

1 **COL6A3-derived endotrophin mediates the effect of obesity on coronary artery disease:**
2 **an integrative proteogenomics analysis**

3
4 Satoshi Yoshiji^{1,2,3,4}, Tianyuan Lu^{5,6}, Guillaume Butler-Laporte^{1,7,8}, Julia Carrasco-Zanini-Sanchez⁹,
5 Yiheng Chen^{1,2}, Kevin Liang^{1,10}, Julian Daniel Sunday Willett^{1,10}, Chen-Yang Su¹⁰, Shidong
6 Wang¹¹, Darin Adra¹, Yann Ilboudo¹, Takayoshi Sasako¹, Vincenzo Forgetta^{1,6}, Yossi Farjoun^{1,12},
7 Hugo Zeberg^{13,14}, Sirui Zhou^{2,15}, Michael Hultström^{1,6,16,17}, Mitchell Machiela¹⁸, Nicholas J.
8 Wareham⁹, Vincent Mooser^{2,15}, Nicholas J. Timpson^{19,20}, Claudia Langenberg^{9,21,22}, J. Brent
9 Richards^{1,2,6,7,23*}

10
11 ¹Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada

12 ²Department of Human Genetics, McGill University, Montréal, Québec, Canada

13 ³Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate School of Medicine,
14 Kyoto University, Kyoto, Japan

15 ⁴Japan Society for the Promotion of Science, Japan

16 ⁵Department of Statistical Sciences, University of Toronto, Canada

17 ⁶Prime Sciences, Montréal, Québec, Canada

18 ⁷Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec,
19 Canada

20 ⁸Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

21 ⁹MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK

22 ¹⁰Quantitative Life Sciences Program, McGill University, Montréal, Québec, Canada

23 ¹¹SomaLogic, Boulder, Colorado, USA

24 ¹²Fulcrum Genomics, Colorado, USA

25 ¹³Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden

26 ¹⁴Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

27 ¹⁵McGill Genome Centre, McGill University, Montréal, Québec, Canada

28 ¹⁶Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University,
29 Uppsala, Sweden

30 ¹⁷Integrative Physiology, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden

31 ¹⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA

32 ¹⁹Integrative Epidemiology Unit, University of Bristol, Bristol, UK

33 ²⁰Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

34 ²¹Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Germany

35 ²²Precision Healthcare University Research Institute, Queen Mary University of London, London, UK

36 ²³Department of Twin Research, King's College London, London, UK

37

38 ***Correspondence:**

39 J. Brent Richards

40 Professor of Medicine, McGill University, Senior Lecturer, King's College London (Honorary),
41 Pavilion H-413, Jewish General Hospital, 3755 Côte-Ste-Catherine Montréal, Québec, H3T 1E2,
42 Canada.

43 CEO, 5 Prime Sciences, Montreal, Québec, Canada

44 Email: brent.richards@mcgill.ca

45 Tel: +1-514-340-8222 Fax: +1-514-340-7529

46 **Abstract**

47 Obesity strongly increases the risk of cardiometabolic diseases, yet the underlying mediators of
48 this relationship are not fully understood. Given that obesity has broad effects on circulating
49 protein levels, we investigated circulating proteins that mediate the effects of obesity on coronary
50 artery disease (CAD), stroke, and type 2 diabetes—since doing so may prioritize targets for
51 therapeutic intervention. By integrating proteome-wide Mendelian randomization (MR) screening
52 4,907 plasma proteins, colocalization, and mediation analyses, we identified seven plasma
53 proteins, including collagen type VI $\alpha 3$ (COL6A3). COL6A3 was strongly increased by body mass
54 index (BMI) ($\beta = 0.32$, 95% CI: 0.26–0.38, $P = 3.7 \times 10^{-8}$ per s.d. increase in BMI) and increased
55 the risk of CAD (OR = 1.47, 95% CI: 1.26–1.70, $P = 4.5 \times 10^{-7}$ per s.d. increase in COL6A3).
56 Notably, COL6A3 is cleaved at its C-terminus to produce endotrophin, which was found to
57 mediate this effect on CAD. In single-cell RNA sequencing of adipose tissues and coronary
58 arteries, COL6A3 was highly expressed in cell types involved in metabolic dysfunction and fibrosis.
59 Finally, we found that body fat reduction can reduce plasma levels of COL6A3-derived
60 endotrophin, thereby highlighting a tractable way to modify endotrophin levels. In summary, we
61 provide actionable insights into how circulating proteins mediate the effect of obesity on
62 cardiometabolic diseases and prioritize endotrophin as a potential therapeutic target.

63 **Background**

64 Over 1.9 billion people worldwide have obesity, which is strongly linked to the risk of many
65 cardiometabolic diseases, including coronary artery disease (CAD), stroke, and type 2 diabetes^{1,2}.
66 There are many biological mechanisms whereby obesity causes disease, including metabolic
67 dysfunction, inflammation, and endothelial damage³. However, most of the factors mediating this
68 relationship are not yet fully understood. Therefore, identifying modifiable mediators of this
69 relationship could yield potential therapeutic targets, which may be targeted pharmaceutically or
70 non-pharmaceutically, for example with lifestyle interventions. Circulating proteins are potential
71 candidates because obesity strongly influences the level of plasma proteins^{4,5}, and they play a
72 critical role in disease development and progression. Moreover, circulating proteins can be
73 measured and sometimes modulated⁶, and their levels can be used as a surrogate measure of
74 target engagement in drug development programs. Therefore, understanding their role in disease
75 could provide multiple avenues to lessen the impact of obesity on cardiometabolic disease.

76
77 One way to understand the role of circulating proteins in disease has been through observational
78 epidemiology studies. However, such studies are not ideal for identifying causal mediators of
79 disease because they are prone to bias from unmeasured confounders and reverse causation^{7,8},
80 wherein the disease itself influences the protein level. What is therefore needed is a method to
81 understand mechanisms of disease, while reducing such biases.

82
83 Mendelian randomization (MR) is a genetic epidemiology approach that can contribute to the
84 understanding of the causal relationship between exposures and outcomes while minimizing the
85 bias from confounding and avoiding reverse causation⁶⁻¹². MR can be described as a natural
86 experiment somewhat analogous to randomized controlled trials (RCTs)¹³ because both rely upon
87 randomization to reduce bias from confounding. In MR studies randomization is achieved through
88 the random allocation of alleles at conception. Moreover, reverse causation can be theoretically
89 avoided because genotype is always assigned prior to the onset of disease.

90
91 Despite these advantages, MR relies on three key assumptions^{7,8}: there exist genetic variants
92 that: (I) are associated with the risk factor of interest; (II) are not correlated with confounders of
93 the exposure-outcome relationship; (III) affect the outcome only through the exposure (also known
94 as lack of horizontal pleiotropy). Of these, the third assumption is the most problematic and can
95 be a source of potential bias in MR. Nevertheless, when these main assumptions are met, MR
96 can be a powerful tool to describe causal relationships in humans—free of model systems.

97
98 Advancements in large-scale proteomics have facilitated the discovery of genetic variants that
99 influence plasma protein levels on a proteome-wide scale¹⁴⁻¹⁶. These genetic variants, referred to
100 as protein quantitative trait loci (pQTLs), can be utilized in MR to estimate the causal effect of
101 circulating protein levels on disease. Such methods have been successfully leveraged to prioritize
102 therapeutic targets, including OAS1 for COVID-19^{9,17} and IL6R for both COVID-19^{18,19} and CAD²⁰,
103 and ANGPTL3 for CAD²¹. As drug discovery is costly and prone to failure²², proteo-genomics-
104 based MR could play an important role since such studies could provide causal targets, which
105 can be measured, thereby providing proximal read-out of drug target engagement, but also
106 providing biomarkers for recruitment into clinical trials. Indeed, drugs with human genetics

107 evidence are more likely to be successful in Phase II and III trials, and two-thirds of FDA-approved
108 drugs in 2021 were supported by human genetics evidence^{23,24}.

109
110 Furthermore, MR methods can be leveraged to understand mediators of the biological pathways
111 connecting obesity with cardiometabolic disease when deployed in a two-step study design^{25,26}.
112 Step 1 begins by estimating the effect of BMI on protein mediators. Step 2 estimates the effect of
113 the identified mediators on the outcome of interest (in this case, cardiometabolic diseases).
114 Previously, we have successfully used this approach to identify a circulating protein, nephronectin,
115 that mediates the impact of obesity on COVID-19 severity²⁷.

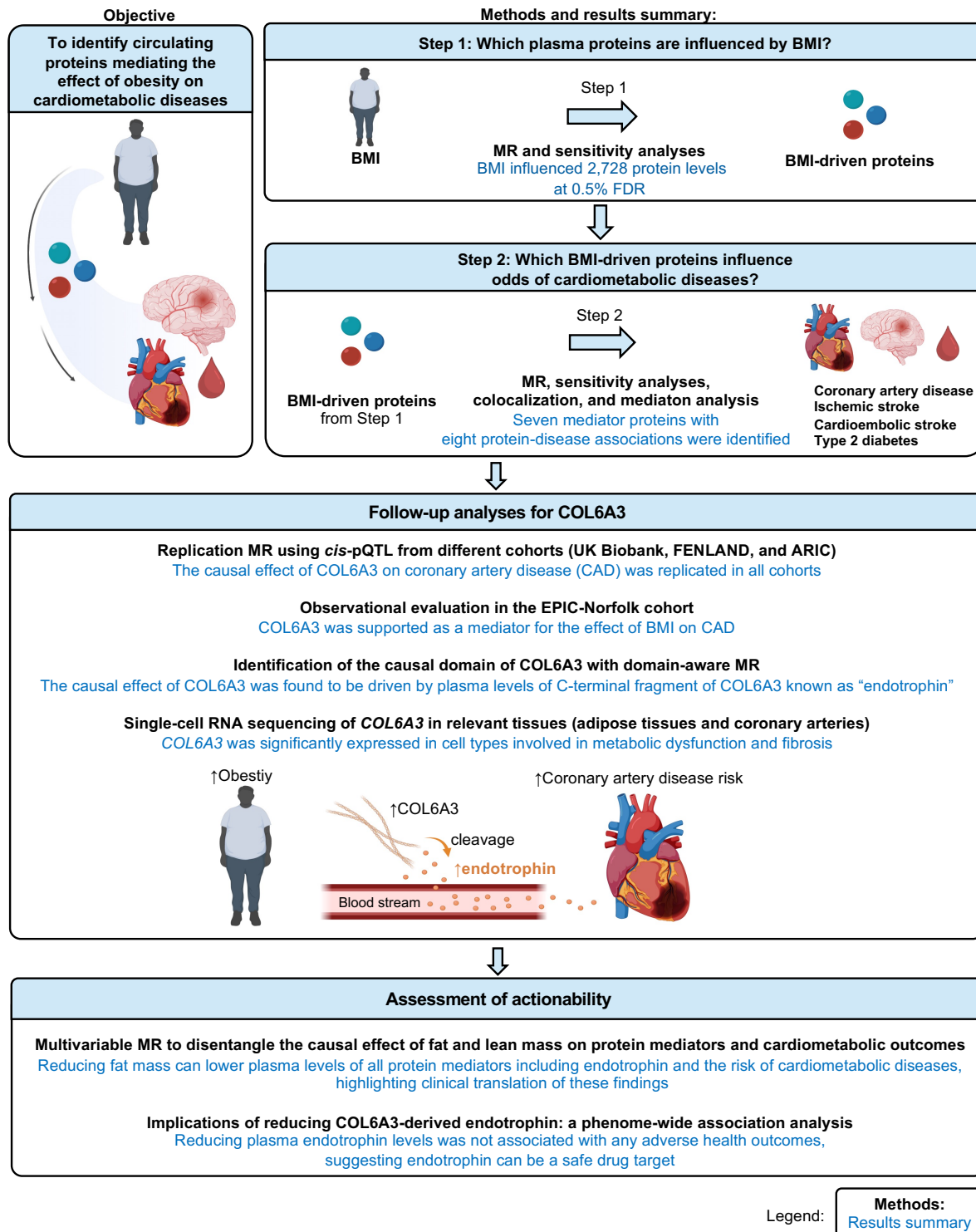
116
117 In the present study, we conducted an integrative MR analysis of on a proteome-wide scale,
118 screening 4,907 proteins, statistical colocalization, and mediation analysis to identify circulating
119 proteins that mediate the effects of obesity on CAD, ischemic stroke, cardioembolic stroke, and
120 type 2 diabetes. We then focused on collagen type VI $\alpha 3$ (COL6A3) as a potential target,
121 performing multiple follow-up analyses, including replication and single-cell sequencing analysis.
122 Additionally, we evaluated the actionability of COL6A3 by assessing the effect of reducing body
123 fat on its circulating protein level in multivariable MR and also assessed the implication of reducing
124 the identified proteins on a phenome-wide association study.

125
126 **Results**

127 The overall study design and a summary of the results are illustrated in **Fig. 1**. The study consisted
128 of four main sections:

- 129 1) Step 1 MR, which evaluated the causal effect of body mass index (BMI) on the levels of
130 circulating plasma proteins. We also evaluated the consistency of MR findings when BMI and
131 body fat percentage were used as the exposures.
- 132 2) Step 2 MR, which assessed the causal effects of BMI-driven proteins on four cardiometabolic
133 outcomes (CAD, ischemic stroke, cardioembolic stroke, and type 2 diabetes).
- 134 3) Follow-up analyses for COL6A3 and its cleavage product, known as endotrophin, which
135 assessed its role in CAD.
- 136 4) Assessment of clinical actionability for COL6A3-derived endotrophin and other protein
137 mediators by reducing body fat mass.

