Dissecting the genetic and proteomic risk factors for delirium

Vasilis Raptis[1,2†], Youngjune Bhak[1], Timothy I Cannings[3], Alasdair M. J. MacLullich[4], Albert Tenesa[1,5†]

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, UK

[2]Advanced Care Research Centre, School of Engineering, College of Science and Engineering, The University of Edinburgh, Edinburgh, UK

[3]School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, UK

[4]Edinburgh Delirium Research Group, Ageing and Health, Usher Institute, University of Edinburgh, Edinburgh, Scotland, UK

[5]Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh, UK

[†]Corresponding authors:

Vasilis Raptis

V.Raptis@sms.ed.ac.uk

Albert Tenesa

albert.tenesa@ed.ac.uk

Abstract

Delirium is an acute change in cognition, common in hospitalised older adults, and associated with high healthcare and human cost. In this work we shed light on the currently poorly understood genetic and proteomic background of delirium. We conducted the largest to date multi-ancestry analysis of genetic variants associated with delirium (1,059,130 individuals, 11,931 cases), yielding the *Apolipoprotein E* (*APOE*) gene as a strong risk factor with possible population and age-varying effects. *APOE* genetic effect on delirium remained significant after adjusting for dementia and Alzheimer disease (AD) and in dementia-free cohorts. A multi-trait analysis of delirium with AD identified 5 delirium genetic risk loci. Investigation of plasma proteins associated with up to 16-years incident delirium (32,652 individuals, 541 cases) revealed known and novel protein biomarkers, implicating brain vulnerability, inflammation and immune response processes. Incorporating proteomic and genetic evidence via mendelian randomisation, colocalisation and druggability analyses, we indicate putatively useful drug target proteins for delirium. Integrating proteins and *APOE* genetic risk with demographics significantly improved incident delirium prediction compared to demographics alone. Our results pave the way to better understanding delirium's aetiology and guide further research on clinically relevant biomarkers.

57 Delirium is a complex neurocognitive condition affecting nearly 25% of hospitalised older adults[1]. It
58 has a rapid onset, usually lasting from hours to days, and is characterised by an acute and often
59 reversible disturbance of attention, cognitive ability and awareness[2]. It is triggered mainly by acute
60 illness, surgery or injury. Delirium is strongly associated with multiple adverse outcomes, including
61 increased mortality, prolonged hospitalisation[3], and increased healthcare costs, with one study
62 estimating an annual cost of $182 billion dollars per year to European healthcare systems[4].

63 Delirium has a complex bidirectional relationship with dementia, in that people with dementia are
64 more likely to experience delirium in the context of an acute trigger, and delirium episodes are strongly
65 associated with future dementia risk[3,5]. There is some overlap in the clinical features of these
66 conditions, notably with respect to cognitive dysfunction, perceptual disturbances and sleep-wake
67 cycle changes[6]. However, the course of delirium is markedly different with respect to onset and
68 duration. The pathophysiology of delirium is complex and likely involves multiple potential
69 mechanisms[3]. Animal model and human biomarker studies implicate systemic inflammation triggering
70 neuroinflammation, disruption of the blood-brain barrier, and neuronal injury in delirium[3,7–9]. Aspects
71 of these mechanisms are also present in dementia, leading to the proposal that there may be shared
72 and potentially bidirectional pathology between delirium and dementia, such as neuroinflammation
73 leading to neurodegeneration[9,10].

74 Despite its high healthcare burden, however, current understanding of genetic and biological
75 mechanisms underlying delirium's pathophysiology is still limited, hindering personalised medicine
76 efforts to predict, prevent and treat the condition. Given delirium's increasing presence in the global
77 ageing population[6], alleviating its human and economic cost[11] through personalised medicine is all the
78 more important.

