

1 **Multi-ancestry proteome-phenome-wide Mendelian randomization offers a comprehensive**
2 **protein-disease atlas and potential therapeutic targets**
3

4 Chen-Yang Su^{1,2}, Adriaan van der Graaf³, Wenmin Zhang⁴, Dong-Keun Jang⁵, Susannah
5 Selber-Hnatiw^{1,6}, Ta-Yu Yang⁷, Guillaume Butler-Laporte^{2,8}, Kevin Y. H. Liang^{2,8}, Fumihiko
6 Matsuda⁷, Maria C. Costanzo⁴, Noel P. Burt⁵, Jason Flannick^{5,9}, Sirui Zhou^{1,2,6}, Vincent
7 Mooser^{1,6}, Tianyuan Lu^{10,11,12,13,*}, Satoshi Yoshiji^{1,2,5,6,8,*}
8

9 ¹McGill Genome Centre, McGill University, Montréal, Québec, Canada

10 ²Quantitative Life Sciences, McGill University, Montreal, Québec, Canada

11 ³Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

12 ⁴Montreal Heart Institute, Université de Montréal, Montreal, Québec, Canada

13 ⁵Programs in Metabolism and Medical & Population Genetics, The Broad Institute of MIT and Harvard,
14 Cambridge, MA, USA

15 ⁶Department of Human Genetics, McGill University, Montreal, Québec, Canada

16 ⁷Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

17 ⁸Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada

18 ⁹Harvard Medical School, Boston, MA, USA

19 ¹⁰Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, USA

20 ¹¹Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI,
21 USA

22 ¹²Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI, USA

23 ¹³Center for Demography of Health and Aging, University of Wisconsin-Madison, Madison, WI, USA
24

25 ***Correspondence and equal contribution:**

26 Tianyuan Lu (tianyuan.lu@wisc.edu) and Satoshi Yoshiji (satoshi.yoshiji@mcgill.ca)

27 **Abstract**

28 Circulating proteins influence disease risk and are valuable drug targets. To enhance the
29 discovery of protein-phenotype associations and identify potential therapeutic targets across
30 diverse populations, we conducted proteome-phenome-wide Mendelian randomization in three
31 ancestries, followed by comprehensive sensitivity analyses. We tested the potential causal effects
32 of up to 2,265 unique proteins on a curated list of 355 distinct phenotypes, assessing 726,035
33 protein-phenotype pairs in European, 33,078 in African, and 115,352 in East Asian ancestries.
34 Notably, 119 proteins were instrumentable only in African ancestry and 17 proteins only in East
35 Asian ancestry due to allele frequency differences that are common in these ancestries but rare
36 in European ancestry. We identified 3,949, 56, and 325 unique protein-phenotype pairs in
37 European, African, and East Asian ancestries, respectively, and assessed their druggability using
38 multiple databases. We highlighted the causal role of IL1RL1 in inflammatory bowel diseases,
39 supported by multiple orthogonal lines of evidence. Taken together, this study underscores the
40 importance of multi-ancestry inclusion and offers a comprehensive atlas of protein-phenotype
41 associations across three ancestries, enhancing our understanding of proteins involved in disease
42 etiology and potential therapeutic targets. Results are available at the Common Metabolic
43 Diseases Knowledge Portal (https://broad.io/protein_mr_atlas).
44

45 Introduction

46 Circulating proteins play a major role in a multitude of biological pathways¹⁻³, are important
47 biomarkers for disease diagnosis, prognosis, and prevention⁴⁻⁶, and serve as valuable drug
48 targets⁷⁻¹⁰. Current high-throughput proteomics platforms measure thousands of circulating
49 plasma proteins. With measurements in large cohorts, recent studies have conducted genome-
50 wide association studies (GWAS) to evaluate genetic variants associated with the abundances of
51 thousands of proteins¹¹⁻¹⁴. We can leverage these proteomic GWAS to find causal proteins for
52 diseases and prioritize drug targets through Mendelian randomization (MR)^{15,16}. In this case, MR
53 uses genetic variants associated with variation in plasma protein levels (known as protein
54 quantitative trait loci, pQTLs) as instruments to measure the causal effect of an exposure (protein
55 abundance) on an outcome (a complex trait). This is shown to reduce confounding and reverse
56 causation biases affecting many epidemiological studies, provided that three key assumptions
57 are met: (1) the genetic variant is associated with the exposure, (2) there is no confounding of the
58 instrument-outcome association, and (3) the genetic variant influences the outcome solely
59 through the exposure.

60
61 Despite applications of proteogenomics in elucidating disease mechanisms and identifying
62 potential therapeutic targets¹⁷⁻²³, previous pQTL studies^{18,19} have been based on smaller sample
63 sizes and measured fewer proteins, assessed limited outcomes, and most importantly, have
64 predominantly focused on individuals of European ancestry. African ancestry proteomics cohorts
65 have emerged in recent years^{13,14,24}, yet existing studies have assessed a limited number of
66 outcomes and comparisons with other ancestries have also been limited¹⁹. Similarly, in East Asian
67 ancestries, few large-scale proteome-phenome wide MR studies exist^{25,26}, largely due to the lack
68 of publicly available pQTLs, despite the presence of existing East Asian ancestry biobanks
69 providing hundreds of publicly available GWAS outcomes^{27,28}.

70
71 The inclusion of multiple ancestries in a proteome-phenome wide atlas of associations can
72 potentially offer significant benefits. Diverse ancestry inclusion in MR leverages the natural
73 variations in genetic architecture present across populations, which allows analyses on otherwise
74 unseen genetic variation, increases in statistical power due to allele frequency increases, and
75 differentiation of causal effect magnitude across populations^{29,30}. Combined, this approach may
76 be able to identify a greater number of instrumentable proteins (i.e., proteins that can be tested
77 in MR), which may lead to an increase in discoveries. By leveraging insights obtained through
78 instrumentable proteins in one ancestry, the identified protein-phenotype associations could
79 contribute to our understanding of disease mechanisms, which may be generalizable across
80 different ancestries and potentially benefit all populations. For instance, PCSK9 loss-of-function
81 variants Y142X and C679X predispose individuals to naturally lower LDL cholesterol levels.
82 These mutations were found to be common in African Americans but rare in European
83 Americans³¹ and inspired the development of PCSK9 inhibitors mimicking these variants to
84 effectively reduce LDL cholesterol levels and risk of cardiovascular events^{32,33}. In addition to
85 providing insights into novel therapeutic targets, enhancing diversity in genomic studies is crucial
86 to ensure equitable health outcomes and address imbalances in health disparities across
87 populations³⁴.

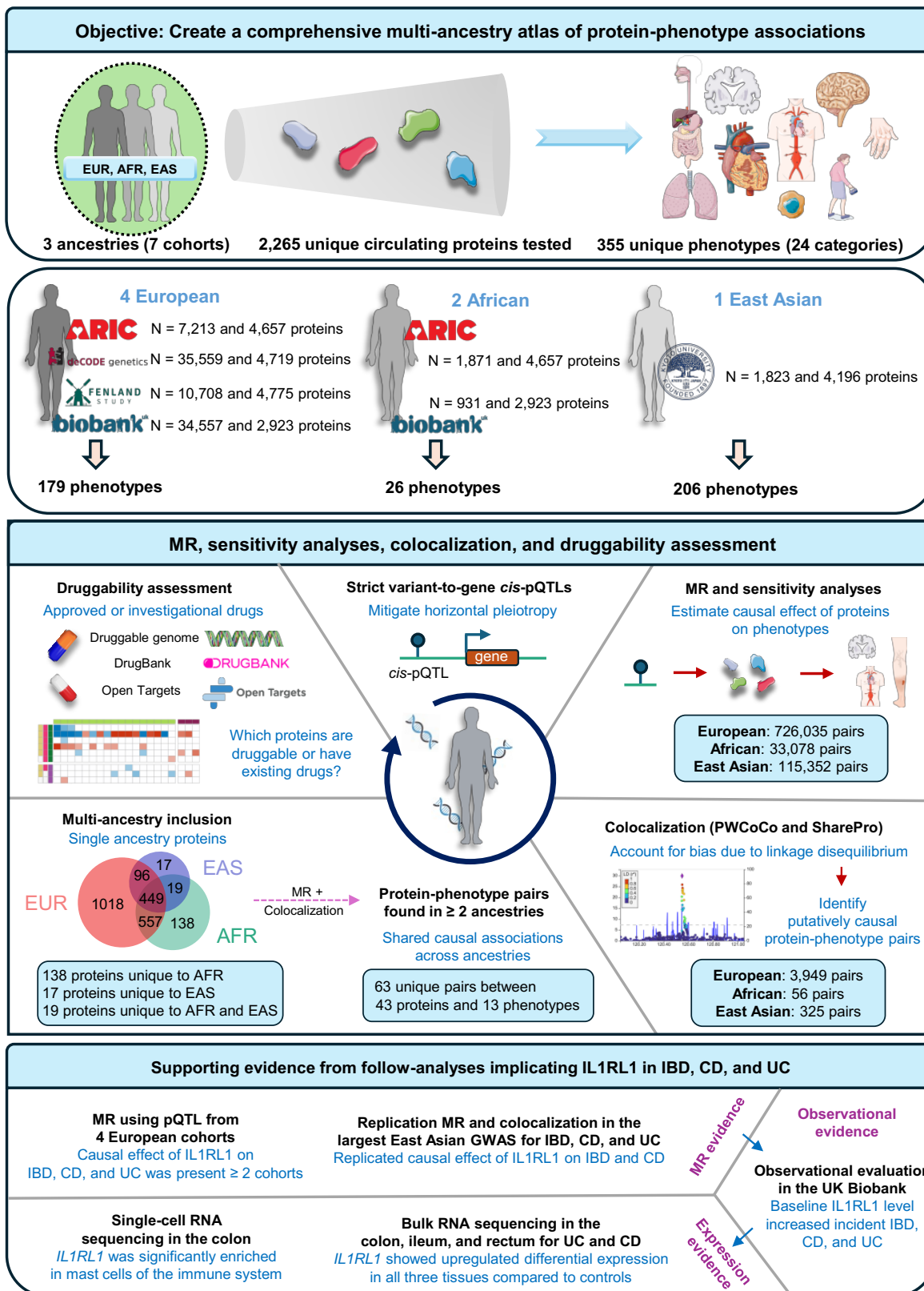
88
89 Here, we combined four of the largest European ancestry proteomics cohorts ($n =$ up to 35,559),
90 two of the largest African proteomics cohorts ($n =$ up to 1,871), and a new East Asian proteomics
91 cohort ($n =$ 1,823). We then performed MR and colocalization analyses on a curated list of the
92 most recent and largest ancestry-specific outcome GWAS to date for 179 European and 26
93 African ancestry outcomes, as well as 206 East Asian ancestry outcomes from Biobank Japan to
94 construct an atlas of protein-phenotype associations. We integrated our findings with multiple
95 drug databases to assess the druggability of the associations and highlighted novel targets.

96 Overall, our study supports the prioritization of thousands of protein-phenotype associations and
97 provides a comprehensive, updated resource for the community, significantly expanding our
98 understanding of these associations. Results are publicly available at the Common Metabolic
99 Diseases Knowledge Portal (https://broad.io/protein_mr_atlas).
100

101 **Results**

102 The overall study design is shown in **Fig. 1**.

103



104

105 **Figure 1. Study design.**

106 We assessed the causal role of 2,265 circulating plasma proteins across three ancestries (four
107 European, two African, and a new East Asian ancestry cohort) on up to 355 phenotypes/outcomes
108 with extensively curated GWAS in each ancestry. For European and African ancestry outcomes,
109 we collected the largest and most recent GWAS available as of February 2024 for 179 and 26
110 outcomes, respectively, while 206 GWAS from BioBank Japan were used for East Asian ancestry
111 outcomes. We implemented a unique approach for defining *cis*-pQTLs using a multi-step process
112 combining a strict *cis*-pQTL definition with Open Targets Genetics variant-to-gene score³⁵ filtering
113 to minimize risk of horizontal pleiotropy, which we term “strict variant-to-gene (V2G) *cis*-pQTLs”.
114 Then, we performed multi-ancestry proteome-wide MR and colocalization analyses using
115 PWCoCo^{18,36} and SharePro³⁷, a new colocalization method we developed, on the curated
116 phenotypes to identify causal protein-phenotype associations. Next, we overlapped prioritized
117 protein-phenotype pairs with drug databases such as the druggable genome³⁸, DrugBank³⁹, and
118 the Open Targets platform⁴⁰. Finally, as an illustrative example of the value of our atlas, we
119 pinpoint IL1RL1 and explore its role in inflammatory bowel disease using multiple lines of
120 evidence. EUR: European, AFR: African, EAS: East Asian; MR: Mendelian randomization; IBD:
121 Inflammatory bowel disease; CD: Crohn’s disease; UC: Ulcerative colitis.

122
123
124

125 **1. Genetic instrument selection for proteins**

126 We summarize the selected genetic instruments in **Supplementary Table 1**. Briefly, we used
127 proteomic GWAS from four European ancestry cohorts: ARIC (4,657 proteins measured in up to
128 7,213 individuals)¹³, deCODE (4,719 proteins measured in up to 35,559 individuals)¹², Fenland
129 (4,775 proteins measured in up to 10,708 individuals)¹¹, and UKB-PPP (2,923 proteins measured
130 in up to 34,557 individuals)¹⁴. For African ancestry, we analyzed proteomic GWAS from two
131 cohorts: ARIC (4657 proteins measured in up to 1,871 individuals)¹³ and UKB-PPP (2923 proteins
132 measured in up to 931 individuals)¹⁴. Additionally, we included a new East Asian ancestry cohort,
133 the Kyoto University Nagahama cohort which measured 4,196 proteins in up to 1,823 individuals.

134
135

136 **1.1. Unifying *cis*-pQTL across cohorts**

137 We re-defined *cis*-pQTLs across proteomics cohorts as pQTLs within 500 kb of the transcription
138 start site (TSS) of the protein-coding gene, with independence and significance defined with
139 linkage disequilibrium (LD) $r^2 < 0.001$ and $P < 5 \times 10^{-8}$, respectively. Independent pQTLs outside
140 the *cis*-region were labeled as *trans*-pQTLs. We analyzed *cis*-pQTLs because they are more likely
141 to have direct biological effects on the proteins of interest^{41,42}. We verified that newly defined *cis*-
142 pQTLs had high concordance within ancestries (**Supplementary Note 1A**).

143
144

145 **1.2. Identifying strict V2G *cis*-pQTLs**

146 We mitigated the risk of horizontal pleiotropy by using a unique approach to select genetic
147 instruments which we term strict V2G *cis*-pQTLs (**Supplementary Note 1B**). Strict V2G *cis*-
148 pQTLs are associated with a single protein-coding gene (“strict”) and have the strongest link to
149 the corresponding protein-coding gene based on the Open Targets Genetics V2G score, which
150 uses multiple sources of evidence to map variants to genes (“V2G”)³⁵ (**Extended Data Fig. 1** and
151 **Supplementary Tables 2–8**). We also assessed if these strict V2G *cis*-pQTLs were protein
152 altering variants (PAVs), as PAVs may affect protein structure and lead to bias in effect size
153 estimation¹. We found that only 70 of 7,399 (0.95%), 30 of 1,684 (1.8%), and 9 of 663 (1.4%)
154 strict V2G *cis*-pQTLs in European, African, and East Asian ancestries, respectively, were PAVs
155 or in high LD with PAVs of high impact. Notably, strict V2G *cis*-pQTLs were more enriched for *cis*-
eQTLs, indicating their role in local gene regulation impacting protein levels (**Extended Data Fig. 2a**). They had significantly larger effects on protein levels (**Extended Data Fig. 2b**) and were

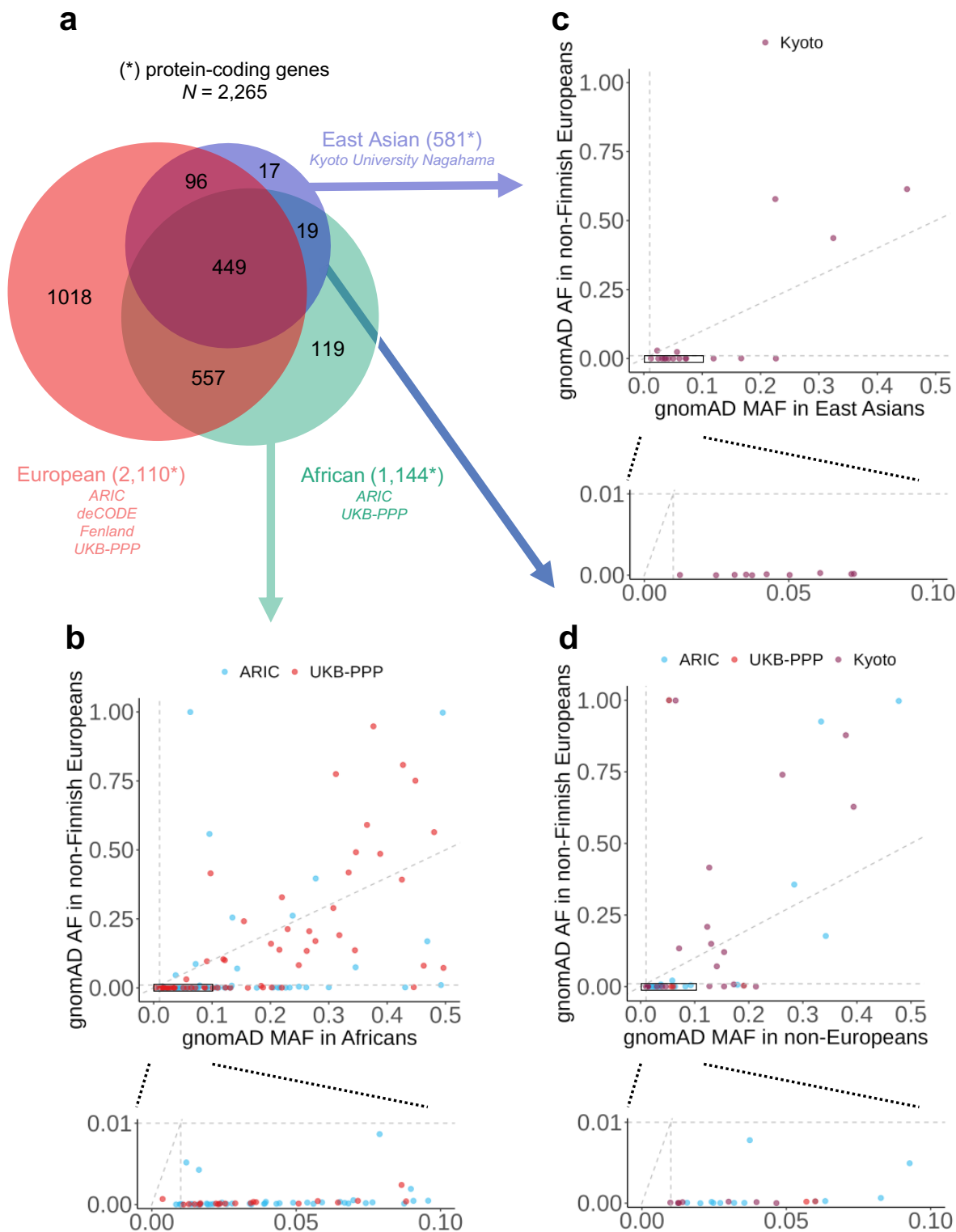
156 significantly closer to the TSS of the corresponding protein-coding gene (**Extended Data Fig. 3**).
157 Hence, strict V2G selection of pQTLs limits the risks of violating MR assumptions (specifically no
158 horizontal pleiotropy) while still retaining adequate statistical power. Specifically, all proteins in
159 each cohort had F-statistics above 10, suggesting that the risk of weak instrument bias is limited⁴³
160 (**Supplementary Table 9**).

