

11 Then subsequent to clumping at an LD R^2 threshold of 0.4 to obtain IVs for each exposure
12 (plasma protein) and calculated the IV strength for each protein to ensure that these IVs were a
13 good genetic proxy for the corresponding plasma protein level (F-statistics were all greater than
14 20), we validate 3156 IVs associated with 1487 plasma proteins. Take IV-outcome dataset into
15 consideration, we harmonize the IV-exposure data and IV-outcome data, we have 2965 IVs
16 associated with 1422 plasma proteins. See Supplementary Material(IV selection and
17 validation.xlsx) for more information.

18

19 Proteome-wide MR analysis identified 15 plasma proteins 20 significantly associated with MI risk

21 After primary MR analysis, 15 of these 1,422 proteins showed a significant genetically predicted
22 causal association with MI risk (P -value $< 3.510 \times 10^{-5}$ (0.05/1,422))(Table 1). Considering that
23 protein levels were rank-inverse normal transformed in the protein GWAS, for MR results, we
24 paid more attention to the direction of the estimates than the magnitude of the estimates. Among
25 these 15 significant proteins(Table 1), genetically predicted plasma levels of
26 Plasminogen(Plasmin), Hepatocyte growth factor-like protein(MSP), Apolipoprotein B(Apo B),
27 Transgelin-2(TAGLN2), Low-density lipoprotein receptor-related protein 4(LRP4), Complement
28 C1s subcomponent(C1s), Angiostatin, Apolipoprotein C-III(Apo C-III), Proprotein convertase
29 subtilisin/kexin type 9(PCSK9), Angiotensin-converting enzyme 2(ACE2), Angiopoietin-related protein 4(ANGPT4) were associated with
30 higher MI risk, while the genetically predicted plasma levels of Fibronectin Fragment 4(FN1.4),
31 Apolipoprotein B(Apo B), Interleukin-6 receptor subunit alpha(IL-6 sRa), Switch-associated
32 protein 70(SWAP70), Fibronectin, Fibronectin Fragment 3(FN1.3) were associated with lower MI
33 risk.

34

1 **Sensitivity Analysis demonstrated the robustness of 7** 2 **candidate potential drug targets**

3 **Horizontal pleiotropy test**

4 First, we performed a horizontal pleiotropy test to judge whether several SNPs associated with one
5 protein has horizontal pleiotropy. According to Table 2, all the selected proteins passed the
6 horizontal pleiotropy test with egger intercept apart from zero, indicating there exists no horizontal
7 pleiotropy in the selected SNPs, fitting the assumption of Mendelian randomization.

8 9 **Heterogeneity test**

10 Secondly, we conducted a heterogeneity test sensitivity analysis. Our results revealed that proteins
11 Plasmin and Angiostatin had Cochran's Q P-values lower than 0.05. This suggests significant
12 heterogeneity among these proteins, violating the assumption of Mendelian randomization. This
13 warrants special consideration in subsequent analyses.

14 15 **Reverse Causality Analysis**

16 Thirdly, a reverse causality analysis was conducted, indicating that proteins MSP, Apo B,
17 Angiostatin demonstrated reverse causality effects on MI. This could mean that myocardial
18 infarction (MI) caused the expression of these plasma proteins, which, as originally thought, are
19 not our founding potential targets for MI.

20 21 **Colocalization Analysis**

22 Lastly, in the colocalization analysis, at a threshold of 0.8, several proteins were found to share the
23 same variants with MI(Table 3), including Apo B (rs12713844; rs563290),
24 Transgelin-2(rs2789422), LRP4(rs2306029; rs4752946), C1s(rs12146727), Apo C-III (rs964184),
25 PCSK9(rs11591147; rs472495; rs505151; rs58255540) and ANGL4(rs116843064). When the
26 threshold was lowered to 0.7, additional variants in SWAP70 (rs10743116; rs12271923;
27 rs2645017; rs415895) were also deemed to share the same variants with MI.

28
29 Through this comprehensive set of analyses, we identified 7 candidate potential drug targets
30 related to myocardial infarction (MI) more accurately, including SWP70, TAGLN2, LRP4, C1s,
31 Apo C-III, PCSK9 and ANGL4, while also shedding light on possible false positives and false
32 negatives. These findings will help deepen our understanding of the molecular mechanisms
33 underlying MI and could have a positive impact on future drug development targeting this
34 condition.

35 36 **Drugability assessment of potential drug targets**

37 We searched DrugBank, PharmGKB, and TTD databases to fully evaluate the Drugability of our
38 selected potential drug targets. According to the 3 databases, potential drug targets were classified
39 into four categories: Successful Target (drugs targeting this protein have been successfully

1 approved), Clinical trial Target (drugs targeting this protein are undergoing clinical trials),
2 Literature-reported Target (this protein has been reported in the literature as a potential drug
3 target), Not currently listed as druggable (no research using this protein as a drug target has been
4 retrieved).(Table 4) In general, C1s is a successful target for Hereditary angioedema and Cold type
5 autoimmune haemolytic anemia; Apo C-III, PCSK9 are Clinical trial Targets ;TAGLN2 and
6 ANGL4, are Literature-reported Target ; LRP4 and SWP70 is Not currently listed as druggable.
7

8 Discussion

9 In the present study, we used proteome-wide GWAS data to conduct a Mendelian randomization
10 analysis. After a series of stringent quality control measures to validate IVs, we obtained IVs that
11 satisfied the three core assumptions of MR. Using these IVs, we identified 15 candidate plasma
12 proteins (Plasminogen(Plasmin), Hepatocyte growth factor-like protein(MSP), Apolipoprotein
13 B(Apo B), Transgelin-2(TAGLN2) , Low-density lipoprotein receptor-related protein 4(LRP4),
14 Complement C1s subcomponent(C1s), Angiostatin, Apolipoprotein C-III(Apo C-III), Proprotein
15 convertase subtilisin/kexin type 9(PCSK9), Angiopoietin-related protein 4(ANGL4),Fibronectin
16 Fragment 4(FN1.4), Apolipoprotein B(Apo B), Interleukin-6 receptor subunit alpha(IL-6 sRa) ,
17 Switch-associated protein 70(SWAP70), Fibronectin, Fibronectin Fragment 3(FN1.3)) casually
18 associated with MI. A full range of sensitivity analyses confirmed the robustness in the primary
19 analysis, selecting 7 candidate plasma proteins including PCSK9,LRP4,APO
20 C-III,ANGL4,SWP70,TAGLN2 and C1s. Finally, by searching multiple drug databases, we have a
21 full understanding of the current status of drug development targeting these proteins, and suggest
22 the possibility of multiple uses for approved drugs and new MI therapeutic targets.
23