79 Previous studies on the genetic determinants of delirium have been small in scale and inconclusive[3,6,12],
80 largely focusing on a single or small sets of candidate genes[13–22]. The apolipoprotein E gene (*APOE*),
81 specifically its ε4 haplotype, is the most intensively studied gene[15,17,19–21], though clear conclusions on
82 its relationship with delirium remain lacking[6,23]. Advances in genome-wide association studies (GWAS)
83 over the last decades[24] have allowed researchers to search across the full spectrum of the human
84 genome for genetic risk factors involved in neurocognitive disorders[25–27], offering invaluable new
85 insights into disease mechanisms. However, research on delirium using these methods has lagged
86 behind other neurocognitive disorders, with only a few, relatively underpowered delirium GWAS
87 conducted to date[28–30]. Moreover, those studies did not address comprehensively the biological
88 implications of the potential gene-disease associations and were conducted primarily on individuals
89 of European descent.

Research on protein biomarkers for delirium has received increased attention recently, identifying potential blood plasma, serum and cerebrospinal fluid (CSF) canditates[6,31]. Those biomarkers include inflammatory proteins, such as interleukin-6 (IL-6) and C-reactive protein (CRP)[6,31,32], Alzheimer disease pathology markers[6], and markers of neuronal injury, such as the neurofilament light chain (NEFL)[9] and glial fibrillary acidic protein (GFAP)[8]. Previous studies however are limited in terms of sample size and, often, the range of proteins tested.

In the current study we aim to upscale the efforts in identifying genetic and proteomic determinants of delirium risk. To achieve this, we: (1) conducted the largest so far meta-analysis of delirium GWAS datasets, the first to include individuals from diverse ancestries; (2) tested for plasma proteome signatures of incident delirium for up to 16 years of follow-up in UK Biobank (UKB)[33] and triangulated the results using genetically supported evidence from mendelian randomisation, colocalisation and druggability analyses; (3) conducted a multi-trait meta-analysis between delirium and Alzheimer disease, leveraging the shared genetic basis of the two conditions[6] and boosting the power to detect genetic associations for delirium[34].

Results

APOE gene as delirium genetic risk factor

To identify genetic variants associated with delirium, a multi-ancestry genome-wide association meta-analysis (GWAMA) was conducted on eight sub-cohorts (Supplementary Table 1) from four global ancestries: European (EUR, $n_{cases} = 7{,}988$, $n_{controls} = 549{,}568$, 52.6% of total sample size), Finnish (FIN, $n_{cases} = 3{,}371$, $n_{controls} = 388{,}560$, 37%), African (AFR, $n_{cases} = 348$; $n_{controls} = 59{,}780$, 5.7%), South Asian (SAS, $n_{cases} = 107$; $n_{controls} = 9{,}356$, 0.9%) and admixed American / Hispanic (AMR, $n_{cases} = 117$; $n_{controls} = 39{,}977$, 3.8%). Age distributions for the UK Biobank (UKB) and All of Us Research Program (AoU) cohorts are presented in Supplementary Figure 1. In total, the GWAMA comprised of up to 1,059,130 individuals and 11,931 delirium cases, yielding results for 24,951,028 single nucleotide polymorphism (SNP) genetic variants. The overall SNP heritability (h^2_{SNP}) of delirium was estimated on the observed scale as $h^2_{SNP} = 0.0055$ ($standard\ error$: $9x10^{-4}$; $p = 2.5x10^{-9}$), significantly different from zero, and as $h^2_{SNP,l} = 0.029$ on the liability scale.

Variants at the *Apolipoprotein E* (*APOE*) gene and within its close genomic region on chromosome 19 (Figure 1) were significantly associated with delirium. The lead variant rs429358 (T>C, odds ratio [95% Confidence Intervals]: 1.60 [1.55 – 1.65]; p = $9.7x10^{-177}$) is an *APOE* missense variant, which together

with the rs7412 C>T variant forms the *APOE*-ε4 haplotype (the rs429358-C and rs7412-C alleles), an established risk factor for Alzheimer disease (AD)[35]. rs7412 also significantly associated with delirium in our GWAMA (C>T, OR [95% CI]: 0.84 [0.79 – 0.88]; p = 1.8×10^{-11}).