161 162 2. Multi-ancestry inclusion is important in instrumenting proteins and reveals population-specific 163 variants

164 Across the three ancestries, we were able to instrument 2,265 unique proteins with 449 of these
165 being shared across all three ancestries (**Fig. 2a**). Specifically, we instrumented 2,110 proteins in
166 European, 1,144 in African, and 581 in East Asian ancestries (see **Supplementary Note 1C** for
167 cohort level description and **Extended Data Fig. 4**). We identified 1,018 proteins that were unique
168 to individuals of European ancestry. Including African ancestries allowed for an additional 138
169 proteins, with 119 unique to African ancestry and 19 shared with East Asian ancestries. Including
170 an East Asian ancestry cohort allowed an additional 17 unique proteins (**Fig. 2a**). Altogether,
171 these findings underscore the value of including multiple ancestries in proteomic analyses.

172
173 Next, to better understand the unique proteins in non-European ancestries, we compared the
174 allele frequencies of their genetic instruments using gnomAD⁴⁴. Of the 130 genetic instruments
175 for the 119 proteins unique to African ancestries, 89 (68.5%) had a minor allele frequency (MAF)
176 below 0.01 in European ancestries (**Fig. 2b**). Similarly, for the 18 genetic instruments for the 17
177 proteins unique to East Asian ancestry, 13 (72.2%) had a MAF below 0.01 in European ancestry
178 (**Fig. 2c**). The majority (29, 63.0%) of the 46 unique genetic instruments for the 19 proteins
179 instrumentable by both African and East Asian but not by European ancestries, were rare (MAF
180 < 0.01) in European ancestry (**Fig. 2d**). These results suggest that populational allele frequency
181 differences allow for more proteins to be included in MR analyses. We refer to them as uniquely
182 instrumentable proteins.

183



184
185
186
187
188
189
190

Figure 2. Cross ancestry comparison of overlapping instrumentable protein-coding genes.
(a) Overlapping instrumentable protein-coding genes across three ancestries. Protein-coding genes were quantified based on Ensembl gene IDs. Red: Unique protein-coding genes across four European ancestry proteomics cohorts; green: unique protein-coding genes across two African ancestry proteomics cohorts; blue: East Asian ancestry proteomics cohort protein-coding genes.

- 191 (b) gnomAD minor allele frequencies (MAF) of *cis*-pQTLs instrumenting the 119 proteins
192 uniquely instrumentable in African ancestry when plotted against corresponding gnomAD
193 non-Finnish European allele frequency (AF). The black box in the top plot is zoomed in
194 and shown in the bottom plot. Blue: ARIC African *cis*-pQTLs. Red: UKB-PPP African
195 ancestry cohort *cis*-pQTLs.
- 196 (c) gnomAD minor allele frequencies (MAF) of *cis*-pQTLs instrumenting the 17 proteins
197 uniquely instrumentable in East Asian ancestry when plotted against corresponding
198 gnomAD non-Finnish European AF. The black box in the top plot is zoomed in and shown
199 in the bottom plot. Purple: Kyoto University Nagahama East Asian ancestry cohort *cis*-
200 pQTLs.
- 201 (d) gnomAD MAF minor allele frequencies (MAF) of *cis*-pQTLs instrumenting the 19 proteins
202 common to African and East Asian but not European ancestries when plotted against
203 corresponding gnomAD non-Finnish AF. The black box in the top plot is zoomed in and
204 shown in the bottom plot. Blue: ARIC African *cis*-pQTLs. Red: UKB-PPP African *cis*-
205 pQTLs. Purple: Kyoto University Nagahama East Asian ancestry cohort *cis*-pQTLs.

206
207
208
209
210
211
212
213
214
215
216
217

3. Two-sample Mendelian randomization

We performed two-sample MR using strict V2G *cis*-pQTLs as instrumental variables to determine causal proteins implicated in human complex traits and diseases. Across three ancestries, we considered 355 unique outcomes (**Fig. 3a**) pertaining to 24 phenotype categories (**Fig. 3b**). These phenotypes/outcomes (as of February 2024) included 179 of the most up to date and largest phenotypic GWAS available for European ancestries (**Supplementary Table 10**), 26 of the largest available GWAS for African ancestries (**Supplementary Table 11**), and 206 from BioBank Japan for East Asian ancestries (**Supplementary Table 12**).

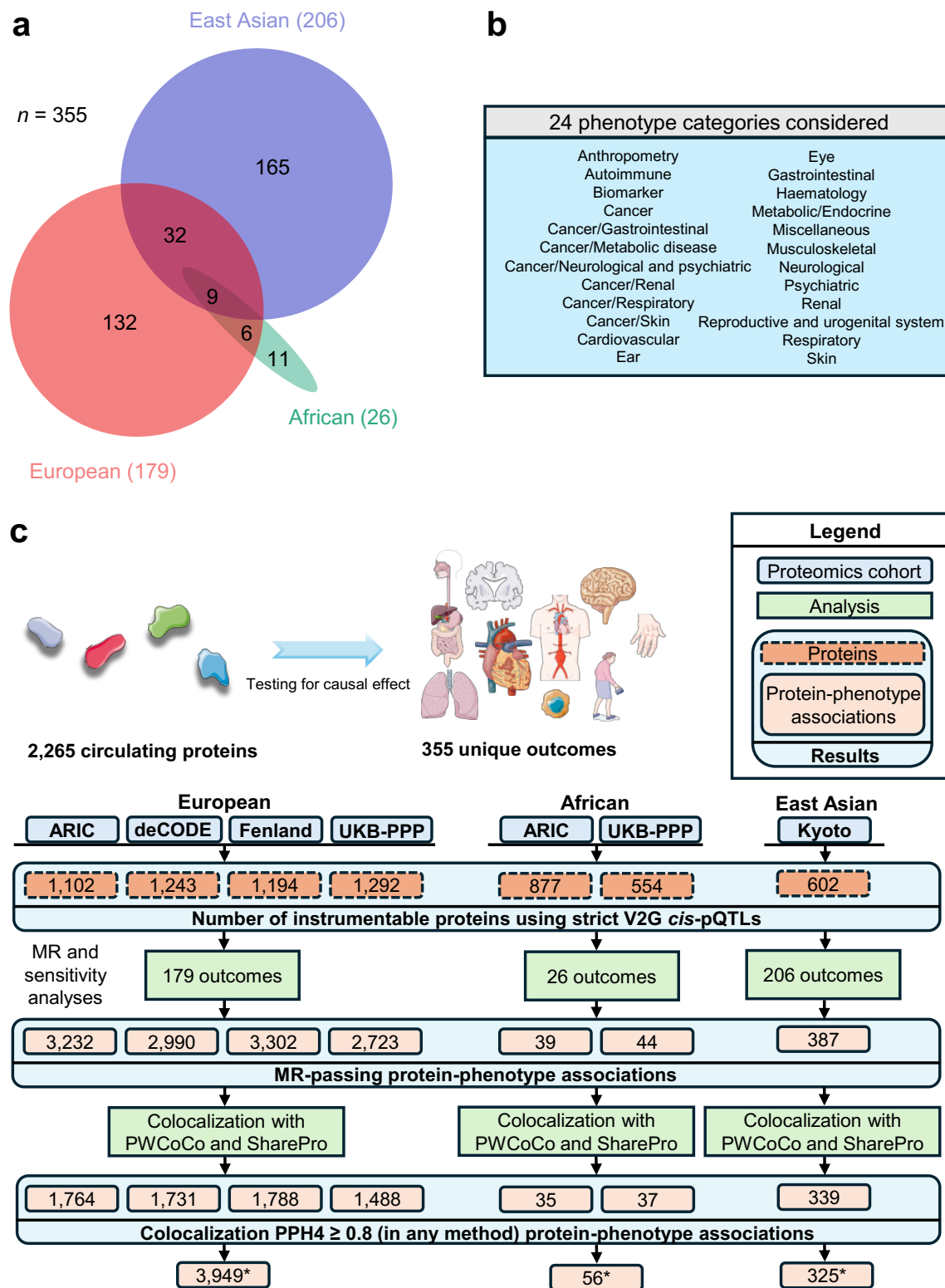


Figure 3. MR analyses to determine the effects of proteins on phenotypes.

(a) Curated GWAS phenotypes common between different ancestries. European (red), African (green), and East Asian (blue) ancestries.

218
219
220
221

- 222 (b) Phenotype categories for outcomes used in all three ancestries.
- 223 (c) Flowchart summary of MR and colocalization analyses to identify protein-phenotype
- 224 associations for each cohort.

225
226
227
228 We tested a total of 726,035 protein-phenotype pairs in European ancestries (173,748 for ARIC,
229 182,433 for deCODE, 180,018 for Fenland, and 189,836 for UKB-PPP), 33,078 in African
230 ancestries (20,442 for ARIC and 12,636 for UKB-PPP), and 115,352 in East Asian ancestries
231 (Kyoto University Nagahama cohort) (see **Data availability**). To control for false positives, we
232 applied a Benjamini-Hochberg-corrected P value (false discovery rate, FDR)⁴⁵ threshold of 0.05
233 (5%) per cohort in each ancestry¹⁹. We note that Bonferroni correction is overly stringent given
234 that (i) proteins are correlated with one another, and (ii) we tested the same protein-phenotype
235 associations across cohorts, making these tests not independent. Nevertheless, we also provide
236 the most stringent Bonferroni corrected associations in section 5. Results were further filtered
237 based on multiple sensitivity analyses robust to MR assumption violations and retained
238 associations are now termed "MR-passing". Assessment of sample overlap between European
239 ancestry GWAS outcomes that were based on the UK Biobank and proteomics from the European
240 ancestry UKB-PPP cohort was performed (**Supplementary Note 2**). Of the tested associations,
241 a total of 12,247 associations were considered MR-passing in European, 83 in African, and 387
242 in East Asian ancestries (**Supplementary Table 13**).

243
244 **4. Colocalization analyses**
245 MR results may be confounded by independent causal variants in LD^{46,47}. To guard against such
246 bias, for each MR-passing protein-phenotype pair, we performed colocalization analysis using
247 two methods, PWCoCo^{18,36} and SharePro³⁷, to verify that the protein abundance and the tested
248 outcome share the same genetic signals (**Fig. 3c**). An MR-passing protein-phenotype pair was
249 considered putatively causal if it was supported by at least one method with a posterior
250 colocalization probability (PP_{\max}) ≥ 0.8 . Across all cohorts and all three ancestries, 56.5% (7,182
251 out of 12,717) of MR-passing associations were supported by colocalization evidence
252 (**Supplementary Table 13**).

253
254 **5. Putatively causal protein-phenotype associations**
255 Upon MR, sensitivity analyses, and colocalization, we identified 3,949 unique putatively causal
256 protein-phenotype pairs in European (**Extended Data Fig. 5a** and **Supplementary Table 14**), 56
257 in African (**Extended Data Fig. 5b** and **Supplementary Table 15**), and 325 in East Asian
258 ancestries (**Extended Data Fig. 5c** and **Supplementary Table 16**). Here, we use protein to refer
259 to protein-coding genes to harmonize across SomaScan and Olink platforms. Results are also
260 hosted at https://broad.io/protein_mr_atlas. We described cohort level associations and proteins
261 implicated in a multitude of phenotypes in **Supplementary Notes 3 and 4**. Particularly, 1,617, 30,
262 and 135 unique associations in European, African, and East Asian ancestries further withstood
263 the Bonferroni correction accounting for the total number of tests across all cohorts and all
264 ancestries ($P < 0.05 / 874,465 = 5.7 \times 10^{-8}$), however, this threshold is likely overly conservative
265 due to many proteins being correlated with one another as well as non-independent tests from
266 the same protein-phenotypes associations being tested across cohorts (**Supplementary Tables**
267 **14 – 16**).

268
269 In European ancestries, the 3,949 significant unique protein-phenotype pairs identified in
270 European ancestries involved putatively causal effects between 995 proteins and 146 phenotypes,
271 of which only 56 (1.4%) showed discordant MR estimates in one of the tested cohorts
272 (**Supplementary Table 14**). These discrepancies could be attributed to population differences or

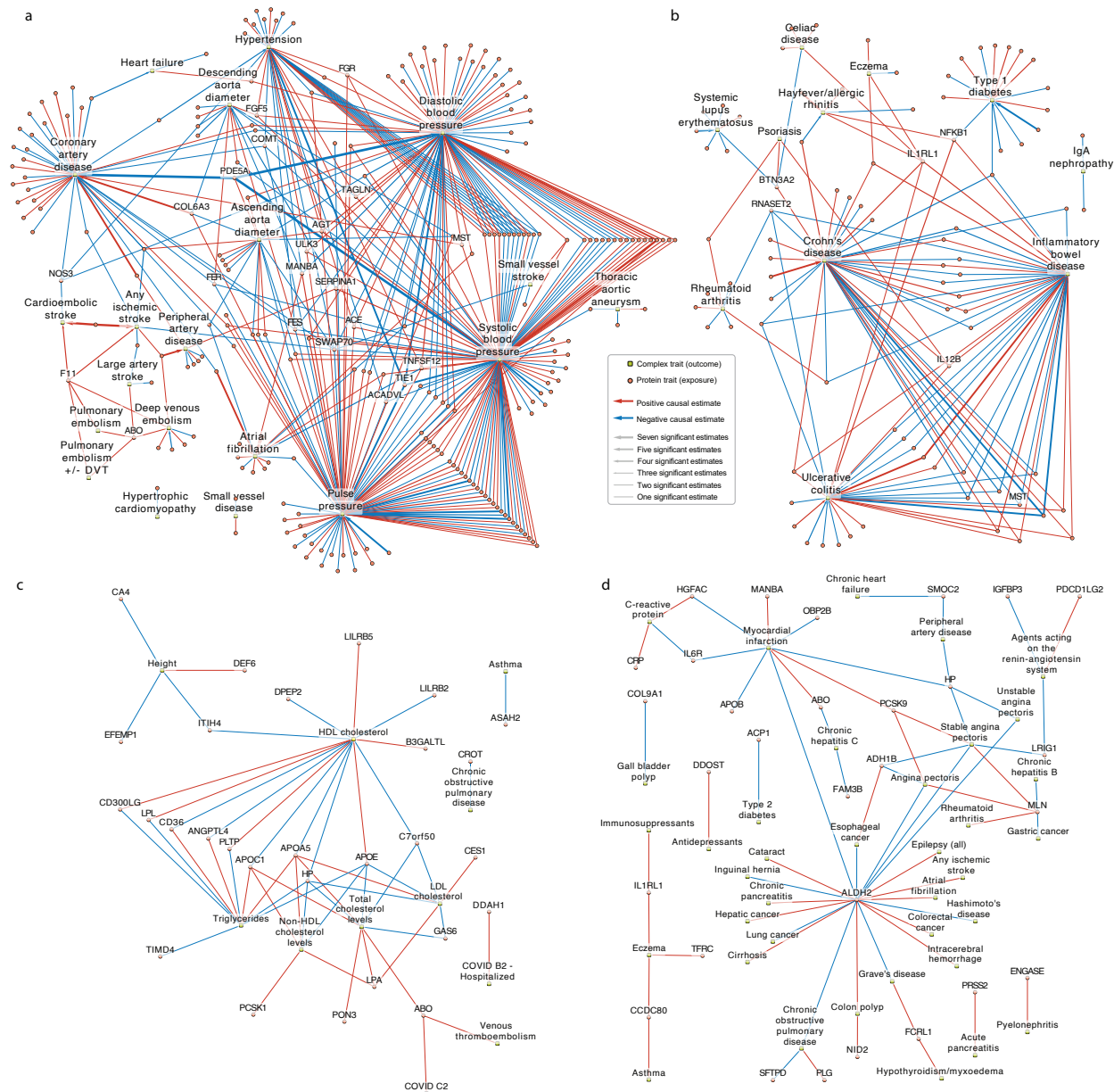
273 variations in proteomic assays, such as differential effects from SomaScan aptamers targeting
274 different domains compared to Olink assays. Of the 3,893 remaining putatively causal European
275 ancestry pairs between 991 proteins and 146 outcomes, 1,692 (43.5%) of the identified putatively
276 causal associations were from 452 proteins uniquely instrumentable by European ancestries.
277 Further, 3,853 (99.0%) associations have not been previously reported by earlier proteome-
278 phenome wide MR studies from Zheng et al.¹⁸ and Zhao et al.¹⁹.

279 *5.1. Cardiovascular and autoimmune diseases*

281 We demonstrate the highly interconnected nature between proteins and outcomes by highlighting
282 cardiovascular (**Fig. 4a** and **Supplementary Note 5A**) and autoimmune phenotypes (**Fig. 4b**) in
283 European ancestries given their significant impact on health. Cardiovascular outcomes were
284 influenced by up to 103 proteins (median: 7) while some proteins influenced up to 8 cardiovascular
285 outcomes (median: 1) (**Fig. 4a**). For instance, we found that an s.d. increase in genetically
286 predicted ULK3 levels increases systolic and diastolic blood pressure, pulse pressure, and
287 hypertension. ULK3 is a nuclear kinase which may contribute to vascular disease by mediating
288 autophagy dysregulation⁴⁸. In concordance, a recent study showed that functional splicing effects
289 of ULK3 can contribute to coronary artery disease (CAD)⁴⁹ suggesting effective modulation of
290 ULK3 may be beneficial for reducing cardiovascular risk.

291
292 Similarly, we found high interconnectedness between autoimmune phenotypes and proteins (**Fig.**
293 **4b**). Autoimmune phenotypes were influenced by a median of 7 proteins; the highest number of
294 associations were for Crohn's disease (CD) (39 proteins), inflammatory bowel disease (IBD) (35
295 proteins), and ulcerative colitis (UC) (28 proteins). Proteins, on the other hand, influenced a
296 median of 2 autoimmune phenotypes. IL1RL1 and IL12B influenced the largest number of
297 outcomes (6 and 4 outcomes, respectively) (**Fig. 4b**) with both increasing risk of IBD, CD, and
298 UC. IL12B is targeted by a commercially available drug ustekinumab for CD⁵⁰; however, there are
299 currently no approved drugs targeting IL1RL1, although there has been increasing interest in
300 modulating IL1RL1 for treating various conditions^{51,52}. For example, tozorakimab targets IL-33,
301 the interleukin that binds IL1RL1 (ST2)⁵³ and anakinra targets IL-1⁵⁴, a closely related interleukin.