24 Protein targets associated with decrease of LDL-C

25 Elevated cholesterol levels, especially high levels of low-density lipoprotein cholesterol (LDL-C),
26 increase the risk of atherosclerosis (also known as coronary heart disease). In this condition,
27 cholesterol and other lipids accumulate on the walls of the blood vessels, forming “plaques”.
28 These plaques can lead to arterial narrowing and hardening, restricting blood flow. When the
29 coronary arteries (the main vessels supplying blood to the heart) are severely narrowed or blocked,
30 the rupture of an unstable plaque followed by subsequent blood clot formation can trigger a
31 myocardial infarction (heart attack). High LDL-C levels increase the risk of the formation and
32 rupture of such unstable plaques. [35]

34 PCSK9

35 Proprotein convertase subtilisin/kexin type 9 (PCSK9) is A successful target for Familial
36 hypercholesterolemia and Chronic obstructive pulmonary disease. The Low- Density Lipoprotein
37 Receptor (LDLR) modulates blood cholesterol levels by binding to and clearing circulating LDL.
38 It acts on low-density lipoprotein receptors (LDLR) and targets them for degradation, thus
39 reducing the liver’s ability to remove LDL- cholesterol from the bloodstream, playing a crucial
40 role in cholesterol metabolism. Inhibitors of PCSK9 are used to treat high levels of
41 LDL-cholesterol. [36]

1

2 Several research reveals that the use of Proprotein Convertase Subtilisin/Kexin type 9 (PCSK9)
3 inhibitors (monoclonal antibodies) can reduce LDL-C levels. Fourier et al.'s experiment [37]
4 revealed that treatment with Evolocumab significantly reduced the risk of the composite outcome
5 of cardiovascular death, MI, stroke, unstable angina, or coronary revascularization by 19%. The
6 ODYSSEY outcome trial [38, 39] also showed similar results. Alirocumab significantly reduced
7 the risk of the primary endpoint by 15% (HR 0.85, 95%CI 0.78–0.93, P = 0.003). In a recent
8 meta-analysis of 67 randomized controlled trials (RCTs) [40] including 259,429 participants,
9 PCSK9 inhibitors plus statin significantly reduced the risk of non-fatal MI (RR 0.82,95% CI 0.72–
10 0.93, p=0.003) or stroke (RR 0.74,95% CI 0.65–0.85, p<<0.001) . This provides strong support for
11 future drug research aiming at improving myocardial infarction outcomes by targeting this
12 mechanism.

13

14 *LRP4*

15 LRP4 is not currently listed as druggable in the drug database, however its super family LRP5 and
16 LRP6 are clinical trial Targets for solid tumour/cancer through Wnt/ β -catenin signaling pathway
17 [41, 42],where LRP4 also holds a pivotal position. Wnt/ β -catenin signaling, a complex and highly
18 conserved cell signaling system, plays a crucial role in the patterning, morphogenesis and growth
19 of a variety of tissues and organs during development and in homeostasis in the adult, While
20 Aberrant Wnt signaling activity is causally linked to degenerative diseases and cancers [43, 44],
21 MI included. Three main phases, inflammation, angiogenesis and fibrosis [45, 46], widely
22 recognized following MI, may exacerbate Wnt/ β -catenin signaling and interfere with the balanced
23 cardiac remodeling, thus these events could aggravate tissue damage and cause heart failure.

24

25 Recent research has found new therapeutic strategy to treat MI, using inhibitors of the canonical
26 Wnt/ β -catenin signaling pathway [47, 48],like anti-cancer drugs in patients after MI to prevent
27 heart failure. Early using of these drugs in the first days after myocardial infarction can could
28 inhibit Wnt signaling pathway, limit the loss of cardiomyocytes and fibrosis, thus promote
29 myocardial regeneration, prevent adverse cardiac remodeling and improve LV function. [49] Dkk1
30 is one representative of the inhibitors which can enhance ischemic injury by promoting the
31 internalization and elimination of basal LRP4/5.[50]

32

33 **Protein targets associated with balance of TG**

34 *Apo C-III*

35 Apolipoprotein C-III (ApoCIII) is a clinical trial target, while its messenger RNA (APOC3 mRNA)
36 has become a successful target for Hypertriglyceridemia and Hyperlipoproteinemia. This protein
37 has its unique proatherogenic effect on metabolism dysfunction of TGs and other various lipid
38 molecules, as well as its influence on acceleration of pro-inflammatory effects between monocytes
39 and endothelial cells, causing the high risk of MI. [51] In a follow-up randomized controlled trial,
40 Mendivil et al. [52] highlighted that the coronary heart disease risk posed by LDL is primarily due
41 to the presence of apoC-III within the LDL.

42

43 In a retrospective case–control study, Gerber et al[53] investigated 90 consecutive patients with
44 CHD and compared them with 209 hospital-based controls, finding an significant odds ratio in a

1 cardiovascular event(OR=1.82,95% CI 1.18–2.79, p=0.003),in a non-HDL apoC-III (OR=2.44,95%
2 CI 1.58–3.78, p<0.001), HDL apoC-III(OR=0.64,95% CI 0.42–0.96, p=0.09), with each SD
3 increase in total serum apoC-III.

4 *ANGL4*

5 ANGL4 is a Literature-reported Target. Lipoprotein lipase (LPL) hydrolyzes fatty acids from
6 triglyceride-rich lipoproteins and regulates their distribution to peripheral tissues. ANGL4
7 mediates the inactivation of the lipoprotein lipase LPL [54, 55], thus playing a significant role in
8 regulating serum triglyceride clearance and lipid metabolism. This has been confirmed by several
9 human genome-wide association studies [56–59]. Numerous genetic studies indicate that elevated
10 plasma triglycerides are associated with increased risk factors for cardiovascular disease [54, 55,
11 60]. Given ANGL4’s role in lipid regulation, ANGL4 is widely considered a potential drug target
12 for cardiovascular diseases. It has been reported that plasma ANGL4 levels can predict future
13 cardiovascular events, suggesting that ANGL4 might serve as a biomarker for coronary artery
14 disease, including myocardial infarction [61], and an effective drug target.