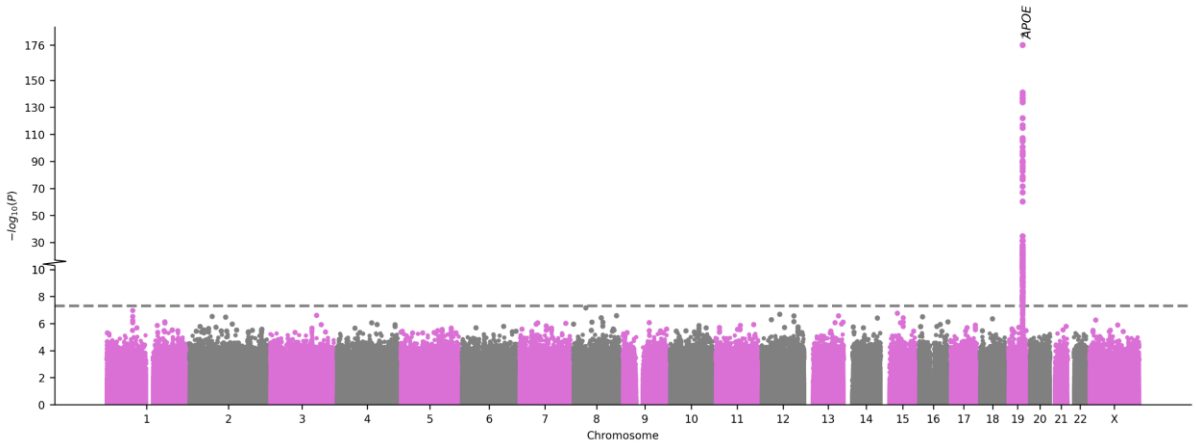


Figure 1: Manhattan plot of the delirium multi-ancestry GWAMA. *Each point represents a genetic variant. The x-axis denotes the variant's genomic position and the y-axis the two-sided meta-analysis association p-value. The grey dashed line denotes the genome-wide significance p-value threshold* 5×10^{-8}*. The gene in which the lead significant variant is located is annotated. GWAMA: genome-wide association meta-analysis; APOE: Apolipoprotein E gene. The plot was created using the GWASLab python package.*

The lead variant rs429358 showed population specific genetic effects for delirium in the contributing sub-cohorts (Figure 2). Significant associations were observed in all, except European (Michigan Genomics Initiative (MGI), p = 0.27) and admixed American (AoU AMR, p = 0.12) populations from the USA, with USA-based populations generally showing smaller effect sizes.

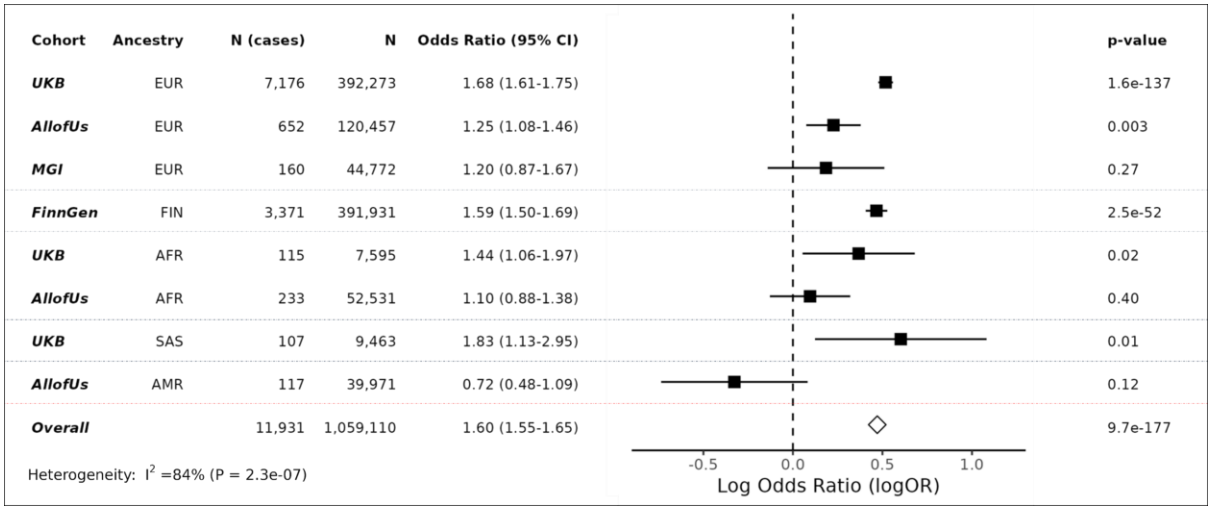


Figure 2: Associations between the GWAMA lead variant and delirium risk for each cohort.