138 Each of these four steps and their results is described in detail below.



139 **Figure 1. Study design.**

140 To identify proteins that mediates the effect of obesity on cardiometabolic diseases, we used a
 141 two-step approach. In Step 1 Mendelian randomization (MR), we assessed the effect of body
 142 mass index (BMI) on 4,907 plasma proteins, which led to the identification of 2,714 proteins
 143 influenced by BMI (referred to as “BMI-driven proteins”) using two-sample MR.

144 In Step 2 MR, we assessed the effect of these BMI-driven proteins on cardiometabolic diseases,
145 again using two-sample MR.

146 In the subsequent sections, we conducted follow-up analyses of COL6A3 and evaluated the
147 potential for actionability of this protein and other mediators we identified.

148 BMI: body mass index, *cis*-pQTL: *cis*-acting quantitative trait loci.

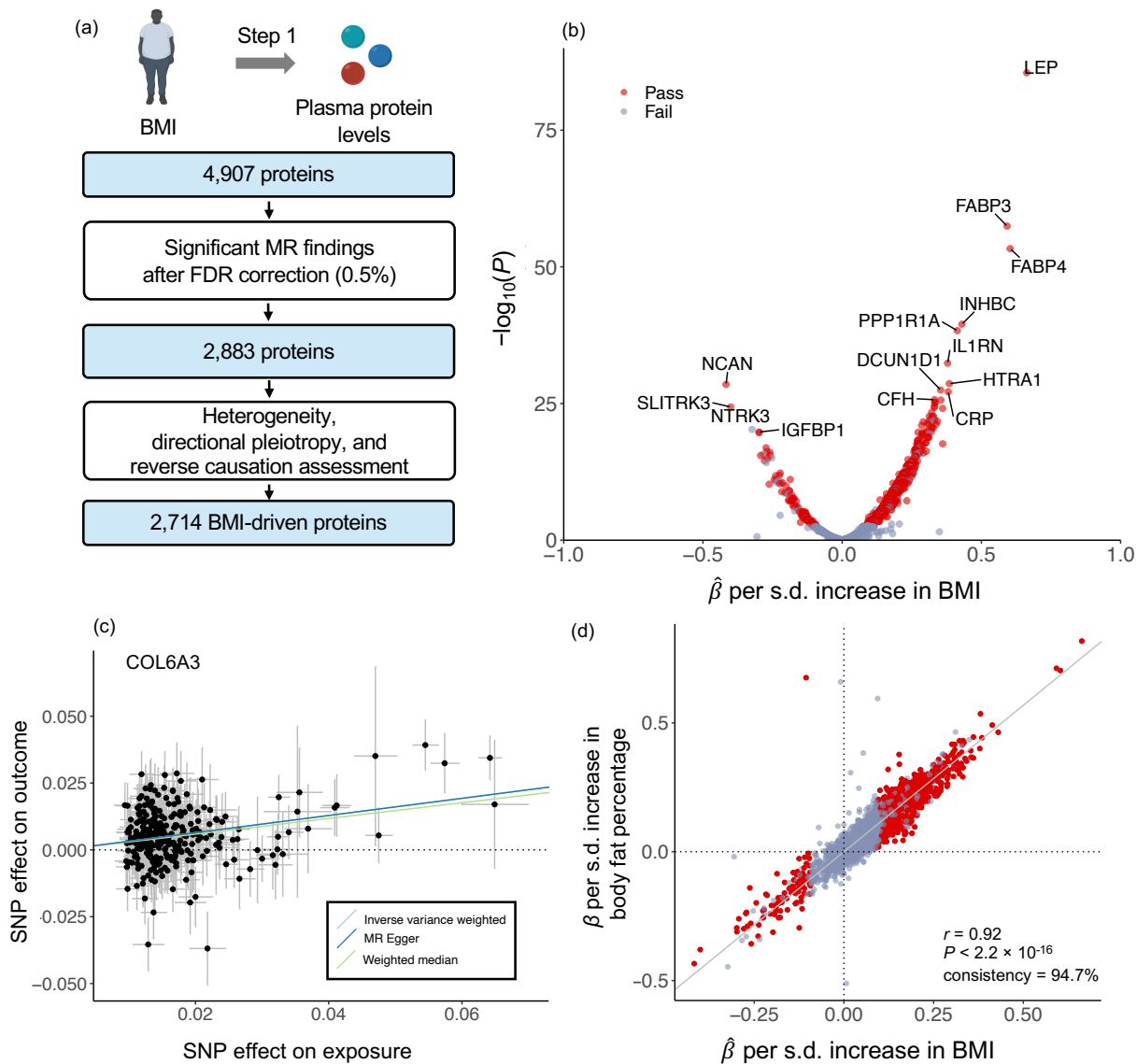
149
150
151

152 **1) Step 1 MR: Identification of the causal effect of BMI on plasma protein levels**

153 We evaluated the causal effect of BMI on 4,907 circulating proteins using the SomaScan v4
154 aptamer binding assay (SomaLogic, Boulder, CO). For clarity, we will refer to protein-targeting
155 aptamers as “proteins” unless otherwise specified. We performed causal inference using two-
156 sample MR, to estimate the effect of an exposure on an outcome of interest using two separate
157 genome-wide association studies (GWAS); one for the BMI and the second for circulating proteins
158 ¹³ (**Methods**). Specifically, we used the GWAS of BMI from the GIANT and UK Biobank
159 consortia²⁸ ($n = 681,275$ individuals) and circulating protein levels from the deCODE study¹⁵ ($n =$
160 $35,559$ individuals). In both studies we included only participants of European genetic ancestry
161 (**Supplementary Table 1**). We performed two-sample MR, using the inverse variance weighted
162 method as the primary analysis and then filtered these results dependent upon sensitivity
163 analyses, including tests for heterogeneity, directional horizontal pleiotropy, and reverse
164 causation. We used false discovery rate (FDR) correction with 0.5% as a stringent threshold for
165 significance, given that many protein levels are correlated with each other and therefore a
166 Bonferroni correction would be overly conservative (see **Methods**). No evidence of weak
167 instrumental variables (suspected when F -statistics < 10) were found (**Supplementary Table 2**).

168
169 We found that BMI influenced 2,728 proteins, passing tests of significance, heterogeneity, and
170 directional pleiotropy (**Supplementary Table 3**). However, among them, 14 showed evidence of
171 reverse causation, wherein the protein influenced BMI (**Supplementary Table 4**), and these 14
172 proteins were removed from further analyses. Thus, we identified a total of 2,714 plasma proteins
173 that are influenced by BMI. Hereafter, these 2,714 proteins are referred to as BMI-driven proteins
174 (**Fig. 2a, 2b, and 2c**).

175
176 Additionally, we performed MR to evaluate the effect of body fat percentage on the same 4,907
177 plasma proteins (**Methods**). We did this because body fat percentage is considered to be a more
178 direct proxy of obesity, whereas BMI is an easy-to-measure, clinically relevant proxy.²⁹ However,
179 the sample size available to assess the genetic determinants of BMI is larger than that of body fat
180 percentage, provide more precise estimates. We found that body fat percentage influenced 94.7%
181 of all BMI-driven proteins with the same direction of effect as BMI (**Fig. 2d**), illustrating a high
182 concordance of results between the two different measures of obesity ($r = 0.93$; $P < 2.2 \times 10^{-16}$).
183 Given the high concordance between MR results from BMI and body fat percentage, we proceed
184 to Step 2 MR with BMI-driven protein results.



185 **Figure 2. MR analyses for the effect of BMI on plasma protein levels.**
 186 (a) Flow diagram outlines Step 1 Mendelian randomization (MR).
 187 (b) A volcano plot illustrates the effect of BMI on each plasma protein from MR analyses using
 188 the inverse variance weighted method. The x-axis represents beta estimates, and the y-axis
 189 represents $-\log_{10}(P)$ values from MR results. Red dots represent proteins that passed all tests,
 190 including significance with a false discovery rate (FDR) $< 0.5\%$, as well as tests for heterogeneity,
 191 directional pleiotropy, and reverse causation. Grey dots represent proteins that failed any of these
 192 tests.
 193 (c) MR scatter plot shows the effect of BMI on plasma levels of COL6A3 using the inverse-
 194 variance weighted method (primary analysis), weighted median, or MR-Egger slope methods.
 195 (d) Directional consistency between MR results for the effect of BMI on plasma proteins and MR
 196 results for the effect of body fat percentage on plasma protein levels using the inverse variance
 197 weighted method.
 198 The x-axis denotes beta estimates from MR results, and r denotes Pearson's correlation.

199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242

2) Step 2 MR: Identification of the causal effect of BMI-driven proteins on cardiometabolic diseases

Next, we estimated the causal effect of these BMI-driven proteins on CAD, ischemic stroke, cardioembolic stroke, and type 2 diabetes, again using two-sample MR (**Fig. 3a**). We used the BMI-driven protein levels identified in Step 1 MR as exposures. The outcomes were CAD, ischemic stroke, cardioembolic stroke, and type 2 diabetes (see **Methods**). To minimize the risk of bias from horizontal pleiotropy, we used *cis*-acting protein quantitative trait loci (*cis*-pQTLs) identified from 35,559 individuals from the deCODE study¹⁵ as instrumental variables. In this context, instrumental variables are genetic variants that influence the exposure (i.e., circulating protein levels). We have defined *cis*-pQTLs as pQTLs that reside within a ± 1 Mb region around a transcription start site of a protein-coding gene. Since such *cis*-pQTLs would be likely to directly influence the circulating protein level by influencing the transcription or translation of mRNA from the gene that encodes the protein, they are less prone to bias from horizontal pleiotropy. Horizontal pleiotropy produces bias from the genetic variant influences the outcome independently of the circulating protein level.

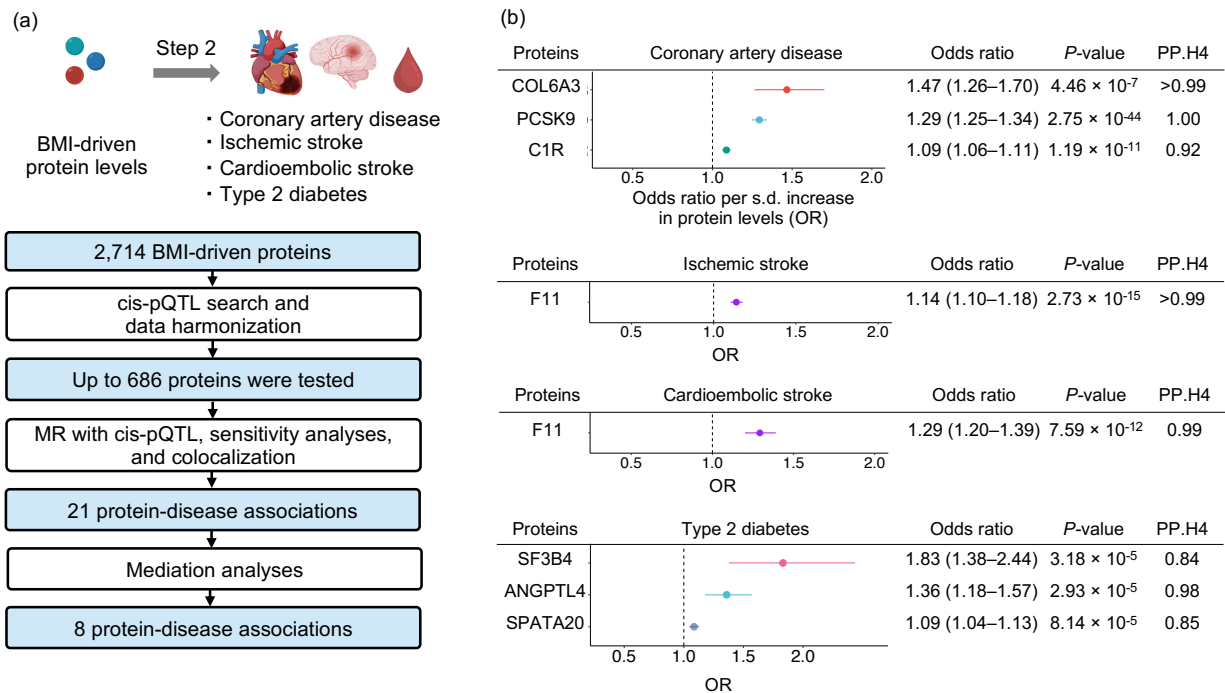
To further reduce the risk of horizontal pleiotropy, we restricted instrumental variables to genetic variants that were *cis*-pQTLs to only one protein. To do so, we removed variants associated with more than two proteins in a *cis*-acting manner (**Fig. 3a; see Methods**). For the outcomes, we used the largest available GWAS for CAD³⁰ (181,522 cases and 1,165,690 controls), ischemic stroke, and cardioembolic stroke³¹ (34,217 ischemic stroke cases, 7,193 cardioembolic stroke cases, and up to 2,703,029 controls), and type 2 diabetes³² (80,154 cases and 853,816 controls).

