

1
2 Deciphering causal proteins in Alzheimer's disease: A novel Mendelian randomization
3 method integrated with AlphaFold3 for 3D structure prediction

4
5 Minhao Yao¹, Gary W. Miller², Badri N. Vardarajan³, Andrea A. Baccarelli⁴,
6 Zijian Guo^{5*}, Zhonghua Liu^{6*}

7
8 ¹ *Department of Statistics and Actuarial Science, The University of Hong Kong, Hong*
9 *Kong*

10 ² *Department of Environmental Health Sciences, Columbia University, New York, NY,*
11 *USA*

12 ³ *Taub Institute on Alzheimer's Disease and the Aging Brain, Department of*
13 *Neurology, Columbia University,*
14 *New York, NY, USA.*

15 ⁴ *Department of Environmental Health, Harvard T.H. Chan School of Public Health,*
16 *Boston, MA, USA*

17 ⁵ *Department of Statistics, Rutgers University, Piscataway, NJ, USA.*

18 ⁶ *Department of Biostatistics, Columbia University, New York, NY, USA.*

19
20 * *Correspondence to: Zijian Guo (zijguo@stat.rutgers.edu) or Zhonghua Liu (Lead*
21 *contact, zl2509@cumc.columbia.edu)*

22
23 **Summary**

24 Hidden confounding biases hinder identifying causal protein biomarkers for
25 Alzheimer's disease in non-randomized studies. While Mendelian randomization (MR)
26 can mitigate these biases using protein quantitative trait loci (pQTLs) as instrumental
27 variables, some pQTLs violate core assumptions, leading to biased conclusions. To
28 address this, we propose MR-SPI, a novel MR method that selects valid pQTL
29 instruments using the Anna Karenina Principle and performs robust post-selection
30 inference. Integrating MR-SPI with AlphaFold3, we developed a computational
31 pipeline to identify causal protein biomarkers and predict 3D structural changes.
32 Applied to genome-wide proteomics data from 54,306 UK Biobank participants and
33 455,258 subjects (71,880 cases and 383,378 controls) for a genome-wide association
34 study of Alzheimer's disease, we identified seven proteins (TREM2, PILRB, PILRA,
35 EPHA1, CD33, RET, and CD55) with structural alterations due to missense mutations.
36 These findings offer insights into the etiology and potential drug targets for Alzheimer's
37 disease.

41 1. Introduction

42 Alzheimer's disease (AD) stands as the primary cause of dementia globally, exerting a
43 considerable strain on healthcare resources^{1,2}. Despite extensive efforts, the etiology
44 and pathogenesis of AD are still unclear, and strategies aimed at impeding or delaying
45 its clinical advancement have largely remained challenging to achieve^{1,3,4}. The amyloid
46 cascade hypothesis posits that AD begins with the accumulation and aggregation of
47 amyloid-beta (A β) peptides in the brain, culminating in the formation of β -amyloid
48 fibrils, leading to tau hyperphosphorylation, neurofibrillary tangle formation and
49 neurodegeneration^{5,6}. However, current AD therapies targeting A β production and
50 amyloid formation offer only transient symptomatic relief and fail to halt disease
51 progression, resulting in a lack of effective drugs for AD^{1,7}. Therefore, it is imperative
52 and urgent to identify causal protein biomarkers to elucidate the underlying
53 mechanisms of AD, and to expedite the development of effective therapeutic
54 interventions for AD.

55

56 In causal inference, randomized controlled trials (RCTs) serve as the gold standard for
57 evaluating the causal effect of an exposure on the health outcome of interest. However,
58 it might be neither feasible nor ethical to perform RCTs where protein levels are
59 considered as the exposures. Mendelian randomization (MR) leverages the random
60 assortment of genes from parents to offspring to mimic RCTs to establish causality in
61 non-randomized studies⁸⁻¹⁰. MR uses genetic variants, typically single-nucleotide
62 polymorphisms (SNPs), as instrumental variables (IVs) to assess the causal association
63 between an exposure and a health outcome¹¹. Recently, many MR methods have been
64 developed to investigate causal relationships using genome-wide association study
65 (GWAS) summary statistics data that consist of effect estimates of SNP-exposure and
66 SNP-outcome associations from two sets of samples, which are commonly referred to
67 as the two-sample MR designs¹²⁻¹⁵. Since summary statistics are often publicly
68 available and provide abundant information of associations between genetic variants
69 and complex traits/diseases, two-sample MR methods become increasingly popular^{14,16-}
70 ¹⁸. In particular, recent studies with large-scale proteomics data have unveiled numerous
71 protein quantitative traits loci (pQTLs) associated with thousands of proteins^{19,20},
72 facilitating the application of two-sample MR methods, where pQTLs serve as IVs and
73 protein levels serve as exposures, to identify proteins as causal biomarkers for complex
74 traits and diseases.

75

76 To employ MR for identifying causal protein biomarkers, conventional MR methods
77 require the pQTLs included in the analysis to be valid IVs for reliable causal inference.
78 A pQTL is called a valid IV if the following three core IV assumptions hold^{9,21}:

- 79 (A1). **Relevance:** The pQTL is associated with the protein;
80 (A2). **Effective Random Assignment:** The pQTL is not associated with any
81 unmeasured confounder of the protein-outcome relationship; and
82 (A3). **Exclusion Restriction:** The pQTL affects the outcome only through the
83 protein in view.

84 Among the three core IV assumptions (A1) - (A3), only the first assumption (A1) can
85 be tested empirically by selecting pQTLs significantly associated with the protein.
86 However, assumptions (A2) and (A3) cannot be empirically verified in general and may
87 be violated in practice, which may lead to a biased estimate of the causal effect. For
88 example, violation of (A2) may occur due to the presence of population stratification^{9,22};
89 and violation of (A3) may occur in the presence of horizontal pleiotropy^{9,23}, which is a
90 widespread biological phenomenon that the pQTL IV affects the outcome through other
91 biological pathways that do not involve the protein in view, for example, through
92 alternative splicing or micro-RNA effects²⁴⁻²⁶.

93
94 Recently, several MR methods have been proposed to handle invalid IVs under certain
95 assumptions, as summarized in Table 1. Some of these additional assumptions for the
96 identification of the causal effect in the presence of invalid IVs are listed below:

- 97 (i) The Instrument Strength Independent of Direct Effect (InSIDE) assumption: the
98 pQTL-protein effect is asymptotically independent of the horizontal pleiotropic
99 effect when the number of pQTL IVs goes to infinity. For example, the random-
100 effects inverse-variance weighted (IVW) method²⁷, MR-Egger²⁸, MR-RAPS
101 (Robust Adjusted Profile Score)¹⁶, and the Mendelian randomization pleiotropy
102 residual sum and outlier (MR-PRESSO) test²⁹.
- 103 (ii) Majority rule condition: up to 50% of the candidate pQTL IVs are invalid. For
104 example, the weighted median method³⁰ and MR-PRESSO.
- 105 (iii) Plurality rule condition or the ZERo Modal Pleiotropy Assumption (ZEMPA)^{15,31}:
106 a plurality of the candidate IVs are valid, which is weaker than the majority rule
107 condition. For example, the mode-based estimation³¹, MRMix³² and the
108 contamination mixture method³³.
- 109 (iv) Other distributional assumptions. For example, MRMix and the contamination
110 mixture method impose normal mixture distribution assumption on the genetic

111 associations and the ratio estimates, respectively.

112

113 Despite many existing efforts, current MR methods still face new challenges when
114 dealing with pQTLs IVs for analyzing proteomics data. First, it's worth noting that the
115 number of pQTLs for each protein tends to be small. For example, in two proteomics
116 studies, the median number of pQTLs per protein is 4^{20,34}. With such a limited number
117 of IVs, those MR methods based on the InSIDE assumption which requires a large
118 number of IVs or other distributional assumptions might yield unreliable results in the
119 presence of invalid IVs^{15,28}. Second, current MR methods require an ad-hoc set of pre-
120 determined genetic IVs, which is often obtained by selecting genetic variants with
121 strong pQTL-protein associations in proteomics data³⁵. Since such traditional way of
122 selecting IVs only requires the proteomics data, hence the same set of selected IVs is
123 used for assessing the causal relationships between the protein in view and different
124 health outcomes. Obviously, this one-size-fits-all strategy for selecting IVs might not
125 work well for different outcomes because the underlying genetic architecture may vary
126 across outcomes. For example, the pattern of horizontal pleiotropy might vary across
127 different outcomes. Therefore, it is desirable to develop an automatic and
128 computationally efficient algorithm to select a set of valid genetic IVs for a specific
129 protein-outcome pair to perform reliable causal inference, especially when the number
130 of candidate pQTL IVs is small.

131

132 In this paper, we develop a novel all-in-one pipeline for causal protein biomarker
133 identification and 3D structural alteration prediction using large-scale genetics,
134 proteomics and phenotype/disease data, as illustrated in Figure 1. Specifically, we
135 propose a two-sample MR method and algorithm that can automatically Select valid
136 pQTL IVs and then performs robust Post-selection Inference (MR-SPI) for the causal
137 effect of proteins on the health outcome of interest. The key idea of MR-SPI is based
138 on the Anna Karenina Principle which states that all valid instruments are alike, while
139 each invalid instrument is invalid in its own way – paralleling Leo Tolstoy's dictum that
140 "all happy families are alike; each unhappy family is unhappy in its own way"³⁶. In
141 other words, valid instruments will form a group and should provide similar ratio
142 estimates of the causal effect, while the ratio estimates of invalid instruments are more
143 likely to be different from each other. With the application of MR-SPI, we can not only
144 identify the causal protein biomarkers associated with disease outcomes, but can also
145 obtain missense genetic variations (used as pQTL IVs) for those identified causal

146 proteins. These missense pQTL IVs will induce changes of amino acids, leading to 3D
147 structural changes of these proteins. A classic example of a missense mutation was
148 found in sickle cell disease, where the mutation at SNP rs334, located on chromosome
149 11 (11p15.4), results in the change of codon 6 of the beta globin chain from [GAA] to
150 [GTA]³⁷⁻³⁹. This substitution leads to the replacement of glutamic acid with valine at
151 position 6 of the beta chain of the hemoglobin protein, altering the structure and
152 function of hemoglobin protein. Consequently, red blood cells assume a crescent or
153 sickle-shaped morphology, impairing blood flow to various parts of the body^{40,41}.
154 Moreover, to further offer novel biological insights into the interpretation of the causal
155 effect at the molecular level, we incorporate AlphaFold3⁴²⁻⁴⁵ into our pipeline to predict
156 the 3D structural alteration resulting from the corresponding missense pQTL IVs for
157 the causal proteins identified by MR-SPI. Our pipeline can elucidate the mechanistic
158 underpinnings of how missense genetic variations translate into 3D structural
159 alterations at the protein level, thereby advancing our understanding of disease etiology
160 and potentially informing targeted therapeutic interventions.

161

162 Our pioneering pipeline for the first time integrates the identification of causal protein
163 biomarkers for health outcomes and the subsequent analysis of their 3D structural
164 alterations into a unified framework, leveraging increasingly publicly available GWAS
165 summary statistics for health research. Within our framework, the proposed MR-SPI
166 serves dual purposes: (1) identifying causal protein biomarkers; and (2) selecting valid
167 missense pQTL IVs for subsequent 3D structural analysis. Compared to existing two-
168 sample MR methods, MR-SPI is the first MR method that utilizes both exposure and
169 outcome data to automatically select a set of valid IVs, especially when the number of
170 candidate IVs is small in proteomics data, which is a prominent challenge with no
171 satisfactory solution up to date. We note that while MR-PRESSO also selects valid IVs
172 for MR analysis, it requires both the stronger majority rule condition and the InSIDE
173 assumption, as well as a minimum number of four candidate IVs for implementation²⁹.
174 In contrast, our proposed MR-SPI does not require the InSIDE assumption and only
175 requires the plurality rule condition that is weaker than the majority rule condition, and
176 only requires a minimum number of three IVs for the proposed voting procedure.
177 Therefore, MR-SPI is more suitable for analyzing proteomics data. Extensive
178 simulations show that our MR-SPI method outperforms other competing MR methods
179 under the plurality rule condition. We employ MR-SPI to perform omics MR (xMR)
180 with 912 plasma proteins using the large-scale UK Biobank proteomics data in 54,306

181 UK Biobank participants²⁰ and find 7 proteins significantly associated with the risk of
182 Alzheimer's disease. We further use AlphaFold3⁴²⁻⁴⁵ to predict the 3D structural
183 alterations of these 7 proteins due to missense genetic variations, and then illustrate the
184 structural alterations graphically using the PyMOL software (<https://pymol.org>),
185 providing new biological insights into their functional roles in AD development and
186 may aid in identifying potential drug targets.