Forrest plot showing the cohort-specific association results for the rs429358 T>C lead variant with delirium. Effect sizes for the C allele and corresponding two-sided p-values are shown for each contributing sub-cohort and the overall meta-analysis. Grey dotted lines separate results for each ancestry group. GWAMA: genome-wide association meta-analysis; UKB: UK Biobank; MGI: Michigan Genomics Initiative; N: sample size; CI: 95% confidence intervals for the odds ratio; EUR: European; FIN: Finnish; AFR: African; SAS: south Asian; AMR: admixed American).

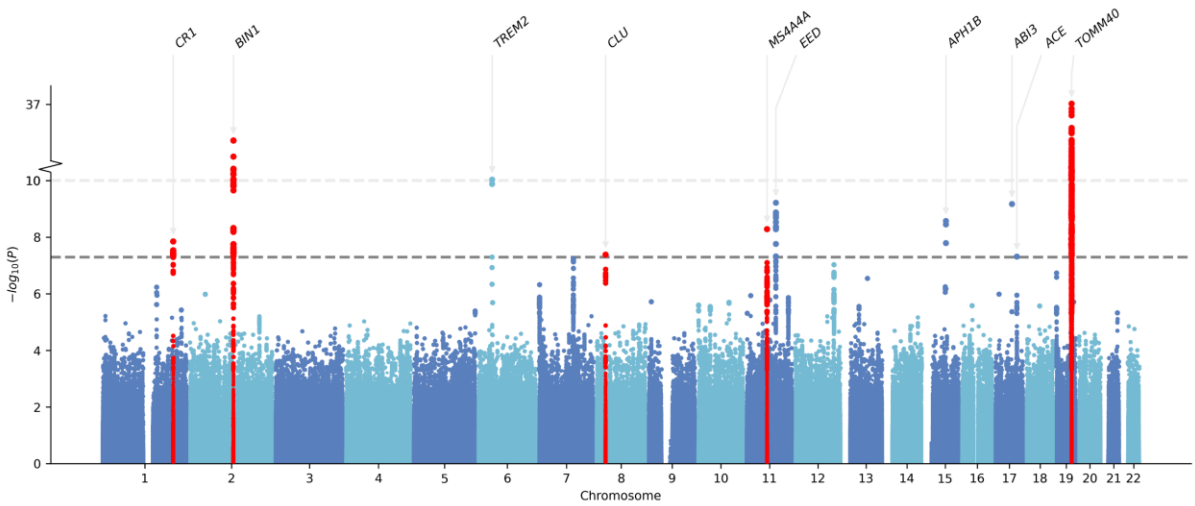
Conditional GWAS analysis on the *APOE*-ε4 haplotype resulted in all variants on the *APOE* region losing significance (Supplementary Figure 2), suggesting that *APOE*-ε4 is the sole independent genetic risk factor in the region. On the same analysis, an intronic variant within the *ADAM32* gene on chromosome 8 gained significance (rs531178459; $p = 3.3x10^{-8}$).

Furthermore, using a series of sensitivity analyses in UKB (see Methods and Supplementary Table 2), we tested to what extent the observed *APOE* association with delirium is driven by underlying dementia. Variants in the *APOE* region remained significant after adjusting for dementia status in the delirium GWAS (Supplementary Figure 3). Specifically, the lead variant from the unadjusted GWAMA, rs429358, showed again a strong association (OR [95% CI]: 1.20 [1.12 – 1.28]; $p = 3.7x10^{-15}$). Variants on the *SEC14L1* gene on chromosome 17 were also significant in the dementia-adjusted analysis. rs429358 remained significant in GWAS conducted on dementia-free (OR [95% CI]: 1.27 [1.2 – 1.34]; $p = 4.7x10^{-18}$; Supplementary Figure 4) and AD-free (OR [95% CI]: 1.45 [1.38 – 1.52]; $p = 3.9x10^{-56}$; Supplementary Figure 5) stratified cohorts. rs429358 genetic signals in dementia-adjusted and dementia/AD-free GWAS were replicated in AoU (Supplementary Table 2). The *APOE* genetic signal

161 remained highly significant in an age-stratified GWAS, on participants with age 60 years or more (OR
162 [95% CI]: 1.7 [1.63 – 1.77]; p = $2.7x10^{-141}$; Supplementary Figure 6 and Supplementary Table 2).