15

16 **Protein targets associated with decrease of LDL-C**

17 *SWP70*

18 SWAP 70 is not currently listed as a therapeutic target. Preliminary evidence from animal studies
19 have shown that SWAP 70 may inhibit the development of non-alcoholic fatty liver disease
20 (NAFLD) by suppressing lipid accumulation, inflammatory responses, and fibrosis. An ELISA
21 study [35, 36] suggests that the SWAP-70 antibody may serve as a biomarker for the recurrence
22 and prognosis of multiple sclerosis (MS). In the context of cardiovascular diseases, a gene
23 knockout experiment [37] has shown that SWAP70 could inhibit the progression of cardiac
24 hypertrophy, possibly through TAK1-dependent suppression of the mitogen-activated protein
25 kinase signaling pathway. Lysosomes may also be involved in regulating the expression levels of
26 SWAP70. Though its potential role in the prevention and treatment of myocardial infarction
27 requires further investigation, these findings indicate a promising avenue for future research,
28 especially in the realm of therapeutic applications.

29

30 **Protein targets with other mechanism**

31 *TAGLN2*

32 TAGLN2 is a literature reported target. TAGLN2 is an actin-binding protein expressed in smooth
33 muscle cells, related to the formation of the cytoskeletal structure, and plays a role in cell
34 migration and tissue repair. Since smooth muscle cells and vascular repair have certain roles in the
35 development of cardiovascular diseases, including myocardial infarction, it is theoretically
36 possible that TAGLN2 may have an indirect association with cardiovascular diseases. Most
37 studies on TAGLN2 focus on its relationship with non-cardiovascular diseases such as cancer,
38 inflammation, and autoimmune diseases. Further research is needed to determine whether it can
39 serve as a reliable biomarker and potential drug target for myocardial infarction.

40

41 *C1s*

42 C1s is a successful target. Complement system and C1s plays a critical role in maintaining
43 homeostasis, which is associated with dysfunction and many disorders. The C1 complex,
44 composed of C1q, C1r, and C1s (C1qr2s2), is the initiator of the classical complement activation

1 pathway. At molecular level, C1r activates C1s so that it can, in turn, activate C2 and C4. C1s B
2 chain is a serine protease that combines with C1q and C1r to form C1, which the first component
3 of the classical pathway of the complement system. As C1s is activated under various pathological
4 conditions and associated with the development of inflammation, autoimmunity and cancer, it has
5 become an outstanding biomarker or target for diagnosis and treatment to large quantities of
6 disease.

7
8 C1s has been a successful target for Hereditary angioedema and Cold type autoimmune
9 haemolytic anemia, but remains literature-reported for MI. Complement system and C1s plays a
10 critical role in maintaining homeostasis, which is associated with dysfunction and many disorders,
11 including cardiovascular artery diseases like MI. [62] C1s is a member of the SCUBE1 family
12 [63], which is a platelet granule protein that is exposed on the surface of platelets upon activation.
13 Activated platelets are key factors in the development of arteriosclerosis. [64] Lin's research [65]
14 suggests that SCUBE may be a crucial platelet-endothelial adhesion molecule in the
15 pathophysiology of cardiovascular disease development. As C1s is activated under various
16 pathological conditions and associated with the development of inflammation, autoimmunity and
17 cancer, it has become an outstanding biomarker or target for diagnosis and treatment to large
18 quantities of disease. More researches can be conducted to use C1s as a target for MI.

19
20 Our study has a lot of advantages.

- 21 1. Comprehensive Data Analysis: By using an expansive dataset from the Icelandic Cancer
22 Project and deCODE genetics, we were able to capture associations between a broad array of
23 genetic variants and plasma proteins.
- 24 2. Robustness through MR Analysis: Mendelian Randomization allows for the identification of
25 causal relationships, minimizing the impact of confounding factors typically encountered in
26 observational studies.
- 27 3. Rigorous Sensitivity Analyses: Through heterogeneity tests, reverse causality analyses, and
28 colocalization analyses, we added layers of validation to our findings, making our conclusions
29 more robust.
- 30 4. Practical Implications: Our study does not merely identify genetic associations; it has
31 tangible implications for the identification of potential drug targets, speeding up the typically long
32 drug development process.

33
34 However, our study still has some limitations.

- 35 1. Potential Population Bias: Our study heavily relies on data from Icelandic individuals. While
36 it provides a comprehensive dataset, the findings might not be generalizable to populations with
37 different genetic backgrounds.
- 38 2. Incomplete Drugability Assessment: Though we explored drugability using prominent
39 databases, emerging research or less-established potential drug targets might have been
40 overlooked.
- 41 3. Complexity of MI Pathogenesis: MI is a multifactorial condition, and while our study sheds
42 light on certain causal factors, there might be other significant variables or interactions not
43 captured in our analysis.
- 44 4. Limitations of MR: Despite the strengths of MR, it's based on several key assumptions, such

1 as the relevance and exclusion criteria. If any of these are violated, it might introduce biases.
2

3 **Conclusion**

4 Our study represents a significant stride in understanding the causal relationships between plasma
5 proteins and MI risk. By harnessing the power of Mendelian Randomization and extensive genetic
6 datasets, we've pinpointed promising drug targets for MI, offering a beacon of hope for
7 streamlining future therapeutic interventions. However, as with all research, our findings have
8 certain limitations. Further research in diverse populations and incorporating more extensive
9 drugability datasets will be essential to validate and expand upon our conclusions.
10