187

188 **2. Results**

189 **2.1 Overview of the pipeline**

190 Our proposed all-in-one pipeline for the identification and 3D structural alteration
191 prediction of causal protein biomarkers consists of three primary steps, as illustrated in
192 Figure 1. First, for each protein biomarker, we employ our proposed MR-SPI to select
193 valid pQTL IVs by incorporating the proteomics GWAS and disease outcome GWAS
194 summary data together, and then estimate the causal effect of each protein on the
195 outcome using the selected valid pQTL IVs. The main idea and more detailed
196 implementation steps for MR-SPI is described in Section 2.2. Second, we perform
197 Bonferroni correction⁴⁶ for the p -values of the estimated causal effects to identify
198 putative causal protein biomarkers associated with the outcome. Third, for each
199 identified protein, we apply AlphaFold3 to predict and compare the 3D structures of
200 both the wild-type protein and mutated protein resulting from missense pQTL IVs.

201

202 **2.2 MR-SPI selects valid genetic instruments by a voting procedure**

203 MR-SPI is an automatic procedure to select valid pQTL instruments and perform robust
204 causal inference using two-sample GWAS and proteomics data. In summary, MR-SPI
205 consists of the following four steps, as illustrated in Figure 2:

- 206 (1). select relevant pQTL IVs that are strongly associated with the protein;
- 207 (2). each relevant pQTL IV provides a ratio estimate of the causal effect, and then
208 all the other relevant pQTL IVs votes for it to be a valid IV if their degrees of
209 violation of assumptions (A2) and (A3) are smaller than a data-dependent
210 threshold as in equation (4);
- 211 (3). select valid pQTL IVs by majority/plurality voting or by finding the maximum
212 clique of the voting matrix that encodes whether two relevant pQTL IVs
213 mutually vote for each other to be valid (the voting matrix is defined in
214 equation (6) in STAR (structured, transparent, accessible reporting) Methods);
- 215 (4). estimate the causal effect using the selected valid pQTL IVs and construct a

216 robust confidence interval with guaranteed nominal coverage even if in the
217 presence of possible IV selection error in finite samples.

218

219 Most current two-sample MR methods only use step (1) to select (relevant) pQTL
220 instruments for downstream MR analysis, while the selected pQTL instruments might
221 violate assumptions (A2) and (A3), leading to possibly unreliable scientific findings.
222 To address this issue, MR-SPI automatically select valid pQTL instruments for a
223 specific protein-outcome pair by further incorporating the outcome GWAS data. Our
224 key idea of selecting valid pQTL instruments is that, under the plurality rule condition,
225 valid IVs will form the largest group and should give “similar” ratio estimates according
226 to the Anna Karenina Principle (see STAR Methods). More specifically, we propose the
227 following two criteria to measure the similarity between the ratio estimates of two
228 pQTLs j and k in step (2):

229 **C1** We say the k th pQTL “votes for” the j th pQTL to be a valid IV if, by
230 assuming the j th pQTL is valid, the k th pQTL's degree of violation of
231 assumptions (A2) and (A3) is smaller than a data-dependent threshold as in
232 equation (4);

233 **C2** We say the ratio estimates of two pQTLs j and k are “similar” if they
234 mutually vote for each other to be valid.

235

236 In step (3), we construct a symmetric binary voting matrix to encode the votes that each
237 relevant pQTL receives from other relevant pQTLs: the (k, j) entry of the voting
238 matrix is 1 if pQTLs j and k mutually vote for each other to be valid, and 0 otherwise.

239 We propose two ways to select valid pQTL IVs based on the voting matrix (see STAR
240 Methods): (1) select relevant pQTLs who receive majority voting or plurality voting as
241 valid IVs; and (2) use pQTLs in the maximum clique of the voting matrix as valid IVs⁴⁷.

242 Our simulation studies show that the maximum clique method can empirically offer
243 lower false discovery rate (FDR)⁴⁸ and higher true positive proportion (TPP) as shown
244 in Table S4 and Supplementary Section S6.

245

246 In step (4), we estimate the causal effect by fitting a zero-intercept ordinary least
247 squares regression of pQTL-outcome associations on pQTL-protein associations using
248 the set of selected valid pQTL IVs, and then construct a standard confidence interval
249 for the causal effect using standard linear regression theory. In finite samples, some
250 invalid IVs with small (but still nonzero) degrees of violation of assumptions (A2) and

251 (A3) might be incorrectly selected as valid IVs, commonly referred to as “locally
252 invalid IVs”⁴⁹. To address this possible issue, we propose to construct a robust
253 confidence interval with a guaranteed nominal coverage even in the presence of IV
254 selection error in finite-sample settings using a searching and sampling method⁴⁹, as
255 described in Supplementary Figure S17 and STAR Methods.

256

257 **2.3 Comparing MR-SPI to other competing MR methods in simulation studies**

258 We conduct extensive simulations to evaluate the performance of MR-SPI in the
259 presence of invalid IVs. We simulate data in a two-sample setting under four setups:
260 (S1) majority rule condition holds, and no locally invalid IVs exist; (S2) plurality rule
261 condition holds, and no locally invalid IVs exist; (S3) majority rule condition holds,
262 and locally invalid IVs exist; (S4) plurality rule condition holds, and locally invalid IVs
263 exist. More detailed simulation settings are described in STAR Methods. We compare
264 MR-SPI to the following competing MR methods: (1) the random-effects IVW
265 method²⁷, (2) MR-RAPS¹⁶, (3) MR-PRESSO²⁹, (4) the weighted median method³⁰, (5)
266 the mode-based estimation³¹, (6) MRMix³², and (7) the contamination mixture
267 method³³. We exclude MR-Egger in this simulation since it is heavily biased in our
268 simulation settings. For simplicity, we shall use IVW to represent the random-effects
269 IVW method hereafter.

270

271 In Figure 3, we present the percent bias, empirical coverage, and average lengths of 95%
272 confidence intervals of those MR methods in simulated data with a sample size of 5,000
273 for both the exposure and the outcome. Additional simulation results under a range of
274 sample sizes ($n=5,000, 10,000, 20,000, 40,000, 80,000$) can be found in Supplementary
275 Figure S1 and Tables S1-S3. When the plurality rule condition holds and no locally
276 invalid IVs exist, MR-SPI has small bias and short confidence interval, and the
277 empirical coverage can attain the nominal level. When locally invalid IVs exist, the
278 standard confidence interval might suffer from finite-sample IV selection error, and
279 thus the empirical coverage is lower than 95% if the sample sizes are not large (e.g.,
280 5,000). In practice, we can perform sensitivity analysis of the causal effect estimate by
281 changing the threshold in the voting step (see STAR Methods and Supplementary
282 Figure S14). If the causal effect estimate is sensitive to the choice of the threshold, then
283 there might exist finite-sample IV selection error. In such cases, the proposed robust
284 confidence interval of MR-SPI can still attain the 95% coverage level and thus is
285 recommended for use. We also examine the performance of MR-SPI in overlapped

286 samples mimicking real data settings with simulation set-up and results given in
287 Supplementary Section S8, and we find that our MR-SPI can still provide valid
288 statistical inference.

289

290 **2.4 Identifying plasma proteins associated with the risk of Alzheimer's disease**

291 Omics MR (xMR) aims to identify omics biomarkers (e.g., proteins) causally associated
292 with complex traits and diseases. In particular, xMR with proteomics data enables the
293 identification of disease-associated proteins, facilitating crucial advancements in
294 disease diagnosis, monitoring, and novel drug target discovery. In this section, we apply
295 MR-SPI to identify putative causal plasma protein biomarkers associated with the risk
296 of Alzheimer's disease (AD). The proteomics data used in our analysis comprises
297 54,306 participants from the UK Biobank Pharma Proteomics Project (UKB-PPP)²⁰. In
298 the UKB-PPP consortium, up to 22.6 million imputed autosomal variants across 1,463
299 proteins post quality control were analyzed, discovering 10,248 primary associations
300 through LD (Linkage Disequilibrium) clumping $\pm 1\text{Mb}$ around the significant variants,
301 including 1,163 in the *cis* region and 9,085 in the *trans* region²⁰. As described in Sun,
302 et al.²⁰, the following filtering steps are used to retain pQTLs in the UKB-PPP summary
303 level proteomics data: (1) genome-wide significant ($p\text{-value} < 3.40 \times 10^{-11}$), after
304 Bonferroni correction; and (2) independent pQTLs using LD clumping ($r^2 < 0.01$).
305 Thus, all these candidates pQTL IVs are independent and strongly associated with the
306 proteins. Summary statistics for AD are obtained from a meta-analysis of GWAS
307 studies for clinically diagnosed AD and AD-by-proxy, comprising 455,258 samples in
308 total⁵⁰. For MR method comparison, we analyze 912 plasma proteins that share four or
309 more candidate pQTLs within the summary statistics for AD, because the
310 implementation of MR-PRESSO requires a minimum of four candidate IVs²⁹.

311

312 As presented in Figure 4(A), MR-SPI identifies 7 proteins that are significantly
313 associated with AD after Bonferroni correction, including CD33, CD55, EPHA1,
314 PILRA, PILRB, RET, and TREM2. The detailed information of the selected pQTL IVs
315 for these 7 proteins can be found in Supplementary Table S6. Among them, four
316 proteins (CD33, PILRA, PILRB, and RET) are positively associated with the risk of
317 AD while the other three proteins (CD55, EPHA1, and TREM2) are negatively
318 associated with the risk of AD. We also note that some competing MR methods may
319 detect additional proteins, which are likely spurious due to invalid pQTL IVs, as
320 demonstrated in Supplementary Section 11. Previous studies have revealed that some

321 of those 7 proteins and the corresponding protein-coding genes might contribute to the
322 pathogenesis of AD⁵¹⁻⁵⁶, as shown in Supplementary Table S8. For example, it has been
323 found that CD33 plays a key role in modulating microglial pathology in AD, with
324 TREM2 acting downstream in this regulatory pathway⁵³. Besides, a recent study has
325 shown that a higher level of soluble TREM2 is associated with protection against the
326 progression of AD pathology⁵⁷. Additionally, RET at mitochondrial complex I is
327 activated during ageing, which might contribute to an increased risk of ageing-related
328 diseases including AD⁵⁵. Using the UniProt database⁵⁸, we also find that genes encoding
329 these 7 proteins are overexpressed in tissues including hemopoietic tissues and brain,
330 as well as cell types including microglial, macrophages and dendritic cells. These
331 findings highlight the potential therapeutic opportunities that target these proteins for
332 the treatment of AD. Furthermore, in the Therapeutic Target Database (TTD)⁵⁹ and
333 DrugBank database⁶⁰, we find existing US Food and Drug Administration (FDA)-
334 approved drugs that target these proteins identified by MR-SPI. For example,
335 gemtuzumab ozogamicin is a drug that targets CD33 and has been approved by FDA
336 for acute myeloid leukemia therapy^{61,62}. Besides, pralsetinib and selpercatinib are two
337 RET inhibitors that have been FDA-approved for the treatment of non-small-cell lung
338 cancers^{63,64}. Therefore, these drugs might be potential drug repurposing candidates for
339 the treatment of AD.

340

341 In Figure 4(B), we present the 3D structural alterations of CD33 due to missense genetic
342 variation of pQTL rs2455069, as predicted by AlphaFold3^{42,43,45}. The 3D structures are
343 shown in blue when the allele is A, and in red when the allele is G at pQTL rs2455069
344 A/G, which is a cis-SNP located on chromosome 19 (19q13.41) and is selected as a
345 valid IV by MR-SPI. The presence of the G allele at pQTL rs2455069 results in the
346 substitution of the 69th amino acid of CD33, changing it from Arginine (colored in
347 green if the allele is A) to Glycine (colored in yellow if the allele is G), consequently
348 causing a local change in the structure of CD33 (R69G). Previous studies have found
349 that CD33 is overexpressed in microglial cells in the brain⁶⁵, and the substitution of
350 Arginine to Glycine in the 69th amino acid of CD33 might lead to the accumulation of
351 amyloid plaques in the brain⁶⁶, thus the presence of the G allele at pQTL rs2455069
352 might contribute to an increased risk of AD. We also apply AlphaFold3 to predict the
353 3D structures of the other proteins that are detected to be significantly associated with
354 AD by MR-SPI, which are presented in Supplementary Figure S16.