163 **Multi-trait analysis between delirium and Alzheimer disease**

164 We found a significant genetic correlation (r_g) between delirium and Alzheimer disease (AD), $r_g =$
165 $0.38\ (SE{:}\ 0.12;\ p = 1.9x10^{-3})$, pointing out to a partially shared genetic architecture. Given this, we
166 further conducted a multi-trait analysis of GWAS summary statistics (MTAG) between delirium and
167 AD. MTAG can increase statistical power to detect new genetic associations, by leveraging the shared
168 genetic information between related traits[34]. Our MTAG analysis identified 10 independent genetic
169 loci associated with delirium, of which 5 replicated in the held-out set (Figure 3 and Supplementary
170 Table 3). The closest genes mapped to the lead replicated variants included: *CR1* (rs4844610 A>C; OR
171 [95% CI]: 1.01 [1.006 - 1.014]; p = $1.4x10^{-8}$), *BIN1* (rs6733839 T>C; OR [95% CI]: 1.015 [1.012 - 1.018];
172 p = $7x10^{-25}$), *CLU* (rs2279590 T>C; OR [95% CI]: 0.992 [0.989 – 0.995]; $4.2x10^{-8}$), *MS4A4A* (rs1582763
173 A>G; OR [95% CI]: 0.991 [0.988 - 0.994]) and *TOMM40* (rs117310449 T>C; OR [95% CI]: 1.09 [1.07 -
174 1.1];p =$7x10^{-38}$).



175

176

177 ***Figure 3: Manhattan plot of the multi trait analysis of GWAS between delirium and Alzheimer***
178 ***disease.*** *Delirium-specific MTAG summary statistics are shown. Each point represents a genetic*
179 *variant. The x-axis denotes the variant's genomic position and the y-axis the MTAG association p-value*
180 *(two-sided) in the discovery set. The grey dashed line denotes the genome-wide significance p-value*
181 *threshold $5x10^{-8}$. The closest gene in which the lead significant variant is located is annotated. Red*
182 *colour highlights loci ±1,000kb around lead variants that replicated. MTAG: Multi-trait analysis of*
183 *GWAS. The plot was created using the GWASLab python package.*

Four out of five of the lead replicated MTAG signals were also significant in the original delirium GWAMA (Supplementary Table 3), with rs2279590 (*CLU* gene) being marginally non-significant (p = 0.058). Effect sizes were consistent between MTAG and delirium GWAMA results. Variants in the *APOE* genomic region that were significant in our delirium GWAMA were also significant in MTAG.

Mediation analysis

We implemented a mediation analysis in order to estimate how much of the *APOE*-ε4 genetic effect on delirium is mediated by dementia, adjusted for age and sex (Figure 4). We found that *APOE*-ε4 exerts a significant direct effect on delirium in a dose-dependent way: $OR_{\text{direct effect (0vs1)}}$ [95% CI]: 1.14 [1.08 - 1.2]; p = 4.5×10^{-7}; for one *APOE*-ε4 copy, and $OR_{\text{direct effect (0vs2)}}$ [95% CI]: 1.29 [1.15 - 1.45]; p = 1.7×10^{-5} for two *APOE*-ε4 copies. The *APOE*-ε4 total effect on delirium was partially mediated by dementia ($OR_{\text{indirect effect (0vs1)}}$ [95% CI]: 1.39 [1.36 - 1.42]; p = 1.1×10^{-237}; and $OR_{\text{indirect effect (0vs2)}}$ [95% CI]: 2.59 [2.47 – 2.77]; p = 2.1×10^{-233}), with the direct effect accounting for 29% and 21% of the total effect for one and two *APOE*-ε4 copies respectively. These results further support the role of *APOE* gene on delirium, independently of dementia.