11 **Supplementary information**

12 IV selection and validation.xlsx is result of IV selection and validation.
13

14 **Sources of Funding**

15 This work was supported by Peking University.
16

17 **Conflict of Interest Disclosures**

18 None.

References

- [1] Pollard, T.J.: The acute myocardial infarction. *Primary Care: Clinics in Office Practice* 27(3), 631–649 (2000)
- [2] Salari, N., Morddarvanjoghi, F., Abdolmaleki, A., Rasoulpoor, S., Khaleghi, A.A., Hezarkhani, L.A., Shohaimi, S., Mohammadi, M.: The global prevalence of myocardial infarction: a systematic review and meta-analysis. *BMC Cardiovascular Disorders* 23(1), 206 (2023)
- [3] Dundar, Y., Hill, R., Dickson, R., Walley, T.: Comparative efficacy of thrombolytics in acute myocardial infarction: a systematic review. *Qjm* 96(2), 103–113 (2003)
- [4] Imming, P., Sinning, C., Meyer, A.: Drugs, their targets and the nature and number of drug targets. *Nature reviews Drug discovery* 5(10), 821–834 (2006)
- [5] Gashaw, I., Ellinghaus, P., Sommer, A., Asadullah, K.: What makes a good drug target? *Drug discovery today* 16(23-24), 1037–1043 (2011)
- [6] Lam, M.P., Ping, P., Murphy, E.: Proteomics research in cardiovascular medicine and biomarker discovery. *Journal of the American College of Cardiology* 68(25), 2819–2830 (2016)
- [7] Hartiala, J.A., Han, Y., Jia, Q., Hilser, J.R., Huang, P., Gukasyan, J., Schwartzman, W.S., Cai, Z., Biswas, S., Tregouet, D.-A., et al.: Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *European heart journal* 42(9), 919–933 (2021)
- [8] Gao, H., Li, L., Rao, S., Shen, G., Xi, Q., Chen, S., Zhang, Z., Wang, K., Ellis, S.G., Chen, Q., et al.: Genome-wide linkage scan identifies two novel genetic loci for coronary artery disease: in gene-poor families. *PLoS One* 9(12), 113935 (2014)
- [9] Wang, Z., He, Z., Xuan, Q., Zhang, Y., Xu, J., Lin, J., Li, H., Chen, W., Jiang, T.: Analysis of the potential ferroptosis mechanism and multitemporal expression change of central ferroptosis-related genes in cardiac ischemia–reperfusion injury. *Frontiers in Physiology* 13, 934901 (2022)
- [10] Grimes, D.A., Schulz, K.F.: Bias and causal associations in observational research. *The Lancet* 359(9302), 248–252 (2002)
- [11] Hernán, M.A.: Methods of public health research—strengthening causal inference from observational data. *New England Journal of Medicine* 385(15), 1345–1348 (2021)
- [12] Zakynthinos, E., Pappa, N.: Inflammatory biomarkers in coronary artery disease. *Journal of Cardiology* 53(3), 317–333 (2009)

- [13] Sun, X., Jia, Z.: A brief review of biomarkers for preventing and treating cardiovascular diseases. *Journal of cardiovascular disease research* 3(4), 251 (2012)
- [14] Bisoendial, R.J., Kastelein, J.J., Levels, J.H., Zwavinga, J.J., Bogaard, B., Reitsma, P.H., Meijers, J.C., Hartman, D., Levi, M., Stroes, E.S.: Activation of inflammation and coagulation after infusion of c-reactive protein in humans. *Circulation Research* 96(7), 714–716 (2005)
- [15] Tall, A.: An overview of reverse cholesterol transport. *European heart journal* 19, 31–5 (1998)
- [16] Barter, P.J., Caulfield, M., Eriksson, M., Grundy, S.M., Kastelein, J.J., Komajda, M., Lopez-Sendon, J., Mosca, L., Tardif, J.-C., Waters, D.D., et al.: Effects of torcetrapib in patients at high risk for coronary events. *New England journal of medicine* 357(21), 2109–2122 (2007)
- [17] Kannel, W.B., CASTELLI, W.P., GORDON, T.: Cholesterol in the prediction of atherosclerotic disease: new perspectives based on the framingham study. *Annals of internal Medicine* 90(1), 85–91 (1979)
- [18] Xue, H., Shen, X., Pan, W.: Constrained maximum likelihood-based mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *The American Journal of Human Genetics* 108(7), 1251–1269 (2021)
- [19] Thanassoulis, G., O'Donnell, C.J.: Mendelian randomization: nature's randomized trial in the post-genome era. *Jama* 301(22), 2386–2388 (2009)
- [20] Bowden, J., Holmes, M.V.: Meta-analysis and mendelian randomization: A review. *Research synthesis methods* 10(4), 486–496 (2019)
- [21] Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmiðdóttir, E.L., Gunnarsdóttir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., et al.: Large-scale integration of the plasma proteome with genetics and disease. *Nature genetics* 53(12), 1712–1721 (2021)
- [22] Henry, A., Gordillo-Marañón, M., Finan, C., Schmidt, A.F., Ferreira, J.P., Karra, R., Sundström, J., Lind, L., Arnlöv, J., Zannad, F., et al.: Therapeutic targets for heart failure identified using proteomics and mendelian randomization. *Circulation* 145(16), 1205–1217 (2022)
- [23] 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(8), 714–725 (2017)

[24] Hartwig, F.P., Davey Smith, G., Bowden, J.: Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *International journal of epidemiology* 46(6), 1985–1998 (2017)

[25] Burgess, S., Bowden, J.: Integrating summarized data from multiple genetic variants in mendelian randomization: bias and coverage properties of inverse-variance weighted methods. *arXiv preprint arXiv:1512.04486* (2015)

[26] Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., 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. *elife* 7, 34408 (2018)

[27] Bowden, J., Davey Smith, G., Haycock, P.C., Burgess, S.: Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology* 40(4), 304–314 (2016)

[28] Bowden, J., Davey Smith, G., Burgess, S.: Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology* 44(2), 512–525 (2015)

[29] Burgess, S., Bowden, J., Fall, T., Ingelsson, E., Thompson, S.G.: Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology (Cambridge, Mass.)* 28(1), 30 (2017)

[30] Burgess, S., Swanson, S.A., Labrecque, J.A.: Are mendelian randomization investigations immune from bias due to reverse causation? *European Journal of Epidemiology* 36, 253–257 (2021)

[31] Hemani, G., Tilling, K., Davey Smith, G.: Orienting the causal relationship between imprecisely measured traits using gwas summary data. *PLoS genetics* 13(11), 1007081 (2017)

[32] Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al.: Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research* 46(D1), 1074–1082 (2018)

[33] Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C.F., Whaley, R., Klein, T.E.: An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics* 110(3), 563–572 (2021)

[34] Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., Qiu, Y., Chen, Y.: Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research* 50(D1), 1398–1407 (2022)