Following MR with *cis*-pQTLs and sensitivity analyses (heterogeneity, pleiotropy, and reverse causation assessment), we performed colocalization to evaluate whether the pQTL of the protein of interest and the disease outcome shared a single causal variant around a 1-Mb (± 500 kb) region surrounding the lead *cis*-pQTL. As different linkage disequilibrium (LD) structures across different study populations may lead to bias in the MR estimates, the presence of a shared single causal variant between the pQTL and the disease outcome can increase the robustness of MR findings (see **Methods**).

After MR with *cis*-pQTLs, sensitivity analyses, and colocalization, we identified 21 protein-disease associations that passed both step 1 and step 2 MR (**Supplementary Table 5**), including collagen type VI $\alpha 3$ (COL6A3) and PCSK9 for CAD, F11 for ischemic and cardioembolic stroke, and SF3B4 for type 2 diabetes. Among these proteins, COL6A3 was associated with the highest odds of CAD per one standard deviation (s.d.) increase in the protein levels (odds ratio (OR) = 1.47, 95% CI: 1.26–1.70, $P = 4.7 \times 10^{-7}$). We note that the finding of PCSK9 serves as a “positive control” and illustrates the utility of this method as PCSK9 is a well-known drug target, and its inhibition has been shown to reduced cardiovascular outcomes in multiple clinical trials³³⁻³⁵. Full results for CAD, ischemic stroke, cardioembolic stroke, and type 2 diabetes are provided in **Supplementary Table 6–9**.

243
 244 As an additional filtering step, we performed mediation analyses for the identified protein-disease
 245 associations. To do this, we used the product of coefficients method^{27,36-38} (**Methods**). Given that
 246 BMI increases the risk of cardiometabolic diseases ($\beta_{\text{BMI-to-cardiometabolic}} > 0$ in **Extended Fig. 1;**
 247 **Supplementary Table 10**), we restricted the analysis to proteins that increased the risk of
 248 cardiometabolic diseases through their mediation pathway ($\beta_{\text{BMI-to-protein}} \times \beta_{\text{protein-to-cardiometabolic}} > 0$ in
 249 **Extended Fig. 1; Supplementary Fig. 10**). Among the 21 protein-disease associations, 8 met
 250 this condition. Notably, all eight protein-disease associations were supported by mediation
 251 analyses, suggesting that the effect of BMI on the cardiometabolic outcome was mediated, at
 252 least partially, by the circulating protein (**Figure 3b; Supplementary Table 10**).

253
 254
 255



256
 257 **Figure 3. MR analyses for the effect of BMI-driven proteins on cardiometabolic diseases.**
 258 (a) Flow diagram of the Step 2 Mendelian randomization (MR) analyses.
 259 (b) Forest plots for the effect of body mass index (BMI)-driven proteins on four cardiometabolic
 260 diseases (coronary artery disease, ischemic stroke, cardioembolic stroke, type 2 diabetes). The
 261 MR analyses were conducted using the largest available GWAS of coronary artery disease³⁰
 262 (181,522 cases and 1,165,690 controls), ischemic stroke (34,217 cases and 2,703,029 controls),
 263 cardioembolic stroke³¹ (7,193 cases and 2,703,029 controls), and type 2 diabetes³² (80,154 cases
 264 and 853,816 controls).

265
 266
 267

268 **Follow-up analyses of COL6A3 (collagen type VI α 3)**

269 Circulating COL6A3 levels had the strongest effects on CAD across all the mediators of the
270 relationship between BMI and this outcome. We therefore sought to further test the hypothesis
271 that COL6A3 mediates the relationship between obesity and cardiometabolic disease using
272 analyses from orthogonal resources.

273

274 **Replication MR using *cis*-pQTL from different cohorts**

275 We evaluated whether the causal relationship between COL6A3 and CAD could be replicated
276 using different sources of *cis*-pQTLs from other cohorts. For this, we conducted two-sample MR
277 using *cis*-pQTLs from three additional cohorts: UK Biobank³⁹ ($n = 35,571$ individuals), Fenland¹⁴
278 ($n = 10,708$ individuals), and ARIC¹⁶ ($n = 7,213$ individuals). MR in all cohorts supported the
279 causal effect of COL6A3 levels on CAD, in the same direction (**Supplementary Table 11**).
280 Specifically, each s.d. increase in COL6A3 was associated with increased odds of CAD in UK
281 Biobank³⁹ (OR = 1.30, 95% CI: 1.17–1.45, $P = 2.4 \times 10^{-6}$), Fenland (OR = 1.23, 95%CI: 1.12–
282 1.35, $P = 8.9 \times 10^{-6}$), and ARIC (OR = 1.09, 95%CI: 1.05–1.13, $P = 1.6 \times 10^{-5}$). Notably, UK
283 Biobank used Olink Explore 3072 assay³⁹, whereas deCODE¹⁵, Fenland¹⁴, and ARIC¹⁶ used
284 SomaScan v4 assay. Hence, concordant MR results using *cis*-pQTLs from the different studies
285 from two different proteomic platforms further strengthened the evidence that COL6A3 partially
286 mediates the relationship between obesity and CAD.

287

288 **Observational epidemiological evaluation in the EPIC-Norfolk cohort**

289 If testing a hypothesis using different designs yields similar results, it is less likely that the results
290 are due to bias specific to one of the study designs. This is because different study designs have
291 different bias architectures and concordant results across study designs strengthens causal
292 inference because it is less likely that a single source of bias generated the results. Such testing
293 has been referred to as a triangulation of evidence.[ref] We therefore performed observational
294 association analysis with a randomly selected sub-cohort of the EPIC-Norfolk study ($n = 872$),
295 which included 207 prevalent or incident cases of CAD (see **Methods**). EPIC-Norfolk is a
296 population-based cohort from the United Kingdom. We found that increased BMI was associated
297 with increased plasma levels of COL6A3 ($\beta = 0.06$, 95% CI: 0.04–0.08, $P = 8.5 \times 10^{-12}$), and a s.d.
298 increase in plasma COL6A3 levels was associated with increased odds of CAD (OR = 1.34, 95%
299 CI: 1.12–1.59, $P = 1.1 \times 10^{-3}$). The mediation analysis supported that plasma COL6A3 levels
300 partially mediated the effect of BMI on CAD (**Supplementary Table 12**).

301

302 Given the robustness of these findings, we then explored the potential mechanism whereby
303 COL6A3 may influence CAD.

304

305 **Identification of the causal domain of COL6A3**

306 Cleavage of proteins can influence their biological mechanism⁴⁰. Previous studies have shown
307 that the C-terminal domain, also known the Kunitz domain, of COL6A3 is proteolytically cleaved
308 to form a biologically active fragment known as “endotrophin”. Endotrophin is produced in multiple
309 tissues, including adipose tissue^{40,41}. Endotrophin strongly induces fibrosis and inflammation, and
310 recent evidence suggests that it is involved in obesity-induced metabolic dysfunction⁴⁰⁻⁴⁵ (**Fig 4a**).
311 Therefore, we evaluated whether this particular domain of COL6A3 is driving its effect on CAD.

312
313 The SomaScan v4 assay measures target protein levels using aptamers, which are short, single-
314 stranded DNA or RNA molecules that can selectively bind to the target protein⁴⁶. SomaScan v4
315 assay has two separate aptamers targeting two domains of COL6A3, the N-terminal and C-
316 terminal (Kunitz domain) (**Methods**). These two separate aptamers thus allowed us to disentangle
317 the effects of the N-terminal and C-terminal containing fragments of COL6A3.

318
319 Intriguingly, we found that the aptamer binding the C-terminal of COL6A3 (**Fig. 4b**) was
320 associated with an increased risk of CAD (OR = 1.46 per s.d. increase in the protein level, 95%
321 CI: 1.37–1.93, $P = 2.7 \times 10^{-8}$), whereas the aptamer binding the N-terminal (i.e., the non-cleaved
322 portion of COL6A3) was not associated with the risk of CAD (OR = 1.06, 95% CI: 0.96–1.18, $P =$
323 0.22) in domain-aware MR (**Supplementary Table 13**). These findings suggest that the C-
324 terminal of COL6A3, which is cleaved into endotrophin, explains the effect of COL6A3 on CAD
325 and the aptamer binding to the C-terminal of COL6A3 may be capturing the plasma levels of
326 endotrophin or endotrophin-containing fragments. In the remainder of the manuscript, we refer to
327 such fragments as endotrophin for clarity.

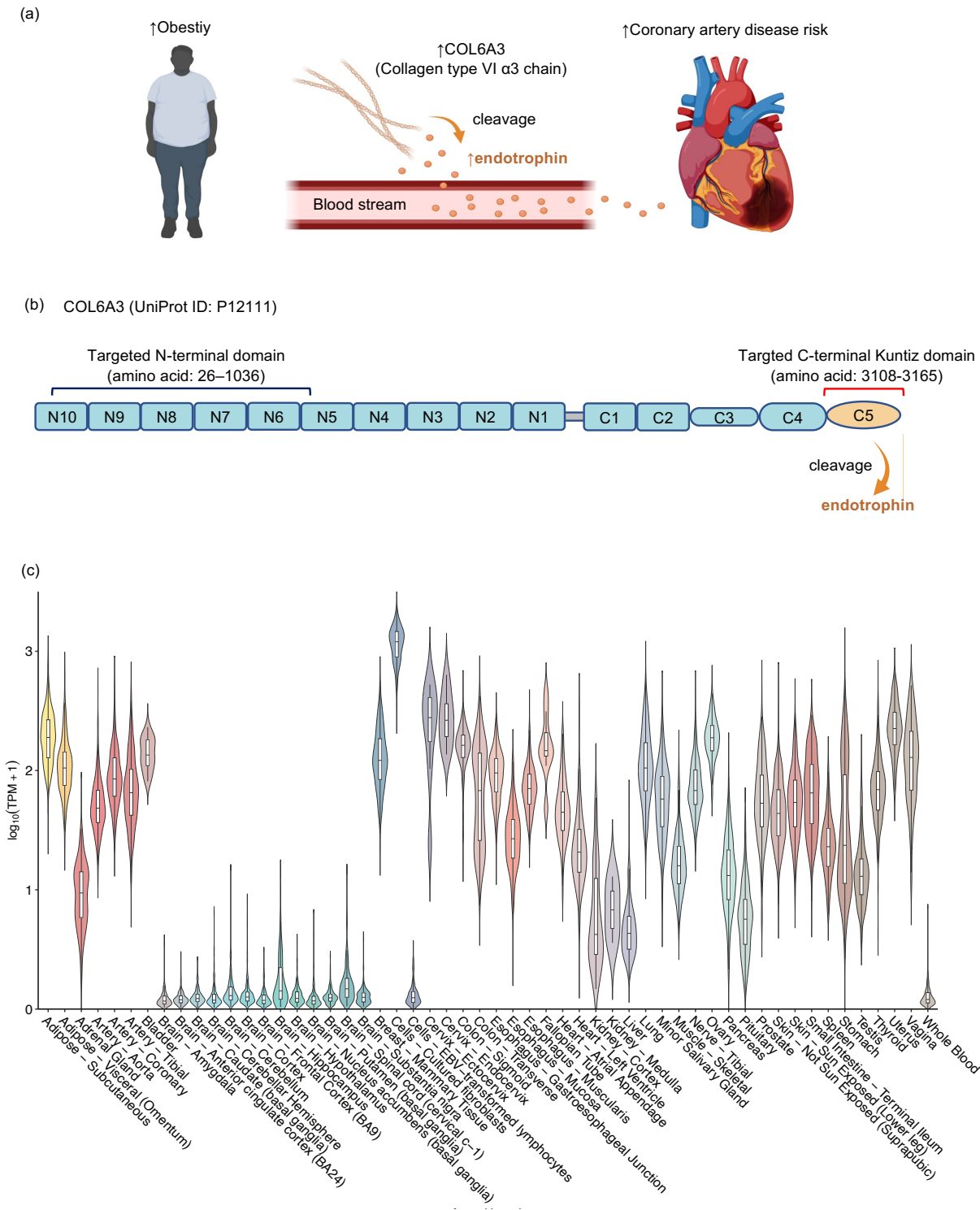
328
329 To further test the hypothesis that endotrophin is responsible for COL6A3's effect upon CAD, we
330 tested whether *cis*-pQTLs from the Olink Explore 3072 assay^{39,47} for COL6A3 were associated
331 with CAD. The Olink Explore 3072 assay uses a polyclonal antibody to target the C-terminal
332 (Kuniz domain) of COL6A3. The *cis*-pQTL (rs1050785) from UK-Biobank, which uses the Olink
333 platform, was in high linkage disequilibrium ($R^2 = 0.73$) with *cis*-pQTL (rs11677932) of the C-
334 terminal-targeting aptamer from the deCODE study but not in LD ($R^2 = 0.0$) with the *cis*-pQTL of
335 the N-terminal-targeting aptamer of COL6A3 (rs2646260). We found that the *cis*-pQTL from the
336 Olink platform was strongly associated with increased odds of CAD (OR = 1.32, 95%CI: 1.16–
337 1.50, $P = 1.75 \times 10^{-5}$) (**Supplementary Table 13**), which was consistent with the finding using
338 SomaScan v4 assay's aptamer binding the C-terminal of COL6A3. Taken together, these results
339 provide evidence from orthogonal proteomic assays that circulating levels of C-terminus COL6A3-
340 derived endotrophin likely explain the effect of COL6A3 levels on CAD.