355

356 In Figure 4(C), we present the point estimates and 95% confidence intervals of the
357 causal effects (on the log odds ratio scale) of these 7 proteins on AD using the other
358 competing MR methods. In Figure 4(C), these proteins are identified by most of the
359 competing MR methods, confirming the robustness of our findings. To the best of our
360 knowledge, there may be two reasons for the differences in results between MRMix
361 and other MR methods for some proteins: (1) MRMix assumes that the pQTL-protein
362 and pQTL-outcome associations follow a bivariate normal-mixture model with four
363 mixture components while the contamination mixture models assume that the ratio
364 estimator follows a normal distribution with two mixture components, and therefore it
365 may be more challenging to obtain reliable causal effect estimates using the MRMix
366 model with a small number of pQTLs per protein; and (2) the default grid search values
367 implemented in the MRMix R package might not be optimal for some proteins. Notably,
368 MR-SPI detects one possibly invalid IV pQTL rs10919543 for TREM2-AD
369 relationship, which is associated with red blood cell count according to PhenoScanner⁶⁷.
370 Red blood cell count is a known risk factor for AD^{68,69}, and thus pQTL rs10919543
371 might exhibit pleiotropy in the relationship of TREM2 on AD. After excluding this
372 potentially invalid IV, MR-SPI suggests that TREM2 is negatively associated with the
373 risk of AD ($\hat{\beta} = -0.04$, $p\text{-value} = 1.20 \times 10^{-18}$). Additionally, we perform the gene
374 ontology (GO) enrichment analysis using the g:Profiler web server⁷⁰
375 (<https://biit.cs.ut.ee/gprofiler/gost>) to gain more biological insights for the 7 proteins
376 identified by MR-SPI, and the results are presented in Figure 4(D) and Supplementary
377 Table S7. After Bonferroni correction, the GO analysis indicates that these 7 proteins
378 are significantly enriched in 20 GO terms, notably, the positive regulation of
379 phosphorus metabolic process and major histocompatibility complex (MHC) class I
380 protein binding. It has been found that increased phosphorus metabolites (e.g.,
381 phosphocreatine) are associated with aging, and that defects in metabolic processes for
382 phospholipid membrane function is involved in the pathological progression of
383 Alzheimer's disease^{71,72}. In addition, MHC class I proteins may play a crucial role in
384 preserving brain integrity during post-developmental stages, and modulation of the
385 stability of MHC class I proteins emerges as a potential therapeutic target for restoring
386 synaptic function in AD⁷³⁻⁷⁵.

387

388 **3. Discussion**

389 In this paper, we develop a novel integrated pipeline that combines our proposed MR-
390 SPI method with AlphaFold3 to identify putative causal protein biomarkers for complex

391 traits/diseases and to predict the 3D structural alterations induced by missense pQTL
392 IVs. Specifically, MR-SPI is an automatic algorithm to select valid pQTL IVs under the
393 plurality rule condition for a specific protein-outcome pair from two-sample GWAS
394 summary statistics. MR-SPI first selects relevant pQTL IVs with strong pQTL-protein
395 associations to minimize weak IV bias, and then applies the proposed voting procedure
396 to select valid pQTL IVs whose ratio estimates are similar to each other. In the possible
397 presence of locally invalid IVs in finite-sample settings, MR-SPI further provides a
398 robust confidence interval constructed by the searching and sampling method⁴⁹, which
399 is immune to finite-sample IV selection error. The valid pQTL IVs selected by MR-SPI
400 serve dual purposes: (1) facilitating more reliable scientific discoveries in identifying
401 putative causal proteins associated with diseases; and (2) shedding new light on the
402 molecular-level mechanism of causal proteins in disease etiology through the 3D
403 structural alterations of mutated proteins induced by missense pQTL IVs. We employ
404 MR-SPI to conduct xMR analysis with 912 plasma proteins using the proteomics data
405 in 54,306 UK Biobank participants and identify 7 proteins significantly associated with
406 the risk of Alzheimer's disease. The 3D structural changes in these proteins, as predicted
407 by AlphaFold3 in response to missense genetic variations of selected pQTL IVs,
408 offering new insights into their biological functions in the etiology of Alzheimer's
409 disease. We also found existing FDA-approved drugs that target some of our identified
410 proteins, which provide opportunities for potential existing drug repurposing for the
411 treatment of Alzheimer's disease. These findings highlight the great potential of our
412 proposed pipeline for identifying protein biomarkers as new therapeutic targets and
413 drug repurposing for disease prevention and treatment.

414

415 We emphasize three main advantages of MR-SPI. First, MR-SPI incorporates both
416 proteomics and outcome data to automatically select a set of valid pQTL instruments
417 in genome-wide studies, and the selection procedure does not rely on any additional
418 distributional assumptions on the genetic effects nor require a large number of candidate
419 IVs. Therefore, MR-SPI is the first method to offer such a practically robust approach
420 to selecting valid pQTL IVs for a specific exposure-outcome pair from GWAS studies
421 for more reliable MR analyses, which is especially advantageous in the presence of
422 wide-spread horizontal pleiotropy and when only a small number of candidate IVs are
423 available in xMR studies. While our real data application specifically focuses on the
424 identification of putative causal protein biomarkers for Alzheimer's disease through the
425 integration of MR-SPI with AlphaFold3, it's important to highlight that MR-SPI holds

426 broader applicability in elucidating causal relationships across complex traits and
427 diseases. For additional data analysis results and insights into the utility of MR-SPI in
428 this context, please refer to Supplementary Sections S9 and S10. Second, we propose a
429 robust confidence interval for the causal effect using the searching and sampling
430 method, which is immune to finite-sample IV selection error. Therefore, when locally
431 invalid IVs are incorrectly selected, MR-SPI can still provide reliable statistical
432 inference for the causal effect using the proposed robust confidence interval. Third,
433 MR-SPI is computationally efficient. The average computation time for constructing
434 the standard CI and the robust CI with 20 candidate IVs is 0.02 seconds and 10.60
435 seconds, respectively, using a server equipped with an Intel Xeon Silver 4116 CPU and
436 64 GB RAM memory.

437

438 In conclusion, MR-SPI is a powerful tool for identifying putative causal protein
439 biomarkers for complex traits and diseases. The integration of MR-SPI with
440 AlphaFold3 as a computationally efficient pipeline can further predict the 3D structural
441 alterations caused by missense pQTL IVs, improving our understanding of molecular-
442 level disease mechanisms. Therefore, our pipeline holds promising implications for
443 drug target discovery, drug repurposing, and therapeutic development.

444

445 **Limitations of the Study**

446 MR-SPI has some limitations. First, MR-SPI uses independent pQTLs as candidate IVs
447 after LD clumping, which might exclude strong and valid pQTL IVs. We plan to extend
448 MR-SPI to include correlated pQTLs with arbitrary LD structure to increase statistical
449 power. Second, the proposed robust confidence interval is slightly more conservative,
450 which is the price to pay for the gained robustness to finite-sample IV selection error.
451 We plan to construct less conservative confidence intervals with improved power to
452 detect more putative causal proteins. Third, we will incorporate colocalization
453 analysis⁷⁶⁻⁷⁹ into our pipeline to better understand the shared genetic architecture
454 between proteins and disease outcomes when unfiltered GWAS summary statistics are
455 available in future studies.

456

457 **4. Resource Availability**

458 **Lead Contact**

459 Further information and requests for resources and reagents should be directed to and
460 will be fulfilled by the lead contact, Zhonghua Liu (zl2509@cumc.columbia.edu)

461

462 **Materials Availability**

463 The materials that support the findings of this study are available from the
464 corresponding authors upon reasonable request. Please contact the lead contact,
465 Zhonghua Liu (zl2509@cumc.columbia.edu) for additional information.

466

467 **Data and Code Availability**

468 All the GWAS data analyzed are publicly available with the following URLs:

- 469 ● GWAS for Alzheimer's disease: https://ctg.cncr.nl/software/summary_statistics;
- 470 ● UK Biobank proteomics data:
471 [https://www.biorxiv.org/content/10.1101/2022.06.17.496443v1.supplementary-](https://www.biorxiv.org/content/10.1101/2022.06.17.496443v1.supplementary-material)
472 [material](https://www.biorxiv.org/content/10.1101/2022.06.17.496443v1.supplementary-material)

473 The R package **MR.SPI** is publicly available at [https://github.com/MinhaoYao00/MR-](https://github.com/MinhaoYao00/MR-SPI)
474 [SPI](https://github.com/MinhaoYao00/MR-SPI).

475

476 **5. Acknowledgments**

477 We thank the UK Biobank Pharma Proteomics Project (<http://ukb-ppp.gwas.eu/>) for
478 providing the proteomics data. We thank the Center for Neurogenomics and Cognitive
479 Research (CNCR, <https://cncr.nl/>) for providing GWAS summary statistics for
480 Alzheimer's disease.

481

482 **6. Author Contributions**

483 Conceptualization, Zijian Guo and Zhonghua Liu; methodology, Minhao Yao, Zijian
484 Guo and Zhonghua Liu; software, Minhao Yao; formal analysis: Minhao Yao; writing
485 (original draft), Minhao Yao, Zijian Guo and Zhonghua Liu; writing (review &
486 editing), Minhao Yao, Gary W. Miller, Badri N. Vardarajan, Andrea A. Baccarelli,
487 Zijian Guo and Zhonghua Liu; supervision, Zijian Guo and Zhonghua Liu.