- [35] Xu, S., Bendeck, M., Gotlieb, A.: Vascular pathobiology: Atherosclerosis and large vessel disease. In: *Cardiovascular Pathology*, pp. 85–124. Elsevier (2016)
- [36] Guo, S., Xia, X.-d., Gu, H.-m., Zhang, D.-w.: Proprotein convertase subtilisin/kexin-type 9 and lipid metabolism. *Lipid Transfer in Lipoprotein Metabolism and Cardiovascular Disease*, 137–156 (2020)
- [37] Gencer, B., Mach, F., Murphy, S.A., De Ferrari, G.M., Huber, K., Lewis, B.S., Ferreira, J., Kurtz, C.E., Wang, H., Honarpour, N., et al.: Efficacy of evolocumab on cardiovascular outcomes in patients with recent myocardial infarction: a pre-specified secondary analysis from the fourier trial. *JAMA cardiology* 5(8), 952–957 (2020)
- [38] Schwartz, G.G., Bessac, L., Berdan, L.G., Bhatt, D.L., Bittner, V., Diaz, R., Goodman, S.G., Hanotin, C., Harrington, R.A., Jukema, J.W., et al.: Effect of alirocumab, a monoclonal antibody to pcsk9, on long-term cardiovascular outcomes following acute coronary syndromes: rationale and design of the odyssey outcomes trial. *American heart journal* 168(5), 682–689 (2014)
- [39] Steg, P.G., Schwartz, G., Szarek, M., et al.: The odyssey outcomes trial: topline results alirocumab in patients after acute coronary syndrome. Presented by Dr. Philippe Steg at the American College of Cardiology Annual Scientific Session (ACC 2018), Orlando, FL 10 (2018)
- [40] Chaiyasothi, T., Nathisuwan, S., Dilokthornsakul, P., Vathesatogkit, P., Thakkinstian, A., Reid, C., Wongcharoen, W., Chaiyakunapruk, N.: Effects of non-statin lipid-modifying agents on cardiovascular morbidity and mortality among statin-treated patients: a systematic review and network meta-analysis. *Frontiers in pharmacology* 10, 547 (2019)
- [41] Li, Y., Bu, G.: *Lrp5/6 in wnt signaling and tumorigenesis* (2005)
- [42] Joiner, D.M., Ke, J., Zhong, Z., Xu, H.E., Williams, B.O.: *Lrp5 and lrp6 in development and disease*. *Trends in Endocrinology & Metabolism* 24(1), 31–39 (2013)
- [43] Clevers, H., Nusse, R.: *Wnt/ β -catenin signaling and disease*. *Cell* 149(6), 1192–1205 (2012)
- [44] Komiyama, Y., Habas, R.: *Wnt signal transduction pathways*. *Organogenesis* 4(2), 68–75 (2008)
- [45] Fang, L., Moore, X.-L., Dart, A.M., Wang, L.-M.: Systemic inflammatory response following acute myocardial infarction. *Journal of geriatric cardiology: JGC* 12(3), 305 (2015)
- [46] Cleutjens, J.P., Blankesteijn, W.M., Daemen, M.J., Smits, J.F.: The infarcted myocardium: simply dead tissue, or a lively target for therapeutic interventions. *Cardiovascular research* 44(2), 232–241 (1999)
- [47] Stangl, V., Baumann, G., Stangl, K., Felix, S.B.: Negative inotropic mediators released from the heart after myocardial ischaemia–reperfusion. *Cardiovascular research* 53(1), 12–30 (2002)

- [48] James, T.N.: The chronotropic action of atp and related compounds studied by direct perfusion of the sinus node. *Journal of Pharmacology and Experimental Therapeutics* 149(2), 233–247 (1965)
- [49] Procopio, M.C., Lauro, R., Nasso, C., Carerj, S., Squadrito, F., Bitto, A., Di Bella, G., Micari, A., Irrera, N., Costa, F.: Role of adenosine and purinergic receptors in myocardial infarction: focus on different signal transduction pathways. *Biomedicines* 9(2), 204 (2021)
- [50] Zheng, J.-S., Boluyt, M.O., O’Neill, L., Crow, M.T., Lakatta, E.G.: Extracellular atp induces immediate-early gene expression but not cellular hypertrophy in neonatal cardiac myocytes. *Circulation research* 74(6), 1034–1041 (1994)
- [51] Jin, J.-L., Guo, Y.-L., Li, J.-J.: Apoprotein c-iii: A review of its clinical implications. *Clinica Chimica Acta* 460, 50–54 (2016)
- [52] Mendivil, C.O., Rimm, E.B., Furtado, J., Chiuve, S.E., Sacks, F.M.: Low-density lipoproteins containing apolipoprotein c-iii and the risk of coronary heart disease. *Circulation* 124(19), 2065–2072 (2011)
- [53] Gerber, Y., Goldbourt, U., Segev, S., Harats, D.: Indices related to apo cii and ciii serum concentrations and coronary heart disease: a case–control study. *Preventive medicine* 37(1), 18–22 (2003)
- [54] Sarwar, N., Danesh, J., Eiriksdottir, G., Sigurdsson, G., Wareham, N., Bingham, S., Boekholdt, S.M., Khaw, K.-T., Gudnason, V.: Triglycerides and the risk of coronary heart disease: 10 158 incident cases among 262 525 participants in 29 western prospective studies. *Circulation* 115(4), 450–458 (2007)
- [55] Jeppesen, J.ø., Hein, H.O., Suadicani, P., Gyntelberg, F.: Triglyceride concentration and ischemic heart disease: an eight-year follow-up in the copenhagen male study. *Circulation* 97(11), 1029–1036 (1998)
- [56] Dewey, F.E., Gusarova, V., O’Dushlaine, C., Gottesman, O., Trejos, J., Hunt, C., Van Hout, C.V., Habegger, L., Buckler, D., Lai, K.-M.V., et al.: Inactivating variants in *angptl4* and risk of coronary artery disease. *New England Journal of Medicine* 374(12), 1123–1133 (2016)
- [57] Genetics, M.I., Investigators, C.E.C.: Coding variation in *angptl4*, *lpl*, and *svep1* and the risk of coronary disease. *New England Journal of Medicine* 374(12), 1134–1144 (2016)
- [58] Helgadottir, A., Gretarsdottir, S., Thorleifsson, G., Hjartarson, E., Sigurdsson, A., Magnusdottir, A., Jonasdottir, A., Kristjansson, H., Sulem, P., Oddsson, A., et al.: Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nature genetics* 48(6), 634–639 (2016)