302
303 Other potentially novel findings involving other disease categories in European ancestries are
304 highlighted in **Supplementary Note 5B**.
305



306
307
308
309
310
311
312
313
314
315
316
317
318
319
320

Figure 4. Protein-phenotype network plots.

Red arrows indicate a positive causal estimate of the protein on the phenotype while blue arrows indicate a negative causal estimate of the protein on the phenotype.

(a) Significant estimates between proteins (orange circles) and cardiovascular traits (green rectangles) both derived from European ancestry individuals. All cardiovascular traits are named, as well as the protein traits that have a significant effect on 4 or more cardiovascular traits. Arrow thickness representing number of significant estimates indicates how often a protein measurement has a causal effect on the outcome trait. The maximum number of significant estimates is 7 which is larger than the total number of European ancestry proteomics studies, 4, due to the presence of more than one SomaScan aptamer for that protein. For simplicity, we only depict protein-phenotype pairs in which all European ancestry cohorts showed concordant direction of effect estimates.

(b) Significant estimates between proteins (orange circles) and autoimmune traits (green rectangles) both derived from European ancestry individuals. All autoimmune traits are

321 named, as well as the protein traits that have a significant effect on 4 or more autoimmune
322 traits. Arrow thickness representing number of significant estimates indicates how often a
323 protein measurement has a causal effect on the outcome trait. For simplicity, we only
324 depict protein-phenotype pairs in which all European ancestry cohorts showed concordant
325 direction of effect estimates.

326 (c) Significant estimates between proteins (orange circles) and all traits (green rectangles)
327 both derived from African ancestry individuals. All arrows are the same thickness to
328 indicate support of the association in a single study.

329 (d) Significant estimates between proteins (orange circles) and binary traits (green
330 rectangles) both derived from East Asian ancestry individuals. All arrows are the same
331 thickness and indicate support of the association in a single study.

332
333
334

335 *5.2. Protein-phenotype associations in African and East Asian ancestries*

336 In African ancestries, we identified 56 unique protein-phenotype associations involving 28
337 proteins and 11 phenotypes (**Fig. 4c** and **Supplementary Table 15**). Of these 56 pairs, 55
338 (98.2%) have not been previously reported by earlier proteome-phenome wide MR studies in
339 African ancestries¹⁹. Notably, 11 (19.6%) protein-phenotype pairs involving four proteins, APOE,
340 C7orf50, CD300LG, and PON3, were uniquely instrumentable in African ancestries (**Extended**
341 **Data Fig. 6a**). For instance, increased PON3 levels was associated with increased total
342 cholesterol ($\beta_{\text{UKB-PPP}} = 0.03$, 95% CI = 0.02–0.05, $P = 1.1 \times 10^{-5}$, $\text{PP}_{\text{max}} = 0.95$). *PON3* is highly
343 expressed in the liver and previously implicated in cholesterol metabolism^{55,56} and atherosclerosis
344 progression. This is notable as *PON3* was not instrumentable in European ancestries, thus,
345 African ancestries are uniquely suitable to identify biologically plausible protein-phenotype
346 associations.

347
348 Furthermore, the Million Veteran Program⁵⁷ recently released summary statistics for additional
349 traits in up to 635,969 individuals. Using this resource, we tested the causal effect of the 119
350 proteins uniquely instrumentable in African ancestries on cardiovascular and autoimmune-related
351 binary outcomes. Using the African ARIC proteomics cohort, we identified 7 associations with
352 cardiovascular outcomes and no associations with autoimmune-related outcomes
353 (**Supplementary Note 5C**). Notably, increased PCYOX1 levels were associated with a reduced
354 risk of coronary atherosclerosis, atrial fibrillation, and flutter, indicating its protective effect.
355 PCYOX1 plays a role in oxidative stress and lipid metabolism and has been implicated in
356 atherosclerosis in rodent studies⁵⁸, supporting our findings. This suggests that as more outcomes
357 with larger sample size, particularly binary disease outcomes, become available in African
358 ancestries, we can better leverage uniquely instrumentable proteins to uncover additional protein-
359 disease associations.

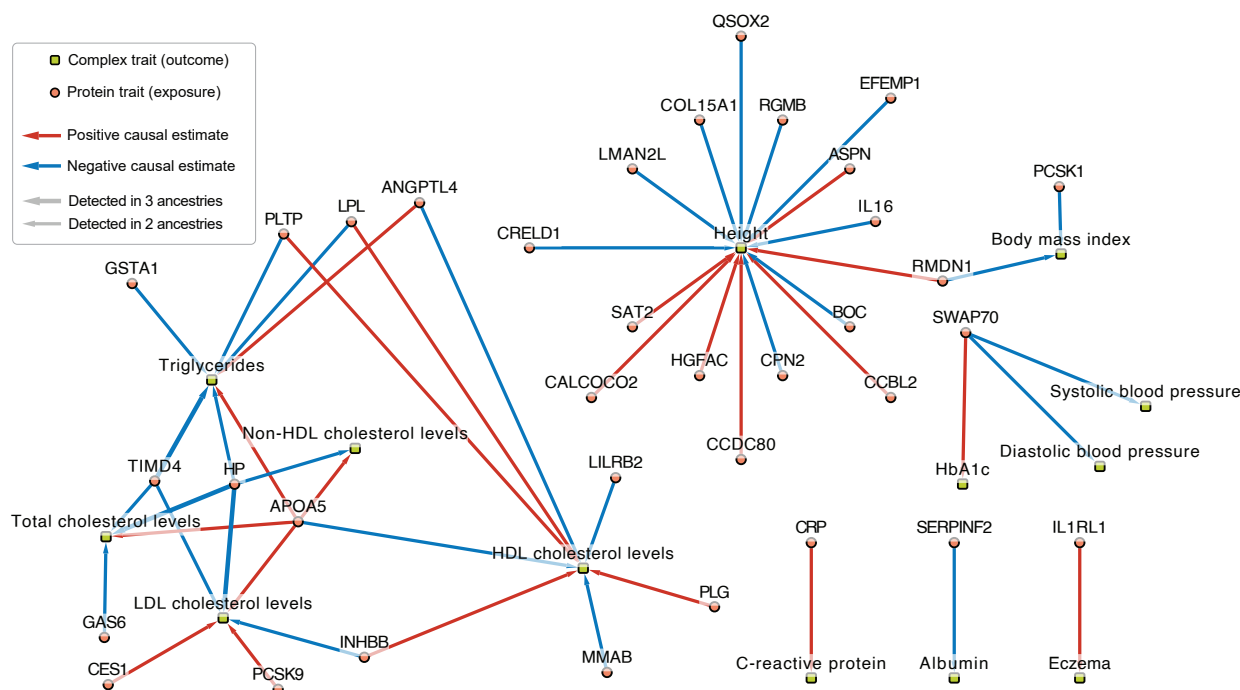
360
361 In East Asian ancestries, 339 putatively causal associations were identified with 325 unique
362 protein-phenotype pairs involving 110 proteins and 86 phenotypes (**Fig. 4d** and **Supplementary**
363 **Table 16**). Among them, increased *SMOC2* level was associated with decreased risk of peripheral
364 artery disease (PAD) (OR = 0.82, 95% CI = 0.74–0.90, $P = 5.5 \times 10^{-5}$, $\text{PP}_{\text{max}} = 0.97$). *SMOC2* is
365 highly expressed in arteries and plays roles in endothelial cell proliferation, angiogenesis, and
366 matrix assembly⁵⁹. Notably, in European ancestries, increased *SMOC2* was associated with
367 decreased pulse pressure, implicating favorable cardiovascular effects across ancestries.

368
369 We also found 8 proteins which were only instrumentable in East Asian ancestry due to lack of
370 genome-wide significant *cis*-pQTLs in other ancestries or stringent strict V2G instrument selection
371 (**Supplementary Note 5D**). These 8 proteins were *ALDH2*, *ANXA7*, *APOA1*, *DDOST*, *GSS*,

372 PLA2G7, PRSS2, and UGT1A1 which together accounted for 67 (20.6%) protein-phenotype
 373 associations (**Extended Data Fig. 6b** and **Supplementary Note 5D**). For instance, increased
 374 PRSS2 levels in East Asian ancestries was associated with an OR of 2.05 for acute pancreatitis
 375 (95% CI: 1.42–2.98, $P = 1.43 \times 10^{-4}$, $PP_{\max} = 0.95$) which has been previously validated in
 376 European only proteome-wide analyses⁶⁰ suggesting concordant effects across ancestries.

377
 378 **6. Concordant effects across ancestries**

379 We evaluated the direction of effect for protein-phenotype associations that passed MR and
 380 colocalization analyses across ancestries since concordance across ancestries may strengthen
 381 the evidence of broader applicability of therapeutic targets. We found 186 total protein-
 382 phenotype/outcome pairs with 63 unique pairs between 43 proteins and 13 outcomes
 383 (**Supplementary Table 17** and **Fig. 5**). Among these, 51 pairs (81.0%) had concordant effects
 384 across ancestries, while 12 pairs (19.0%) had inconsistent effects (**Supplementary Figure 3** and
 385 **Supplementary Note 6**). Notable concordant associations included PCSK9 and LDL cholesterol
 386 in European and East Asian ancestries, and haptoglobin (HP), which binds free hemoglobin to
 387 prevent oxidative damage, with LDL cholesterol and total cholesterol across all three ancestries.
 388 Angiotensin-related protein 4 (ANGPTL4) was negatively associated with HDL cholesterol and
 389 positively with triglycerides in European and African ancestries, while lipoprotein-lipase (LPL)
 390 showed opposite associations in these ancestries. ANGPTL4 has been shown to act as a local
 391 inhibitor of LPL⁶¹ which serves as the rate-limiting enzyme in the degradation of triglycerides⁶²,
 392 concordant with our findings. While no approved drugs exist for ANGPTL4, inhibition of a closely
 393 related protein ANGPTL3 through an RNA interference therapy zodasiran is currently undergoing
 394 clinical trials for cholesterol lowering⁶³. Thus, many of these biological effects were validated in
 395 our multi-ancestry analyses. We note that while concordant effects across ancestries provide
 396 strong evidence of support, discordant effects in MR do not automatically indicate biologically
 397 discordant effects across ancestries, which could be due to epitope-binding effects or other
 398 technical variations⁶⁴.



399
 400

401 **Figure 5. Multi-ancestry network plot for protein-phenotype pairs present in two or more**
402 **ancestries.**

403 All estimates shown had evidence in European ancestry and at least one other ancestry (African
404 or East Asian ancestry). Significant estimates between proteins (orange circles) and traits (green
405 rectangles). Arrow thickness indicates the number of ancestries in which a protein measurement
406 has a causal effect on the phenotype. For simplicity, we only depict estimates when cohorts within
407 the same ancestry and across different ancestries showed concordant direction of effect
408 estimates. Red arrows indicate a positive causal estimate of the protein on the phenotype while
409 blue arrows indicate a negative causal estimate of the protein on the phenotype.

410
411

412
413 **7. Druggability**

414 **7.1. Druggability of the instrumentable protein-coding genes**

415 Across all ancestries, between 60% to 67% of instrumentable protein-coding genes overlapped
416 with the druggable genome³⁸, which classifies genes into Tier 1, 2, or 3 according to druggability,
417 and between 20.5% and 23.6% overlapped with Tiers 1 and 2 (**Supplementary Table 18** and
418 **Supplementary Note 7A**). Druggability tiers for instrumentable proteins are in **Supplementary**
419 **Tables 19–25**.

420

421 Next, we incorporated the druggable genome³⁸, DrugBank³⁹, and Open Targets Platform⁴⁰ to
422 determine which instrumentable proteins overlapped at least one database. Proteins overlapping
423 with DrugBank³⁹ had approved or investigational drugs available while those overlapping Open
424 Targets have information on their clinical development phase and status of protein-drug-disease
425 combinations. Cross-ancestry comparison stratified by proteomics platform for instrumentable
426 proteins overlapping at least one database shows that in SomaScan v4, African ancestry adds 68
427 additional targets beyond European ancestry cohorts (**Extended Data Fig. 7a**), while East Asian
428 ancestry contributes 34 more targets (**Extended Data Fig. 7b**). In Olink Explore 3072, including
429 African ancestry presents 62 additional targets (**Extended Data Fig. 7c**). These findings suggest
430 that data from African and East Asian ancestries could enhance drug development by offering
431 more potential therapeutic targets.

432

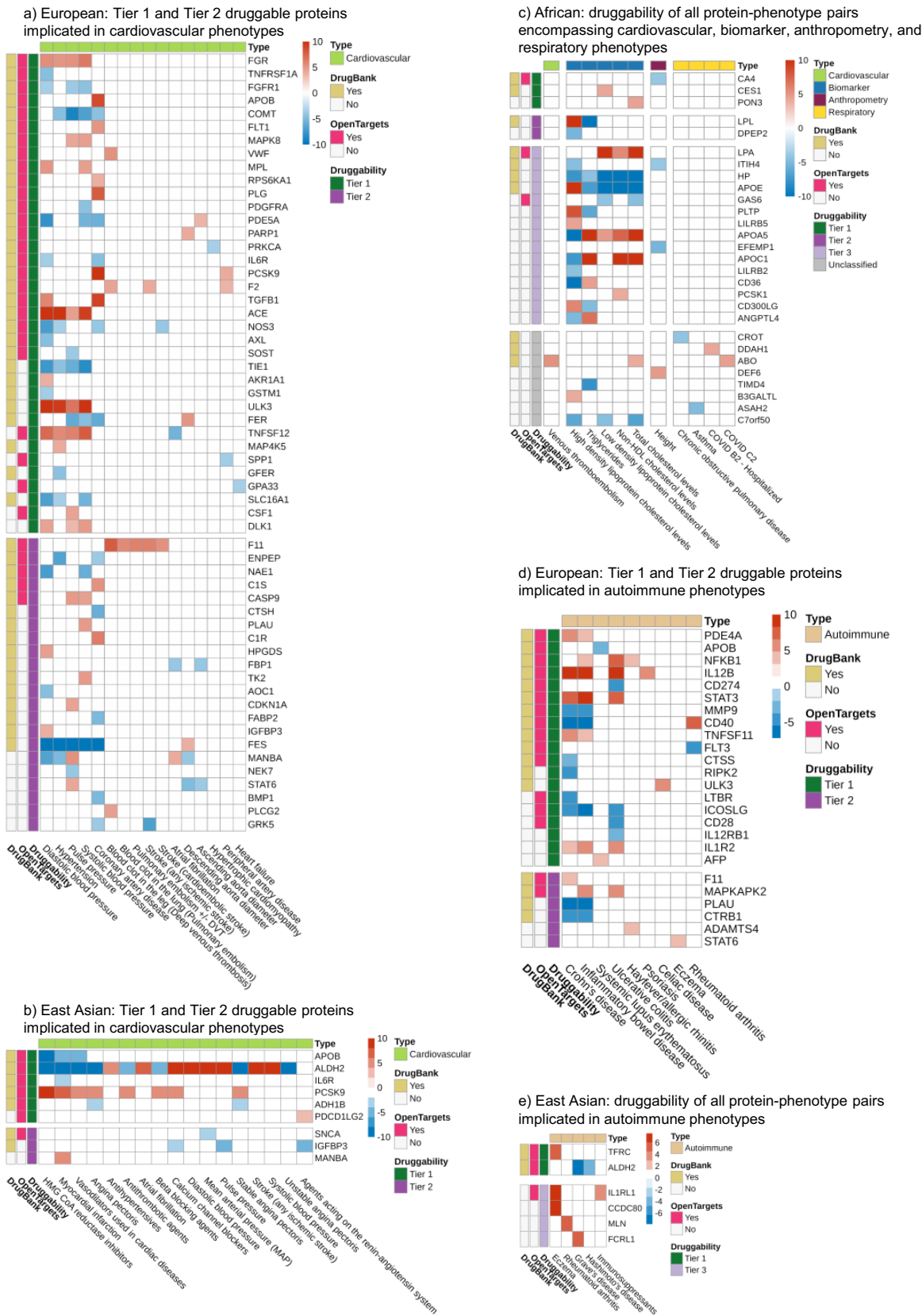
433 **7.2. Druggability of protein-phenotype pairs by integrating the druggable genome, DrugBank,**
434 **and Open Targets Platform**

435 Across three ancestries, among the 1,037 number of protein-coding genes that have at least one
436 protein-phenotype association, 669 (64.5%) were present in at least one database. Specifically,
437 579 (55.8%) protein-coding genes overlapped with the druggable genome³⁸, 350 (33.8%) had
438 approved or investigational drugs in DrugBank³⁹, and 191 (18.4%) overlapped with the Open
439 Targets Platform⁴⁰ (**Supplementary Table 26**). Notably 32 (3.1%) were unique to non-Europeans.
440 This highlights that multi-ancestry inclusion can expand the list of actionable druggable
441 associations. Overlap with the druggable genome and DrugBank stratified by ancestry is
442 presented in **Supplementary Note 7B**.

443

444 We found that higher levels of MANBA, a Tier 2 target, increased risk of atrial fibrillation in
445 European ancestry (**Fig. 6a**) and myocardial infarction in East Asian ancestry (**Fig. 6b**). Currently
446 no drugs exist for MANBA but its role in lysosomal metabolism suggests that modulating its activity
447 could have therapeutic potential. Further, increased ANGPTL4, a Tier 3 target, leads to increased
448 triglycerides and decreased HDL cholesterol levels in African ancestries (**Fig. 6c**) and increased
449 CAD risk in European ancestries (**Supplementary Note 7C**) supporting its potential as a
450 therapeutic target. Notably, we found that increased STAT3, a Tier 1 target, increased risk of IBD

451 and its subtypes, CD, and UC (**Fig. 6d**). Danvatirsen, a STAT-3 mRNA 3'UTR antisense inhibitor
452 has been undergoing phase 1 and 2 clinical trials for multiple cancer types and may be potentially
453 repurposed for IBD. Currently, astegolimab, an inhibitor of IL1RL1, a Tier 3 target, has completed
454 phase 2 trials for eczema and asthma; in our study, we find genetic support where increased
455 IL1RL1 increased risk of eczema in East Asian ancestry (**Fig. 6e**) and was concordant in
456 European ancestries (**Supplementary Table 17**). Druggability visualization for remaining
457 diseases for European and East Asian ancestries are provided in **Supplementary Note 7C**.
458



459
460
461

Figure 6. Integrating proteome-phenome-wide MR with the druggable genome, DrugBank, and clinical trial information from the Open Targets Platform.

462 Legend information:

463 Each cell displays a putatively causal protein-phenotype association. Cell color displays the MR
464 effect estimate based on Z score averaged across cohorts capped at -10 to +10 with red showing
465 a positive Z score indicating a positive MR effect of the protein (displayed as the gene name) on
466 the phenotype and blue showing a negative Z score indicating a negative MR effect of the protein
467 on the phenotype. For simplicity, in European ancestries, we only display protein-phenotypes with
468 consistent effect across European cohorts.