488

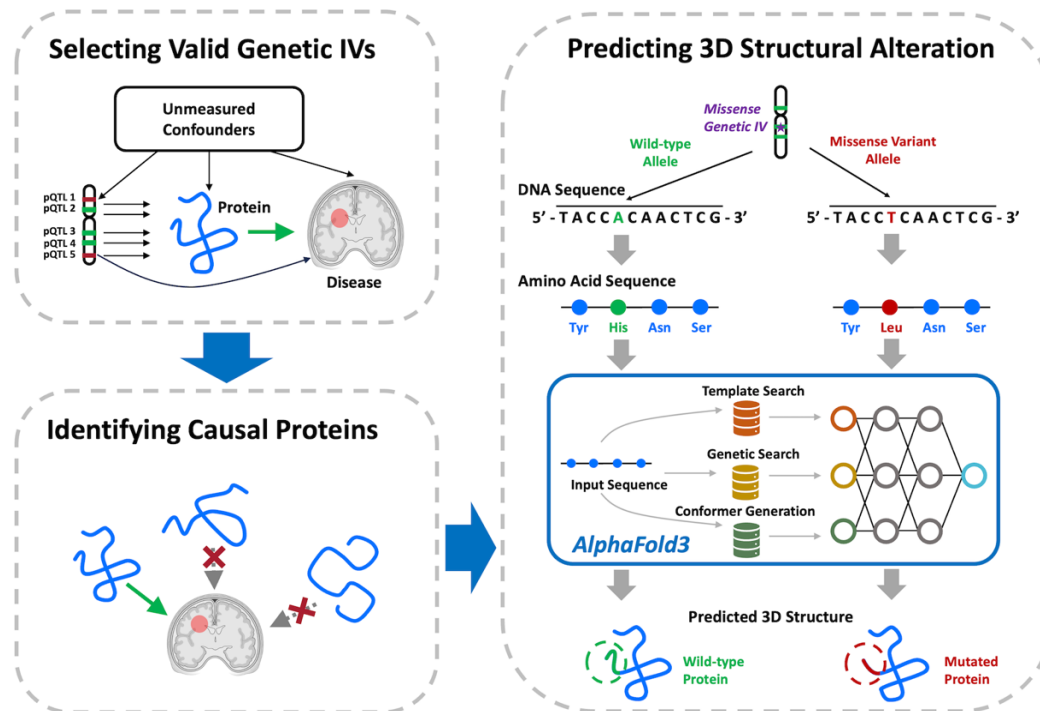
489 **7. Declaration of Interests**

490 The authors declare no competing interests.

491

492 **8. Main Figure Titles and Legends**

493



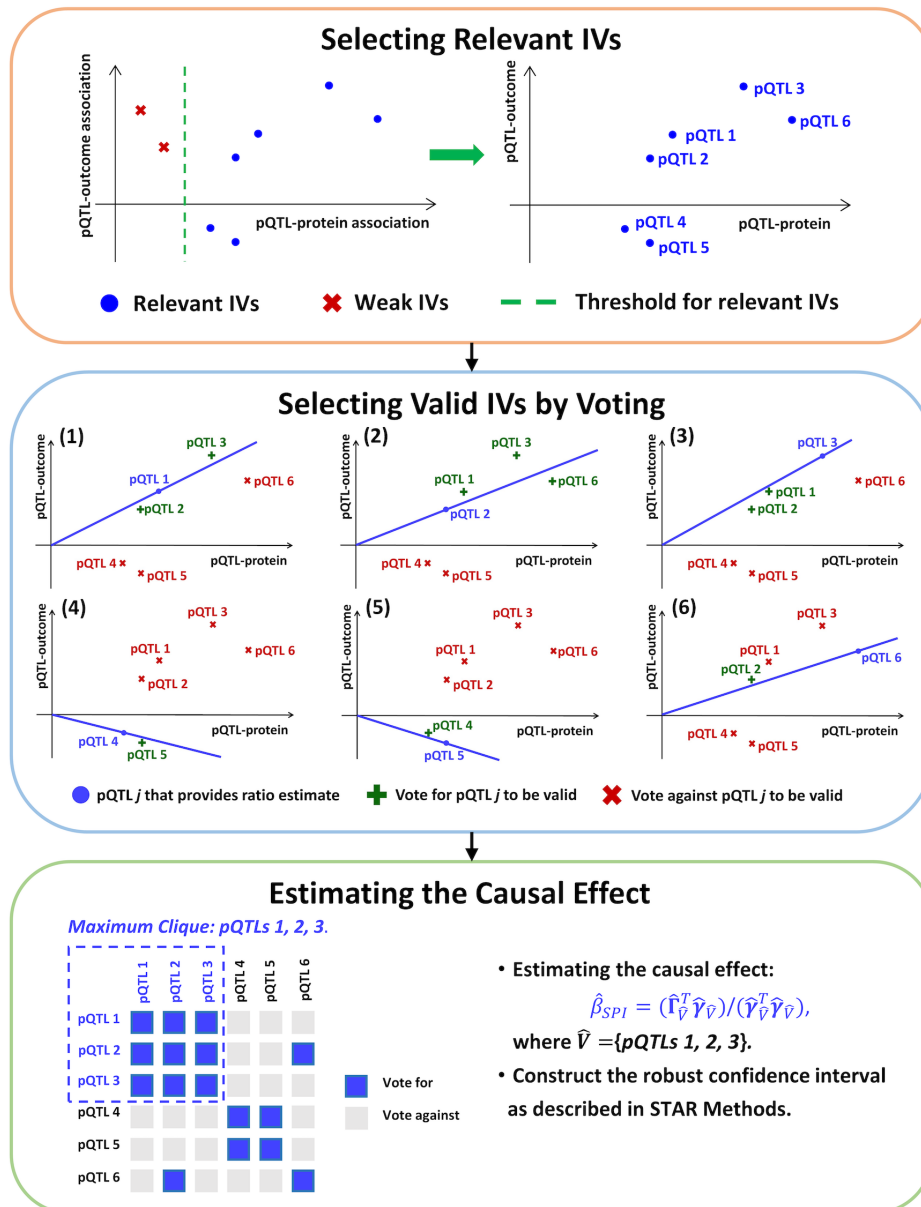
494

495 **Main Figure 1. Overview of the pipeline.**

496 First, we apply MR-SPI for each protein to (1) select valid pQTL IVs under the
497 plurality condition, and (2) estimate the causal effect on the outcome of interest.
498 Second, we perform the Bonferroni correction procedure for causal protein
499 identification. Third, for each causal protein biomarker, we apply AlphaFold3 to
500 predict the 3D structural alterations due to missense pQTL IVs selected by MR-SPI.

501

502

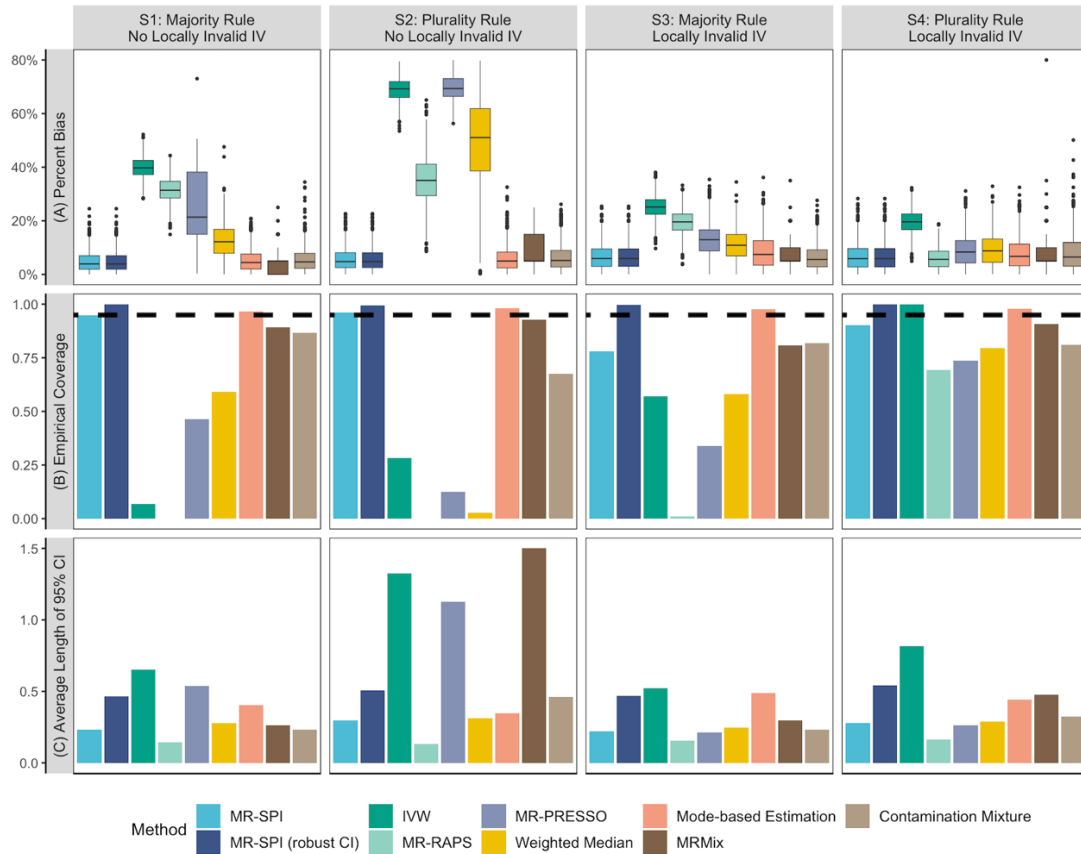


503

504 **Main Figure 2. The MR-SPI framework.**

505 First, MR-SPI selects relevant IVs with strong pQTL-protein associations. Second,
 506 each relevant IV provides a ratio estimate of the causal effect and then receives votes
 507 on itself to be valid from the other relevant IVs whose degrees of violation of (A2)
 508 and (A3) are small under this ratio estimate of causal effect. For example, by
 509 assuming pQTL 1 is valid, the slope of the line connecting pQTL 1 and the origin
 510 represents the ratio estimate of pQTL 1, and pQTLs 2 and 3 vote for pQTL 1 to be
 511 valid because they are close to that line, while pQTLs 4, 5 and 6 vote against it since
 512 they are far away from that line. Third, MR-SPI estimates the causal effect by fitting a
 513 zero-intercept OLS regression of pQTL-outcome associations on pQTL-protein
 514 associations and construct the robust confidence interval using selected valid pQTL
 515 IVs in the maximum clique of the voting matrix, which encodes whether two pQTLs
 516 mutually vote for each other to be valid IVs.

517



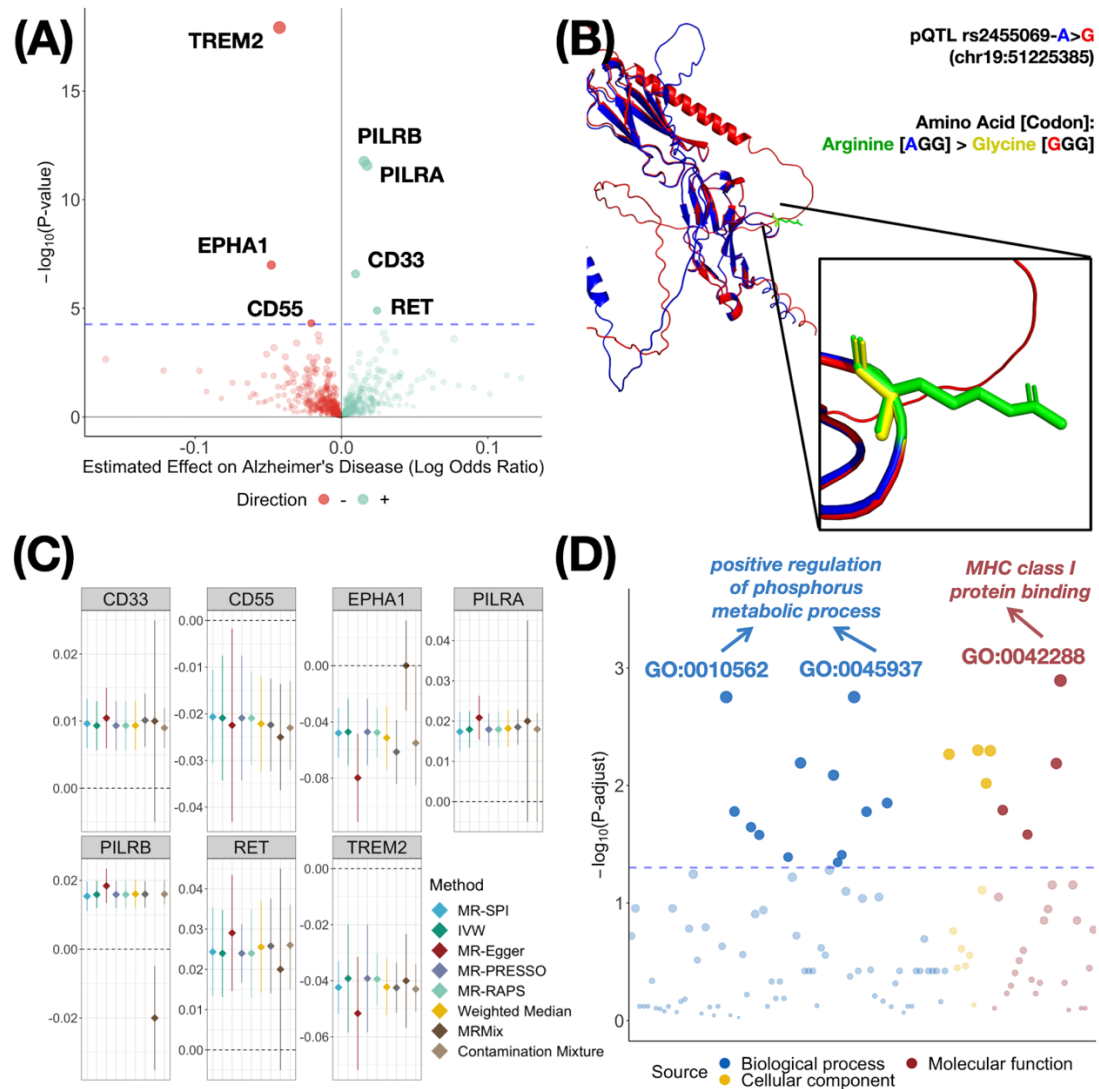
518

519 **Main Figure 3. Empirical performance of MR-SPI and the other competing MR**
 520 **methods in simulated data with sample size 5,000.**

521 (A) Boxplot of the percent bias in causal effect estimates. (B) Empirical coverage of
 522 95% confidence intervals. The black dashed line in (B) represents the nominal level
 523 (95%). (C) Average lengths of 95% confidence intervals.

524

525



526

527 **Main Figure 4. Causal effect estimates, gene ontology analysis and structural**
528 **alteration prediction for putatively causal proteins.**

529 (A) Volcano plot of associations of plasma proteins with Alzheimer's disease using MR-
530 SPI. The horizontal axis represents the estimated effect size (on the log odds ratio scale),
531 and the vertical axis represents the $-\log_{10}(p\text{-value})$. Positive and negative
532 associations are represented by green and red points, respectively. The size of a point is
533 proportional to $-\log_{10}(p\text{-value})$. The blue dashed line represents the significance
534 threshold using Bonferroni correction ($p\text{-value} < 5.48 \times 10^{-5}$). (B) 3D Structural
535 alterations of CD33 predicted by AlphaFold3 due to missense genetic variation of pQTL
536 rs2455069. The ribbon representation of 3D structures of CD33 with Arginine and
537 Glycine at position 69 are colored in blue and red, respectively. The amino acids at
538 position 69 are displayed in stick representation, with Arginine and Glycine colored in
539 green and yellow, respectively. The predicted template modeling (pTM) yields a score
540 of 0.6 for both structures, which suggests that AlphaFold3 provides good predictions
541 for these two 3D structures. (C) Forest plot of significant associations of proteins with
542 Alzheimer's disease identified by MR-SPI. Confidence intervals are clipped to vertical
543 axis limits. (D) Bubble plot of GO analysis results using the 7 significant proteins

544 detected by MR-SPI. The horizontal axis represents the z -score of the enriched GO
 545 term, and the vertical axis represents the $-\log_{10}(p\text{-value})$ after Bonferroni correction.
 546 Each point represents one enriched GO term. The blue dashed line represents the
 547 significance threshold (adjusted $p\text{-value} < 0.05$ after Bonferroni correction).