- [59] Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., Cohen, J.C.: Population-based resequencing of *angptl4* uncovers variations that reduce triglycerides and increase hdl. *Nature genetics* 39(4), 513–516 (2007)
- [60] Reiner, Z.: Hypertriglyceridaemia and risk of coronary artery disease. *Nature Reviews Cardiology* 14(7), 401–411 (2017)
- [61] Muendlein, A., Saely, C.H., Leiberer, A., Fraunberger, P., Kinz, E., Rein, P., Vonbank, A., Zanolin, D., Malin, C., Drexel, H.: Angiopoietin-like protein 4 significantly predicts future cardiovascular events in coronary patients. *Atherosclerosis* 237(2), 632–638 (2014)
- [62] Ye, J., Yang, P., Yang, Y., Xia, S.: Complement *c1s* as a diagnostic marker and therapeutic target: Progress and prospective. *Frontiers in Immunology* 13, 1015128 (2022)
- [63] Tu, C.-F., Su, Y.-H., Huang, Y.-N., Tsai, M.-T., Li, L.-T., Chen, Y.-L., Cheng, C.-J., Dai, D.-F., Yang, R.-B.: Localization and characterization of a novel secreted protein *scube1* in human platelets. *Cardiovascular research* 71(3), 486–495 (2006)
- [64] Huo, Y., Schober, A., Forlow, S.B., Smith, D.F., Hyman, M.C., Jung, S., Littman, D.R., Weber, C., Ley, K.: Circulating activated platelets exacerbate atherosclerosis in mice deficient in apolipoprotein e. *Nature medicine* 9(1), 61–67 (2003)
- [65] Lin, Y.-C., Sahoo, B.K., Gau, S.-S., Yang, R.-B.: The biology of *scube*. *Journal of biomedical science* 30(1), 1–45 (2023)

Table 1 15 Plasma Proteins Identified Causally Associated with MI Risk

exposure	method	nsnp	snp	b	se	b_CI	OR	OR_CI	p-value
TAGLN2	Wald ratio	1	rs2789422	0.394	0.087	(0.224, 0.564)	1.483	(1.251, 1.757)	5.46E-06*
SWP70	Weighted median	3	rs12271923,rs2645017,	-0.082	0.018	(-0.116, -0.047)	0.922	(0.89, 0.954)	3.72E-06*
	Inverse variance weighted		rs415895	-0.083	0.017	(-0.117, -0.05)	0.92	(0.889, 0.952)	1.39E-06*
	Simple mode			-0.086	0.029	(-0.142, -0.03)	0.917	(0.867, 0.97)	0.0947
	Weighted mode			-0.081	0.018	(-0.117, -0.045)	0.922	(0.89, 0.956)	0.0474
	MR Egger			-0.07	0.035	(-0.138, -0.002)	0.932	(0.871, 0.998)	0.291
IL-6 sRa	Weighted median	4	rs12083537,rs12126142,	-0.036	0.007	(-0.05, -0.021)	0.965	(0.951, 0.979)	9.39E-07*
	Inverse variance weighted		rs140604580,rs61812626	-0.035	0.009	(-0.054, -0.017)	0.965	(0.948, 0.983)	1.68E-03
	Simple mode			-0.034	0.01	(-0.054, -0.014)	0.967	(0.948, 0.986)	0.0452
	Weighted mode			-0.035	0.008	(-0.051, -0.019)	0.965	(0.95, 0.981)	0.0222
	MR Egger			-0.039	0.021	(-0.079, 0.002)	0.962	(0.924, 1.002)	0.202
LRP4	Inverse variance weighted	2	rs2306029,rs4752946	0.075	0.017	(0.041, 0.109)	1.078	(1.042, 1.115)	1.31E-05*
Apo B	Wald ratio	1	rs563290	0.516	0.109	(0.303, 0.729)	1.675	(1.354, 2.073)	2.08E-06*
FN1.3	Wald ratio	1	rs1250258	-0.266	0.053	(-0.369, -0.163)	0.767	(0.692, 0.85)	4.25E-07*
FN1.4	Wald ratio	1	rs1250259	-0.271	0.052	(-0.374, -0.169)	0.762	(0.688, 0.845)	2.21E-07*
Angiostatin	Weighted median	4	rs10945684,rs11751347,	0.167	0.037	(0.095, 0.239)	1.182	(1.1, 1.27)	5.28E-06*
	Inverse variance weighted		rs1652492,rs4252185	-0.017	0.322	(-0.648, 0.614)	0.984	(0.523, 1.848)	0.959

	Simple mode			0.235	0.039	(0.158, 0.312)	1.265	(1.171, 1.366)	9.42E-03
	Weighted mode			0.2	0.04	(0.123, 0.278)	1.222	(1.13, 1.321)	0.015
	MR Egger			-0.089	0.644	(-1.351, 1.173)	0.915	(0.259, 3.233)	0.903
ANGL4	Wald ratio	1	rs116843064	0.449	0.095	(0.262, 0.636)	1.566	(1.3, 1.888)	2.48E-06*
Fibronectin	Wald ratio	1	rs1250258	-0.236	0.047	(-0.328, -0.145)	0.79	(0.721, 0.865)	4.25E-07*
Plasminogen	Weighted median	4	rs11751347,rs4252159,	0.222	0.037	(0.149, 0.296)	1.249	(1.161, 1.344)	2.48E-09*
	Inverse variance weighted		rs4252185,rs6938647	-0.038	0.341	(-0.707, 0.63)	0.963	(0.493, 1.878)	0.911
	Simple mode			0.088	0.052	(-0.015, 0.19)	1.092	(0.985, 1.209)	0.192
	Weighted mode			0.231	0.036	(0.161, 0.3)	1.259	(1.174, 1.35)	7.48E-03
	MR Egger			0.514	0.665	(-0.789, 1.817)	1.672	(0.454, 6.152)	0.52
MSP	Wald ratio	1	rs1131095	0.032	0.007	(0.019, 0.046)	1.033	(1.019, 1.047)	1.42E-06*
PCSK9	Weighted median	4	rs11591147,rs472495,	0.257	0.04	(0.179, 0.335)	1.293	(1.196, 1.398)	1.03E-10*
	Inverse variance weighted		rs505151,rs58255540	0.268	0.041	(0.188, 0.348)	1.308	(1.207, 1.417)	5.78E-11*
	Simple mode			0.182	0.086	(0.014, 0.35)	1.2	(1.014, 1.42)	0.124
	Weighted mode			0.33	0.041	(0.249, 0.41)	1.39	(1.283, 1.507)	3.99E-03
	MR Egger			0.358	0.048	(0.263, 0.453)	1.43	(1.301, 1.572)	0.0178
Apo C-III	Wald ratio	1	rs964184	0.262	0.052	(0.161, 0.364)	1.3	(1.175, 1.439)	4.01E-07*
C1s	Wald ratio	1	rs12146727	0.06	0.014	(0.032, 0.088)	1.062	(1.033, 1.092)	2.28E-05*

*: significant P-value. For a particular plasma protein, although p-values for some methods is not extremely significant, the OR range is similar to those methods with significant P-values.