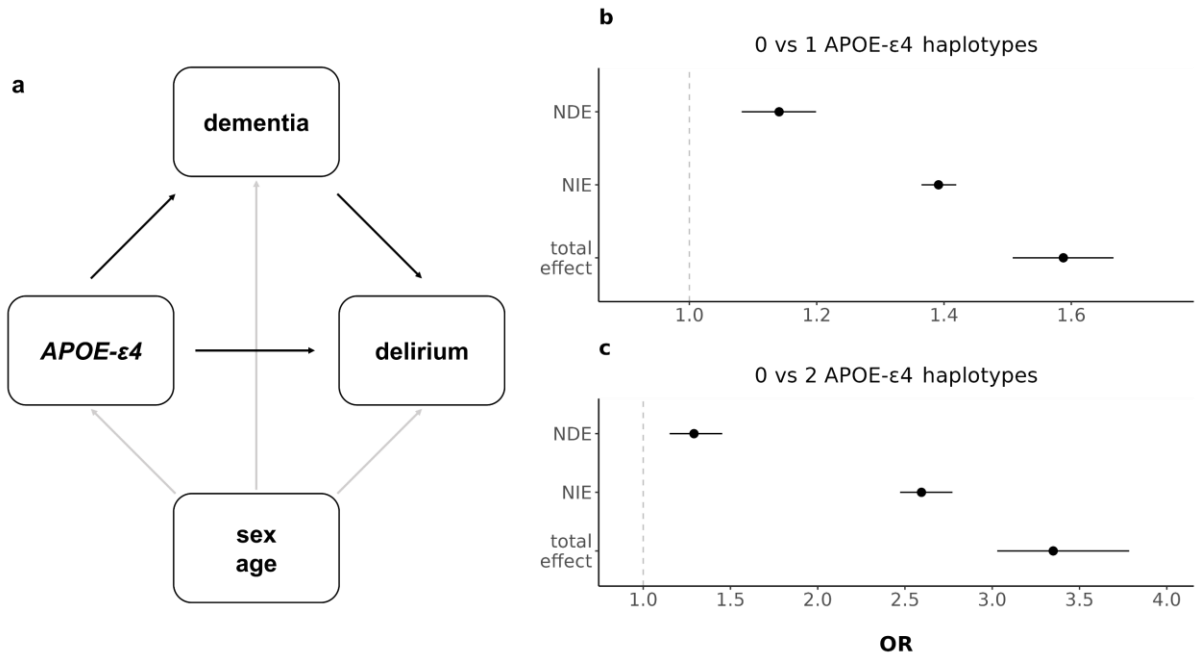


Figure 4: Mediation analysis of APOE-ε4 on delirium, mediated by all-cause dementia. (a) *Hypothesised directed acyclic graph, showing all-cause dementia as mediator of APOE-ε4 genetic effect on delirium (black arrows), adjusted for baseline covariates: age and sex (grey arrows).* ***(b,c)*** *Effect decomposition of the total APOE-ε4 effect, into natural direct (independent of dementia) and natural indirect (mediated through dementia) effects.* ***(b)*** *shows the effect of having one APOE-ε4*

*haplotype, and **(c)** the effect of having two APOE-ε4 haplotypes, compared to having none. OR: odds ratio. NDE: natural direct effect; NIE: natural indirect effect.*

Our sensitivity analysis regarding the role of unmeasured confounding suggests that, an unmeasured confounder associated with both dementia and delirium with approximate effect sizes of 1.54 each, over and above the measured covariates, could suffice to completely explain away the observed direct effect of one *APOE*-ε4 copy on delirium, but weaker confounding could not[36]. Respectively, the minimum unmeasured confounder effect size would need to be 1.9 to completely explain away the direct effect of two *APOE*-ε4 copies.

Finally, a mediation analysis further adjusting for cognitive function, socioeconomic deprivation and chronic disease burden, on top of age and sex, still detected a significant direct effect of *APOE*-ε4 on delirium: $OR_{\text{direct effect (0vs1)}}$ [95% CI]: 1.12 [1.01 - 1.25]; p = 0.038 (32% of the total effect) for one *APOE*-ε4 copy, and $OR_{\text{direct effect (0vs2)}}$ [95% CI]: 1.55 [1.18 - 2.06]; p = $2x10^{-3}$ (34% of the total effect) for two *APOE*-ε4 copies respectively (Supplementary Figure 7).