548

549

550 9. Main Tables and Legends

551

552 Main Table 1: Comparison of MR methods and the underlying assumptions for
 553 handling invalid IVs. Balanced pleiotropy means on average the pleiotropic effects
 554 have zero mean. NOME assumption refers to NO Measurement Error in the exposure
 555 data.

MR Method	InSIDE assumption	Majority/plurality rule condition	Other distributional assumptions
Random-effects IVW	Required	Not required	Balanced pleiotropy, NOME assumption
MR-Egger	Required	Not required	NOME assumption
MR-RAPS	Required	Not required	Balanced pleiotropy
MR-PRESSO	Required	Majority rule condition	Not required
Weighed median	Not required	Majority rule condition	Not required
Mode-based estimation	Not required	Plurality rule condition	Not required
MRMix	Not required	Plurality rule condition	Direct effects of a SNP on the exposure and outcome follow a mixture of normal distributions
Contamination mixture method	Not required	Plurality rule condition	Ratio estimate of a SNP follow a mixture of normal distributions, NOME assumption
MR-SPI	Not required	Plurality rule condition	Not required

556

557

558

559

560

561

562 **STAR Methods**

563 **Key resources table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UK Biobank proteomics data	Sun, et al. ²⁰	http://ukb-ppp.gwas.eu/
Genome-wide association studies of Alzheimer's disease	Jansen, et al. ⁵⁰	https://www.nature.com/articles/s41588-018-0311-9
Software and algorithms		
AlphaFold3	Abramson, et al. ⁴⁵	https://alphafoldserver.com
g:Profiler	Raudvere, et al. ⁷⁰	https://biit.cs.ut.ee/gprofiler/gost
MendelianRandomization	Yavorska and Burgess ⁸⁰	https://github.com/cran/MendelianRandomization
MRMix	Qi and Chatterjee ³²	https://github.com/gqi/MRMix
MR-PRESSO	Verbanck, et al. ²⁹	https://github.com/rondolab/MR-PRESSO
MR-RAPS	Zhao, et al. ¹⁶	https://github.com/qingyuanzhao/mr.raps
MR-SPI	This study	https://github.com/MinhaoYaooo/MR-SPI
PLINK	Purcell, et al. ⁸¹	https://www.cog-genomics.org/plink/1.9
PyMol	PyMOL Molecular Graphics	https://pymol.org

	System	
R	The R Foundation for Statistical Computing	https://www.r-project.org/

564

565

566 **Method details**

567 *Two-sample GWAS summary statistics*

568 Suppose that we obtain p independent pQTLs $\mathbf{Z} = (Z_1, \dots, Z_p)^\top$ by using LD
569 clumping that retains one representative pQTL per LD region⁸¹. We also assume that
570 the pQTLs are standardized⁸² such that $\mathbb{E}Z_j = 0$ and $\text{Var}(Z_j) = 1$ for $1 \leq j \leq p$.

571 Let D denote the exposure and Y denote the outcome. We assume that D and Y
572 follow the exposure model $D = \mathbf{Z}^\top \boldsymbol{\gamma} + \delta$ and the outcome model $Y = D\beta + \mathbf{Z}^\top \boldsymbol{\pi} +$

573 e , respectively, where β represents the causal effect of interest, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^\top$

574 represents the IV strength, and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)^\top$ encodes the violation of

575 assumptions (A2) and (A3)^{83,84}. If assumptions (A2) and (A3) hold for pQTL j , then

576 $\pi_j = 0$ and otherwise $\pi_j \neq 0$ (see Supplementary Section S1 for details). The error

577 terms δ and e with respective variances σ_δ^2 and σ_e^2 are possibly correlated due to

578 unmeasured confounding factors. By plugging the exposure model into the outcome

579 model, we obtain the reduced-form outcome model $Y = \mathbf{Z}^\top (\beta\boldsymbol{\gamma} + \boldsymbol{\pi}) + \epsilon$, where $\epsilon =$

580 $\beta\delta + e$. Let $\boldsymbol{\Gamma} = (\Gamma_1, \dots, \Gamma_p)^\top$ denote the pQTL-outcome associations, then we have

581 $\boldsymbol{\Gamma} = \beta\boldsymbol{\gamma} + \boldsymbol{\pi}$. If $\gamma_j \neq 0$, then pQTL j is called a relevant IV. If both $\gamma_j \neq 0$ and $\pi_j =$

582 0 , then pQTL j is called a valid IV. Let $\mathcal{S} = \{j: \gamma_j \neq 0, 1 \leq j \leq p\}$ denote the set of

583 all relevant IVs, and $\mathcal{V} = \{j: \gamma_j \neq 0 \text{ and } \pi_j = 0, 1 \leq j \leq p\}$ denote the set of all valid

584 IVs. The majority rule condition can be expressed as $|\mathcal{V}| > \frac{1}{2}|\mathcal{S}|$ ⁸⁴, and the plurality

585 rule condition can be expressed as $|\mathcal{V}| > \max_{c \neq 0} |\{j \in \mathcal{S}: \pi_j/\gamma_j = c\}|$ ⁸³. If the

586 plurality rule condition holds, then valid IVs with the same ratio of pQTL-outcome

587 effect to pQTL-protein effect will form a plurality. Based on this key observation, our

588 proposed MR-SPI selects the largest group of pQTLs as valid IVs with similar ratio

589 estimates of the causal effect using a voting procedure described in detail in the next
590 subsection.

591 Let $\hat{\gamma}_j$ and $\hat{\Gamma}_j$ be the estimated marginal effects of pQTL j on the protein and the
592 outcome, and $\hat{\sigma}_{\gamma_j}$ and $\hat{\sigma}_{\Gamma_j}$ be the corresponding estimated standard errors
593 respectively. Let $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)^\top$ and $\hat{\boldsymbol{\Gamma}} = (\hat{\Gamma}_1, \dots, \hat{\Gamma}_p)^\top$ denote the vector of
594 estimated pQTL-protein and pQTL-outcome associations, respectively. In the two-
595 sample setting, the summary statistics $\{\hat{\gamma}_j, \hat{\sigma}_{\gamma_j}\}_{1 \leq j \leq p}$ and $\{\hat{\Gamma}_j, \hat{\sigma}_{\Gamma_j}\}_{1 \leq j \leq p}$ are
596 calculated from two non-overlapping samples with sample sizes n_1 and n_2
597 respectively. When all the pQTLs are independent of each other, the joint asymptotic
598 distribution of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Gamma}}$ is

$$599 \begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \\ \hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma} \end{pmatrix} \xrightarrow{d} N \left[\mathbf{0}, \begin{pmatrix} \frac{1}{n_1} \mathbf{V}_\gamma & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} \mathbf{V}_\Gamma \end{pmatrix} \right]$$

600 where the diagonal entries of \mathbf{V}_γ and \mathbf{V}_Γ are $\mathbf{V}_{\gamma,jj} = \text{Var}(Z_{ij}^2)\gamma_j^2 + \sum_{l \neq j} \gamma_l^2 + \sigma_\delta^2$
601 and $\mathbf{V}_{\Gamma,jj} = \text{Var}(Z_{ij}^2)\Gamma_j^2 + \sum_{l \neq j} \Gamma_l^2 + \sigma_\epsilon^2$, respectively, and the off-diagonal entries of
602 \mathbf{V}_γ and \mathbf{V}_Γ are $\mathbf{V}_{\gamma,j_1j_2} = \gamma_{j_1}\gamma_{j_2}$ and $\mathbf{V}_{\Gamma,j_1j_2} = \Gamma_{j_1}\Gamma_{j_2}$ ($j_1 \neq j_2$), respectively. The
603 derivation of the limit distribution can be found in Supplementary Section S2. Therefore,
604 with the summary statistics of the protein and the outcome, we estimate the covariance
605 matrices $\frac{1}{n_1} \mathbf{V}_\gamma$ and $\frac{1}{n_2} \mathbf{V}_\Gamma$ as:

$$606 \frac{1}{n_1} \hat{\mathbf{V}}_{\gamma,j_1j_2} = \begin{cases} \hat{\sigma}_{\gamma_{j_1}}^2 & \text{if } j_1 = j_2, \\ \frac{1}{n_1} \hat{\gamma}_{j_1} \hat{\gamma}_{j_2} & \text{if } j_1 \neq j_2. \end{cases} \text{ and } \frac{1}{n_2} \hat{\mathbf{V}}_{\Gamma,j_1j_2} = \begin{cases} \hat{\sigma}_{\Gamma_{j_1}}^2 & \text{if } j_1 = j_2 \\ \frac{1}{n_2} \hat{\Gamma}_{j_1} \hat{\Gamma}_{j_2} & \text{if } j_1 \neq j_2 \end{cases} \quad (1)$$

607 After obtaining $\{\hat{\boldsymbol{\gamma}}, \hat{\mathbf{V}}_\gamma, \hat{\boldsymbol{\Gamma}}, \hat{\mathbf{V}}_\Gamma\}$, we then perform the proposed IV selection procedure as
608 illustrated in Figure 2 in the main text.

609

610 **Selecting valid instruments by voting**

611 The first step of MR-SPI is to select relevant pQTLs with large IV strength using

612 proteomics data. Specifically, we estimate the set of relevant IVs \mathcal{S} by:

$$613 \quad \hat{\mathcal{S}} = \left\{ 1 \leq j \leq p: \frac{|\hat{\gamma}_j|}{\hat{\sigma}_{\gamma_j}} > \Phi^{-1} \left(1 - \frac{\alpha^*}{2} \right) \right\} \quad (2)$$

614 where $\hat{\sigma}_{\gamma_j}$ is the standard error of $\hat{\gamma}_j$ in the summary statistics, $\Phi^{-1}(\cdot)$ is the
615 quantile function of the standard normal distribution, and α^* is the user-specified
616 threshold with the default value of 1×10^{-6} . This step is equivalent to filtering the
617 pQTLs in the proteomics data with p -value $< \alpha^*$, and is adopted by most of the
618 current two-sample MR methods to select (relevant) genetic instruments for
619 downstream MR analysis. Note that the selected pQTL instruments may not satisfy the
620 IV independence and exclusion restriction assumptions and thus maybe invalid. In
621 contrast, our proposed MR-SPI further incorporates the outcome data to automatically
622 select a set of valid genetic instruments from $\hat{\mathcal{S}}$ for a specific protein-outcome pair.

623 Under the plurality rule condition, valid pQTL instruments with the same ratio of
624 pQTL-outcome effect to pQTL-protein effect (i.e., Γ_j/γ_j) will form a plurality and
625 yield “similar” ratio estimates of the causal effect. Based on this key observation, MR-
626 SPI selects a plurality of relevant IVs whose ratio estimates are “similar” to each other
627 as valid IVs. Specifically, we propose the following two criteria to measure the
628 similarity between the ratio estimates of two pQTLs j and k :

629 **C1:** We say the k th pQTL “votes for” the j th pQTL to be a valid IV if, by assuming
630 the j th pQTL is valid, the k th pQTL’s degree of violation of assumptions (A2) and
631 (A3) is smaller than a threshold as in equation (4);

632 **C2:** We say the ratio estimates of two pQTLs j and k are “similar” if they mutually
633 vote for each other to be valid IVs.