Abbreviation of plasma proteins(exposure): TAGLN2:Transgelin-2; SWP70:Switch-associated protein 70; IL-6 sRa:Interleukin-6 receptor subunit alpha;

LRP4:Low-density lipoprotein receptor-related protein 4; Apo B:Apolipoprotein B; FN1.3:Fibronectin Fragment 3; FN1.4:Fibronectin Fragment 4;
ANGL4:Angiotensin-converting enzyme 2-related protein 4; Plasmin: Plasminogen; MSP: Hepatocyte growth factor-like protein; PCSK9:Proprotein convertase subtilisin/kexin type 9;
Apo C-III: Apolipoprotein C-III; C1s:Complement C1s subcomponent

Table 2 Sensitivity test of 15 candidate proteins

exposure	nsnp	snp	Horizontal pleiotropy test			Heterogeneity test			Reverse
			Egger intercept	se	p-value	Q	Q_df	Q_p-value	
TAGLN2	1	rs2789422							
SWP70	3	rs12271923,rs2645017, rs415895	-0.0043	0.01	0.7395	0.3623	2	0.8343	
IL-6 sRa	4	rs12083537,rs12126142, rs140604580,rs61812626	0.0028	0.0145	0.8655	5.0337	3	0.1693	
LRP4	2	rs2306029,rs4752946				0.0046	1	0.9461	
Apo B	1	rs563290							✓
FN1.3	1	rs1250258							
FN1.4	1	rs1250259							
Angiostatin	4	rs10945684,rs11751347, rs1652492,rs4252185	0.0153	0.108	0.9006	334.103	3	4.13E-72	✓
ANGL4	1	rs116843064							
Fibronectin	1	rs1250258							
Plasminogen	4	rs11751347,rs4252159, rs4252185,rs6938647	-0.1266	0.1304	0.4341	332.7267	3	8.20E-72	
MSP	1	rs1131095							✓

PCSK9	4	rs11591147,rs472495, rs505151,rs58255540	-0.0257	0.0109	0.1427	5.624	3	0.1314
Apo C-III	1	rs964184						
C1s	1	rs12146727						

* Cochran's Q P-values < 0.05, indicating significant heterogeneity

Proteins associated with single snp share no pleiotropy. All the selected proteins passed the horizontal pleiotropy test. Plasmin and Angiostatin didn't pass the heterogeneity test. MSP, Apo B, Angiostatin has a reverse causality.

Table 3 Colocalization Analysis

variant	protein	outcome	coloc-prob
rs10743116	SWP70	MI	71.60%
rs12271923			71.66%
rs2645017			72.02%
rs415895			71.73%
rs12083537; rs12126142; rs140604580; rs61812626	IL-6 sRa	MI	0.94%
rs2306029	LRP4	MI	94.89%
rs4752946			94.90%
rs12713844; rs563290	Apo B	MI	92.00%
rs1250258	FN1.3	MI	1.24%
rs1250259	FN1.4	MI	0.60%
rs139078629			1.46%
rs10945684; rs11751347; rs1652492; rs4252185	Angiostatin	MI	0.00%
rs116843064	ANGL4	MI	99.81%
rs1250258	Fibronectin	MI	1.08%
rs11751347; rs4252159; rs4252185; rs6938647	Plasminogen	MI	0.00%
rs1131095; rs142690032; rs3197999	MSP	MI	9.75%
rs11591147; rs472495; rs505151; rs58255540	PCSK9	MI	100.00%
rs964184	Apo C-III	MI	99.80%
rs12146727	C1s	MI	96.86%
rs2789422	TAGLN2	MI	99.11%