469 The y-axis shows the three drug databases. DrugBank (yellow square): DrugBank³⁹ shows
470 whether the protein has an available drug in the database. Open Targets (pink square): Open
471 Targets Platform⁴⁰ shows whether the protein has available clinical trial information. Druggability:
472 The druggable genome from Finan et al.³⁸ is shown for Tiers 1 (dark green, representing direct
473 targets of approved small molecules and biotherapeutic drugs), Tier 2 (dark purple, representing
474 proteins closely related to approved drug targets or which have associated drug-like compounds),
475 Tier 3 (light purple, representing secreted or extracellular proteins, those distantly related to
476 approved drug targets, and members of important druggable gene families not covered in Tier 1
477 or Tier 2), and Unclassified (gray, all other proteins not in Tiers 1 to 3). Proteins on the y-axis
478 within each tier are sorted based on the number of supported databases.

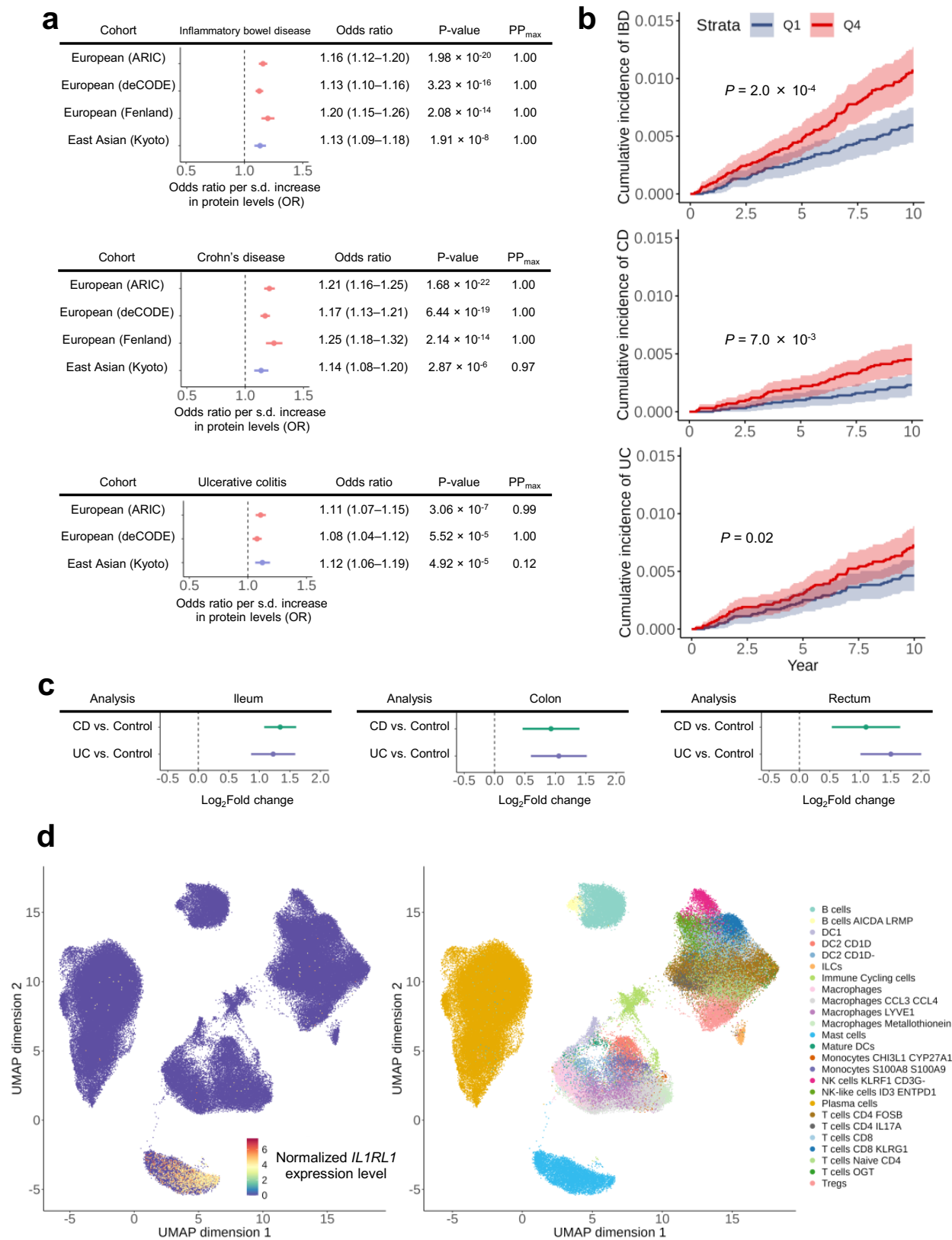
- 479 (a) European ancestry putatively causal protein-phenotype pairs stratified to Tier 1 and Tier
480 2 druggable proteins and cardiovascular phenotypes.
481 (b) East Asian ancestry putatively causal protein-phenotype pairs stratified to Tier 1 and Tier
482 2 druggable proteins and cardiovascular phenotypes.
483 (c) All of the African ancestry putatively causal protein-phenotype pairs with no druggability
484 stratification.
485 (d) European ancestry putatively causal protein-phenotype pairs stratified to Tier 1 and Tier
486 2 druggable proteins and autoimmune phenotypes.
487 (e) East Asian ancestry putatively causal protein-phenotype pairs stratified to autoimmune
488 phenotypes.

489
490
491

492 8. Converging evidence of the causal effect of IL1RL1 on IBD

493 Upon further stringent filtering to find protein-phenotype pairs with strong evidence (see **Methods**),
494 we found that increased IL1RL1 levels was associated with increased risk of IBD, CD, and UC in
495 European ancestries across three cohorts for IBD and CD and two cohorts for UC with consistent
496 directions, which we validated using the largest available East Asian ancestry GWAS for IBD
497 (14,393 cases and 15,456 controls), CD (7,372 cases and 15,456 controls), and UC (6,862 cases
498 and 15,456 controls) from Liu et al.⁶⁵ (**Fig. 7a**). In African ancestries, the number of cases were
499 limited in the largest publicly available GWAS for IBD (1,285 cases and 119,314 controls)⁵⁷ and
500 UC (857 cases and 119,909 controls)⁵⁷, thus estimates were not significant likely due to
501 insufficient power.

502



503
504

Figure 7. *IL1RL1* in IBD, CD, and UC.

- 505 a) MR for the effect of *IL1RL1* and IBD (top), CD (middle), and UC (bottom) in European and
506 East Asian ancestries. PP_{\max} is the maximum colocalization posterior probability between
507 PWCoCo and SharePro.
508 b) Kaplan Meier estimates for cumulative incident of IBD (top), CD (middle), and UC (bottom)
509 by baseline *IL1RL1* level quantiles in the UK Biobank. P values were computed using the
510 log-rank test.
511 c) Bulk RNA sequencing of *IL1RL1* in the ileum (left), colon (middle), and rectum (right).
512 d) Single-cell RNA sequencing analyses of *IL1RL1*. *IL1RL1* expression patterns showing
513 720,633 cells collected from the terminal ileum and colon of 71 donors with different levels
514 of inflammation. Single-cell transcriptomic data was obtained from Kong et al.⁶⁶ (SCP1884
515 <https://singlecell.broadinstitute.org/>).
516
517
518

519 To triangulate the evidence, we performed supplementary observational analysis in the UK
520 Biobank using Cox proportional hazards models. We adjusted for age, sex, recruitment center,
521 Olink measurement batch, Olink processing time, and the first 10 genetic principal components.
522 Notably, none of the genetic principal components were significant, suggesting that the
523 association is not specific to ancestry. Over 10 years of follow-up, a one s.d. increase in *IL1RL1*
524 was associated with elevated risk of IBD (hazard ratio, HR = 1.18; 95% CI: 1.05–1.33; $P = 6.6 \times$
525 10^{-3}), CD (HR = 1.21; 95% CI: 1.00–1.47; $P = 0.047$), and UC (HR = 1.15; 95% CI: 1.00–1.33; P
526 = 0.048), consistent with MR findings (**Supplementary Table 27**). Kaplan-Meier estimates for
527 cumulative incidence of disease stratified by baseline *IL1RL1* level (lowest 25% versus highest
528 25% in the UK Biobank population) also showed differences in IBD (log-rank test $P = 2.0 \times 10^{-4}$),
529 CD (log-rank test $P = 7.0 \times 10^{-3}$), and UC (log-rank test $P = 0.02$) (**Fig. 7b**). We also performed
530 an alternative, less stringent filter and prioritized proteins involved in CAD and type 2 diabetes
531 which we provide in **Supplementary Note 8**.
532

533 8.1. *IL1RL1* expression analyses

534 To further assess the role of *IL1RL1* in IBD, we used the IBD Transcriptome and
535 Metatranscriptome Meta-Analysis (IBD TaMMA) platform⁶⁷ to compare expression of *IL1RL1*
536 transcripts between IBD patients and healthy controls in the ileum, colon, and rectum. We found
537 significantly higher *IL1RL1* gene expression in all three tissues (**Fig. 7c**), suggesting that
538 increased *IL1RL1* expression is a consistent feature of IBD regardless of the specific location
539 within the gastrointestinal tract.
540

541 To gain further insights into the role of *IL1RL1* in IBD, CD and UC, we analyzed single-cell *IL1RL1*
542 expression in 720,633 cells from the terminal ileum and colon of 71 participants with different
543 levels of inflammation status from Kong et al.⁶⁶ (SCP1884 <https://singlecell.broadinstitute.org/>). In
544 single-cell RNA sequencing, *IL1RL1* showed significant enrichment in mast cells compared to 24
545 other cell types (permutation $P < 2.0 \times 10^{-4}$) (**Fig. 7d**). Mast cells, key players in allergic reactions
546 and inflammation and a key cell type involved in the pathogenesis of IBD^{68,69}, may contribute to
547 chronic IBD by releasing inflammatory mediators like histamine and cytokines when activated by
548 *IL1RL1* in inflamed tissues. These findings align with our MR analyses showing that increased
549 *IL1RL1* leads to increased risk of IBD, CD and UC.
550

551 Discussion

552 In this study, we conducted comprehensive multi-ancestry proteome-phenome analyses across
553 three ancestries. Using seven large proteomics cohorts including European, African, and East
554 Asian ancestries, we analyzed 355 complex traits or diseases, and identified 3,949, 56, and 325
555 putative causal effects of protein abundance on diseases and traits, respectively. By integrating
556 data from druggable genomes and drug databases, we prioritized potential protein targets for
557 drug development. Our findings offer a comprehensive atlas of protein-phenotype associations
558 and an evidence-based resource to support drug discovery and development, expand insight into
559 disease, and highlight potential targets for therapeutic intervention.

560
561 Our study provides an updated map to earlier phenome-wide MR studies of the human plasma
562 proteome on complex diseases which were either limited to European ancestries¹⁸ or considered
563 only a few diseases in European and African ancestries¹⁹. The significance of incorporating
564 multiple ancestries is underscored by our identification of several proteins that are uniquely
565 instrumentable by each ancestry due to allele frequency differences. Specifically, we
566 instrumented an additional 119 proteins exclusively in African ancestry, 17 in East Asian ancestry,
567 and 19 shared between African and East Asian ancestries. Moreover, the finding that a significant
568 proportion of population-specific genetic variants—68.5% in African and 72.2% in East Asian—
569 have a MAF below 0.01 in European ancestries highlights the potential for missed genetic
570 discoveries when studies focus on a single ancestry. Research on the genetic architecture across
571 different ancestries reveals both commonalities and differences, influenced by evolutionary
572 history, genetic diversity, and population-specific factors^{26,30}. Thus, by including diverse African
573 and East Asian ancestries in proteomic analyses, we were able to instrument more proteins by
574 leveraging common genetic variants in these underrepresented populations which were mostly
575 rare in European ancestries, enhancing the potential for novel discoveries and exemplifying the
576 value of including non-European individuals for comprehensive and inclusive proteomic and
577 genetic analyses.

578
579 The inclusion of African and East Asian ancestries allowed discovery of protein-phenotype
580 associations. We emphasize that these findings were attributable to uniquely instrumentable
581 proteins in each ancestry and do not necessarily indicate ancestry-specific biological mechanisms.

582
583 As an illustrative example of the value of our atlas, we found that increased circulating
584 abundances of IL1RL1 was causal for IBD, CD, and UC in European and East Asian ancestries,
585 which was supported by observational analyses and gene expression analyses. IL1RL1 could be
586 a promising therapeutic approach for IBD, potentially reducing mast cell-driven inflammation.
587 Potential drugs targeting IL1RL1 include astegolimab which has completed phase 2 trials for
588 eczema and asthma, tozorakimab which neutralizes IL-33, the interleukin that binds IL1RL1 (ST2),
589 and anakinra which targets IL-1, a closely related interleukin. However, further research is
590 required to assess the safety and efficacy of potential IL1RL1 inhibition. Many such findings may
591 exist, and this atlas may be used as a tool to facilitate the selection of targets during primary or
592 pre-clinical drug development, exploring drug repurposing opportunities, and improve
593 understanding of proteins implicated in complex traits and diseases.

594
595 Our study has several key strengths. We curated GWAS for a wide range of complex traits and
596 diseases, decreasing the overlap and redundancy and increasing the power for discovery in MR.
597 Second, our study included diverse proteomics cohorts, including African ancestry and a new
598 East Asian ancestry cohort, the Kyoto University Nagahama cohort. This latter cohort, which used
599 the SomaScan v4 platform, is the largest to date aside from the China Kadoorie Biobank^{25,26}.
600 Further, our study combined both ARIC SomaScan and UKB-PPP Olink African proteomics
601 cohorts at a phenome-wide scale and to include three ancestries. Notably, we identified uniquely

602 instrumentable proteins in African and East Asian ancestries that were not found in European
603 ancestries, highlighting the value of including cohorts from diverse ancestral backgrounds. Third,
604 we performed extensive stringent filtering on genetic instruments with strict V2G criteria. Fourth,
605 we harnessed novel state-of-the-art colocalization methods to reduce the risk of confounding from
606 LD while increasing the statistical power to support more protein-phenotype associations with
607 colocalization evidence³⁷.

608
609 This study has several limitations. First, while two separate proteomics cohorts were included for
610 African ancestries, the number of outcomes considered was still limited. Additionally, African
611 ancestry phenotypes were curated from various cohorts with potentially finer genetic architecture
612 differences than controlled for by using continental ancestries, thereby potentially biasing our
613 analyses. Second, differences in measurement units make direct comparison of MR effect size
614 estimates for continuous outcomes difficult. Nevertheless, direction of effect should be robust to
615 this limitation. Third, while we reduced the risk of horizontal pleiotropy by using strict V2G *cis*-
616 pQTLs, this resulted in many proteins being instrumented by a single *cis*-pQTL, limiting the ability
617 to perform MR sensitivity analyses. Nonetheless, we used robust colocalization methods to
618 mitigate risk of reporting false positives. Fourth, although we used assays measuring nearly 5,000
619 proteins from SomaScan and 3,000 proteins from Olink, coverage is still limited with regard to the
620 entire proteome. Lastly, while we analyzed three diverse ancestries, the sample sizes for both
621 proteomic GWAS and outcome GWAS were much larger for European ancestry, leading to
622 differences in the number of associations. Greater coverage of proteins and larger sample sizes
623 in non-European ancestries are needed.

624
625 In conclusion, through integrative multi-ancestry plasma proteome-phenome MR and extensive
626 sensitivity analyses, we provided a comprehensive atlas of protein-phenotype associations
627 across three ancestries and highlighted the value of multi-ancestry inclusion, as illustrated by
628 uniquely instrumentable proteins in non-European ancestries. This study serves as a valuable
629 resource for understanding disease mechanisms and prioritizing potential new targets.
630

631 **Methods**

632 **1. Proteomics cohorts**

633 We analyzed proteomics cohorts from three ancestries consisting of European (four cohorts:
634 ARIC, deCODE, Fenland, and UKB-PPP), African (two cohorts: ARIC and UKB-PPP), and East
635 Asian (one cohort: Kyoto University Nagahama East Asian cohort). All cohorts had proteomics
636 measured on the aptamer-based SomaScan assay v4 except for the UKB-PPP study, which used
637 the antibody-based Olink Explore 3072 platform.

638 **1.1. European ancestry cohorts**

640 We analyzed the GWAS of protein levels in individuals of European ancestry using four different
641 studies. Three of these four studies (ARIC, deCODE, and Fenland described below) had
642 proteomics measurements from the aptamer-based SomaScan assay v4 from SomaLogic
643 (Boulder, Colorado, USA). In brief, SomaScan assay v4 uses aptamers, which are single-
644 stranded oligonucleotides that have specific binding affinities to protein targets and can measure
645 up to 5,000 unique proteins. The UKB-PPP study had proteomics measurements from the
646 antibody-based Olink Explore 3072 platform, which measures up to 3,000 proteins. Briefly, Olink
647 (Uppsala, Sweden) uses proximity extension assay (PEA) technology, which detects proteins
648 through the binding of two separate antibodies carrying complementary oligonucleotide tags,
649 which hybridize to the protein target. We restricted our analyses to proteins encoded by autosomal
650 genes and analyzed a list of 4,687 proteins from SomaLogic and 2,823 proteins from Olink.

651 **1.1.1. ARIC**

653 The Atherosclerosis Risk in Communities (ARIC)¹³ measured protein levels from 9,084 American
654 participants of European and African ancestries using the SomaScan assay v4. Of these
655 participants, 4,657 plasma proteins were measured for 7,213 European American individuals.

656 **1.1.2. deCODE**

658 The deCODE study¹² provided 4,907 aptamers that measure 4,719 proteins in 35,559 Icelandic
659 individuals of European ancestry using SomaScan assay v4.

660 **1.1.3. Fenland**

662 The Fenland study¹¹ measured 4,775 proteins in 10,708 individuals of European ancestry using
663 SomaScan assay v4.

664 **1.1.4. UKB-PPP**

666 The UK Biobank Pharma Proteomics Project (UKB-PPP)¹⁴ conducted proteomic profiling on
667 54,219 individuals of multiple genetic ancestries in the UK Biobank using the Olink Explore 3072
668 platform. From this cohort, 34,557 European individuals, each with 2,923 unique proteins
669 measured, were utilized in the UKB-PPP as the discovery cohort, and we used this discovery
670 cohort in our study.

671 **1.2. African ancestry cohorts**

673 For African ancestries, we used GWAS of protein levels from two different studies (ARIC and
674 UKB-PPP) measured on SomaScan assay v4 and Olink Explore 3072, respectively.

675 **1.2.1. ARIC**

677 The ARIC cohort was previously described in the European cohort section and consists of 9,084
678 European American and African American individuals. Of these participants, 4,657 proteins from
679 the SomaScan v4 assay were measured for 1,871 African American individuals.

680 **1.2.2. UKB-PPP**

682 We used the African ancestry individuals from the UKB-PPP study, which consists of 931
683 individuals with 2,923 unique proteins measured with the Olink Explore 3072 platform.