634 The ratio estimate of the j th pQTL is defined as $\hat{\beta}^{[j]} = \hat{\Gamma}_j/\hat{\gamma}_j$. By assuming the j th
635 pQTL is valid, the plug-in estimate of the k th pQTL’s degree of violation of (A2) and
636 (A3) can be obtained by

$$637 \quad \hat{\pi}_k^{[j]} = \hat{\Gamma}_k - \hat{\beta}^{[j]}\hat{\gamma}_k = (\hat{\beta}^{[k]} - \hat{\beta}^{[j]})\hat{\gamma}_k \quad (3)$$

638 as we have $\Gamma_k = \beta\gamma_k + \pi_k$ for the true causal effect β , and $\hat{\Gamma}_k = \hat{\beta}^{[k]}\hat{\gamma}_k$ for the ratio

639 estimate $\hat{\beta}^{[k]}$ of the k th pQTL. From equation (3), $\hat{\pi}_k^{[j]}$ has two noteworthy
 640 implications. First, $\hat{\pi}_k^{[j]}$ measures the difference between the ratio estimates of pQTLs
 641 j and k (multiplied by the k th pQTL-protein effect estimate $\hat{\gamma}_k$), and a small $\hat{\pi}_k^{[j]}$
 642 implies that the difference scaled by $\hat{\gamma}_k$ is small. Second, $\hat{\pi}_k^{[j]}$ represents the k th IV's
 643 degree of violation of assumptions (A2) and (A3) by regarding the j th pQTL's ratio
 644 estimate $\hat{\beta}^{[j]}$ as the true causal effect, thus a small $\hat{\pi}_k^{[j]}$ implies a strong evidence that
 645 the k th IV supports the j th IV to be valid. Therefore, we say the k th IV votes for the
 646 j th IV to be valid if:

$$647 \quad \frac{|\hat{\pi}_k^{[j]}|}{\widehat{\text{SE}}(\hat{\pi}_k^{[j]})} \leq \sqrt{\log \min(n_1, n_2)} \quad (4)$$

648 where $\widehat{\text{SE}}(\hat{\pi}_k^{[j]})$ is the standard error of $\hat{\pi}_k^{[j]}$, which is given by:

$$649 \quad \widehat{\text{SE}}(\hat{\pi}_k^{[j]}) \\
 650 = \sqrt{\frac{1}{n_2} \left(\hat{\mathbf{V}}_{\Gamma, kk} + \left(\frac{\hat{\gamma}_k}{\hat{\gamma}_j} \right)^2 \hat{\mathbf{V}}_{\Gamma, jj} - 2 \frac{\hat{\gamma}_k}{\hat{\gamma}_j} \hat{\mathbf{V}}_{\Gamma, jk} \right) + \frac{1}{n_1} (\hat{\beta}^{[j]})^2 \left(\hat{\mathbf{V}}_{\gamma, kk} + \left(\frac{\hat{\gamma}_k}{\hat{\gamma}_j} \right)^2 \hat{\mathbf{V}}_{\gamma, jj} - 2 \frac{\hat{\gamma}_k}{\hat{\gamma}_j} \hat{\mathbf{V}}_{\gamma, jk} \right)} \quad (5)$$

651 and the term $\sqrt{\log \min(n_1, n_2)}$ in equation (4) ensures that the violation of (A2) and
 652 (A3) can be correctly detected with probability one as the sample sizes go to infinity,
 653 as shown in Supplementary Section S3.

654 For each relevant IV in $\hat{\mathcal{S}}$, we collect all relevant IVs' votes on whether it is a valid IV
 655 according to equation (4). Then we construct a voting matrix $\hat{\mathbf{\Pi}} \in \mathbb{R}^{|\hat{\mathcal{S}}| \times |\hat{\mathcal{S}}|}$ to
 656 summarize the voting results and evaluate the similarity of two pQTLs' ratio estimates
 657 according to criterion C2. Specifically, we define the (k, j) entry of $\hat{\mathbf{\Pi}}$ as:

$$658 \quad \hat{\Pi}_{k,j} = I \left(\max \left\{ \frac{|\hat{\pi}_k^{[j]}|}{\widehat{\text{SE}}(\hat{\pi}_k^{[j]})}, \frac{|\hat{\pi}_j^{[k]}|}{\widehat{\text{SE}}(\hat{\pi}_j^{[k]})} \right\} \leq \sqrt{\log \min(n_1, n_2)} \right) \quad (6)$$

659 where $I(\cdot)$ is the indicator function such that $I(A) = 1$ if event A happens and
 660 $I(A) = 0$ otherwise. From equation (6), we can see that the voting matrix $\hat{\mathbf{\Pi}}$ is

661 symmetric, and the entries of $\hat{\Pi}$ are binary: $\hat{\Pi}_{k,j} = 1$ represents pQTLs j and k
 662 vote for each other to be a valid IV, i.e., the ratio estimates of these two pQTLs are close
 663 to each other; $\hat{\Pi}_{k,j} = 0$ represents that they do not. For example, in Figure 2, $\hat{\Pi}_{1,2} =$
 664 1 since the ratio estimates of pQTLs 1 and 2 are similar, while $\hat{\Pi}_{1,4} = 0$ because the
 665 ratio estimates of pQTLs 1 and 4 differ substantially, as pQTLs 1 and 4 mutually “vote
 666 against” each other to be valid according to equation (4).

667 After constructing the voting matrix $\hat{\Pi}$, we select the valid IVs by applying
 668 majority/plurality voting or finding the maximum clique of the voting matrix⁴⁷. Let
 669 $\mathbf{VM}_k = \sum_{j \in \mathcal{S}} \hat{\Pi}_{k,j}$ be the total number of pQTLs whose ratio estimates are similar to
 670 pQTL k . For example, $\mathbf{VM}_1 = 3$ in Figure 2, since three pQTLs (including pQTL 1
 671 itself) yield similar ratio estimates to pQTL 1 according to criterion **C2**. A large \mathbf{VM}_k
 672 implies strong evidence that pQTL k is a valid IV, since we assume that valid IVs form
 673 a plurality of the relevant IVs. Let $\hat{\mathcal{V}}_M = \{k \in \hat{\mathcal{S}}: \mathbf{VM}_k > |\hat{\mathcal{S}}|/2\}$ denote the set of IVs
 674 with majority voting, and $\hat{\mathcal{V}}_P = \{k \in \hat{\mathcal{S}}: \mathbf{VM}_k = \max_{l \in \hat{\mathcal{S}}} \mathbf{VM}_l\}$ denote the set of IVs
 675 with plurality voting, then the union $\hat{\mathcal{V}} = \hat{\mathcal{V}}_M \cup \hat{\mathcal{V}}_P$ can be a robust estimate of \mathcal{V} in
 676 practice. Alternatively, we can also find the maximum clique in the voting matrix as an
 677 estimate of \mathcal{V} . A clique in the voting matrix is a group of IVs who mutually vote for
 678 each other to be valid, and the maximum clique is the clique with the largest possible
 679 number of IVs⁴⁷.

680 *Estimation and inference of the causal effect*

681 After selecting the set of valid pQTL instruments $\hat{\mathcal{V}}$, the causal effect β is estimated
 682 by

$$683 \quad \hat{\beta}_{\text{SPI}} = \frac{\hat{\mathbf{F}}_{\hat{\mathcal{V}}}^T \hat{\mathcal{Y}}_{\hat{\mathcal{V}}}}{\hat{\mathcal{Y}}_{\hat{\mathcal{V}}}^T \hat{\mathcal{Y}}_{\hat{\mathcal{V}}}} \quad (7)$$

684 where $\hat{\mathcal{Y}}_{\hat{\mathcal{V}}}$ and $\hat{\mathbf{F}}_{\hat{\mathcal{V}}}$ are the estimates of pQTL-protein associations and pQTL-outcome
 685 associations of the selected valid IVs in $\hat{\mathcal{V}}$, respectively. The MR-SPI estimator in
 686 equation (7) is the regression coefficient obtained by fitting a zero-intercept ordinary
 687 least squares regression of $\hat{\mathbf{F}}_{\hat{\mathcal{V}}}$ on $\hat{\mathcal{Y}}_{\hat{\mathcal{V}}}$. Since the pQTLs are standardized, the genetic

688 associations $\hat{\gamma}_j$ and $\hat{\Gamma}_j$ are scaled by $\sqrt{2f_j(1-f_j)}$ (compared to the genetic
 689 associations calculated using the unstandardized pQTLs, denoted by $\check{\gamma}_j$ and $\check{\Gamma}_j$),
 690 where f_j is the minor allele frequency of pQTL j . As $f_j(1-f_j)$ is approximately
 691 proportional to the inverse variance of $\check{\Gamma}_j$ when each pQTL IV explains only a small
 692 proportion of variance in the outcome⁸⁵, the MR-SPI estimator of the causal effect in
 693 equation (7) is approximately equal to the inverse-variance weighted estimator²⁷
 694 calculated with $\{\check{\gamma}_j, \check{\Gamma}_j\}_{j \in \hat{\mathcal{V}}}$.

695 Let $\alpha \in (0,1)$ be the significance level and $z_{1-\alpha/2}$ be the $(1-\alpha/2)$ -quantile of the
 696 standard normal distribution, then the $(1-\alpha)$ confidence interval for β is given by:

$$697 \quad \text{CI} = \left(\hat{\beta}_{\text{SPI}} - z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{SPI}})}, \hat{\beta}_{\text{SPI}} + z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{SPI}})} \right) \quad (8)$$

698 where $\widehat{\text{Var}}(\hat{\beta}_{\text{SPI}})$ is the estimated variance of $\hat{\beta}_{\text{SPI}}$, which can be found in
 699 Supplementary Section S4. As $\min\{n_1, n_2\} \rightarrow \infty$, we have $\mathbb{P} \left\{ \beta \in \left(\hat{\beta}_{\text{SPI}} - \right.$
 700 $\left. z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{SPI}})}, \hat{\beta}_{\text{SPI}} + z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{SPI}})} \right) \right\} \rightarrow 1 - \alpha$ under the plurality rule
 701 condition, as shown in Supplementary Section S5. Hence, MR-SPI provides a
 702 theoretical guarantee for the asymptotic coverage probability of the confidence interval
 703 under the plurality rule condition.

704 We summarize the proposed procedure of selecting valid IVs and constructing the
 705 corresponding confidence interval by MR-SPI in Algorithm 1.

706

707 ***A robust confidence interval via searching and sampling***

708 In finite-sample settings, the selected set of relevant IVs $\hat{\mathcal{S}}$ might include some invalid
 709 IVs whose degrees of violation of (A2) and (A3) are small but nonzero, and we refer to
 710 them as “locally invalid IVs”⁴⁹. When locally invalid IVs exist and are incorrectly
 711 selected into $\hat{\mathcal{V}}$, the confidence interval in equation (8) becomes unreliable, since its

712 validity (i.e., the coverage probability attains the nominal level) requires that the invalid
 713 IVs are correctly filtered out. In practice, we can multiply the threshold
 714 $\sqrt{\log \min(n_1, n_2)}$ in the right-hand side of equation (4) by a scaling factor η to
 715 examine whether the confidence interval calculated by equation (8) is sensitive to the
 716 choice of the threshold. If the confidence interval varies substantially to the choice of
 717 the scaling factor η , then there might exist finite-sample IV selection error especially
 718 with locally invalid IVs. We demonstrate this issue with two numerical examples
 719 presented in Supplementary Figure S14. Supplementary Figure S14(a) shows an
 720 example in which MR-SPI provides robust inference across different values of the
 721 scaling factor, while Supplementary Figure S14(b) shows an example that MR-SPI
 722 might suffer from finite-sample IV selection error, as the causal effect estimate and the
 723 corresponding confidence interval are sensitive to the choice of the scaling factor η .
 724 This issue motivates us to develop a more robust confidence interval.

725 To construct a confidence interval that is robust to finite-sample IV selection error, we
 726 borrow the idea of searching and sampling⁴⁹, with main steps described in
 727 Supplementary Figure S17. The key idea is to sample the estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$
 728 repeatedly from the following distribution:

$$729 \quad \begin{pmatrix} \hat{\boldsymbol{\gamma}}^{(m)} \\ \hat{\boldsymbol{\Gamma}}^{(m)} \end{pmatrix} \sim N \left[\begin{pmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\Gamma}} \end{pmatrix}, \begin{pmatrix} \frac{1}{n_1} \hat{\mathbf{V}}_{\boldsymbol{\gamma}} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} \hat{\mathbf{V}}_{\boldsymbol{\Gamma}} \end{pmatrix} \right], m = 1, \dots, M \quad (9)$$

730 where M is the number of sampling times (by default, we set $M = 1,000$). Since $\hat{\boldsymbol{\gamma}}$
 731 and $\hat{\boldsymbol{\Gamma}}$ follow distributions centered at $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, there exists m^* such that $\hat{\boldsymbol{\gamma}}^{(m^*)}$
 732 and $\hat{\boldsymbol{\Gamma}}^{(m^*)}$ are close enough to the true values $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$ when the number of sampling
 733 times M is sufficiently large, and thus the confidence interval obtained by using $\hat{\boldsymbol{\gamma}}^{(m^*)}$
 734 and $\hat{\boldsymbol{\Gamma}}^{(m^*)}$ instead of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Gamma}}$ might have a larger probability of covering β .