341
342 Moreover, domain-aware MR analysis revealed that the aptamer targeting the C-terminal of
343 COL6A3 (cleaved portion) was more strongly increased by an increase in BMI ($\beta = 0.32$, 95% CI:
344 0.26–0.38, $P = 3.7 \times 10^{-24}$) than the aptamer targeting N-terminal (uncleaved portion) ($\beta = 0.10$,
345 95% CI: 0.04–0.16, $P = 2.1 \times 10^{-3}$), as shown by non-overlapping confidence intervals. These
346 findings indicate that an increase in BMI could increase both the expression of COL6A3 and its
347 cleavage, but has a preferential effect on the cleavage of COL6A3 into endotrophin.

348 349 **COL6A3 expression analyses**

350 We next explored the tissues in which COL6A3 is expressed using GTEx v8, which is a
351 compendium of expression data from 49 tissues across 838 individuals⁴⁸. In GTEx v8
352 (<https://gtexportal.org/>), COL6A3 was significantly expressed in multiple tissues, including
353 adipose tissue and coronary arteries when compared to the whole blood ($P < 0.001$) (**Fig. 4c**).
354 Therefore, it is possible that these tissues may locally produce COL6A3 and consequently its
355 cleavage product, endotrophin. While tissue-level examination of expression is helpful, such

356 methods do not permit resolution to the cellular level. Considering that the adipose tissue is
357 reported to be the primary source of COL6A3⁴⁵ and that the coronary artery is the location of
358 primary lesions in CAD⁴⁹, to better understand the cell type of origin of COL6A3 we analyzed
359 single-cell COL6A3 expression in human white adipose tissues⁵⁰ (SCP1376 at
360 <https://singlecell.broadinstitute.org/>) and coronary arteries in patients with CAD⁴⁹ (GSE131780 at
361 <https://www.ncbi.nlm.nih.gov/geo/>).



362
363 **Figure 4. Follow-up analyses for collagen type VI α 3 (COL6A3).**

364 (a) Schematic illustration of proposed relationship between obesity, COL6A3 (Collagen type VI
365 α 3 chain), endotrophin, and coronary artery disease. Obesity leads to increased production of
366 COL6A3, whose C-terminal is cleaved into an active form termed endotrophin, which increases
367 the risk of coronary artery disease.

368 (b) Schematic diagram of COL6A3 (UniProt ID: P12111). COLA3 consists of a short collagenous
369 region flanked by multiple von Willebrand factor type A (vWF-A) modules (N1–N10 in the N-
370 terminal and C1,2 in the C-terminal). There are three additional C-terminal domains unique to
371 COL6A3 (C3–C5), which are not present in other collagen type VI families. The most C-terminal
372 domain (C5) is cleaved into a soluble protein termed endotrophin.

373 The two amino acid sequences targeted by the aptamers are as follows: the N-terminal-binding
374 aptamer targets the amino acid sequence 26–1036 (uncleaved section), while the C-terminal
375 aptamer targets the amino acid sequence 3108–3165 (cleaved section). The figure has been
376 modified from ref^{81,82}.

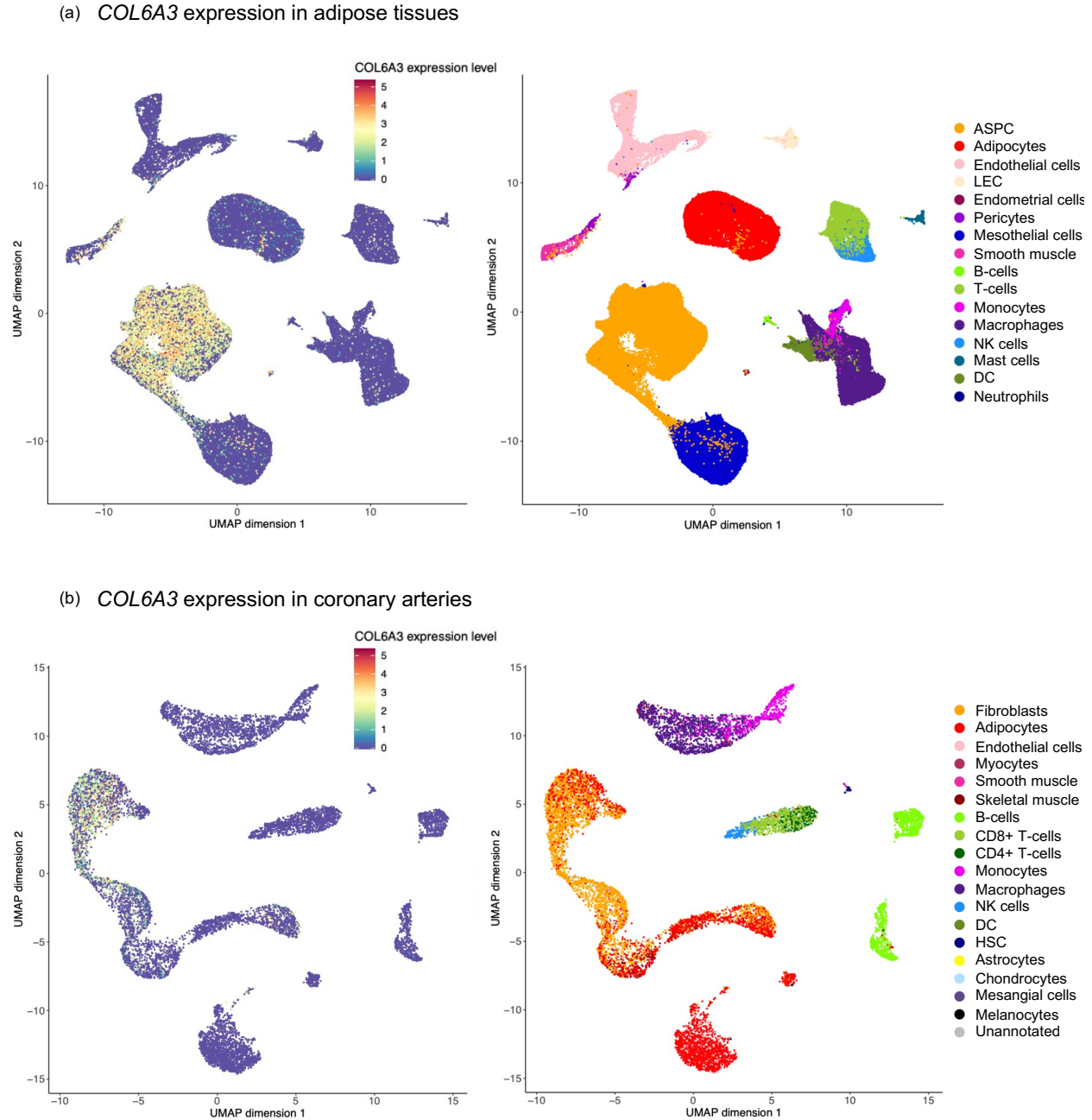
377 (c) COL6A3 expression profile in human tissues in GTEx v8⁴⁸. COL6A3 expression levels were
378 represented on a log transcript per 10 thousand plus one (TPM + 1) scale.

379

380

381

382 In single-cell sequencing, COL6A3 was significantly enriched in adipose progenitor/stem cells of
383 adipose tissues when compared to other cell types in adipose tissues (permutation $P < 0.001$;
384 see **Methods**) (**Fig. 5a**). Given that these cell populations play critical roles in maintaining adipose
385 tissue and metabolic function^{51,52}, the findings indicate that metabolic dysfunction may be an
386 underlying biological mechanism whereby COL6A3 influences CAD. Additionally, we found that
387 COL6A3 was significantly expressed in fibroblasts, which plays a key role in the atherosclerosis
388 of the coronary artery⁵³, when compared to other cell types in the coronary artery (permutation
389 $P < 0.001$; see **Methods**) (**Fig. 5b**). Taken together, these findings suggested that these cell types
390 may be responsible for the local production of COL6A3 in these tissues.



391 **Figure 5. Single-cell sequencing analyses of *COL6A3*.**

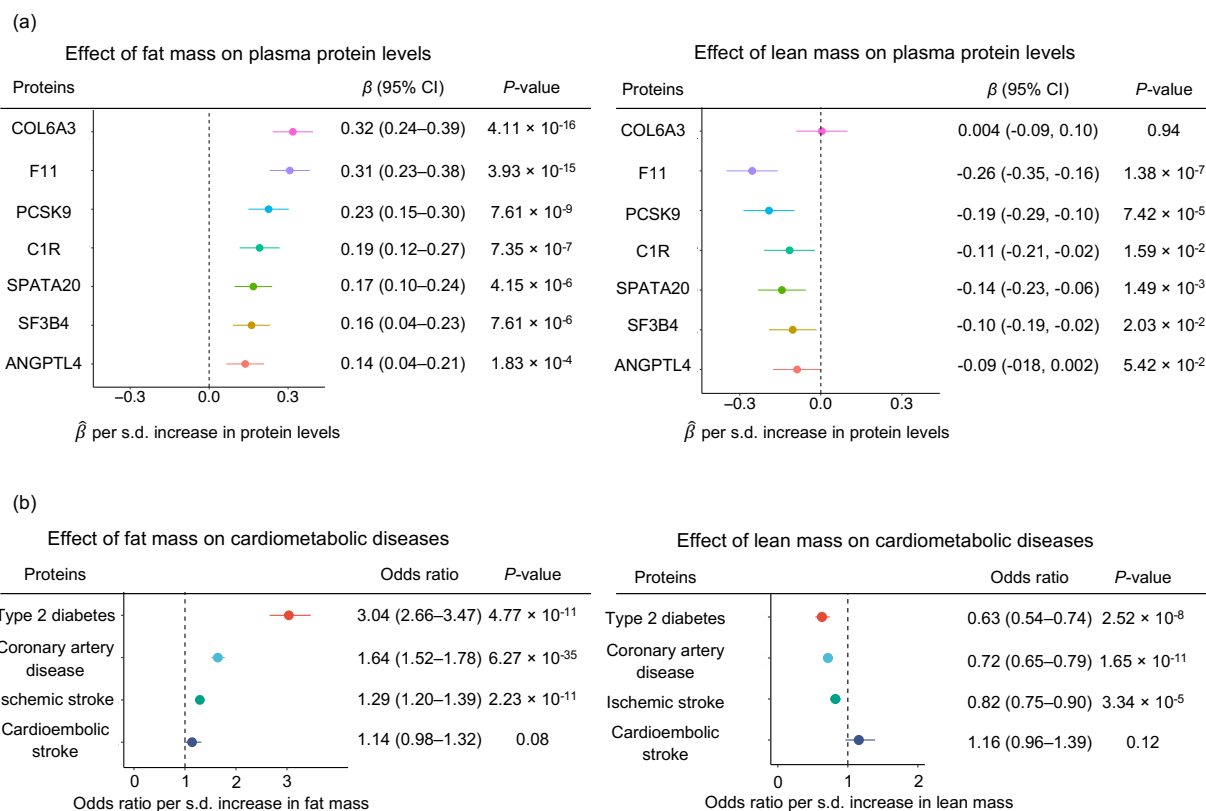
392 *COL6A3* expression patterns in the adipose tissues (a) and coronary arteries (b). We obtained
 393 single-cell transcriptomic data of human adipose tissue from Emont et al.⁵⁰ (SCP1376 at
 394 <https://singlecell.broadinstitute.org/>) and the data of coronary arteries from Wirka et al.⁴⁹
 395 (GSE131780 at the Gene Expression Omnibus database <https://www.ncbi.nlm.nih.gov/geo/>).
 396 ASPC: adipose stem and progenitor cells, LEC: lymphatic endothelial cells, NK: natural killer cells,
 397 DC: dendritic cells.

398
 399
 400

401 **Assessment of clinical actionability**

402 While identifying mediators of the effect of obesity on cardio-metabolic disease is relevant, such
403 targets could become clinically relevant if their modification through weight loss or other methods
404 influenced disease outcomes. We therefore explored whether reducing fat mass and/or increasing
405 lean mass could improve plasma COL6A3-derived endotrophin and other protein levels, thereby
406 reducing the risk of cardiometabolic diseases. For this, we used multivariable MR to evaluate the
407 independent effects of body fat and lean mass (i.e., body fat-free mass) on the protein mediators
408 and cardiometabolic disease outcomes (**Methods**).

409
410 We found that an s.d. increase in fat mass was independently associated with increased plasma
411 levels of all protein mediators (COL6A3-derived endotrophin, F11, PCSK9, C1R, SPATA20,
412 SF3B4, and ANGPTL4) (**Fig. 6b and Supplementary Table 14**) and increased odds of type 2
413 diabetes, CAD, and ischemic stroke. On the contrary, an s.d. increase in lean mass was
414 independently associated with decreased plasma levels of some protein mediators including F11
415 and PCSK9 (**Fig. 6b and Supplementary Table 15**).



416 **Figure 6. Multivariable MR analysis for evaluating the independent effects of fat mass and**
 417 **lean mass on plasma NPNT levels (a) and cardiometabolic diseases (b).**

418 We performed multivariable Mendelian randomization (MR) using fat mass and lean mass as
 419 exposures and plasma protein levels of the seven protein mediators or cardiometabolic diseases
 420 as outcomes.