684

685 [1.3. East Asian ancestry cohort](#)

686 *Kyoto University Nagahama cohort*

687 The Nagahama Primary Prevention Cohort Project (Kyoto-Nagahama cohort) is a joint project
688 between the Kyoto University Graduate School of Medicine and Nagahama City, Shiga Prefecture
689 that involved 10,000 residents of Nagahama ([https://w3.genome.med.kyoto-](https://w3.genome.med.kyoto-u.ac.jp/en/nagahama-project/)
690 [u.ac.jp/en/nagahama-project/](https://w3.genome.med.kyoto-u.ac.jp/en/nagahama-project/)). Data generation was performed at the Kyoto University Center for
691 Genome Medicine, where 1,823 Japanese individuals of East Asian ancestry were whole
692 genome-sequenced and had 4,196 proteins measured using the SomaScan assay v4. Further
693 details can be found in **Supplementary Note 9**.

694

695 [2. Identification of strict variant-to-gene \(v2g\) cis-protein quantitative trait loci \(pQTLs\)](#)

696 Each cohort had had different *cis*-pQTL definitions. For example, the ARIC¹³ study defined *cis*-
697 pQTL as those within ± 500 kb of the TSS of the protein-coding gene with $FDR < 5\%$. The
698 deCODE study¹² defined *cis*-pQTL as those within ± 1 Mb of the TSS of the protein-coding gene
699 with $P < 1.8 \times 10^{-9}$. The Fenland Study¹¹ defined *cis*-pQTL as those within ± 500 kb of the
700 protein-coding gene with $P < 1.004 \times 10^{-11}$. The UKB-PPP¹⁴ noted that of their identified pQTLs,
701 66.9% of proteins tested (1,954 of 2,922 proteins) had a *cis*-pQTL within ± 1 Mb of the protein-
702 coding gene with $P < 1.7 \times 10^{-11}$. Thus, we created a common *cis* definition as follows.

703

704 [2.1. Linkage disequilibrium \(LD\) clumping](#)

705 We performed LD clumping (clumping window of 1 Mb, significance level of 5×10^{-8} , and clumping
706 r^2 threshold of 0.001) on each proteomic GWAS in each ancestry cohort. For European proteomic
707 GWAS, we used a reference panel composed of 50,000 randomly sampled unrelated UK
708 Biobank⁷⁰ individuals of European ancestry (UKB 50k). For African proteomic GWAS, we curated
709 an LD reference panel from the Human Genome Diversity Project and 1000 Genomes Project
710 (HGDP + 1kGP) reference panel⁷¹ for 994 African ancestry individuals. In East Asian ancestries,
711 we used the 1000 Genomes East Asian (1kGP EAS) reference panel. We retained variants with
712 a MAF > 0.01 in all reference panels.

713

714 [2.2. Identification of cis- and trans-pQTLs](#)

715 To determine *cis*-pQTLs, we used the Ensembl BioMart⁷² package version Ensembl 105: Dec
716 2021 (Genome Reference Consortium Human Build 38, GRCh38.p13) to generate a human
717 protein-coding genes file. We considered relevant attributes such as the canonical TSS, gene
718 start, and gene end. Since Fenland proteomic GWAS were in GRCh37 coordinates, we
719 separately curated a protein-coding genes file using BioMart on the GRCh37 assembly in
720 Ensembl using version 110. For analysis, we excluded protein-coding genes located on sex
721 chromosomes and those located within the major histocompatibility complex (MHC) (GRCh38:
722 chr6 28,510,020– 33,480,577) due to the complex LD structure, high allelic diversity, and strong
723 pleiotropy in this region⁷³. We defined a *cis*-pQTL as a pQTL within ± 500 kb of the TSS of the
724 protein-coding gene. All other pQTLs were *trans*. We used *cis*-pQTLs since they are more likely
725 to directly impact the transcription and translation of the protein of interest.

726

727 [2.3. Strict variant-to-gene cis-pQTL definition](#)

728 To minimize potential horizontal pleiotropic effects, we defined a unique strict V2G definition for
729 *cis*-pQTLs whereby a *cis*-pQTL is a *strict V2G cis-pQTL* if it is a *cis*-pQTL for only one protein-
730 coding gene (*strict*) and it has the strongest link to the corresponding protein-coding gene based
731 on multiple sources of evidence and concomitantly has the highest Open Targets Genetics V2G
732 score (**Extended Data Fig. 1**). We outline these steps in the following two sections.

733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783

2.3.1. Strict *cis*-pQTL

For genome-wide significant independent *cis*-pQTLs in each cohort, we retained those associated with a single protein-coding gene. We defined this as a “strict” *cis*-pQTL definition. Here, we used *protein-coding genes* instead of *proteins (aptamers)* for the strict *cis* definition due to SomaScan assay v4 having multiple instances where two or more aptamers target a single protein, which would result in the unwarranted scenario where pQTLs that were *cis* for more than one aptamer of the same protein were removed.

2.3.2. Open Targets Genetics Variant-to-Gene (V2G)

We used the Open Targets Genetics³⁵ database (<https://genetics.opentargets.org/>) to determine whether each strict *cis*-pQTL held the highest V2G score for its corresponding associated protein-coding gene. This ensures that the variant is a suitable proxy for the plasma levels of the protein-coding gene. For instance, the variant may directly impact the protein-coding gene, potentially by altering its transcription, thereby influencing its plasma abundance. Briefly, Open Targets Genetics V2G scores are generated through a model trained on molecular QTLs (eQTL, sQTL, pQTL), chromatin interaction experiments such as promoter capture Hi-C (PCHi-C), *in silico* functional predictions such as Ensembl Variant Effect Predictor (VEP), and the distance between variants and genes' canonical transcription start sites. This composition of evidence enables accurate assignment of variants to genes. We note that the pQTL datasets utilized by Open Targets Genetics for model training are from earlier studies and do not overlap with the proteomics cohorts we analyzed in this study, mitigating the risk of overfitting.

Together, we combined the strict *cis* definition and V2G score from Open Targets Genetics to curate strict V2G *cis*-pQTLs, which decreases the chance of horizontal pleiotropy in MR.

3. Protein altering variant (PAV) and expression quantitative trait loci (eQTL)

Since pQTLs altering the binding epitope of a protein may reflect assay specificity instead of the true biological function of the protein^{1,64,74}, we annotated strict V2G *cis*-pQTLs as protein-altering variants (PAVs) of moderate or high impact if they or any variants in high LD ($r^2 > 0.8$) were identified as PAVs using VEP⁷⁵ (**Supplementary Tables 2–8**). VEP annotates the impact of variants into four categories: Modifier, Low, Moderate, or High (https://useast.ensembl.org/info/genome/variation/prediction/predicted_data.html). “Modifier” impact refers to variants in non-coding regions or affecting non-coding genes in which evidence of impact is hard to predict or limited. “Low” impact refers to variants that may not change protein behavior. “Moderate” impact variants are non-disruptive and may change protein effectiveness and include missense variants. “High” impact variants are disruptive and can cause truncation of proteins, loss of function, or trigger nonsense-mediated decay. If a strict V2G *cis*-pQTL and all of its LD proxies were labelled as Modifier or Low impact, we considered this strict V2G *cis*-pQTL to have no PAV. If a strict V2G *cis*-pQTL or any of its LD proxies were labeled Moderate or High, we considered this strict V2G *cis*-pQTL to be a PAV of -Moderate or -High impact, respectively.

We also conducted a *cis*-eQTL enrichment analysis using 49 tissues from GTEx v8 in European ancestries since if a *cis*-pQTL overlaps with the *cis*-eQTL of the same gene, it strengthens the evidence that the *cis*-pQTL acts directly on the gene products, reducing the risk of horizontal pleiotropy. To do so, we determined whether the strict V2G *cis*-pQTL was a *cis*-eQTL for the protein-coding gene of interest with $P < 1 \times 10^{-5}$ by querying 49 tissues in GTEx v8 European eQTLs. Querying was performed based on CHR:POS:EA:NEA and CHR:POS:NEA:EA (CHR: chromosome; POS: position; EA: effect allele; NEA: non-effect or other allele). In order to compare effect sizes across different cohorts, we aligned the strict V2G *cis*-pQTL effect allele in each cohort to the corresponding ancestry-specific reference panel's alternative allele (UKB 50k

784 for European, HGDP + 1kGP for African, and 1kGP EAS for East Asian ancestries). We note that
785 pQTLs in the Fenland cohort were based on GRCh37 coordinates and were lifted to GRCh38
786 prior to querying for *cis*-eQTLs across GTEx v8 tissues. After synchronizing the genome assembly,
787 7 out of 3,045 strict V2G *cis*-pQTLs in Fenland were automatically labeled as having no eQTL
788 since no corresponding GRCh38 coordinate existed for these variants, and 3,038 *cis*-pQTLs were
789 analyzed.

790 791 [4. GWAS outcome curation and selection](#)

792 We manually curated the latest and largest GWAS (as of February 2024) for European and African
793 ancestry outcomes. East Asian ancestry outcomes were selected from BioBank Japan¹⁵. All
794 GWAS in GRCh37 were lifted to GRCh38 with the liftOver tool. We outline the curation steps for
795 each ancestry's GWAS outcomes in detail:

796 797 [4.1. European ancestry outcomes](#)

798 During the curation, we considered 510 outcomes downloaded from 50 studies and one database.
799 We removed outcomes that were duplicated and had a larger GWAS available, had ambiguous
800 or broad definitions, were likely heterogeneous, were sex-specific, had missing relevant columns,
801 had no download link available, and were not relevant to our outcomes of interest. We retained
802 179 outcomes for analysis (**Supplementary Table 10**). We labeled rsids, chromosome, and
803 position if they were missing. Cases and sample sizes for each outcome were manually extracted
804 from the original manuscript or the supplementary tables of each corresponding study, and we
805 further categorized each outcome into one of 23 "Type" categories pertaining to the human
806 system the outcome was based on or most likely to fall under.

807 808 [4.2. African ancestry outcomes](#)

809 We manually curated 26 of the most up-to-date and publicly available African ancestry GWAS
810 summary statistics (**Supplementary Table 11**). Restricted access GWAS from dbGap were not
811 considered due to data access difficulties. Missing sample sizes were manually annotated with
812 the sample size by inspecting the original manuscript and Supplementary Tables, while missing
813 rsids were labeled using the HGDP + 1kGP reference panel. We annotated missing chromosomes
814 and positions and used VEP to annotate variants missing effect allele frequency with
815 gnomADg_AFR_AF (gnomAD genomes for African/American populations). We categorized each
816 outcome into one of 8 "Type" categories, including Respiratory, Musculoskeletal, Cardiovascular,
817 Eye, Anthropometry, Biomarker, Psychiatric, and Metabolic/endocrine.

818
819 As exploratory analyses, we included 114 binary cardiovascular and 9 binary autoimmune-related
820 outcomes from the Million Veteran Program ($n = 635,969$) for African individuals.

821 822 [4.3. East Asian ancestry outcomes](#)

823 We curated 220 outcomes from Biobank Japan (<https://pheweb.jp/>)²⁷. We excluded 14 sex-
824 specific outcomes, including Abortion, Breast cancer, Cervical cancer, Cesarean section, Ectopic
825 pregnancy, Endometriosis, Endometrial cancer, Mastopathy, Ovarian cancer, Ovarian cyst, Pre-
826 eclampsia, Prostate cancer, Uterine fibroid, and Uterine prolapse. We analyzed 206 outcomes
827 (**Supplementary Table 12**). We also provide "Type" labels for each outcome, denoting the human
828 system the outcome was based on or was most likely related to.

829 830 [5. Two-sample Mendelian randomization](#)

831 To assess the putative causal effect of protein abundance on outcomes in European, African, and
832 East Asian ancestries, we performed two-sample MR using TwoSampleMR v.0.5.7⁷⁶. To mitigate
833 horizontal pleiotropy, we used strict V2G *cis*-pQTLs as instrumental variables to proxy protein-
834 level exposures, as defined in the earlier sections of the methods. We harmonized exposure and

835 outcome GWAS using the `harmonise_data()` function and performed a proxy search if an
836 instrument was absent in the outcome GWAS. For European, African, and East Asian ancestries,
837 we used the UKB 50k, HGDP + 1kGP, and 1kGP East Asian reference panels that were
838 previously used for LD clumping for the proxy search, respectively. We searched for proxies using
839 PLINK v.1.9⁷⁷ parameters `--ld-window=5000`, `--ld-window-kb=5000`, `--ld-window-r2=0.8` and
840 retained proxies with minor allele frequencies ≤ 0.42 .

841
842 MR analyses were performed using the `mr()` function. For proteins with a single genetic instrument,
843 the association between the protein and outcome was evaluated using a Wald ratio estimate. For
844 proteins with ≥ 2 genetic instruments, we used an inverse variance weighted random effects
845 estimate. We determined whether genetic instrumental variables had F-statistics $> 10^{43,78}$,
846 indicating strong associations with the exposure and thus less chance of weak instrument bias,
847 which may bias the causal effect estimates towards the null in two-sample MR. F-statistics are
848 shown in **Supplementary Table 9**. We corrected for multiple testing per cohort in each ancestry
849 by applying a Benjamini-Hochberg-corrected P threshold (FDR)⁴⁵ of 0.05 (5%) as done
850 previously¹⁹.

851
852 We note that beta estimates from MR for continuous traits are not directly comparable across
853 different outcomes because the units used in GWAS vary. For instance, some GWAS use clinical
854 units, whereas others use standardized and/or residualized values.

855 856 **6. MR sensitivity analyses**

857 To increase the robustness of MR findings, we further filtered MR results based on multiple
858 sensitivity analyses, including heterogeneity tests, MR using alternative approaches (including
859 weighted median, weighted mode, MR-Egger), and Steiger directionality test⁷⁹ to assess reverse
860 causality. Following these filtering steps, we refer to retained protein-phenotype associations as
861 “MR-passing”. The sensitivity analyses are described for proteins with ≥ 2 instruments and
862 proteins with ≥ 3 instruments below:

863 864 **6.1. Sensitivity analyses for proteins with two or more instruments**

865 The heterogeneity test was performed for proteins with ≥ 2 instruments and describes whether
866 strict V2G *cis*-pQTLs of the same protein are likely to show comparable effects on the tested
867 outcomes. For heterogeneity testing, we used the `mr_heterogeneity()` function to compute a
868 heterogeneity P value (`Q_pval`), and we calculated I^2 statistics using the `lsq()` function. If an
869 association had an I^2 threshold ≥ 0.5 and a heterogeneity P value (`Q_pval`) < 0.05 , this indicated
870 considerable heterogeneity.

871 872 **6.2. Sensitivity analyses for proteins with three or more instruments**

873 For proteins with ≥ 3 genetic instruments, we performed additional sensitivity analyses with
874 alternative MR methods such as MR weighted median, MR weighted mode, and MR-Egger
875 methods, as well as Steiger directionality testing. To check the consistency of MR estimates, we
876 required the MR estimate, as well as the sensitivity analyses estimates from the MR weighted
877 median, MR weighted mode, and MR-Egger approaches, to all have the same sign. For
878 directional pleiotropy, we used the `mr_pleiotropy_test()` function to perform the MR-Egger
879 intercept test and considered a P value < 0.05 as a statistically significant deviation from the null
880 and an indication of directional pleiotropy. We also performed Steiger filtering on all proteins
881 (`directionality_test()` function). Any pQTLs that explain more variance in the outcome than in the
882 exposure potentially indicate reverse causation and were removed from further analysis.

883 884 **7. Colocalization of proteomic GWAS with outcome GWAS**

885 To assess whether plasma protein levels share the same causal variant with GWAS outcomes,
886 we employed two colocalization methods to ensure the robustness of our findings. We performed
887 PWCoCo^{18,36} and SharePro³⁷ for MR associations that passed all MR sensitivity analyses
888 described in the previous section. PWCoCo and SharePro are recent methods allowing multiple
889 independent associations to be assessed. Both improve on the original coloc method⁸⁰, which
890 was limited by the assumption that a single variant exists per GWAS, wherein the method only
891 considers the strongest of these distinct association signals when multiple independent
892 associations exist. A detailed description of both methods is provided in **Supplementary Note**
893 **10**. Colocalization analyses were performed around a 1-Mb region centered on the lead (lowest
894 *P* value) *cis*-pQTL. We set a colocalization posterior probability (PP) of a shared causal variant \geq
895 0.8 in any of PWCoCo or SharePro as evidence of colocalization. For simplicity, we report the
896 maximum PP between PWCoCo and SharePro (PP_{max}) in the main text. We reported putatively
897 causal associations (**Supplementary Tables 14–16**) as associations which pass all MR
898 sensitivity analyses, Steiger filtering, and also colocalized with a PP \geq 0.8 in any one of PWCoCo
899 or SharePro. Due to the difficulty in verifying the corresponding Olink assay target for each
900 SomaScan aptamer, we counted unique protein-phenotype associations using protein-coding
901 genes to harmonize between proteomics platforms. **Supplementary Tables 14–16** contain MR
902 and colocalization summaries, summaries of effect direction consistency across cohorts, flags of
903 proteins instrumented by PAVs of high impact, and whether the protein-phenotype association
904 came from a protein uniquely instrumentable in that ancestry. We also annotate whether protein-
905 phenotype associations passed the most stringent Bonferroni correction for the total number of
906 MR tests across three ancestries ($P < 0.05 / 874,465$).

907 8. Distinguishing between previously reported and unreported protein-phenotype pairs

908 8.1. Comparing against earlier studies identifying putatively causal protein-phenotype pairs with 909 MR and colocalization evidence

910 To identify the status of protein-phenotype associations as reported or unreported (not found from
911 pre-existing proteome-phenome-wide MR studies) in European ancestries, we overlapped our
912 associations with recent proteome-phenome-wide MR analyses from Zheng et al. 2020¹⁸ and
913 multi-ancestry proteome-wide MR analyses from Zhao et al. 2022¹⁹. We used the 111 identified
914 putatively causal associations (65 proteins on 52 phenotypes) from “Table S7” of Zheng et al.¹⁸
915 and the 45 associations from “ST7A” of Zhao et al.¹⁹.

916 We note that Zhao et al. used three colocalization methods and a more relaxed threshold of PP
917 > 0.7 as evidence of colocalization. To do so, we harmonized the outcomes from Zheng and Zhao
918 to our outcomes and matched protein-phenotype pairs using ensembl ID and outcome name for
919 the Zheng study and UniProt ID and outcome name for the Zhao study. Any of our identified
920 putatively causal European ancestry association pairs not from these two studies were identified
921 as unreported.

922 For African ancestries, we overlapped our protein-phenotype pairs with the single protein-
923 phenotype pair passing FDR correction identified in “ST8A” of Zhao et al.¹⁹. Any pairs that did not
924 overlap were considered unreported.

925 9. Protein-phenotype network plots

926 We used Python igraph (v.0.10.8) to generate networks. Placement of nodes was generated
927 in Cytoscape (v.3.10.2) and manual editing was performed in Adobe Illustrator (v.28.2).