735 For each sampling, we construct the confidence interval by searching over a grid of β
 736 values such that more than half of the selected IVs in $\hat{\mathcal{V}}$ are detected as valid. As for
 737 the choice of grid, we start with the smallest interval $[L, U]$ that contains all the
 738 following intervals:

$$739 \quad \left(\hat{\beta}^{[j]} - \sqrt{\log \min(n_1, n_2) \widehat{\text{Var}}(\hat{\beta}^{[j]})}, \hat{\beta}^{[j]} + \sqrt{\log \min(n_1, n_2) \widehat{\text{Var}}(\hat{\beta}^{[j]})} \right) \text{ for } j \in \hat{\mathcal{V}} \quad (10)$$

740 where $\hat{\beta}^{[j]}$ is the ratio estimate of the j th pQTL IV, $\widehat{\text{Var}}(\hat{\beta}^{[j]}) = (\widehat{\mathbf{V}}_{\Gamma,jj}/n_2 +$
 741 $(\hat{\beta}^{[j]})^2 \widehat{\mathbf{V}}_{\gamma,jj}/n_1) / \hat{\gamma}_j^2$ is the variance of $\hat{\beta}^{[j]}$, and $\sqrt{\log \min(n_1, n_2)}$ serves the same
 742 purpose as in equation (4). Then we discretize $[L, U]$ into $\mathcal{B} = \{b_1, b_2, \dots, b_K\}$ as the
 743 grid set such that $b_1 = L, b_K = U$ and $|b_{k+1} - b_k| = n_{\min}^{-0.6}$ for $1 \leq k \leq K - 2$,
 744 where $n_{\min} = \min(n_1, n_2)$. We set the grid size $n_{\min}^{-0.6}$ so that the error caused by
 745 discretization is smaller than the parametric rate $n_{\min}^{-1/2}$.

746 For each grid value $b \in \mathcal{B}$ and sampling index $1 \leq m \leq M$, we propose an estimate
 747 of π_j by $\hat{\pi}_j^{(m)}(b) = (\hat{\Gamma}_j^{(m)} - b\hat{\gamma}_j^{(m)}) \cdot I(|\hat{\Gamma}_j^{(m)} - b\hat{\gamma}_j^{(m)}| \geq \lambda \hat{\rho}_j(b, \alpha))$ for $j \in \hat{\mathcal{V}}$,
 748 where $\hat{\rho}_j(b, \alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2|\hat{\mathcal{V}}|}\right) \sqrt{(\widehat{\mathbf{V}}_{\Gamma,jj}/n_2 + b^2 \widehat{\mathbf{V}}_{\gamma,jj}/n_1)}$ is a data-dependent
 749 threshold, $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard
 750 normal distribution, $\alpha \in (0, 1)$ is the significance level, and $\lambda = (\log \min(n_1, n_2) /$
 751 $M)^{\frac{1}{2|\hat{\mathcal{V}}|}}$ ($\lambda < 1$ when M is sufficiently large) is a scaling factor to make the
 752 thresholding more stringent so that the confidence interval in each sampling is shorter,
 753 as we will show shortly. Here, $\hat{\pi}_j^{(m)}(b) = 0$ indicates that the j th pQTL is detected as
 754 a valid IV in the m th sampling if we take $\{\hat{\gamma}^{(m)}, \hat{\Gamma}^{(m)}\}$ as the estimates of genetic
 755 associations and b as the true causal effect. Let $\hat{\boldsymbol{\pi}}_{\hat{\mathcal{V}}}^{(m)}(b) = (\hat{\pi}_j^{(m)}(b))_{j \in \hat{\mathcal{V}}}$, then we
 756 construct the m th sampling's pseudo confidence interval $\text{pCI}^{(m)}$ by searching for the
 757 smallest and largest $b \in \mathcal{B}$ such that more than half of pQTLs in $\hat{\mathcal{V}}$ are detected to be
 758 valid. Define $\beta_{\min}^{(m)} = \min\{b \in \mathcal{B}: \|\hat{\boldsymbol{\pi}}_{\hat{\mathcal{V}}}^{(m)}(b)\|_0 < |\hat{\mathcal{V}}|/2\}$ and $\beta_{\max}^{(m)} = \max\{b \in$
 759 $\mathcal{B}: \|\hat{\boldsymbol{\pi}}_{\hat{\mathcal{V}}}^{(m)}(b)\|_0 < |\hat{\mathcal{V}}|/2\}$, then the m th sampling's pseudo confidence interval is
 760 constructed as $\text{pCI}^{(m)} = (\beta_{\min}^{(m)}, \beta_{\max}^{(m)})$.

761 From the definitions of $\hat{\pi}_j^{(m)}(b)$ and $\text{pCI}^{(m)}$, we can see that, when λ is smaller,
 762 there will be fewer pQTLs in $\hat{\mathcal{V}}$ being detected as valid for a given $b \in \mathcal{B}$, which leads
 763 to fewer $b \in \mathcal{B}$ satisfying $\|\hat{\boldsymbol{\pi}}_{\hat{\mathcal{V}}}^{(m)}(b)\|_0 < |\hat{\mathcal{V}}|/2$, thus the pseudo confidence interval
 764 in each sampling will be shorter. If there does not exist $b \in \mathcal{B}$ such that the majority

765 of IVs in $\hat{\mathcal{V}}$ are detected as valid, we set $\text{pCI}^{(m)} = \emptyset$. Let $\mathcal{M} = \{1 \leq m \leq$
766 $M: \text{pCI}^{(m)} \neq \emptyset\}$ denote the set of all sampling indexes corresponding to non-empty
767 searching confidence intervals, then the proposed robust confidence interval is given
768 by:

$$769 \quad \text{CI}^{\text{robust}} = \left(\min_{m \in \mathcal{M}} \beta_{\min}^{(m)}, \max_{m \in \mathcal{M}} \beta_{\max}^{(m)} \right) \quad (11)$$

770 We summarize the procedure of constructing the proposed robust confidence interval
771 in Algorithm 2.

772

773 *Simulation settings*

774 We set the number of candidate IVs $p = 10$, as the average number of candidate pQTL
775 IVs for the plasma proteins in the UK Biobank proteomics data is around 7.4. We set
776 the sample sizes $n_1 = n_2 = 5,000, 10,000, 20,000, 40,000, \text{ or } 80,000$. We generate
777 the j th genetic instruments Z_j and X_j independently from a binomial distribution
778 $\text{Bin}(2, f_j)$, where $f_j \sim U(0.05, 0.50)$ is the minor allele frequency of pQTL j . Then
779 we generate the protein level $\mathbf{D} = (D_1, \dots, D_{n_1})^\top$ and the outcome $\mathbf{Y} = (Y_1, \dots, Y_{n_2})^\top$
780 according to the exposure model and the outcome model, respectively. Finally, we
781 calculate the genetic associations and their corresponding standard errors for the protein
782 and the outcome, respectively. As for the parameters, we fix the causal effect $\beta = 1$,
783 and we consider 4 settings for $\boldsymbol{\gamma} \in \mathbb{R}^p$ and $\boldsymbol{\pi} \in \mathbb{R}^p$:

784 **(S1)**: set $\boldsymbol{\gamma} = 0.2 \cdot (\mathbf{1}_5, -\mathbf{1}_5)^\top$ and $\boldsymbol{\pi} = 0.2 \cdot (\mathbf{0}_6, \mathbf{1}_4)^\top$.

785 **(S2)**: set $\boldsymbol{\gamma} = 0.2 \cdot (\mathbf{1}_5, -\mathbf{1}_5)^\top$ and $\boldsymbol{\pi} = 0.2 \cdot (\mathbf{0}_4, \mathbf{1}_3, -\mathbf{1}_3)^\top$.

786 **(S3)**: set $\boldsymbol{\gamma} = 0.2 \cdot (\mathbf{1}_5, -\mathbf{1}_5)^\top$ and $\boldsymbol{\pi} = 0.2 \cdot (\mathbf{0}_6, \mathbf{1}_2, 0.25, 0.25)^\top$.

787 **(S4)**: set $\boldsymbol{\gamma} = 0.2 \cdot (\mathbf{1}_5, -\mathbf{1}_5)^\top$ and $\boldsymbol{\pi} = 0.2 \cdot (\mathbf{0}_4, \mathbf{1}_2, 0.25, \mathbf{1}_2, -0.25)^\top$.

788 Settings **(S1)** and **(S3)** satisfy the majority rule condition, while **(S2)** and **(S4)** only
789 satisfy the plurality rule condition. In addition, **(S3)** and **(S4)** simulate the cases where
790 locally invalid IVs exist, as we shrink some of the pQTLs' violation degrees of
791 assumptions (A2) and (A3) down to 0.25 times in these two settings. In total, we run
792 1,000 replications in each setting.

793

794 ***Implementation of existing MR methods***

795 We compare the performance of MR-SPI with eight other MR methods in simulation
796 studies and real data analyses. These methods are implemented as follows:

- 797 • Random-effects IVW, MR-Egger, the weighted median method, the mode-
798 based estimation and the contamination mixture method are implemented in the
799 R package “MendelianRandomization” ([https://github.com/cran/](https://github.com/cran/MendelianRandomization)
800 [MendelianRandomization](https://github.com/cran/MendelianRandomization)). The mode-based estimation is run with “iteration
801 = 1000”. All other methods are run with the default parameters.
- 802 • MR-PRESSO is implemented in the R package “MR-PRESSO”
803 ([https://github.com/](https://github.com/rondolab/MR-PRESSO) rondolab/MR-PRESSO) with outlier test and distortion test.
- 804 • MR-RAPS is performed using the R package “mr.raps”
805 ([https://github.com/qingyuanzhao/](https://github.com/qingyuanzhao/mr.raps) mr.raps) with the default options.
- 806 • MRMix is run with the R package “MRMix” (<https://github.com/gqi/MRMix>)
807 using the default options.

808

Algorithm 1: Selecting pQTL IVs and Performing Causal Inference by MR-SPI

- Input:** Summary statistics of independent pQTLs $\{\hat{\gamma}_j, \hat{\sigma}_{\gamma_j}, \hat{\Gamma}_j, \hat{\sigma}_{\Gamma_j}\}_{1 \leq j \leq p}$;
Sample sizes n_1 for the proteomics data and n_2 for the outcome;
Threshold α^* for selecting relevant IVs; Significance level $\alpha \in (0,1)$.
- Output:** An estimate of the set of valid IVs $\hat{\mathcal{V}}$, the causal effect estimate $\hat{\beta}_{\text{SPI}}$ and the corresponding confidence interval CI.
- 1 Estimate the variance-covariance matrices $\hat{\mathbf{V}}_{\gamma}$ and $\hat{\mathbf{V}}_{\Gamma}$ as in equation (1);
 - 2 Select the set of relevant IVs $\hat{\mathcal{S}}$ as in equation (2);
 - 3 **for** $j \in \hat{\mathcal{S}}$ **do**
 - 4 Calculate $\hat{\beta}^{[j]} = \hat{\Gamma}_j / \hat{\gamma}_j$ and $\hat{\pi}_k^{[j]} = \hat{\Gamma}_k - \hat{\beta}^{[j]} \hat{\gamma}_k$ for $k \in \hat{\mathcal{S}}$;
 - 5 Each relevant IV $k \in \hat{\mathcal{S}}$ votes for the j th IV to be valid if $|\hat{\pi}_k^{[j]}| / \widehat{\text{SE}}(\hat{\pi}_k^{[j]}) \leq \sqrt{\log \min(n_1, n_2)}$;
 - 6 **end**
 - 7 Construct the symmetric voting matrix $\hat{\mathbf{\Pi}} \in \mathbb{R}^{|\hat{\mathcal{S}}| \times |\hat{\mathcal{S}}|}$ as in equation (6);
 - 8 Select the set of valid IVs $\hat{\mathcal{V}}$ by majority voting, plurality voting or finding the maximum clique in the voting matrix;
 - 9 Estimate the causal effect as in equation (7), and construct the corresponding confidence interval as in equation (8) using the selected valid IVs in $\hat{\mathcal{V}}$.
-

Algorithm 2: Constructing A Robust Confidence Interval via Searching and Sampling

Input: Summary statistics of independent pQTLs $\{\hat{\gamma}_j, \hat{\sigma}_{\gamma_j}, \hat{\Gamma}_j, \hat{\sigma}_{\Gamma_j}\}_{1 \leq j \leq p}$;
Sample sizes n_1 for the proteomics data and n_2 for the outcome;
Threshold α^* for selecting relevant IVs; Significance level $\alpha \in (0,1)$; Sampling number M .