421
 422
 423
 424 This has important clinical implications for actionability because interventions such as exercise,
 425 appropriate diet, or weight loss drugs such as the GLP-1 receptor agonist semaglutide and GLP-
 426 1/GIP co-agonist tirzepatide, which reduces body fat mass more than lean mass^{54,55}, could be
 427 effective in improving these protein levels and subsequently decreasing the risk of
 428 cardiometabolic diseases. However, future clinical trials are needed to confirm this hypothesis.

429
 430 Lastly, we evaluated whether reducing COL6A3-derived endotrophin is associated with any
 431 adverse health outcomes using a phenome-wide association analysis in the UK Biobank, FinnGen,
 432 and the GWAS catalog. We did this because clinical trials for some drug candidates have been
 433 terminated due to unexpected adverse events in later stages of the trials^{22,56}; thus, understanding
 434 the potential effects of perturbing the target on a phenome-wide level may to anticipate possible
 435 adverse events. Therefore, we assessed whether reducing COL6A3-derived endotrophin levels
 436 may have any implications on other traits. For this, we queried traits associated with the lead *cis*-
 437 pQTL of COL6A3 (rs11677932) from the deCODE study (a proxy for the COL6A3's C-terminal-
 438 derived endotrophin) in data from UK Biobank, FinnGen, and GWAS catalog using the Open

439 Target Genetics (<https://genetics.opentargets.org/>) at $P < 1.0 \times 10^{-5}$. The phenome-wide
440 association analysis revealed that decreased plasma levels of COL6A3-derived endotrophin (A-
441 allele of rs11677932; $\beta = -0.07$, $P = 1.5 \times 10^{-14}$) was associated with decreased risk of coronary
442 atherosclerosis ($\beta = -0.05$, $P = 1.0 \times 10^{-5}$), increased heel bone mineral density ($\beta = 0.02$, $P = 2.9$
443 $\times 10^{-12}$), and increased lung function (FEV1/FVC) ($\beta = 0.02$, $P = 5.2 \times 10^{-13}$) in addition to reduced
444 risk of CAD ($\beta = -0.03$, $P = 2.9 \times 10^{-12}$) (**Supplementary Table 16**). This suggests that decreasing
445 COL6A3-derived endotrophin may decrease the risk of multiple morbidities, including coronary
446 atherosclerosis, CAD, bone mineral density, and lung function, offering COL6A3-derived
447 endotrophin as an attractive therapeutic target.

448

449 Discussion

450 Obesity is a major risk factor of multiple diseases, and therapies are required that reduce its
451 clinical consequences. Here, we identified seven protein mediators (from eight protein-disease
452 associations) that partially mediate the effect of obesity on cardiometabolic diseases in humans.
453 All of these protein levels, including COL6A3, could potentially be improved through body fat
454 reduction, illustrating their possible clinical actionability. Furthermore, triangulation of evidence
455 with multiple follow-up analyses indicated that endotrophin, which is derived from the cleavage of
456 COL6A3, drives a part of the effect of obesity on CAD. These findings provide insights into how
457 obesity causes cardiometabolic disease and provide circulating proteins that could be
458 investigated as potential drug targets to lessen the public health burden of obesity.

459

460 The major finding of this study is the mediating role of endotrophin in the effect of obesity on CAD
461 in humans. Previous studies reported endotrophin as an important hormone that induces
462 metabolic dysfunction, fibrosis, and inflammation in rodent models^{40-42,57}, and cross-sectional
463 studies in humans have found that increased circulating endotrophin level was observationally
464 associated with cardiovascular events and all-cause mortality^{43-45,58}. However, cross-sectional
465 observational studies cannot disentangle cause and consequence. Therefore, our study, which
466 utilized MR to make causal inferences, provides evidence that endotrophin acts as a causal
467 mediator for the relationship between obesity and CAD in humans. Considering our findings that
468 reducing COL6A3 and its cleaved product, endotrophin, can reduce the risk of CAD without
469 apparent adverse health outcomes, directly targeting endotrophin can be an attractive therapeutic
470 approach, and it may be particularly effective in individuals with obesity.

471

472 Notably, we found that the aptamer targeting C-terminal of COL6A3 (also called the Kunitz domain,
473 which is cleaved into endotrophin) was more strongly affected by an increase in BMI than the
474 aptamer targeting the N-terminal. This indicates that obesity may increase both COL6A3
475 expression and the cleavage of COL6A3, but with a preferential influence on the cleavage of
476 COL6A3 into endotrophin, leading to an increase in endotrophin levels. Several studies using
477 mice models have shown that the bone morphogenetic protein 1 (BMP1)⁴¹, matrix
478 metalloproteinase 14 (MMP14)⁵⁹, and other MMPs⁶⁰ can release the C-terminal of COL6A3 as
479 endotrophin after proteolytic cleavage. However, as these studies were conducted using rodent
480 models, further research is needed to establish whether the same applies to humans. Despite
481 this, inhibition of BMP1 reduces scar formation and supports the survival of cardiomyocytes⁶¹,
482 which may be partly due to lower levels of endotrophin. Nevertheless, BMP1 also cleaves other

483 procollagens into mature collagens, which introduces pleiotropy. Therefore, more research is
484 necessary to determine how to selectively inhibit the cleavage of the C-terminal of COL6A3 to
485 reduce endotrophin levels.

486
487 Our study also illuminated other proteins, such as ANGPTL4, which mediate the relationship
488 between obesity and type 2 diabetes. Previous studies have shown that ANGPTL4 inhibits
489 lipoprotein lipase⁶², thereby reducing triglyceride levels⁶³. Additionally, ANGPTL4 has also been
490 implicated as an important player in obesity-induced glucose intolerance⁶²⁻⁶⁶, consistent with our
491 findings. Currently, an ANGPTL4 inhibitor, which is hepatocyte-targeting GalNAc-conjugated
492 antisense oligonucleotides that downregulate ANGPTL4 levels in liver and adipose tissue, is in
493 phase 1 clinical trial for hypertriglyceridemia⁶⁷. Our research indicates that this drug may be tested
494 for the prevention of type 2 diabetes, and further clinical trials are required to evaluate the safety
495 and efficacy of ANGPTL4 inhibition in humans. Another notable finding is F11 (coagulation factor
496 XI) as a mediator of the effect of obesity on cardiometabolic disease. F11 is a critical player in the
497 coagulation pathway and has been identified as causal for stroke by multiple studies^{6,68}. However,
498 few studies highlighted its role as a mediator. Currently, the F11 inhibitor, abelacimab⁶⁹, is in
499 phase III clinical trial for venous thromboembolism (NCT05171049 at
500 <https://www.clinicaltrials.gov/>). Our findings suggest that this drug may be effective for reducing
501 the risk of ischemic stroke, especially for individuals with obesity.

502
503 This study has important limitations. First, we focused on analyzing data solely from European-
504 ancestry individuals to prevent confounding by population stratification. While the ARIC cohort
505 reported *cis*-pQTL for individuals of African ancestry¹⁶, the sample size ($n = 1,871$) is still limited
506 when compared to data for those of European ancestry (deCODE study; $n = 35,559$). The same
507 applies to CAD GWAS, with 181,522 CAD cases in European ancestry individuals⁷⁰ compared to
508 only 17,247 cases in African ancestry individuals⁷¹. This limited sample size in African ancestry
509 individuals reduces the statistical power of MR analysis. Therefore, further efforts are needed to
510 increase the sample size of non-European-ancestry pQTL data. Second, we did not perform sex-
511 stratified analysis due to the unavailability of sex-specific datasets. Third, while the mediation
512 analyses results with both MR and observational evaluation in EPIC-Norfolk provided additional
513 evidence supporting COL6A3-derived endotrophin as a causal mediator, it should be noted that
514 the mediation analyses are based on additional assumptions⁷². Therefore, we used them as one
515 of several orthogonal validation methods. Fourth, we did not explore the molecular mechanism
516 whereby these proteins mediated the effect. Finally, although we triangulated multiple lines of
517 evidence to propose several promising therapeutic targets that mediate an important proportion
518 of the effect of obesity on cardiometabolic diseases (e.g., COL6A-derived endotrophin and
519 ANGPTL4), future clinical trials are required to explore the effect of pharmacologically influencing
520 these protein levels.

521 522 **Conclusions**

523 These results provide actionable insights into how circulating proteins mediate the effect of
524 obesity on cardiometabolic diseases. Our study highlights the importance of body fat reduction to
525 reduce the risk of cardiometabolic diseases and offers potential therapeutic targets, including
526 COL6A3-derived endotrophin, which may be prioritized for drug development.

527 **Methods**

528 **Step 1 MR**

529 ***MR to evaluate the effect of BMI on plasma protein levels***

530 We performed two-sample MR using BMI as exposure and circulating protein levels as outcomes.
531 The BMI exposure data came from a meta-analysis GWAS of UK Biobank and GIANT involving
532 693,529 European-ancestry individuals²⁸ (**Supplementary Table 1**). For the outcomes, we used
533 a GWAS of protein levels from the deCODE study¹⁵, measuring 4,907 proteins in 35,559
534 individuals of European ancestry using the SomaScan assay v4 from SomaLogic (Boulder,
535 Colorado, USA).

536

537 We performed two-sample MR using genome-wide significant and independent single nucleotide
538 polymorphisms (SNPs) with $P < 5 \times 10^{-8}$ and $r^2 < 0.001$ as instrumental variables. We excluded
539 SNPs in the human major histocompatibility complex region because of their complex linkage
540 disequilibrium structures. Clumping was performed using PLINK v1.9 ([https://www.cog-
541 genomics.org/plink/](https://www.cog-genomics.org/plink/)) with 10-Mb window. When the instrumental variable SNPs were not present
542 in the outcome GWAS, we identified proxy SNPs with $r^2 \geq 0.8$ using snappy v1.0
543 (<https://gitlab.com/richards-lab/vince.forgetta/snappy/>). To reduce the risk of weak instrument
544 bias, we calculated F-statistics and evaluated whether they were above ten^{73,74} (**Supplementary
545 Table 2**).

546

547 After harmonizing the exposure and outcome GWAS, we performed two-sample MR analysis
548 using the inverse variance weighted method with a random-effects model as the primary analysis,
549 implemented using TwoSampleMR v0.5.6. We set FDR < 0.005 (0.5%) as a stringent threshold
550 for significance. We used FDR correction, given that many proteins are correlated with each other
551 and that a Bonferroni correction can be overly conservative in such situations. However, we used
552 a strict threshold of 0.5% instead of a conventional threshold of 5% to reduce false positive
553 findings, as our intention was not to generate a complete list of potential associations, but rather
554 to generate a smaller set of high-confidence findings. We used weighted median, weighted mode,
555 and MR-Egger slope as supplementary analyses to evaluate the directional concordance of the
556 effect. Heterogeneity was tested using the I^2 statistic with results of $I^2 > 50\%$ and heterogeneity
557 $P < 0.05$ considered as substantial heterogeneity. Directional horizontal pleiotropy was tested
558 using the MR-Egger intercept test, and results with $P < 0.05$ were considered to indicate the
559 presence of directional horizontal pleiotropy.

560

561 For reverse MR, wherein we examined the effect of plasma protein levels on BMI, we performed
562 two-sample MR using *cis*-pQTLs variants from the deCODE study as exposures and BMI GWAS
563 from UK Biobank as an outcome. We used the inverse variance weighted method or the Wald
564 ratio method when only one SNP was available. We used FDR < 0.5% as a threshold for
565 significance. For BMI GWAS, we used data from the UK Biobank instead of the meta-analysis
566 GWAS of UK Biobank and GIANT because a number of *cis*-pQTL SNPs were not available in the
567 latter due to the stringent quality control process of the meta-analysis.