928 10. Effect concordance within-ancestry

929 To determine the concordance of MR effect estimates within European and African ancestries,
930 which both included more than one proteomics cohort, we compared protein-coding genes to

936 harmonize assay names across SomaScan assay v4 and Olink 3072 Explore platforms. However,
937 we could not verify whether Olink assays for a particular protein targeted the same domain as its
938 corresponding SomaScan assay, which may lead to differences in MR estimates. Moreover, since
939 some proteins were instrumented by *cis*-pQTLs that may be PAVs of high impact, which could
940 also lead to discordance in effects, we annotated these associations with a flag and advise caution
941 in the interpretation of these flagged results (**Supplementary Tables 14–16**).

942 943 **11. Druggability assessment**

944 We performed a druggability assessment on instrumentable protein-coding genes and protein-
945 phenotype associations that showed putatively causal relationships.

946 947 **11.1. Druggability assessment of instrumentable proteins**

948 For instrumentable protein-coding genes, we determined druggability based on Finan et al.³⁸, as
949 described below.

950 951 **11.1.1. Finan et al.**

952 Finan et al.³⁸ considered 20,300 protein-coding genes annotated using Ensembl version 73 and
953 classified 4,479 (22%) into three tiers (Tier 1, 2, and 3) as drugged or druggable. Tier 1 (1,427
954 genes) encompasses the primary targets of approved small molecules and biotherapeutic drugs,
955 along with those influenced by clinical-phase drug candidates. Tier 2 (682 gene) involves proteins
956 closely associated with drug targets or linked to drug-like compounds. Meanwhile, Tier 3 (2,370
957 genes) comprises secreted or extracellular proteins that are distantly related to approved drug
958 targets, and those in important druggable gene families not covered in Tiers 1 or 2. We denoted
959 all other protein-coding genes that did not fall in Tier 1, 2, or 3 categories as “Unclassified”. To
960 check the overlap of protein-coding genes across all cohorts when stratifying into Finan et al. tiers
961 (**Supplementary Figure 5 and 6**), we used the UpSetR⁸¹ package v.1.4.0
962 (<https://github.com/hms-dbmi/UpSetR>).

963 964 **11.2. Druggability assessment and enrichment of protein-phenotype pairs**

965 We first assessed how many proteins from the identified pairs of putatively causal protein-
966 phenotype associations had existing drugs by querying DrugBank³⁹. We used Ensembl ID to
967 match protein targets in DrugBank. Next, we created heatmaps using the pheatmap package
968 v.1.0.12 in R incorporating the druggable genome, Drugbank, and Open Targets Platform
969 (described below). We overlapped protein-phenotype pairs with the DrugBank database to
970 determine whether any drugs existed for these disease-implicated proteins while Open Targets
971 was used to determine whether any clinical trial information existed for these proteins.

972 973 **11.2.1. DrugBank**

974 We used DrugBank database v.5.1.12 (<https://go.drugbank.com/releases/latest>) and R package
975 dbparser v.2.0.2 to parse the DrugBank database xml file. We aggregated drugs and target
976 information and overlapped this with putatively causal protein-phenotype associations to
977 determine which proteins had an available drug.

978 979 **11.2.2. Open Targets**

980 We used Open Targets v.24.03 (<https://platform.opentargets.org/downloads>). We used the
981 knownDrugsAggregated dataset, which provides information on known drugs for a given disease
982 and contains protein target information. Open Targets also includes information on clinical trials
983 and its phases, which we used to determine the status of a protein target and its corresponding
984 drug. We overlapped this dataset by matching Ensembl ID with our identified protein-phenotype
985 associations.

986

987 12. Follow-up analyses showing evidence for IL1RL1

988 We filtered protein-phenotype associations in European ancestries by ensuring that the protein
989 was instrumented in two or more ancestries, was targeted by both SomaScan and Olink assays,
990 had an MR effect that was concordant across all cohorts, had MR and colocalization evidence in
991 three or more cohorts for the protein-phenotype pair, and was implicated in at least one binary
992 phenotype. Following these filtering steps, we identified 53 candidate protein-phenotype pairs and
993 highlighted IL1RL1 in IBD, CD, and UC as an illustrative example.

994 12.1. MR and colocalization of IL1RL1 in East Asian ancestry with IBD, CD, and UC

995 To validate that IL1RL1 was also putatively causal for IBD, CD, and UC in non-European
996 ancestries, we performed two-sample MR and colocalization with PWCoCo and SharePro using
997 the IL1RL1 strict V2G *cis*-pQTL (rs12712135) identified in the Kyoto University Nagahama East
998 Asian ancestry proteomics cohort. We used the largest East Asian ancestry GWAS for IBD
999 (14,393 cases and 15,456 controls), CD (7,372 cases and 15,456 controls), and UC (6,862 cases
1000 and 15,456 controls) from Liu et al.⁶⁵. Harmonization was performed similarly to the primary
1001 analyses, and we used the Wald ratio to obtain MR effect estimates. Colocalization was
1002 performed as described in earlier sections of the methods, and we used $PP \geq 0.8$ as the threshold
1003 for evidence of colocalization in any of PWCoCo or SharePro.

1004 12.2. MR of IL1RL1 in African ancestry with IBD and UC

1005 We estimated the causal effect of IL1RL1 on IBD (1,285 cases and 119,314 controls) and UC
1006 (857 cases and 119,909 controls) in African ancestry using outcome GWAS from the Million
1007 Veteran Program⁵⁷. We performed MR using the IL1RL1 strict V2G *cis*-pQTL (rs1420101) in the
1008 ARIC and UKB-PPP African ancestry cohorts. The Wald ratio was used to obtain MR effect
1009 estimates.

1010 12.3. Cox regression analysis for 10-year cumulative events of IBD, CD, and UC in the UK 1011 Biobank

1012 We used multivariable Cox proportional hazards regression to determine whether baseline
1013 plasma IL1RL1 protein level was associated with cumulative events of IBD, CD, or UC. We
1014 adjusted for age, sex, recruitment center, Olink measurement batch, Olink processing time, and
1015 the first 10 genetic principal components (UKB field: 22009) to adjust for genetic ancestry while
1016 protein levels were rank-based inverse normal transformed. We used the `coxph()` function from
1017 the survival R package v.3.2.13 and considered $P < 0.05$ as nominal significance of association.
1018 None of the genetic principal components were significantly associated with the IBD, CD, and UC
1019 outcome in each analysis.

1020 We checked the proportional hazards assumption using the `cox.zph()` function for the IL1RL1
1021 association analysis for IBD, CD, and UC for the rank-based inverse normal transformed IL1RL1
1022 covariate. The proportional hazards assumption tests the null hypothesis that each covariate's
1023 effect estimate does not vary with time. We used the "GLOBAL" variable from `cox.zph()` which
1024 tests the null hypothesis of whether all inputted covariates meet the proportional hazards
1025 assumption. We considered $P < 0.05$ as evidence that the proportional hazards assumption was
1026 not fulfilled.

1027 We defined IBD using ICD10 codes K50-K51, CD using K50, and UC using K51. We calculated
1028 the time to event by subtracting the date of event registration from the date of enrollment (data
1029 field: 53), focusing on events occurring within 10 years of enrollment. We excluded cases of
1030 prevalent IBD that met these criteria before enrollment, and those without a recorded event date
1031 for IBD, and performed the same steps for CD and UC. Controls for the IBD analysis were defined
1032 as individuals without an IBD, UC, or CD record based on self-reported medical history. Controls

1038 for the CD analysis were defined as individuals without a CD record, and controls for the UC
1039 analysis were defined as individuals without a UC record, based on self-reported medical history.
1040 We analyzed 333 cases and 40,001 controls for IBD, 130 cases and 40,388 controls for CD, and
1041 240 cases and 40,178 controls for UC.

1042
1043 We plotted Kaplan-Meier curves by stratifying individuals into the bottom 25% and the top 25%
1044 based on baseline plasma *IL1RL1* levels. We performed a log-rank test to assess whether there
1045 is a statistically significant difference in survival between these two groups, with a nominal $P <$
1046 0.05.

1047 1048 [12.4. Bulk RNA-sequencing](#)

1049 We used the IBD Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) platform⁶⁷
1050 (<https://ibd-meta-analysis.herokuapp.com>) to evaluate changes in *IL1RL1* gene expression. IBD
1051 TaMMA encompasses 3,853 RNA-Seq datasets from 26 studies on IBD and control samples
1052 across various tissues. All datasets were processed using a uniform computational pipeline and
1053 underwent batch correction for harmonizing data, enabling consistent comparison across studies.
1054 Differential expression results for ileum, colon, and rectum biopsies from CD and UC patients
1055 versus healthy controls were downloaded from IBD TaMMA. We assessed the log₂ fold change
1056 to create forest plots.

1057 1058 [12.5. Single-cell RNA-sequencing](#)

1059 To gain a better understanding of the enrichment of *IL1RL1* in specific cell types, we obtained
1060 single-cell RNA sequencing data from Kong et al.⁶⁶ (SCP1884 from
1061 <https://singlecell.broadinstitute.org/>), which profiled 720,633 cells from the terminal ileum and
1062 colon of 71 CD individuals with different levels of inflammation. Specific details of sample
1063 collection, data processing, and single-cell profiling have been described previously⁶⁶. We
1064 evaluated the normalized gene expression levels of *IL1RL1* in 25 different cell types and replotted
1065 the first two dimensions of Uniform Manifold Approximation and Projection (UMAP) coordinates
1066 to visualize the cell clusters. To determine if *IL1RL1* was more significantly expressed in certain
1067 cell types, we conducted 5,000 permutations of the cell type labels. We assessed how often a
1068 specific cell type had the same or a higher proportion of cells expressing *IL1RL1* compared to all
1069 the cells in the overall population (permutation P value).

1070 1071 [13. STROBE-MR statement](#)

1072 Our study closely adheres to the STROBE-MR guidelines and the STROBE-MR checklist is
1073 attached in **Supplementary Note 11**.

1074 1075 [14. Ethics declarations](#)

1076 All contributing cohorts obtained ethical approval from their institutional ethics review boards. The
1077 contributing proteomics cohorts include the Atherosclerosis Risk in Communities (ARIC) Study,
1078 deCODE study, Fenland study, UK Biobank, and Kyoto University Nagahama study. The UK
1079 Biobank has approval from the North West Multi-centre Research Ethics Committee as a
1080 Research Tissue Bank.

1081 1082 [15. Data availability](#)

1083 We will provide unfiltered proteome-phenome wide MR results for European (ARIC, deCODE,
1084 Fenland, UKB-PPP), African (ARIC, UKB-PPP), and East Asian (Kyoto University Nagahama
1085 cohort) ancestries on FigShare upon publication. We caution against directly comparing MR effect
1086 estimates across continuous outcomes, as the outcomes were collected from various sources
1087 and may not be scaled to the same units.

1088

1089 [15.1. Proteomic GWAS](#)

1090 ARIC summary statistics (EUR and AFR): <http://nilanjanchatterjeelab.org/pwas/>
1091 deCODE summary summary statistics (EUR): <https://www.deCODE.com/summarydata/>
1092 Fenland summary statistics (EUR): <https://omicscience.org/apps/pgwas/>
1093 UKB-PPP summary statistics (EUR, AFR): <http://ukb-ppp.gwas.eu/>
1094 Kyoto University Nagahama cohort summary statistics (EAS): Available through contacting the
1095 authors of this study.

1097 [15.2. Outcome GWAS summary statistics](#)

1098 Information on the 179 European outcomes used in this study and the link to the original summary
1099 statistics is available in **Supplementary Table 10**.
1100 The 26 African outcomes are available in **Supplementary Table 11**.
1101 The 206 East Asian outcomes are available in **Supplementary Table 12**.
1102 Million Veteran Program outcomes can be found in the original study⁵⁷.
1103 The largest East Asian ancestry IBD, CD, and UC GWAS are publicly available from Liu et al.⁶⁵
1104 and were downloaded from <https://www.ibdgenetics.org/>

1106 [15.3. Variant-to-gene score](#)

1107 Open Targets Genetics³⁵ (<https://genetics-docs.opentargets.org/data-access/data-download>),

1109 [15.4. Reference panels](#)

1110 European ancestries: UKB 50k (<https://www.ukbiobank.ac.uk/>)
1111 East Asian ancestries: 1000 Genomes Project (<https://www.internationalgenome.org/data>)
1112 African ancestries: HGDP+1KG (<https://gnomad.broadinstitute.org/news/2020-10-gnomad-v3-1-new-content-methods-annotations-and-data-availability/#the-gnomad-hgdp-and-1000-genomes-callset>)

1116 [15.5. Druggability](#)

1117 We used Finan et al. 2017³⁸ for the list of 4,479 protein-coding genes in each druggability tier,
1118 DrugBank database v.5.1.12 (<https://go.drugbank.com/releases/latest>) for information on drugs
1119 targeting specific proteins, and Open Targets Platform database v.24.03
1120 (<https://platform.opentargets.org/downloads>) for clinical trial phase and status information for
1121 protein-drug-disease triplets.

1123 [15.6. Expression analyses](#)

1124 For gene expression data, we used data from Kong et al.⁶⁶ (SCP1884 at Single Cell Portal
1125 <https://singlecell.broadinstitute.org/>).

1127 [16. Code availability](#)

1128 We used R v.4.1.2 (<https://www.r-project.org/>),
1129 Python 3.10 (<https://www.python.org/downloads/release/python-3100/>)
1130 PLINK v.1.9⁷⁷ (<http://pngu.mgh.harvard.edu/purcell/plink/>),
1131 TwoSampleMR v.0.5.6 (<https://mrcieu.github.io/TwoSampleMR/>),
1132 coloc v.5.2.3⁸⁰ (<https://chr1swallace.github.io/coloc/>),
1133 PWCoCo^{18,36} (<https://github.com/jwr-git/pwcoco>),
1134 SharePro v.5.0.0³⁷ (https://github.com/zhwm/SharePro_coloc/),
1135 Cytoscape v.3.10.2 (<https://cytoscape.org/>),
1136 LocusZoom⁸² (<https://my.locuszoom.org/>)

1137
1138 Code used in this study will be made available at <https://github.com/chenyangsu/pQTL-MR> upon
1139 publication.

1140

1141 **17. Acknowledgments**

1142 This research has been conducted using the UK Biobank Resource under Application Number
1143 27449. C.-Y.S. is supported by a CIHR Canada Graduate Scholarship Doctoral Award (Funding
1144 Reference Number: 187673), an FRQS doctoral training scholarship, and a Lady Davis
1145 Institute/TD-Bank Scholarship. T.L. has been supported by start-up funding from the Office of the
1146 Vice Chancellor for Research and Graduate Education, School of Medicine and Public Health,
1147 and Department of Population Health Sciences at the University of Wisconsin-Madison. S.Y. is
1148 supported by the Japan Society for the Promotion of Science.

1149

1150 The funders had no role in the study design, data collection and analysis, decision to publish, or
1151 preparation of the manuscript. We acknowledge Servier Medical Art (<https://smart.servier.com/>)
1152 for providing images that were used to create diagrams in this study. We thank the Digital
1153 Research Alliance of Canada for providing computing resources.

1154

1155 **18. Author contributions**

1156 Conception and design: C.-Y.S., T.L., S.Y.

1157 Methodology: C.-Y.S., W. Z., T.L., S.Y.

1158 Data curation: C.-Y.S., S.S.-H., T.-Y.Y., K.Y.H.L., Y.C., F. M., T.L., S.Y.

1159 Data Analysis: C.-Y.S., T.L., S.Y.

1160 Visualization: C.-Y.S., A.v.d.G., T.L., S.Y.

1161 Knowledge portal: C.-Y.S., D.-K.J., M.C., S.Y.

1162 Writing—Original Draft: C.-Y.S.

1163 Writing—Review and Editing: all authors

1164 Supervision: T.L., S.Y.

1165 Project administration: T.L., S.Y.

1166 Funding acquisition: V.M., S.Z., T.L., S.Y.

1167

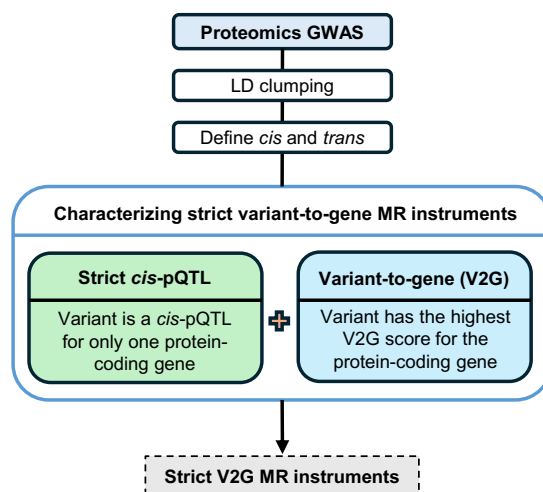
1168 **19. Competing Interests**

1169 W.Z., G.B.-L., and T.L. have been consulting for 5 Prime Sciences. However, this study was
1170 performed separately with no relationship to 5 Prime Sciences. The other authors declare no
1171 conflict of interest.

1172

1173 **Extended Data Figures**

1174



1175

1176

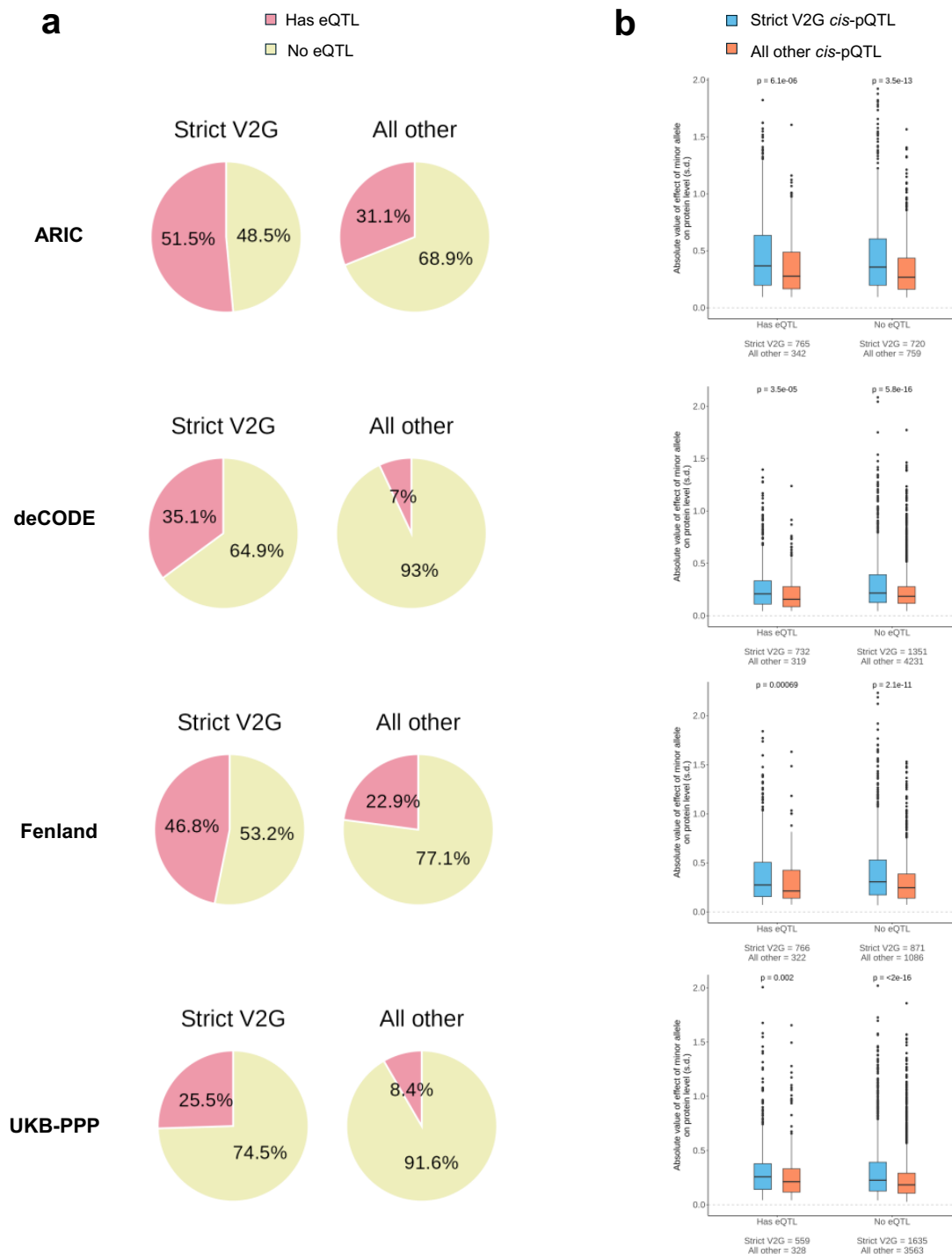
1177

1178

1179

Extended Data Fig. 1. Flow diagram showing the definition of strict variant-to-gene *cis*-pQTLs.