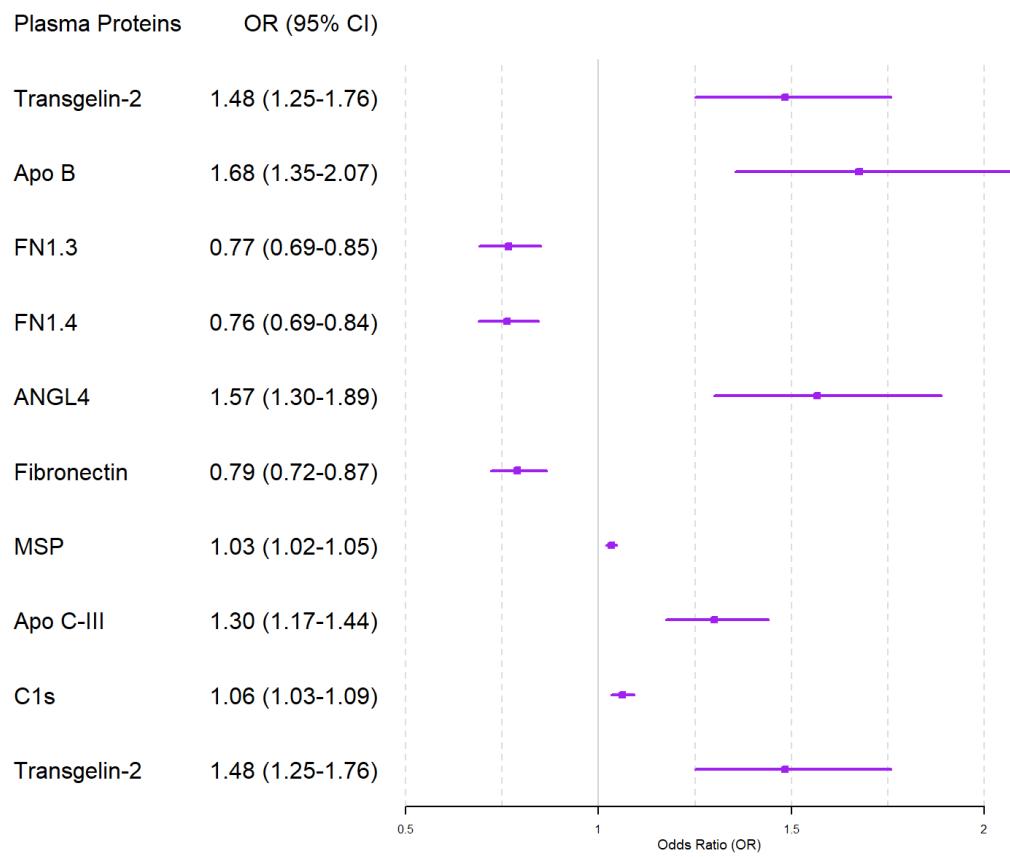


Fig. 2 Results of selected proteins with single snp

Fig. 3 Results of selected proteins with several snps

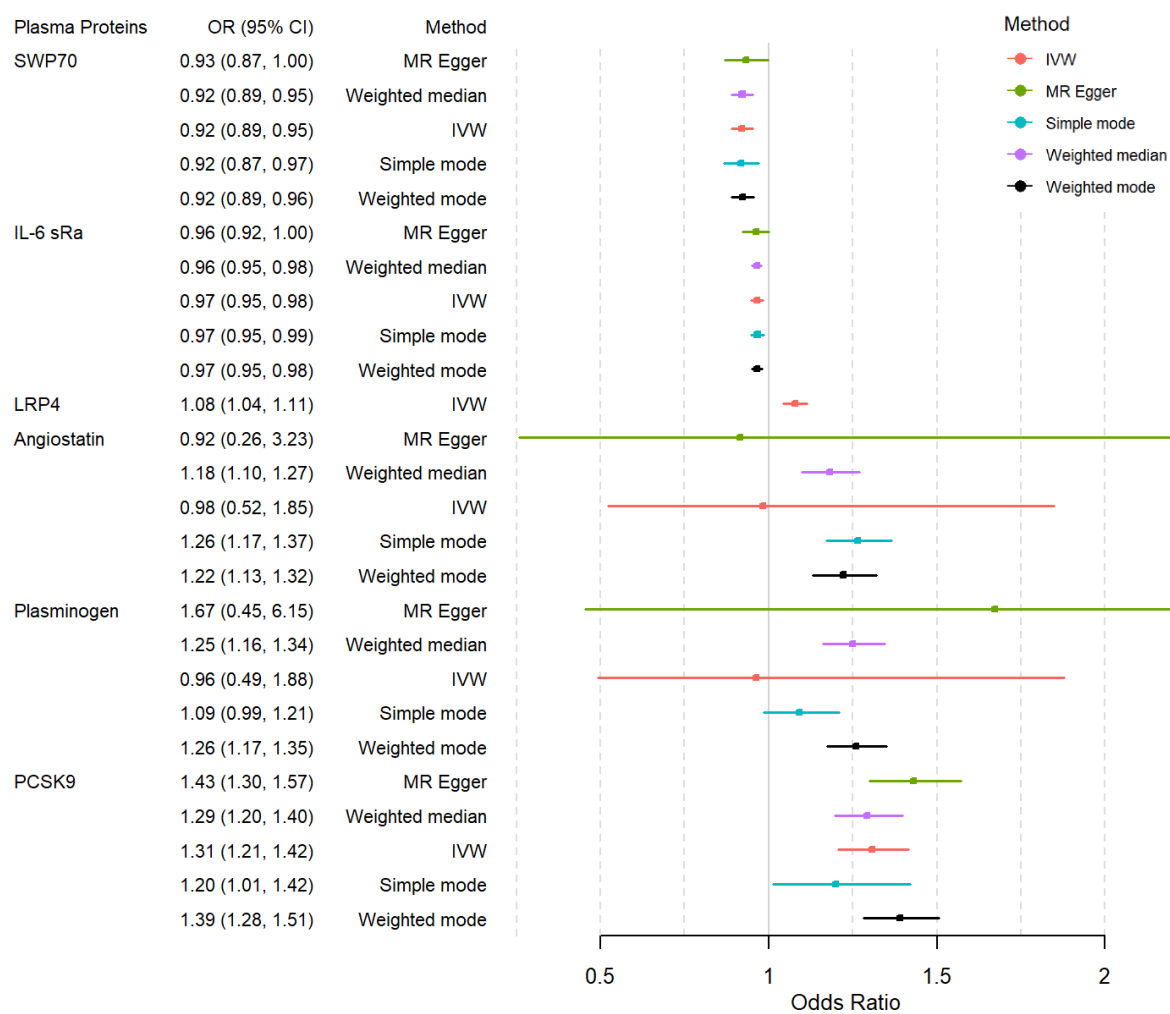


Table 4 Drugability assessment of potential drug targets

Protein	Target Info	Target Name	Target type	Disease	Drugs
TAGLN2	T93147	Transgelin-2 (TAGLN2)	Literature-reported Target	not Available	not Available
SWP70					
LRP4					
Apo B	T07533	Apolipoprotein B-100 (APOB)	Clinical trial Target	High blood cholesterol level	SPC4955
	T09507	APOB messenger RNA (APOB mRNA)	Successful Target	Familial hypercholesterolemia; Hypercholesterolaemia	Mipomersen
ANGL4	T98635	Angiotensin-converting enzyme 2 (ACE2)	Literature-reported Target	Not Available	Not Available
PCSK9	T62206	Proprotein convertase subtilisin/kexin type 9 (PCSK9)	Successful Target	Familial hypercholesterolemia; Chronic obstructive pulmonary disease; COPD	REGN-727
	T86652	PCSK9 messenger RNA (PCSK9 mRNA)	Clinical trial Target	Cardiovascular disease; Metabolic disorder	MPSK3169A
Apo C-III	T86115	ApoC-III messenger RNA (APOC3 mRNA)	Successful Target	Successful Target	Volanesorsen
	T31518	Apolipoprotein C-III (ApoCIII)	Clinical trial Target	Not Available	Not Available
C1s	T76024	Complement C1s component (C1S)	Successful Target	Hereditary angioedema; Cold type autoimmune haemolytic anemia	Cinryze