568

569 ***MR to evaluate the effect of body fat percentage on plasma protein levels***

570 While BMI is an easily measurable, clinically relevant proxy of obesity with the largest GWAS,
571 body fat percentage is considered a more direct measurement of body fat accumulation. Thus, a
572 high concordance between the BMI and body fat accumulation MR results may strengthen the
573 inference from the findings of Step 1 MR for BMI.
574

575 Therefore, we performed two-sample MR using body fat percentage as exposure and plasma
576 protein levels as outcomes. We used GWAS of body fat percentage in 454,633 European-
577 ancestry individuals from UK Biobank (Accession ID: ukb-b-8909 at IEU OpenGWAS project) and
578 the same protein levels for GWAS from the deCODE study as used in Step 1 MR.
579

580 **Step 2 MR**

581 ***MR with cis-pQTL to evaluate the effect of BMI-driven proteins on disease outcomes***

582 Next, we performed two-sample MR using circulating protein levels as exposures and
583 cardiometabolic diseases as outcomes, separately for each disease outcome. We used *cis*-pQTL
584 variants from the deCODE study in 35,559 European-ancestry individuals¹⁵ as the instrumental
585 variables. The *cis*-pQTL was defined as pQTL located within 1 Mb (\pm 1Mb) from the transcription
586 start site of the corresponding protein-coding gene. For the outcome, we used the largest
587 available GWAS of CAD³⁰ (181,522 CAD cases and 1,165,690 controls), ischemic stroke, and
588 cardioembolic stroke³¹ (34,217 ischemic stroke cases, 7,193 cardioembolic stroke cases, and up
589 to 2,703,029 controls), and type 2 diabetes³² (80,154 type 2 diabetes cases and 853,816 controls).
590 After data harmonization, we estimated the effect of each of the BMI-driven proteins on these
591 outcomes. Two-sample MR was performed using TwoSampleMR v0.5.6 with an inverse variance
592 weighted method and a random-effects model or Wald ratio when only one SNP was available as
593 an instrumental variable. FDR < 0.5% was set as the threshold for significance. To minimize the
594 risk of horizontal pleiotropy, we removed the variants associated with more than one protein in a
595 *cis*-acting manner; therefore, we only retained the variants that were *cis*-pQTL for one protein
596 (7008 out of 7572 variants are associated with only one protein in a *cis*-acting manner, and these
597 7008 variants are used as instrumental variables). To further test the absence of directional
598 horizontal pleiotropy, we used the MR-Egger intercept test when applicable (i.e., if there are at
599 least three instrumental variables). Additionally, we used the MR-Steiger test from TwoSampleMR
600 v.0.5.6 to assess reverse causation, whereby cardiometabolic diseases influence plasma levels
601 of proteins.
602

603 ***Colocalization***

604 To ensure that the proteins and cardiometabolic diseases share the same causal genetic signal
605 and avoid false-positive findings, We also performed colocalization using coloc R package
606 v5.1.0⁷⁵. We evaluated whether *cis*-pQTL of the protein shared the same causal variant with
607 cardiometabolic diseases within 1 Mb (\pm 500 kb). We used default prior of $p_1 = 10^{-4}$, $p_2 = 10^{-4}$,
608 and $p_{12} = 10^{-5}$ for coloc, where p_1 is a prior probability of trait 1 having a genetic association in the
609 region, p_2 is a prior probability of trait 2 having a genetic association in the region, and p_{12} is a
610 prior probability of the two traits having a shared genetic association. We considered the posterior
611 probability of a shared causal variant ($PP_{\text{shared}} > 0.8$) as evidence of colocalization.
612

613 ***Mediation analyses***

614 As a validation analysis, we performed mediation analyses using network MR with a product of
615 coefficients method. We did not adjust for the exposure (BMI) when estimating the effect of the
616 mediator on the outcome ($\beta_{\text{mediator-to-cardiometabolic}}$) to avoid weak instrument bias. This approach has
617 been adopted in multiple studies^{26,36-38}.

618 Considering that the proportion mediated can be only estimated when the direction of effects is
619 consistent between total causal effect and causal mediation effect, we restricted the analyses to
620 proteins that meet the following criteria: $\beta_{\text{total}} \times \beta_{\text{mediated}} > 0$

621 where: β_{total} denotes the total effect (i.e., the effect of BMI on cardiometabolic diseases), and
622 β_{mediated} denotes the causal mediation effect (i.e., the effect mediated by the circulating proteins).

623

624 To estimate the causal mediation effects (β_{mediated}), we estimated the effect of BMI on the plasma
625 protein levels ($\beta_{\text{BMI-to-protein}}$) and the effect of the plasma proteins on cardiometabolic diseases
626 ($\beta_{\text{protein-to-cardiometabolic}}$), and then multiplied these values ($\beta_{\text{mediated}} = \beta_{\text{BMI-to-protein}} \times \beta_{\text{protein-to-cardiometabolic}}$).

627 For this, we performed MR using the same instrumental variables as in Steps 1 and 2 of MR.
628 Subsequently, we divided β_{mediated} by β_{total} to estimate the proportion mediated and calculated the
629 *P*-value under the null hypothesis that the protein of interest did not mediate the effect of BMI on
630 the outcome of interest. We considered results with $P < 0.05$ to be significant. Since proteins can
631 be correlated (e.g., in the same biological pathways), we did not apply Bonferroni correction.

632

633 **Follow-up analyses**

634 ***Replication MR using cis-pQTL from different cohorts***

635 To replicate the causal estimates for the effect of COL6A3 on coronary artery disease, we
636 conducted two-sample MR using *cis*-pQTLs from different cohorts: UK Biobank³⁹ ($n = 35,571$
637 individuals), Fenland¹⁴ ($n = 10,708$ individuals), and ARIC ($n = 7,213$ individuals), using the same
638 method as described in Step 2 MR.

639

640 ***Mediation analysis with individual-level data in the EPIC-Norfolk cohort***

641 The EPIC-Norfolk study, a component of the pan-European EPIC Study, is a cohort of 25,639
642 middle-aged individuals from the general population of Norfolk, a county in Eastern England⁷⁶,
643 who attended the baseline assessment between 1993–1998. We performed mediation analysis
644 in a randomly selected subcohort ($n = 872$) of the EPIC-Norfolk study, in which proteomic profiling
645 was performed using the SomaScan v4 assay. Death certificates and hospitalisation data were
646 obtained using National Health Service numbers through linkage with the NHS digital database.
647 Electronic health records were coded by trained nosologists according to the International
648 Statistical Classification of Diseases and Related Health Problems, 9th (ICD-9) or 10th Revision
649 (ICD-10). Participants were identified as CAD cases if the corresponding ICD-codes (ICD-9: 410-
650 414, ICD-10:I20-I25) were registered on the death certificate (as the underlying cause of death or
651 as a contributing factor), or as the cause of hospitalization. The current study is based on follow-
652 up to the 31st March 2018. The case definition included all individuals identified as prevalent (at
653 the baseline study assessment) or incident CAD cases over the follow-up period of over 20-years.
654 The plasma protein levels were normalized with rank-based inverse normal transformation using
655 R package RNOMni v1.01. We used the product of coefficients methods to calculate the
656 proportion mediated, as described above, using the R package mediation⁷⁷ v4.5.0. We used linear
657 regression adjusting for age and sex to estimate the effect of BMI on plasma COL6A3 levels and

658 the effect of BMI, and logistic regression adjusting for age and sex to estimate the effect of BMI
659 on the risk of CAD and the effect of plasma COL6A3 levels on the risk of CAD. Significance of
660 the indirect effect and the proportion mediated was estimated by computing unstandardized
661 effects in 1000 bootstrapped samples, and calculating the corresponding 95% confidence
662 intervals.

663

664 **Identification of the causal domain of COL6A3**

665 **Target region of the SomaScan v4 assay and the Olink Explore 3072 assay**

666 We used SomaScan Menu 7K (<https://menu.somalogic.com/>) to determine the target amino acid
667 sequence of two aptamers for COL6A3 from on SomaScan v4 assay with additional support from
668 SomaLogic (Boulder, Colorado, USA). We also obtained data on the target region of Olink Explore
669 3072 assay from Olink (Uppsala, Sweden). In SomaScan v4 assay, two aptamers target COL6A3:
670 one for the C-terminal of COL6A3, also known as Kunitz domain (UniProt ID: P12111, target
671 amino acid sequence: 3108-3165) and another for the N-terminal (UniProt ID: P12111, target
672 amino acid sequence: 26-1036). In Olink Explore 3072 assay, the assay targets the C-terminal
673 Kunitz domain of COL6A3 with polyclonal antibody (OID20292:v1).

674

675 **Linkage disequilibrium of COL6A3's cis-pQTL from the deCODE study and UK Biobank**

676 We used the LDmatrix tool available at LDlink (<https://ldlink.nci.nih.gov>) with the 1000 genomes
677 European samples as the reference panel⁷⁸ to calculate R^2 values between three SNPs: the cis-
678 pQTL for COL6A3 from UK Biobank (rs1050785), the cis-pQTL of the C-terminal-targeting
679 aptamer (rs11677932) from the deCODE study, and the cis-pQTL of the N-terminal-targeting
680 aptamer of COL6A3 (rs2646260) from the deCODE study.

681

682 **COL6A3 expression analyses**

683 We downloaded bulk gene expression data in human tissues (GTEx_Analysis_2017-06-
684 05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz) from GTEx portal (<https://gtexportal.org/>). We
685 generated the violin plots of COL6A3 expression levels in each tissue using R v4.1.2. We used a
686 two-sided Wilcoxon rank sum test to compare COL6A3 expression in each tissue with its
687 expression in the whole blood.

688

689 **Single-cell RNA sequencing analysis**

690 To investigate COL6A3 expression at single-cell resolution in adipose tissues and coronary
691 arteries, we reanalyzed the published expression matrix data from Emont et al.⁵⁰ (SCP1376 at
692 <https://singlecell.broadinstitute.org/>) and Wirka et al.⁴⁹ (GSE131780 at Gene Expression Omnibus
693 database <https://www.ncbi.nlm.nih.gov/geo/>), focusing on COL6A3 expression. Following Wirka
694 et al.⁴⁹, we removed low-quality cells that expressed < 500 genes or had a mitochondrial content
695 > 7.5%, and genes expressed in < 5 cells. Cells expressing > 3,500 genes were also removed to
696 avoid bias due to doublets. The retained gene expression profiles were normalized to library size.
697 The top 2,000 most variable genes were selected after variance-stabilizing transformation using
698 the FindVariableFeatures function in Seurat v4.0.6. Principal component analysis was performed
699 based on these 2,000 most variable genes after scaling and centering. Nearest-neighbor graph
700 construction was conducted based on the first 10 principal components using the FindNeighbors
701 function in Seurat v4.0.6 with default settings. Cell clusters were identified using the FindClusters

702 function in Seurat v4.0.6 with default settings. Uniform Manifold Approximation and Projection
703 (UMAP) was also performed on the first 10 principal components. Two-dimensional visualization
704 of the cell clusters was based on the first two UMAP dimensions. We used SingleR v2.0.0 to
705 annotate the cell clusters with the Blueprint/ENCODE dataset as the reference using default
706 settings.

707

708 To assess whether certain cell types express *COL6A3* more significantly than others, we
709 performed 1,000 permutations of the cell type labels and calculated the frequency (permutation
710 p-value) of the same cell type containing the same or a larger proportion of cells expressing
711 *COL6A3* compared to all cells.

712

713 **Follow-up analyses for the identified proteins**

714 ***Assessment of actionability***

715 To estimate the independent effects of fat mass and lean mass on plasma protein levels, we
716 performed multivariable MR using fat mass and lean mass as exposures and protein levels as
717 outcomes.

718

719 ***GWAS of fat mass and lean mass***

720 We retrieved the GWAS data for fat mass and lean mass (i.e., fat-free mass) from UK Biobank
721 through the OpenGWAS portal (<https://gwas.mrcieu.ac.uk/>). The data included 454,137
722 individuals of European ancestry for fat mass and 454,850 individuals for lean mass. The
723 accession codes for the datasets were ukb-b-19393 for fat mass and ukb-b-13354 for lean
724 mass. The fat mass and fat-free mass of the UK Biobank participants (second release, 2017) were
725 evaluated by UK biobank with bioelectrical impedance analysis using the Tanita BC418MA body
726 composition analyzer (Tanita, Tokyo, Japan).

727

728 ***Multivariable MR to evaluate the independent effect of fat mass and lean mass on protein 729 levels and cardiometabolic diseases***

730 To obtain instrumental variables, we applied the same selection criteria as in Steps 1 and 2 of
731 MR ($P < 5 \times 10^{-8}$ and $r^2 < 0.001$), excluding those in the MHC region (GRCh37; chr6: 28,477,797–
732 33,448,354). We performed data harmonization in TwoSampleMR v0.56 and multivariable MR
733 with the inverse variance weighted method and a random-effect model in MVMR v0.3⁷⁴. We
734 calculated conditional F-statistics using MVMR v0.3⁷⁴ and evaluated whether they were above
735 $10^{73,74}$ (**Supplementary Table 14 and 15**). The phenotypic correlation matrix was calculated
736 using metaCCA v1.22.0⁷⁹. As additional sensitivity analyses, we performed multivariable MR-
737 Egger analysis using MendelianRandomization v0.6.085⁸⁰.

738

739 ***Phenome-wide association study for rs11677932***

740 We queried traits associated with the lead *cis*-pQTL of *COL6A3* (rs11677932) from the deCODE
741 study in the UK Biobank, FinGen, and GWAS catalog using the Open Target Genetics
742 (<https://genetics.opentargets.org/>)

743

744 **Ethical approval**

745 All contributing cohorts obtained ethical approval from their intuitional ethics review boards. The
746 contributing cohorts include UK Biobank, GIANT consortium, deCODEstudy, Fenland study,
747 AGES Reykjavik study, INTERVAL study, CARDIoGRAMplusC4D, GIGASTROKE, and MAGIC
748 consortium. The study was approved by the Norfolk Research Ethics Committee (no. 05/
749 Q0101/191), and all participants gave their informed written consent.

750

751 **Data availability**

752 We used GWAS summary statistics from the following source:

753 BMI GWAS from GIANT and UK Biobank (<https://portals.broadinstitute.org/collaboration/giant/>).
754 Plasma proteome GWAS from the deCODEstudy (<https://www.deCODE.com/summarydata/>), UK
755 Biobank (<https://doi.org/10.1101/2022.06.17.496443>), Fenland
756 (<https://omicscience.org/apps/pgwas/>), and the AGES Reykjavik study
757 (<https://doi.org/1.1126/science.aaq1327>).