Flow diagram showing the selection of strict variant-to-gene (V2G) *cis*-pQTLs used as instruments for MR starting from the proteomic GWAS.

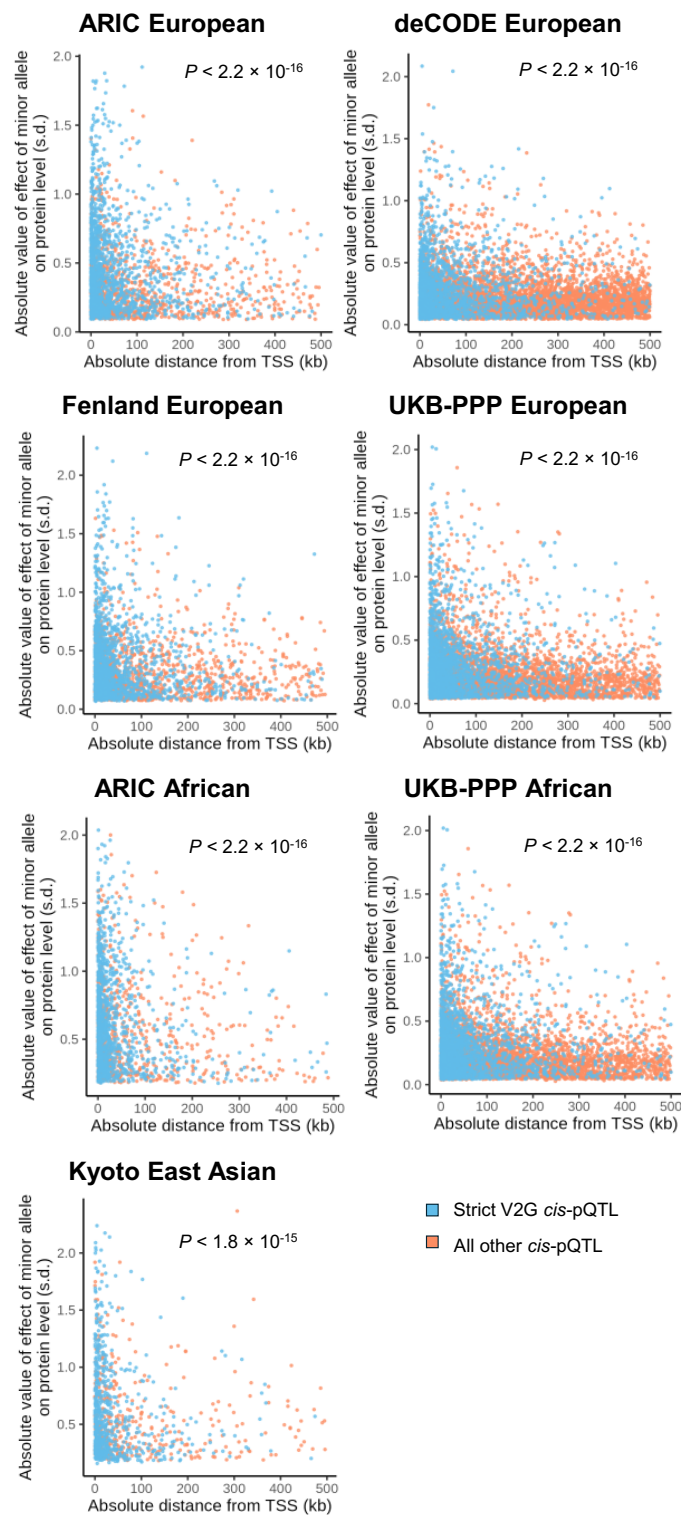


1180
1181
1182
1183
1184
1185
1186
1187
1188

Extended Data Fig. 2. eQTL enrichment analysis comparing strict V2G *cis*-pQTLs against all other *cis*-pQTLs.

- (a) Pie charts showing proportion of eQTL enrichment in the ARIC, deCODE, Fenland, and UKB-PPP European ancestry cohorts for strict V2G *cis*-pQTLs compared to all other *cis*-pQTLs. Red: presence of a *cis*-eQTL; Yellow: absence of a *cis*-eQTL.
- (b) Absolute value of effect of the minor allele on protein level broken down by the presence or absence of *cis*-eQTLs in the ARIC, deCODE, Fenland, and UKB-PPP European

1189 ancestry cohorts. Strict V2G, strict variant-to-gene *cis*-pQTLs; All other, all other *cis*-
1190 pQTLs that were removed due to strict V2G filtering. Boxplots show the median, lower,
1191 and upper quartiles; whiskers end at 1.5 times the interquartile range from the top and
1192 bottom of the box; points outside the whisker boundaries are plotted individually; smaller
1193 black dots represent individual points and are used to show the number of samples
1194 included in each boxplot; significance level *P* value is based on the Mann-Whitney U test.
1195

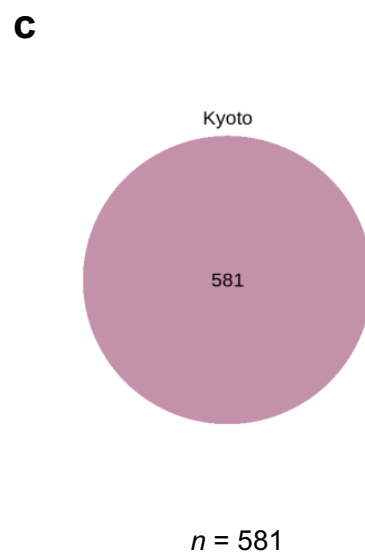
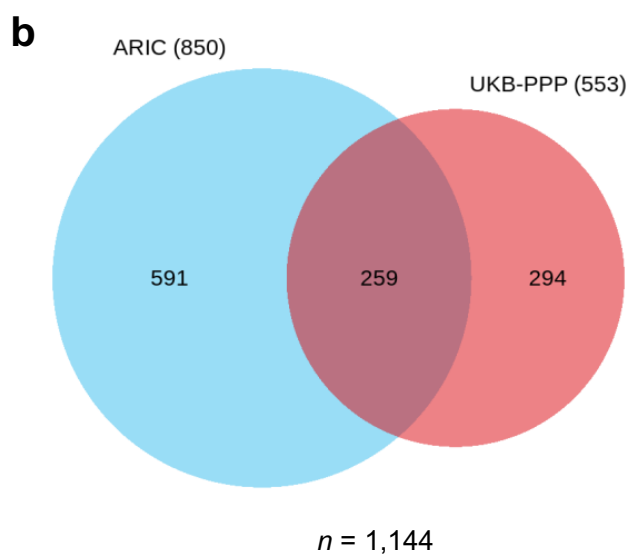
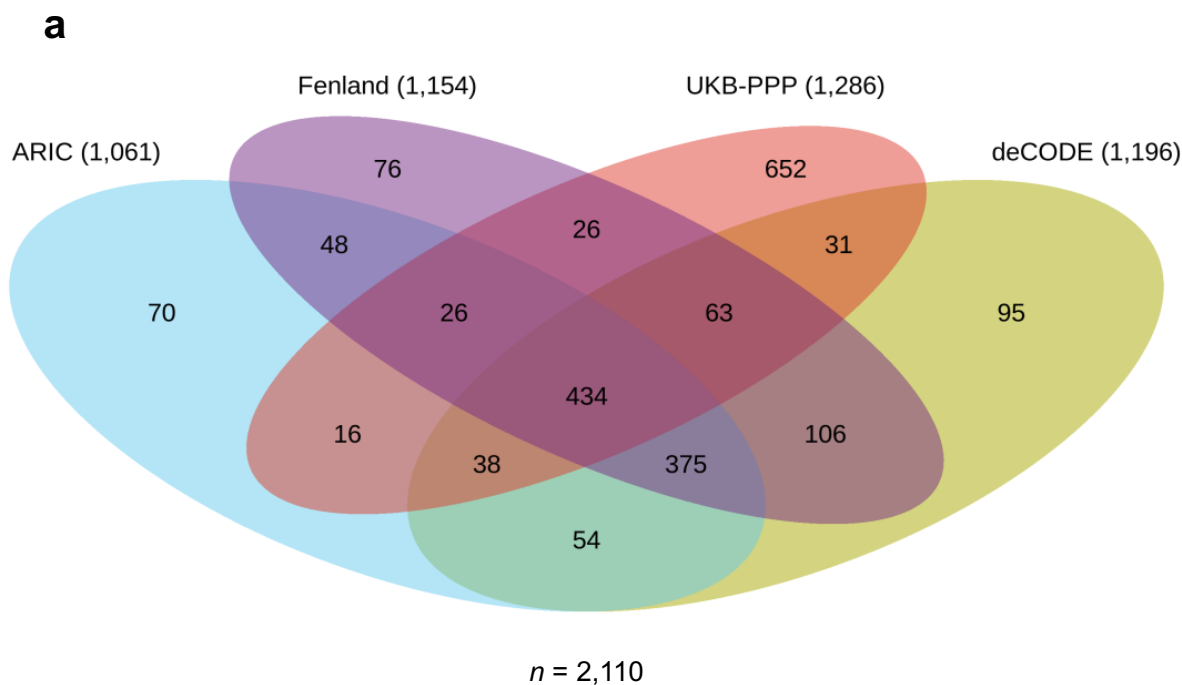


1196
1197 **Extended Data Fig. 3. Absolute distance from transcription start site versus effect size for**
1198 **strict V2G *cis*-pQTLs and all other *cis*-pQTLs.**

1199 The x-axis shows the distance of the pQTL from the canonical transcription start site of the
1200 associated protein-coding gene while the y-axis shows the absolute value of the effect size
1201 estimate of the effect allele aligned to the minor allele of each ancestry's respective reference
1202 panel. (Note, in European, African, and East Asian ancestry proteomics cohorts, the effect allele

1203 of *cis*-pQTLs in each cohort was aligned to the minor allele of the corresponding variant in their
1204 respective reference panels—UKB 50k for European, HGDP+1kGP for African, and 1kGP for
1205 East Asian ancestry—to harmonize alleles across each ancestral cohort for plotting). Strict V2G
1206 *cis*-pQTLs are highlighted in blue while all other *cis*-pQTLs are highlighted in orange. Note that
1207 Fenland European *cis*-pQTLs are presented in GRCh37 coordinates while all other cohorts are
1208 presented in GRCh38 coordinates. *P* values show a one-sided t-test testing whether strict V2G
1209 *cis*-pQTLs have smaller absolute distance to the TSS compared to all other *cis*-pQTLs.
1210
1211

1212



1213
1214
1215
1216
1217
1218
1219

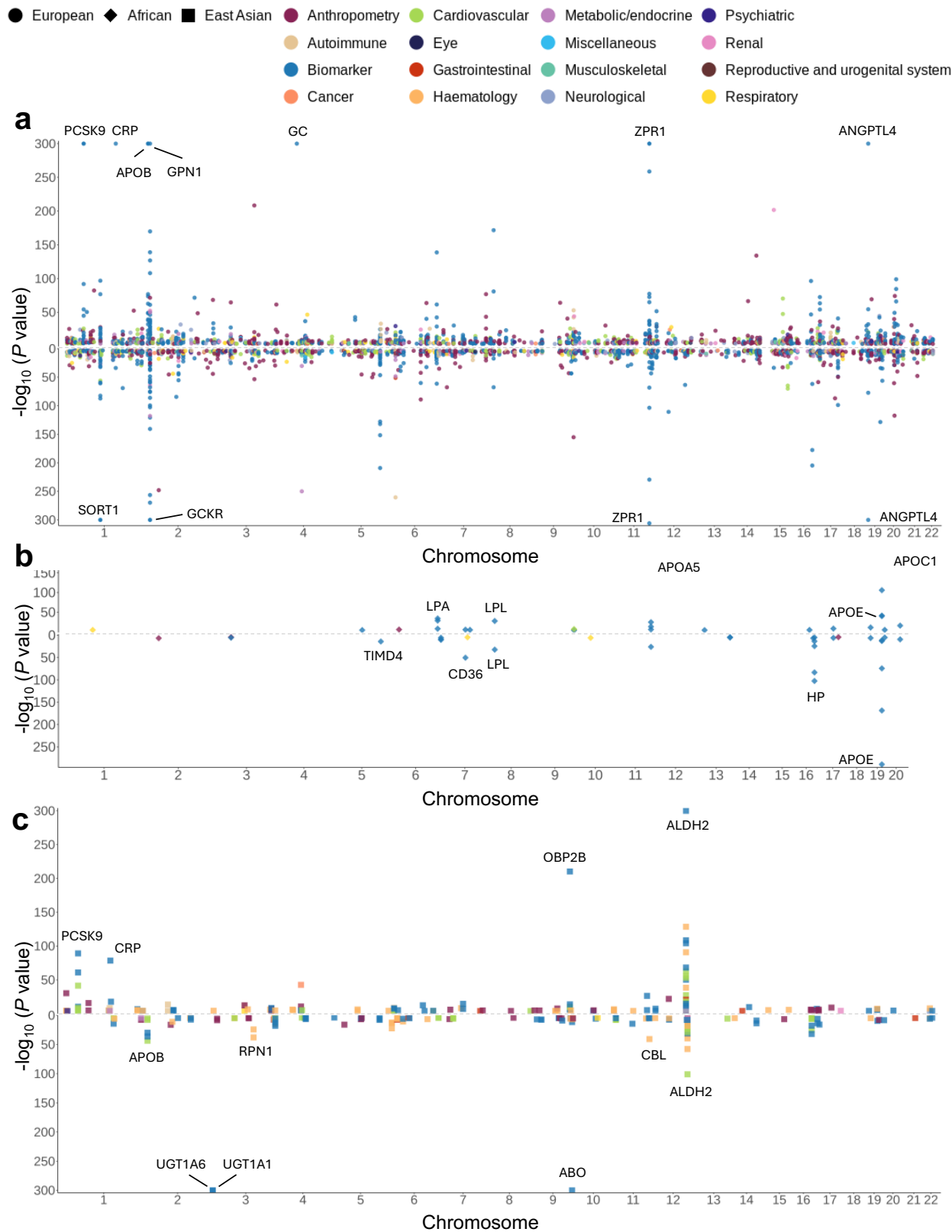
Extended Data Fig. 4. Within ancestry comparison of instrumentable proteins.

Venn diagrams of overlapping instrumentable proteins within ancestries.

(a) European cohorts involving ARIC, deCODE, Fenland, and UKB-PPP (4 cohorts).

(b) African cohorts involving ARIC and UKB-PPP (2 cohorts).

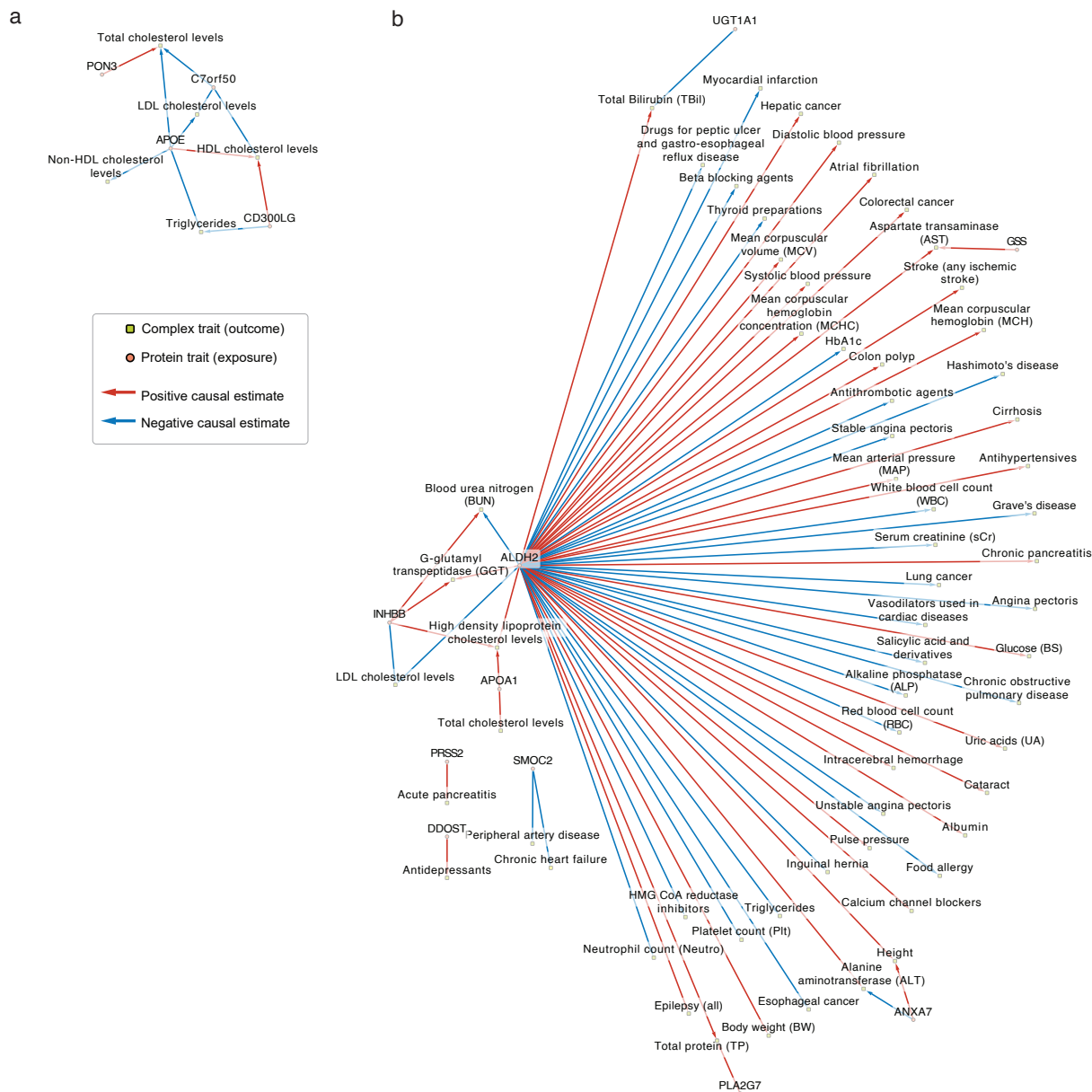
(c) East Asian ancestry cohort from Kyoto University Nagahama (single cohort).