Output: The robust confidence interval CI^{robust} .

- 1 Estimate the set of valid IVs $\hat{\mathcal{V}}$ as in Algorithm 1;
 - 2 Construct the initial interval $[L, U]$ as in equation (10) and obtain the corresponding grid set \mathcal{B} ;
 - 3 **for** $m \leftarrow 1$ to M **do**
 - 4 Sample $\hat{\gamma}^{(m)}$ and $\hat{\Gamma}^{(m)}$ from the distribution in equation (9);
 - 5 Calculate $\{\hat{\pi}_{\hat{\mathcal{V}}}^{(m)}(b)\}_{b \in \mathcal{B}}$ by $\hat{\pi}_j^{(m)}(b) = (\hat{\Gamma}_j^{(m)} - b\hat{\gamma}_j^{(m)}) \cdot I(|\hat{\Gamma}_j^{(m)} - b\hat{\gamma}_j^{(m)}| \geq \lambda \hat{\rho}_j(b, \alpha)), j \in \hat{\mathcal{V}}$;
 - 6 Construct $pCI^{(m)}$ by grid search over \mathcal{B} ;
 - 7 **end**
 - 8 Construct the robust confidence interval CI^{robust} as in equation (11).
-

References

1. Self, W.K. and Holtzman, D.M. (2023) Emerging diagnostics and therapeutics for Alzheimer disease. *Nature medicine* **29**, 2187-2199.
2. Nichols, E. *et al.* (2022) Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health* **7**, e105-e125.
3. Van Dyck, C.H. *et al.* (2023) Lecanemab in early Alzheimer's disease. *New England Journal of Medicine* **388**, 9-21.
4. Sims, J.R. *et al.* (2023) Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *Jama* **330**, 512-527.
5. Hardy, J.A. and Higgins, G.A. (1992) Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184-185.
6. Karran, E. and De Strooper, B. (2022) The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics. *Nature reviews Drug discovery* **21**, 306-318.
7. Khoury, R., Rajamanickam, J. and Grossberg, G.T. (2018) An update on the safety of current therapies for Alzheimer's disease: focus on rivastigmine. *Therapeutic advances in drug safety* **9**, 171-178.
8. Davey Smith, G. and Ebrahim, S. (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32**, 1--22.
9. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N. and Davey Smith, G. (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**, 1133--1163.
10. Davey Smith, G. and Hemani, G. (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* **23**, R89--R98.
11. Davey Smith, G. and Ebrahim, S. (2004) Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* **33**, 30--42.
12. Burgess, S., Butterworth, A. and Thompson, S.G. (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* **37**, 658--665.
13. Pierce, B.L. and Burgess, S. (2013) Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *American Journal of Epidemiology* **178**, 1177--1184.
14. Lawlor, D.A. (2016) Commentary: Two-sample Mendelian randomization: opportunities and challenges. *International Journal of Epidemiology* **45**, 908.
15. Slob, E.A.W. and Burgess, S. (2020) A comparison of robust Mendelian randomization methods using summary data. *Genetic Epidemiology* **44**, 313--329.
16. Zhao, Q., Wang, J., Hemani, G., Bowden, J. and Small, D.S. (2020) Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *The Annals of Statistics* **48**, 1742--1769.
17. Morrison, J., Knoblauch, N., Marcus, J.H., Stephens, M. and He, X. (2020) Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics* **52**, 740--747.

18. Cheng, Q., Zhang, X., Chen, L.S. and Liu, J. (2022) Mendelian randomization accounting for complex correlated horizontal pleiotropy while elucidating shared genetic etiology. *Nature Communications* **13**, 1--13.
19. Yao, C. *et al.* (2018) Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nature communications* **9**, 3268.
20. Sun, B.B. *et al.* (2022) Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. *BioRxiv*, 2022--06.
21. Didelez, V. and Sheehan, N. (2007) Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* **16**, 309--330.
22. Sanderson, E., Richardson, T.G., Hemani, G. and Davey Smith, G. (2021) The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *International Journal of Epidemiology* **50**, 1350--1361.
23. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. and Smoller, J.W. (2013) Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14**, 483--495.
24. Sivakumaran, S. *et al.* (2011) Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics* **89**, 607--618.
25. Parkes, M., Cortes, A., Van Heel, D.A. and Brown, M.A. (2013) Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nature Reviews Genetics* **14**, 661--673.
26. Schmidt, A.F. *et al.* (2020) Genetic drug target validation using Mendelian randomisation. *Nature Communications* **11**, 3255.
27. Bowden, J. *et al.* (2017) A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine* **36**, 1783--1802.
28. Bowden, J., Davey Smith, G. and Burgess, S. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512--525.
29. Verbanck, M., Chen, C.-Y., Neale, B. and Do, R. (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50**, 693--698.
30. Bowden, J., Davey Smith, G., Haycock, P.C. and Burgess, S. (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* **40**, 304--314.
31. Hartwig, F.P., Davey Smith, G. and Bowden, J. (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* **46**, 1985--1998.
32. Qi, G. and Chatterjee, N. (2019) Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nature Communications* **10**, 1--10.
33. Burgess, S., Foley, C.N., Allara, E., Staley, J.R. and Howson, J.M.M. (2020) A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* **11**, 1--11.
34. He, B., Shi, J., Wang, X., Jiang, H. and Zhu, H.-J. (2020) Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. *BMC biology* **18**, 1-16.
35. Swerdlow, D.I. *et al.* (2016) Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *International Journal of Epidemiology* **45**, 1600--

- 1616.
36. Zaneveld, J.R., McMinds, R. and Vega Thurber, R. (2017) Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nature microbiology* **2**, 1-8.
 37. Shriner, D. and Rotimi, C.N. (2018) Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the holocene wet phase. *The American Journal of Human Genetics* **102**, 547-556.
 38. Karki, R., Pandya, D., Elston, R.C. and Ferlini, C. (2015) Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC medical genomics* **8**, 1-7.
 39. Ashley-Koch, A., Yang, Q. and Olney, R.S. (2000) Sickle hemoglobin (Hb S) allele and sickle cell disease: a HuGE review. *American journal of epidemiology* **151**, 839-845.
 40. Rees, D.C., Williams, T.N. and Gladwin, M.T. (2010) Sickle-cell disease. *The Lancet* **376**, 2018-2031.
 41. Kato, G.J. *et al.* (2018) Sickle cell disease. *Nature reviews Disease primers* **4**, 1-22.
 42. Jumper, J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589.
 43. Mirdita, M. *et al.* (2022) ColabFold: making protein folding accessible to all. *Nature methods* **19**, 679-682.
 44. Wayment-Steele, H.K. *et al.* (2023) Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*, 1-3.
 45. Abramson, J. *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3.
 46. Dunn, O.J. (1961) Multiple comparisons among means. *Journal of the American Statistical Association* **56**, 52--64.
 47. Ouyang, Q., Kaplan, P.D., Liu, S. and Libchaber, A. (1997) DNA solution of the maximal clique problem. *Science* **278**, 446--449.
 48. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289--300.
 49. Guo, Z. (2023) Causal inference with invalid instruments: post-selection problems and a solution using searching and sampling. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85**, 959-985.
 50. Jansen, I.E. *et al.* (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* **51**, 404--413.
 51. Naj, A.C. *et al.* (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature genetics* **43**, 436--441.
 52. Rathore, N. *et al.* (2018) Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS genetics* **14**, e1007427.
 53. Griciuc, A. *et al.* (2019) TREM2 acts downstream of CD33 in modulating microglial pathology in Alzheimer's disease. *Neuron* **103**, 820--835.
 54. Helgadottir, H.T. *et al.* (2019) Somatic mutation that affects transcription factor binding upstream of CD55 in the temporal cortex of a late-onset Alzheimer disease patient. *Human Molecular Genetics* **28**, 2675--2685.
 55. Rimal, S. *et al.* (2023) Reverse electron transfer is activated during aging and contributes

- to aging and age-related disease. *EMBO reports* **24**, e55548.
56. Winfree, R.L. *et al.* (2023) TREM2 gene expression associations with Alzheimer's disease neuropathology are region-specific: implications for cortical versus subcortical microglia. *Acta Neuropathologica* **145**, 733-747.
57. Yang, X. *et al.* (2023) Functional characterization of Alzheimer's disease genetic variants in microglia. *Nature Genetics*, 1-10.
58. (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research* **51**, D523-D531.
59. Zhou, Y. *et al.* (2022) Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research* **50**, D1398-D1407.
60. Wishart, D.S. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* **46**, D1074-D1082.
61. Bross, P.F. *et al.* (2001) Approval summary: gemtuzumab ozogamicin in relapsed acute myeloid leukemia. *Clinical cancer research* **7**, 1490-1496.
62. Norsworthy, K.J. *et al.* (2018) FDA approval summary: mylotarg for treatment of patients with relapsed or refractory CD33-positive acute myeloid leukemia. *The oncologist* **23**, 1103-1108.
63. Kim, J. *et al.* (2021) FDA approval summary: pralsetinib for the treatment of lung and thyroid cancers with RET gene mutations or fusions. *Clinical Cancer Research* **27**, 5452-5456.
64. Bradford, D. *et al.* (2021) FDA approval summary: selpercatinib for the treatment of lung and thyroid cancers with RET gene mutations or fusions. *Clinical Cancer Research* **27**, 2130-2135.
65. Griciuc, A. *et al.* (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **78**, 631-643.
66. Tortora, F. *et al.* (2022) CD33 rs2455069 SNP: correlation with alzheimer's disease and hypothesis of functional role. *International Journal of Molecular Sciences* **23**, 3629.
67. Kamat, M.A. *et al.* (2019) PhenoScanner V2: an expanded tool for searching human genotype--phenotype associations. *Bioinformatics* **35**, 4851--4853.
68. Faux, N.G. *et al.* (2014) An anemia of Alzheimer's disease. *Molecular Psychiatry* **19**, 1227-1234.
69. Winchester, L.M., Powell, J., Lovestone, S. and Nevado-Holgado, A.J. (2018) Red blood cell indices and anaemia as causative factors for cognitive function deficits and for Alzheimer's disease. *Genome Medicine* **10**, 1--12.
70. Raudvere, U. *et al.* (2019) g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191--W198.
71. Rijpma, A., van der Graaf, M., Meulenbroek, O., Rikkert, M.G.O. and Heerschap, A. (2018) Altered brain high-energy phosphate metabolism in mild Alzheimer's disease: A 3-dimensional ³¹P MR spectroscopic imaging study. *NeuroImage: Clinical* **18**, 254-261.
72. Parasoglou, P. *et al.* (2022) Phosphorus metabolism in the brain of cognitively normal midlife individuals at risk for Alzheimer's disease. *Neuroimage: Reports* **2**, 100121.
73. Lazarczyk, M.J. *et al.* (2016) Major Histocompatibility Complex class I proteins are critical for maintaining neuronal structural complexity in the aging brain. *Scientific reports* **6**,

- 26199.
74. Kim, M.-S. *et al.* (2023) Neuronal MHC-I complex is destabilized by amyloid- β and its implications in Alzheimer's disease. *Cell & Bioscience* **13**, 181.
 75. Le Guen, Y. *et al.* (2023) Multiancestry analysis of the HLA locus in Alzheimer's and Parkinson's diseases uncovers a shared adaptive immune response mediated by HLA-DRB1*04 subtypes. *Proceedings of the National Academy of Sciences* **120**, e2302720120.
 76. Giambartolomei, C. *et al.* (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383.
 77. Barfield, R. *et al.* (2018) Transcriptome - wide association studies accounting for colocalization using Egger regression. *Genetic epidemiology* **42**, 418-433.
 78. Hukku, A. *et al.* (2021) Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *The American Journal of Human Genetics* **108**, 25-35.
 79. Zuber, V. *et al.* (2022) Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *The American Journal of Human Genetics* **109**, 767-782.
 80. Yavorska, O.O. and Burgess, S. (2017) MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International journal of epidemiology* **46**, 1734-1739.
 81. Purcell, S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559--575.
 82. Bulik-Sullivan, B.K. *et al.* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291--295.
 83. Guo, Z., Kang, H., Tony Cai, T. and Small, D.S. (2018) Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 793--815.
 84. Kang, H., Zhang, A., Cai, T.T. and Small, D.S. (2016) Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American statistical Association* **111**, 132-144.
 85. Burgess, S., Dudbridge, F. and Thompson, S.G. (2016) Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Statistics in medicine* **35**, 1880-1906.