758 We also used coronary artery disease GWAS from CARDIoGRAMplusC4D
759 (<http://www.cardiogramplusc4d.org/>), stroke GWAS from GIGASTROKE (GCST90104534 and
760 GCST90104535, at <https://www.ebi.ac.uk/gwas/studies/>), and type 2 diabetes GWAS from
761 Mahajan *et al.* (<https://doi.org/10.1038/s41588-022-01058-3>).

762 For gene expression data, we used data from Nathan *et al.* (SCP498 at Single Cell Portal
763 <https://singlecell.broadinstitute.org/>) and Wirka *et al.* (GSE131780 at Gene Expression Omnibus
764 database <https://www.ncbi.nlm.nih.gov/geo/>).

765

766 **Code availability**

767 We used R v4.1.2 (<https://www.r-project.org/>), TwoSampleMR v.0.5.6
768 (<https://mrcieu.github.io/TwoSampleMR/>), snappy v1.0 (<https://gitlab.com/richards-lab/vince.forgetta/snappy>), coloc v5.1.0 (<https://chr1swallace.github.io/coloc/>), PLINK v1.9
769 (<http://pngu.mgh.harvard.edu/purcell/plink/>), GCTA fastGWA v1.93.3
770 (<https://yanglab.westlake.edu.cn/software/gcta/>), and Seurat v4.0.6 (<https://satijalab.org/seurat/>).
771 Custom codes will be made available on GitHub (https://github.com/satoshi-yoshiji/cm_proteogenomics/) upon publication of the manuscript.

772

773 **Acknowledgments**

776 The Richards research group is supported by the Canadian Institutes of Health Research (CIHR:
777 365825, 409511, 100558, 169303), the McGill Interdisciplinary Initiative in Infection and
778 Immunity (MI4), the Lady Davis Institute of the Jewish General Hospital, the Jewish General
779 Hospital Foundation, the Canadian Foundation for Innovation, the NIH Foundation, Cancer
780 Research UK, Genome Québec, the Public Health Agency of Canada, McGill University, Cancer
781 Research UK [grant number C18281/A29019] and the Fonds de Recherche Québec Santé
782 (FRQS). J.B.R. is supported by an FRQS Mérite Clinical Research Scholarship. Support from
783 Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Wellcome Trust,
784 Medical Research Council, European Union, the National Institute for Health Research (NIHR)-
785 funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's
786 and St Thomas' NHS Foundation Trust in partnership with King's College London.
787 NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of
788 Parents and Children (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR

789 Biomedical Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit
790 (MC_UU_000111/1) and works within the CRUK Integrative Cancer Epidemiology Programme
791 (C18281/A29019).

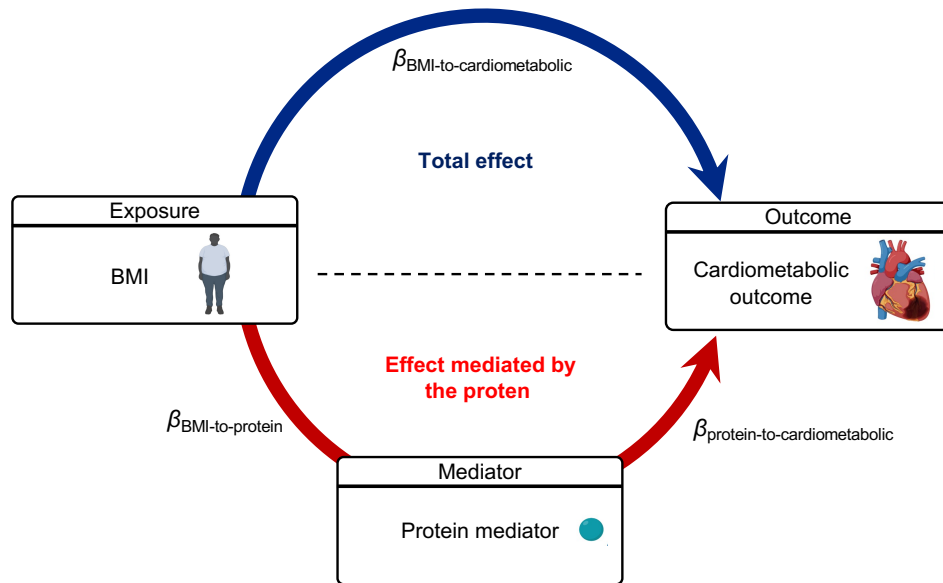
792 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the
793 Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH,
794 and NINDS. The data used for the analyses described in this manuscript were obtained from: the
795 GTEx Portal on March 26, 2023.

796 S.Y. is supported by the Japan Society for the Promotion of Science. T.L. is supported by a
797 Schmidt AI in Science Postdoctoral Fellowship, a Vanier Canada Graduate Scholarship, an FRQS
798 doctoral training fellowship, and a McGill University Faculty of Medicine Studentship. G.B.L. is
799 supported by scholarships from the FRQS, the CIHR, and Québec's ministry of health and social
800 services. Y.C. is supported by an FRQS doctoral training fellowship and the Lady Davis
801 Institute/TD Bank Studentship Award. The funders had no role in study design, data collection
802 and analysis, decision to publish, or preparation of the manuscript. We acknowledge Biorender
803 (biorender.com) for providing materials used to create the illustrative diagram.

804

805 **Competing Interests**

806 J.B.R. has served as an advisor to GlaxoSmithKline and Deerfield Capital. J.B.R.'s institution has
807 received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline, and Biogen for
808 projects unrelated to this research. J.B.R. is the CEO of 5 Prime Sciences
809 (www.5primesciences.com), which provides research services for biotech, pharma, and venture
810 capital companies for projects unrelated to this research. T.L. and V.F. are employees of 5 Prime
811 Sciences. The remaining authors declare no competing interests.



812 **Extended Figure 1. Schematic illustration of the mediation analysis.**

813 The figure demonstrates the causal relationship between BMI, the protein mediator, and
 814 cardiometabolic diseases using directed acyclic graphs. The dark blue arrow represents the total
 815 effect of BMI on cardiometabolic diseases ($\beta_{\text{BMI-to-cardiometabolic}}$), while the red arrow represents the
 816 effect of BMI on cardiometabolic diseases mediated by the protein mediator. To calculate the ratio
 817 mediated, we used the product of coefficients method. This involved multiplying the effect of BMI
 818 on the protein mediator ($\beta_{\text{BMI-to-protein}}$) by the effect of the protein mediator on cardiometabolic
 819 diseases ($\beta_{\text{protein-to-cardiometabolic}}$) to estimate the effect mediated by the protein ($\beta_{\text{mediated}} = \beta_{\text{BMI-to-protein}}$
 820 $\times \beta_{\text{protein-to-cardiometabolic}}$). Subsequently, we divided β_{mediated} by β_{total} to estimate the proportion
 821 mediated and calculated the *P*-value under the null hypothesis that the protein of interest did not
 822 mediate the effect of BMI on the outcome of interest.

823 BMI: body mass index, MR: Mendelian randomization.

824
825
826

827 **References**

828 1 Powell-Wiley, T. M. *et al.* Obesity and Cardiovascular Disease: A Scientific Statement
 829 From the American Heart Association. *Circulation* **143**, e984-e1010 (2021).
 830 <https://doi.org/doi:10.1161/CIR.0000000000000973>
 831 2 Czech, M. P. Insulin action and resistance in obesity and type 2 diabetes. *Nat Med* **23**,
 832 804-814 (2017). <https://doi.org:10.1038/nm.4350>
 833 3 Koenen, M., Hill, M. A., Cohen, P. & Sowers, J. R. Obesity, Adipose Tissue and Vascular
 834 Dysfunction. *Circ Res* **128**, 951-968 (2021).
 835 <https://doi.org:10.1161/CIRCRESAHA.121.318093>
 836 4 Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity
 837 using genetic drivers. *Nat Commun* **12**, 1279 (2021). [https://doi.org:10.1038/s41467-021-](https://doi.org:10.1038/s41467-021-21542-4)
 838 [21542-4](https://doi.org:10.1038/s41467-021-21542-4)
 839 5 Goudswaard, L. J. *et al.* Effects of adiposity on the human plasma proteome: observational
 840 and Mendelian randomisation estimates. *Int. J. Obes. (Lond.)* **45**, 2221-2229 (2021).
 841 <https://doi.org:10.1038/s41366-021-00896-1>

842 6 Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the
843 plasma proteome on complex diseases. *Nat Genet* **52**, 1122-1131 (2020).
844 <https://doi.org:10.1038/s41588-020-0682-6>

845 7 Skrivankova, V. W. *et al.* Strengthening the reporting of observational studies in
846 epidemiology using mendelian randomisation (STROBE-MR): explanation and
847 elaboration. *BMJ* **375**, n2233 (2021). <https://doi.org:10.1136/bmj.n2233>

848 8 Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in
849 Epidemiology Using Mendelian Randomization. *JAMA* **326**, 1614 (2021).
850 <https://doi.org:10.1001/jama.2021.18236>

851 9 Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry
852 against COVID-19 susceptibility and severity. *Nat Med* **27**, 659-667 (2021).
853 <https://doi.org:10.1038/s41591-021-01281-1>

854 10 Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal
855 genes and pathways for cardiovascular disease. *Nat Commun* **9**, 3268 (2018).
856 <https://doi.org:10.1038/s41467-018-05512-x>

857 11 Miller, C. L. *et al.* Integrative functional genomics identifies regulatory mechanisms at
858 coronary artery disease loci. *Nat Commun* **7**, 12092 (2016).
859 <https://doi.org:10.1038/ncomms12092>

860 12 Lu, T., Forgetta, V., Greenwood, C. M. T., Zhou, S. & Richards, J. B. Circulating Proteins
861 Influencing Psychiatric Disease: A Mendelian Randomization Study. *Biol. Psychiatry* **93**,
862 82-91 (2023). <https://doi.org:10.1016/j.biopsych.2022.08.015>

863 13 Burgess, S. *et al.* Using genetic association data to guide drug discovery and
864 development: Review of methods and applications. *Am J Hum Genet* **110**, 195-214 (2023).
865 <https://doi.org:10.1016/j.ajhg.2022.12.017>

866 14 Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science*
867 **374**, eabj1541 (2021). <https://doi.org:10.1126/science.abj1541>

868 15 Ferkingstad, E. *et al.* DECODE: Large-scale integration of the plasma proteome with
869 genetics and disease. *Nat Genet* **53**, 1712-1721 (2021). <https://doi.org:10.1038/s41588-021-00978-w>

870 16 Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African ancestry
871 identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet* (2022).
872 <https://doi.org:10.1038/s41588-022-01051-w>

873 17 Reis, G. *et al.* Early Treatment with Pegylated Interferon Lambda for Covid-19. *N Engl J*
874 *Med* **388**, 518-528 (2023). <https://doi.org:10.1056/NEJMoa2209760>

875 18 Bovijn, J., Lindgren, C. M. & Holmes, M. V. Genetic variants mimicking therapeutic
876 inhibition of IL-6 receptor signaling and risk of COVID-19. *The Lancet Rheumatology* **2**,
877 e658-e659 (2020). [https://doi.org:10.1016/s2665-9913\(20\)30345-3](https://doi.org:10.1016/s2665-9913(20)30345-3)

878 19 Group, R. C. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY):
879 a randomised, controlled, open-label, platform trial. *Lancet* **397**, 1637-1645 (2021).
880 [https://doi.org:10.1016/S0140-6736\(21\)00676-0](https://doi.org:10.1016/S0140-6736(21)00676-0)

881 20 Georgakis, M. K. *et al.* Interleukin-6 Signaling Effects on Ischemic Stroke and Other
882 Cardiovascular Outcomes: A Mendelian Randomization Study. *Circ Genom Precis Med*
883 **13**, e002872 (2020). <https://doi.org:10.1161/CIRCGEN.119.002872>

884 21 Dewey, F. E. *et al.* Genetic and Pharmacologic Inactivation of ANGPTL3 and
885 Cardiovascular Disease. *N Engl J Med* **377**, 211-221 (2017).
886 <https://doi.org:10.1056/NEJMoa1612790>

887 22 Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat Rev*
888 *Genet* (2023). <https://doi.org:10.1038/s41576-022-00572-8>

889 23 Ochoa, D. *et al.* Human genetics evidence supports two-thirds of the 2021 FDA-approved
890 drugs. *Nat Rev Drug Discov* **21**, 551 (2022). <https://doi.org:10.1038/d41573-022-00120-3>

891

892 24 King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as
893 likely to be approved? Revised estimates of the impact of genetic support for drug
894 mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).
895 <https://doi.org:10.1371/journal.pgen.1008489>

896 25 Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes
897 and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**,
898 e1002607 (2012). <https://doi.org:10.1371/journal.pgen.1002607>

899 26 Richardson, T. G., Fang, S., Mitchell, R. E., Holmes, M. V. & Davey Smith, G. Evaluating
900 the effects of cardiometabolic exposures on circulating proteins which may contribute to
901 severe SARS-CoV-2. *EBioMedicine* **64**, 103228 (2021).
902 <https://doi.org:10.1016/j.ebiom.2021.103228>

903 27 Yoshiji, S. *et al.* Proteome-wide Mendelian randomization implicates nephronectin as an
904 actionable mediator of the effect of obesity on COVID-19 severity. *Nat Metab* **5**, 248-264
905 (2023). <https://doi.org:10.1038/s42255-023-00742-w>