1220
1221
1222

Extended Data Fig. 5. Putatively causal protein-phenotype associations across European, African, and East Asian ancestries.

1223 Miami plots displaying chromosomal position (*x*-axis) of significant putatively causal protein-
1224 outcome associations (MR-passing and colocalized with $PP_{\max} \geq 0.8$) in (a) European, (b) African,
1225 (c) East Asian ancestry. The *y*-axis shows *P* values from the MR causal estimates where the
1226 exposure is protein level and outcome is the complex trait or disease. Colors indicate the type of
1227 complex trait or disease. Cancer types were harmonized under a single “Cancer” group. Ancestry
1228 is denoted by filled circle (European), filled diamond (African), and filled square (East Asian). Each
1229 data point is plotted based on the chromosome and transcription start-site of the protein-coding
1230 gene. For simplicity, *Z* scores and *P* values are averaged across cohorts in the European ancestry
1231 plot and the African ancestry plot and shown as a single data point. In European ancestry, only
1232 associations that were consistent across all cohorts are shown.
1233



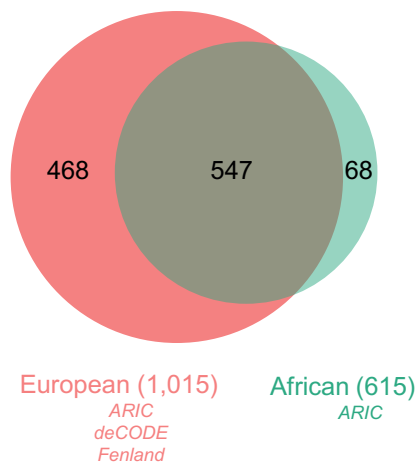
1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244

Extended Data Fig. 6. Uniquely instrumentable protein-phenotype pairs in African and East Asian ancestries.

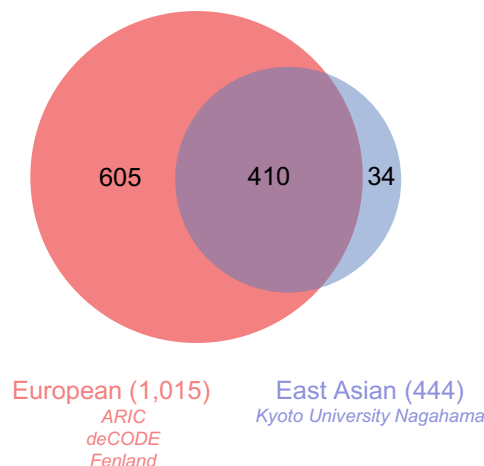
Red arrows indicate a positive causal estimate of the protein on the outcome while blue arrows indicate a negative causal estimate of the protein on the outcome.

- (a) Protein-phenotype pairs from 4 proteins uniquely instrumentable in African ancestry. Significant estimates between proteins (orange circles) and traits (green rectangles).
- (b) Protein-phenotype pairs from 8 proteins uniquely instrumentable in East Asian ancestry. Significant estimates between proteins (orange circles) and traits (green rectangles).

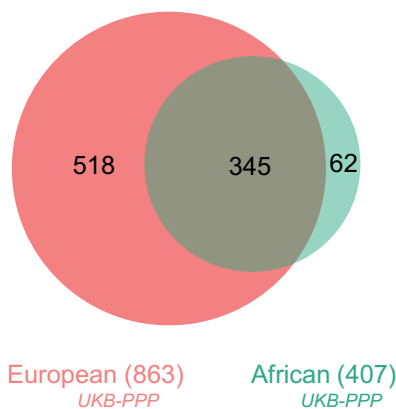
a SomaScan v4 European vs. African ancestry



b SomaScan v4 European vs. East Asian ancestry



c Olink Explore 3072 European vs. African ancestry



1245 **Extended Data Fig. 7. Cross-ancestry comparison stratified by proteomics platform of**
 1246 **instrumentable proteins overlapping at least one drug database (druggable genome,**
 1247 **DrugBank, or Open Targets Platform).**

1249 (a) Comparison of SomaScan v4 platform instrumentable proteins overlapping at least one drug
 1250 database between three European cohorts (ARIC, deCODE, and Fenland) and one African cohort
 1251 (ARIC).

1252 (b) Comparison of SomaScan v4 platform instrumentable proteins overlapping at least one drug
 1253 database between three European cohorts (ARIC, deCODE, and Fenland) and one East Asian
 1254 cohort (Kyoto University Nagahama).

1255 (c) Comparison of Olink Explore 3072 platform instrumentable proteins overlapping at least one
 1256 drug database between one European cohort (UKB-PPP) and one African cohort (UKB-PPP).
 1257

1258 **References**

- 1259 1. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
- 1260 2. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease.
- 1261 *Science* **361**, 769–773 (2018).
- 1262 3. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood
- 1263 plasma proteome. *Nat Commun* **8**, 14357 (2017).
- 1264 4. Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat Med*
- 1265 **25**, 1851–1857 (2019).
- 1266 5. Su, C.-Y. *et al.* Circulating proteins to predict COVID-19 severity. *Sci Rep* **13**, 6236 (2023).
- 1267 6. Carrasco-Zanini, J. *et al.* Proteomic signatures improve risk prediction for common and rare
- 1268 diseases. *Nat Med* **30**, 2489–2498 (2024).
- 1269 7. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat Rev Drug Discov* **1**, 727–730
- 1270 (2002).
- 1271 8. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat Rev*
- 1272 *Drug Discov* **5**, 993–996 (2006).
- 1273 9. Bakheet, T. M. & Doig, A. J. Properties and identification of human protein drug targets.
- 1274 *Bioinformatics* **25**, 451–457 (2009).
- 1275 10. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*
- 1276 **16**, 19–34 (2017).
- 1277 11. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases.
- 1278 *Science* **374**, eabj1541 (2021).
- 1279 12. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and
- 1280 disease. *Nat Genet* **53**, 1712–1721 (2021).
- 1281 13. Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African
- 1282 ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet*
- 1283 **54**, 593–602 (2022).
- 1284 14. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK
- 1285 Biobank. *Nature* **622**, 329–338 (2023).
- 1286 15. Skrivankova, V. W. *et al.* Strengthening the reporting of observational studies in
- 1287 epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration.
- 1288 *BMJ* **375**, n2233 (2021).
- 1289 16. Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in
- 1290 Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**,
- 1291 1614–1621 (2021).
- 1292 17. Chong, M. *et al.* Novel Drug Targets for Ischemic Stroke Identified Through Mendelian
- 1293 Randomization Analysis of the Blood Proteome. *Circulation* **140**, 819–830 (2019).
- 1294 18. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the
- 1295 plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).
- 1296 19. Zhao, H. *et al.* Proteome-wide Mendelian randomization in global biobank meta-analysis
- 1297 reveals multi-ancestry drug targets for common diseases. *Cell Genomics* **2**, 100195 (2022).
- 1298 20. Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry
- 1299 against COVID-19 susceptibility and severity. *Nat Med* **27**, 659–667 (2021).
- 1300 21. Yoshiji, S. *et al.* Proteome-wide Mendelian randomization implicates nephronectin as an
- 1301 actionable mediator of the effect of obesity on COVID-19 severity. *Nat Metab* **5**, 248–264
- 1302 (2023).
- 1303 22. Yoshiji, S. *et al.* COL6A3-derived endotrophin mediates the effect of obesity on coronary
- 1304 artery disease: an integrative proteogenomics analysis. Preprint at *medRxiv*
- 1305 <https://doi.org/10.1101/2023.04.19.23288706> (2023).
- 1306 23. Lu, T., Forgetta, V., Greenwood, C. M. T., Zhou, S. & Richards, J. B. Circulating Proteins
- 1307 Influencing Psychiatric Disease: A Mendelian Randomization Study. *Biological Psychiatry* **93**,
- 1308 82–91 (2023).

- 1309 24. Katz, D. H. *et al.* Whole Genome Sequence Analysis of the Plasma Proteome in Black
1310 Adults Provides Novel Insights Into Cardiovascular Disease. *Circulation* **145**, 357–370
1311 (2022).
- 1312 25. Wang, B. *et al.* Comparative studies of genetic and phenotypic associations for 2,168
1313 plasma proteins measured by two affinity-based platforms in 4,000 Chinese adults. Preprint
1314 at *medRxiv* <https://doi.org/10.1101/2023.12.01.23299236> (2023).
- 1315 26. Said, S. *et al.* Ancestry diversity in the genetic determinants of the human plasma
1316 proteome and associated new drug targets. Preprint at *medRxiv*
1317 <https://doi.org/10.1101/2023.11.13.23298365> (2023).
- 1318 27. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human
1319 phenotypes. *Nat Genet* **53**, 1415–1424 (2021).
- 1320 28. Feng, Y.-C. A. *et al.* Taiwan Biobank: A rich biomedical research database of the
1321 Taiwanese population. *Cell Genomics* **2**, 100197 (2022).
- 1322 29. Carlson, C. S. *et al.* Generalization and Dilution of Association Results from European
1323 GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLOS Biology* **11**,
1324 e1001661 (2013).
- 1325 30. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across
1326 Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).
- 1327 31. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from
1328 frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161–165 (2005).
- 1329 32. Robinson, J. G. *et al.* Efficacy and Safety of Alirocumab in Reducing Lipids and
1330 Cardiovascular Events. *New England Journal of Medicine* **372**, 1489–1499 (2015).
- 1331 33. Sabatine, M. S. *et al.* Evolocumab and Clinical Outcomes in Patients with
1332 Cardiovascular Disease. *New England Journal of Medicine* **376**, 1713–1722 (2017).
- 1333 34. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–
1334 250 (2022).
- 1335 35. Ghousaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated
1336 genes using large-scale genetics and functional genomics. *Nucleic Acids Research* **49**,
1337 D1311–D1320 (2021).
- 1338 36. Robinson, J. W. *et al.* An efficient and robust tool for colocalisation: Pair-wise
1339 Conditional and Colocalisation (PWCoCo). Preprint at *medRxiv*
1340 <https://doi.org/10.1101/2022.08.08.503158> (2022).
- 1341 37. Zhang, W. *et al.* SharePro: an accurate and efficient genetic colocalization method
1342 accounting for multiple causal signals. *Bioinformatics* **40**, btae295 (2024).
- 1343 38. Finan, C. *et al.* The druggable genome and support for target identification and validation
1344 in drug development. *Science Translational Medicine* **9**, eaag1166 (2017).
- 1345 39. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery
1346 and exploration. *Nucleic Acids Research* **34**, D668–D672 (2006).
- 1347 40. Koscielny, G. *et al.* Open Targets: a platform for therapeutic target identification and
1348 validation. *Nucleic Acids Research* **45**, D985–D994 (2017).
- 1349 41. Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat*
1350 *Commun* **11**, 3255 (2020).
- 1351 42. Gkatzionis, A., Burgess, S. & Newcombe, P. J. Statistical methods for cis-Mendelian
1352 randomization with two-sample summary-level data. *Genetic Epidemiology* **47**, 3–25 (2023).
- 1353 43. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G.
1354 Mendelian randomization: Using genes as instruments for making causal inferences in
1355 epidemiology. *Statistics in Medicine* **27**, 1133–1163 (2008).
- 1356 44. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
1357 141,456 humans. *Nature* **581**, 434–443 (2020).

- 1358 45. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
1359 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
1360 *(Methodological)* **57**, 289–300 (1995).
- 1361 46. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization:
1362 Review and comparison of approaches. *The American Journal of Human Genetics* **109**, 767–
1363 782 (2022).
- 1364 47. van der Graaf, A. *et al.* Mendelian randomization while jointly modeling cis genetics
1365 identifies causal relationships between gene expression and lipids. *Nat Commun* **11**, 4930
1366 (2020).
- 1367 48. Goruppi, S. *et al.* The ULK3 Kinase Is Critical for Convergent Control of Cancer-
1368 Associated Fibroblast Activation by CSL and GLI. *Cell Reports* **20**, 2468–2479 (2017).
- 1369 49. Hodonsky, C. J. *et al.* Multi-ancestry genetic analysis of gene regulation in coronary
1370 arteries prioritizes disease risk loci. *Cell Genomics* **4**, 100465 (2024).
- 1371 50. Verstockt, B. *et al.* IL-12 and IL-23 pathway inhibition in inflammatory bowel disease. *Nat*
1372 *Rev Gastroenterol Hepatol* **20**, 433–446 (2023).
- 1373 51. O'Meara, E. *et al.* Independent Prognostic Value of Serum Soluble ST2 Measurements
1374 in Patients With Heart Failure and a Reduced Ejection Fraction in the PARADIGM-HF Trial
1375 (Prospective Comparison of ARNI With ACEI to Determine Impact on Global Mortality and
1376 Morbidity in Heart Failure). *Circulation: Heart Failure* **11**, e004446 (2018).
- 1377 52. Jiang, Y. *et al.* An IL1RL1 genetic variant lowers soluble ST2 levels and the risk effects
1378 of APOE- ϵ 4 in female patients with Alzheimer's disease. *Nat Aging* **2**, 616–634 (2022).
- 1379 53. England, E. *et al.* Tozorakimab (MEDI3506): an anti-IL-33 antibody that inhibits IL-33
1380 signalling via ST2 and RAGE/EGFR to reduce inflammation and epithelial dysfunction. *Sci*
1381 *Rep* **13**, 9825 (2023).
- 1382 54. Cohen, S. B. The use of anakinra, an interleukin-1 receptor antagonist, in the treatment
1383 of rheumatoid arthritis. *Rheumatic Disease Clinics of North America* **30**, 365–380 (2004).
- 1384 55. Aviram, M. *et al.* Paraoxonase inhibits high-density lipoprotein oxidation and preserves
1385 its functions. A possible peroxidative role for paraoxonase. *J Clin Invest* **101**, 1581–1590
1386 (1998).
- 1387 56. Reddy, S. T. *et al.* Human Paraoxonase-3 Is an HDL-Associated Enzyme With Biological
1388 Activity Similar to Paraoxonase-1 Protein but Is Not Regulated by Oxidized Lipids.
1389 *Arteriosclerosis, Thrombosis, and Vascular Biology* **21**, 542–547 (2001).
- 1390 57. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million
1391 Veteran Program. *Science* **385**, eadj1182 (2024).
- 1392 58. Banfi, C. *et al.* Prenylcysteine oxidase 1, an emerging player in atherosclerosis.
1393 *Commun Biol* **4**, 1–17 (2021).
- 1394 59. Rocnik, E. F., Liu, P., Sato, K., Walsh, K. & Vaziri, C. The Novel SPARC Family Member
1395 SMOC-2 Potentiates Angiogenic Growth Factor Activity*. *Journal of Biological Chemistry*
1396 **281**, 22855–22864 (2006).
- 1397 60. Bourgault, J. *et al.* Proteome-Wide Mendelian Randomization Identifies Causal Links
1398 Between Blood Proteins and Acute Pancreatitis. *Gastroenterology* **164**, 953-965.e3 (2023).
- 1399 61. Lafferty, M. J., Bradford, K. C., Erie, D. A. & Neher, S. B. Angiopoietin-like Protein 4
1400 Inhibition of Lipoprotein Lipase: EVIDENCE FOR REVERSIBLE COMPLEX FORMATION*
1401 *This work was supported by a grant from the Pew Foundation (to S. B. N.). *Journal of*
1402 *Biological Chemistry* **288**, 28524–28534 (2013).
- 1403 62. Wang, H. & Eckel, R. H. Lipoprotein lipase: from gene to obesity. *American Journal of*
1404 *Physiology-Endocrinology and Metabolism* **297**, E271–E288 (2009).
- 1405 63. Rosenson Robert S. *et al.* Zodasiran, an RNAi Therapeutic Targeting ANGPTL3, for
1406 Mixed Hyperlipidemia. *New England Journal of Medicine* **0**,.
- 1407 64. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives
1408 for large population-based studies. *Nat Rev Genet* **22**, 19–37 (2021).

- 1409 65. Liu, Z. *et al.* Genetic architecture of the inflammatory bowel diseases across East Asian
1410 and European ancestries. *Nat Genet* **55**, 796–806 (2023).
- 1411 66. Kong, L. *et al.* The landscape of immune dysregulation in Crohn’s disease revealed
1412 through single-cell transcriptomic profiling in the ileum and colon. *Immunity* **56**, 444–458.e5
1413 (2023).
- 1414 67. Massimino, L. *et al.* The Inflammatory Bowel Disease Transcriptome and
1415 Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nat Comput Sci* **1**, 511–515
1416 (2021).
- 1417 68. Hamilton, M. J., Frei, S. M. & Stevens, R. L. The Multifaceted Mast Cell in Inflammatory
1418 Bowel Disease. *Inflammatory Bowel Diseases* **20**, 2364–2378 (2014).
- 1419 69. Boeckxstaens, G. Mast cells and inflammatory bowel disease. *Current Opinion in*
1420 *Pharmacology* **25**, 45–49 (2015).
- 1421 70. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
1422 *Nature* **562**, 203–209 (2018).
- 1423 71. Koenig, Z. *et al.* A harmonized public resource of deeply sequenced diverse human
1424 genomes. Preprint at *medRxiv* <https://doi.org/10.1101/2023.01.23.525248> (2024).
- 1425 72. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological
1426 databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
- 1427 73. Butler-Laporte, G. *et al.* HLA allele-calling using multi-ancestry whole-exome sequencing
1428 from the UK Biobank identifies 129 novel associations in 11 autoimmune diseases. *Commun*
1429 *Biol* **6**, 1–17 (2023).
- 1430 74. Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M. & Davey Smith, G.
1431 Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular
1432 drug development. *Nat Rev Cardiol* **18**, 435–453 (2021).
- 1433 75. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122
1434 (2016).
- 1435 76. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across
1436 the human phenome. *eLife* **7**, e34408 (2018).
- 1437 77. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-
1438 Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
- 1439 78. Pierce, B. L., Ahsan, H. & VanderWeele, T. J. Power and instrument strength
1440 requirements for Mendelian randomization studies using multiple genetic variants.
1441 *International Journal of Epidemiology* **40**, 740–752 (2011).
- 1442 79. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship between
1443 imprecisely measured traits using GWAS summary data. *PLOS Genetics* **13**, e1007081
1444 (2017).
- 1445 80. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
1446 Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).
- 1447 81. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of
1448 intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
- 1449 82. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan
1450 results. *Bioinformatics* **26**, 2336–2337 (2010).
- 1451