906 28 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body
907 mass index in ~700000 individuals of European ancestry. *Human Mol Genet* **27**, 3641-
908 3649 (2018). <https://doi.org:10.1093/hmg/ddy271>

909 29 Peltz, G., Aguirre, M. T., Sanderson, M. & Fadden, M. K. The role of fat mass index in
910 determining obesity. *Am. J. Hum. Biol.* **22**, 639-647 (2010).
911 <https://doi.org:10.1002/ajhb.21056>

912 30 van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an
913 Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res* **122**,
914 433-443 (2018). <https://doi.org:10.1161/CIRCRESAHA.117.312086>

915 31 Mishra, A. *et al.* Stroke genetics informs drug discovery and risk prediction across
916 ancestries. *Nature* **611**, 115-123 (2022). <https://doi.org:10.1038/s41586-022-05165-3>

917 32 Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of
918 diverse populations for discovery and translation. *Nat Genet* **54**, 560-572 (2022).
919 <https://doi.org:10.1038/s41588-022-01058-3>

920 33 Sabatine, M. S. *et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular
921 Disease. *N Engl J Med* **376**, 1713-1722 (2017). <https://doi.org:10.1056/NEJMoa1615664>

922 34 Schwartz, G. G. *et al.* Alirocumab and Cardiovascular Outcomes after Acute Coronary
923 Syndrome. *N Engl J Med* **379**, 2097-2107 (2018).
924 <https://doi.org:10.1056/NEJMoa1801174>

925 35 Ray, K. K. *et al.* Two Phase 3 Trials of Inclisiran in Patients with Elevated LDL Cholesterol.
926 *N Engl J Med* **382**, 1507-1519 (2020). <https://doi.org:10.1056/NEJMoa1912387>

927 36 Woolf, B., Zagkos, L. & Gill, D. TwoStepCisMR: A Novel Method and R Package for
928 Attenuating Bias in cis-Mendelian Randomization Analyses. *Genes* **13** (2022).
929 <https://doi.org:10.3390/genes13091541>

930 37 Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G. & Consortium, E. P.-I.
931 Network Mendelian randomization: using genetic variants as instrumental variables to
932 investigate mediation in causal pathways. *Int. J. Epidemiol.* **44**, 484-495 (2015).
933 <https://doi.org:10.1093/ije/dyu176>

934 38 Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy
935 for establishing the causal role of epigenetic processes in pathways to disease. *Int. J.*
936 *Epidemiol.* **41**, 161-176 (2012). <https://doi.org:10.1093/ije/dyr233>

937 39 Sun, B. B. *et al.* Genetic regulation of the human plasma proteome in 54,306 UK Biobank
938 participants. *bioRxiv*, 2022.2006.2017.496443 (2022).
939 <https://doi.org:10.1101/2022.06.17.496443>

940 40 Williams, L., Layton, T., Yang, N., Feldmann, M. & Nanchahal, J. Collagen VI as a driver
941 and disease biomarker in human fibrosis. *FEBS J* **289**, 3603-3629 (2022).
942 <https://doi.org:10.1111/febs.16039>

943 41 Heumuller, S. E. *et al.* C-terminal proteolysis of the collagen VI alpha3 chain by BMP-1
944 and proprotein convertase(s) releases endotrophin in fragments of different sizes. *J Biol*
945 *Chem* **294**, 13769-13780 (2019). <https://doi.org:10.1074/jbc.RA119.008641>

946 42 Przyklenk, M. *et al.* Lack of evidence for a role of anthrax toxin receptors as surface
947 receptors for collagen VI and for its cleaved-off C5 domain/endotrophin. *iScience* **25**,
948 105116 (2022). <https://doi.org:10.1016/j.isci.2022.105116>

949 43 Staunstrup, L. M. *et al.* Endotrophin is associated with chronic multimorbidity and all-cause
950 mortality in a cohort of elderly women. *EBioMedicine* **68**, 103391 (2021).
951 <https://doi.org:10.1016/j.ebiom.2021.103391>

952 44 Holm Nielsen, S. *et al.* The novel collagen matrikine, endotrophin, is associated with
953 mortality and cardiovascular events in patients with atherosclerosis. *J. Intern. Med.* **290**,
954 179-189 (2021). <https://doi.org:10.1111/joim.13253>

955 45 Sun, K., Park, J., Kim, M. & Scherer, P. E. Endotrophin, a multifaceted player in metabolic
956 dysregulation and cancer progression, is a predictive biomarker for the response to
957 PPARgamma agonist treatment. *Diabetologia* **60**, 24-29 (2017).
958 <https://doi.org:10.1007/s00125-016-4130-1>

959 46 Joshi, A. & Mayr, M. In Aptamers They Trust: The Caveats of the SOMAscan Biomarker
960 Discovery Platform from SomaLogic. *Circulation* **138**, 2482-2485 (2018).
961 <https://doi.org:10.1161/CIRCULATIONAHA.118.036823>

962 47 He, B., Huang, Z., Huang, C. & Nice, E. C. Clinical applications of plasma proteomics and
963 peptidomics: Towards precision medicine. *Proteomics Clin. Appl.* **16**, e2100097 (2022).
964 <https://doi.org:10.1002/prca.202100097>

965 48 Consortium, T. G. The GTEx Consortium atlas of genetic regulatory effects across human
966 tissues. *Science* **369**, 1318-1330 (2020).

967 49 Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation
968 and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med* **25**, 1280-1289
969 (2019). <https://doi.org:10.1038/s41591-019-0512-5>

970 50 Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature*
971 **603**, 926-933 (2022). <https://doi.org:10.1038/s41586-022-04518-2>

972 51 Miklosz, A., Nikitiuk, B. E. & Chabowski, A. Using adipose-derived mesenchymal stem
973 cells to fight the metabolic complications of obesity: Where do we stand? *Obes. Rev.* **23**,
974 e13413 (2022). <https://doi.org:10.1111/obr.13413>

975 52 Liao, X., Zhou, H. & Deng, T. The composition, function, and regulation of adipose stem
976 and progenitor cells. *J Genet Genomics* **49**, 308-315 (2022).
977 <https://doi.org:10.1016/j.jgg.2022.02.014>

978 53 Liberale, L. *et al.* The Role of Adipocytokines in Coronary Atherosclerosis. *Curr*
979 *Atheroscler Rep* **19**, 10 (2017). <https://doi.org:10.1007/s11883-017-0644-3>

980 54 Blundell, J. *et al.* Effects of once-weekly semaglutide on appetite, energy intake, control
981 of eating, food preference and body weight in subjects with obesity. *Diabetes Obes Metab*
982 **19**, 1242-1251 (2017). <https://doi.org:10.1111/dom.12932>

983 55 Jastreboff, A. M. *et al.* Tirzepatide Once Weekly for the Treatment of Obesity. *New*
984 *England Journal of Medicine* **387**, 205-216 (2022).
985 <https://doi.org:10.1056/NEJMoa2206038>

986 56 Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving
987 the likelihood of success: A review. *Contemp Clin Trials Commun* **11**, 156-164 (2018).
988 <https://doi.org:10.1016/j.conctc.2018.08.001>

989 57 Sun, K. *et al.* Endotrophin triggers adipose tissue fibrosis and metabolic dysfunction. *Nat*
990 *Commun* **5**, 3485 (2014). <https://doi.org:10.1038/ncomms4485>

991 58 Chirinos, J. A. *et al.* Endotrophin, a Collagen VI Formation-Derived Peptide, in Heart
992 Failure. *NEJM Evidence* **1**, EVIDoA2200091 (2022).
993 <https://doi.org:doi:10.1056/EVIDoA2200091>

994 59 Li, X. *et al.* Critical Role of Matrix Metalloproteinase 14 in Adipose Tissue Remodeling
995 during Obesity. *Mol. Cell. Biol.* **40**, e00564-00519 (2020).
996 <https://doi.org/doi:10.1128/MCB.00564-19>

997 60 Jo, W. *et al.* MicroRNA-29 Ameliorates Fibro-Inflammation and Insulin Resistance in
998 HIF1alpha-Deficient Obese Adipose Tissue by Inhibiting Endotrophin Generation.
999 *Diabetes* **71**, 1746-1762 (2022). <https://doi.org/10.2337/db21-0801>

1000 61 Vukicevic, S. *et al.* Bone morphogenetic protein 1.3 inhibition decreases scar formation
1001 and supports cardiomyocyte survival after myocardial infarction. *Nat Commun* **13**, 81
1002 (2022). <https://doi.org/10.1038/s41467-021-27622-9>

1003 62 Lim, G. B. Genetics: Polymorphisms in ANGPTL4 link triglycerides with CAD. *Nat. Rev.*
1004 *Cardiol.* **13**, 245 (2016). <https://doi.org/10.1038/nrcardio.2016.46>

1005 63 Dewey, F. E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease.
1006 *N Engl J Med* **374**, 1123-1133 (2016). <https://doi.org/10.1056/NEJMoa1510926>

1007 64 Morris, A. Obesity: ANGPTL4 - the link binding obesity and glucose intolerance. *Nat Rev*
1008 *Endocrinol* **14**, 251 (2018). <https://doi.org/10.1038/nrendo.2018.35>

1009 65 Janssen, A. W. F. *et al.* Loss of angiopoietin-like 4 (ANGPTL4) in mice with diet-induced
1010 obesity uncouples visceral obesity from glucose intolerance partly via the gut microbiota.
1011 *Diabetologia* **61**, 1447-1458 (2018). <https://doi.org/10.1007/s00125-018-4583-5>

1012 66 Gusarova, V. *et al.* Genetic inactivation of ANGPTL4 improves glucose homeostasis and
1013 is associated with reduced risk of diabetes. *Nat Commun* **9**, 2252 (2018).
1014 <https://doi.org/10.1038/s41467-018-04611-z>

1015 67 Deng, M. *et al.* ANGPTL4 silencing via antisense oligonucleotides reduces plasma
1016 triglycerides and glucose in mice without causing lymphadenopathy. *J. Lipid Res.* **63**,
1017 100237 (2022). <https://doi.org/10.1016/j.jlr.2022.100237>

1018 68 Georgakis, M. K. & Gill, D. Mendelian Randomization Studies in Stroke: Exploration of
1019 Risk Factors and Drug Targets With Human Genetic Data. *Stroke* **52**, 2992-3003 (2021).
1020 <https://doi.org/10.1161/STROKEAHA.120.032617>

1021 69 Verhamme, P. *et al.* Abrelcimab for Prevention of Venous Thromboembolism. *N Engl J*
1022 *Med* **385**, 609-617 (2021). <https://doi.org/10.1056/NEJMoa2105872>

1023 70 Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes
1024 for coronary artery disease in over a million participants. *Nature Genetics* (2022).
1025 <https://doi.org/10.1038/s41588-022-01233-6>

1026 71 Tcheandjieu, C. *et al.* Large-scale genome-wide association study of coronary artery
1027 disease in genetically diverse populations. *Nat Med* **28**, 1679-1692 (2022).
1028 <https://doi.org/10.1038/s41591-022-01891-3>

1029 72 Carter, A. R. *et al.* Mendelian randomisation for mediation analysis: current methods and
1030 challenges for implementation. *Eur J Epidemiol* **36**, 465-478 (2021).
1031 <https://doi.org/10.1007/s10654-021-00757-1>

1032 73 Pierce, B. L., Ahsan, H. & Vanderweele, T. J. Power and instrument strength requirements
1033 for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**,
1034 740-752 (2011). <https://doi.org/10.1093/ije/dyq151>

1035 74 Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic
1036 instruments in two-sample multivariable Mendelian randomization. *Stat. Med.* **40**, 5434-
1037 5452 (2021). <https://doi.org/10.1002/sim.9133>

1038 75 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic
1039 association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
1040 <https://doi.org/10.1371/journal.pgen.1004383>

1041 76 Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European
1042 Prospective Investigation of Cancer. *Br. J. Cancer* **80 Suppl 1**, 95-103 (1999).

1043 77 Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R Package for
1044 Causal Mediation Analysis. *Journal of Statistical Software* **59**, 1 - 38 (2014).
1045 <https://doi.org:10.18637/jss.v059.i05>
1046 78 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**,
1047 68-74 (2015). <https://doi.org:10.1038/nature15393>
1048 79 Cichonska, A. *et al.* metaCCA: summary statistics-based multivariate meta-analysis of
1049 genome-wide association studies using canonical correlation analysis. *Bioinformatics* **32**,
1050 1981-1989 (2016). <https://doi.org:10.1093/bioinformatics/btw052>
1051 80 Grant, A. J. & Burgess, S. Pleiotropy robust methods for multivariable Mendelian
1052 randomization. *Stat. Med.* **40**, 5813-5830 (2021). <https://doi.org:10.1002/sim.9156>
1053 81 Wang, J. & Pan, W. The Biological Role of the Collagen Alpha-3 (VI) Chain and Its Cleaved
1054 C5 Domain Fragment Endotrophin in Cancer. *Onco Targets Ther.* **13**, 5779-5793 (2020).
1055 <https://doi.org:10.2147/OTT.S256654>
1056 82 Chen, P., Cescon, M. & Bonaldo, P. Collagen VI in cancer and its biological mechanisms.
1057 *Trends Mol. Med.* **19**, 410-417 (2013). <https://doi.org:10.1016/j.molmed.2013.04.001>