Protein risk factors for incident delirium

A proteome-wide association study (PWAS)on 32,652 European UKB participants (541 cases) revealed 109 out of the 2,919 total proteins, whose plasma levels were significantly associated with incident delirium up to 16-years of follow-up (Figure 5a and Supplementary Table 4) at a Bonferroni adjusted p-value threshold (p = $1.7x10^{-5}$). The APOE protein had a negative effect on incident delirium risk, meaning that higher plasma levels of the protein are associated with reduced future risk. The association was significant at the nominal level (0.05), but not after multiple test correction (OR [95% CI]: 0.86 [0.79 - 0.94], p = $7x10^{-4}$, Supplementary Table 4).

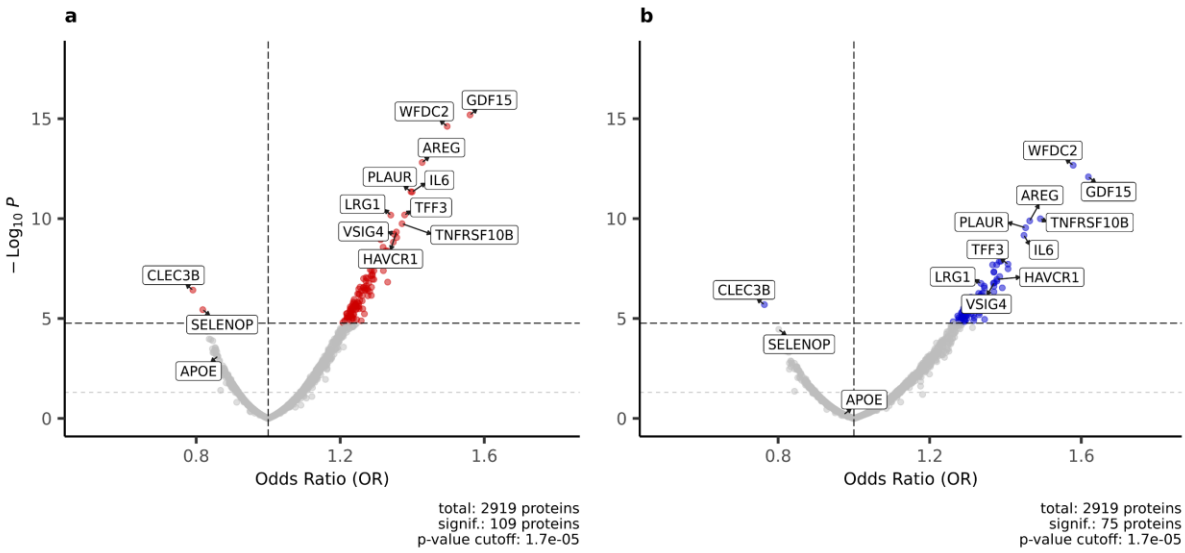


Figure 5: Associations of plasma proteins with incident delirium*. Volcano plot showing the proteome-wide associations of 2,919 plasma proteins with incident delirium in (**a**) the non-stratified, and (**b**) the dementia-free UKB proteomic set. Odds ratios of the logistic regression models are plotted on the x-axis and respective two-sided -log$_{10}$(p-values) on the y-axis. All models are adjusted for age at protein collection, sex and BMI. The black dashed horizontal line indicates the Bonferroni adjusted p-value threshold: 1.7x10^{-5}, with proteins above this line colored red. The grey dashed horizontal line indicates the nominal p-value threshold: 0.05. The plot was created using the EnhancedVolcano R package (v1.18.0).*

Further adjusting the individual protein models for *APOE*-ε4 status did not substantially alter the results (Supplementary Table 5), having highly correlated effect size estimates with the models adjusted only for age, sex and BMI (Pearson's correlation r=0.99). Additionally including a protein* *APOE*-ε4 interaction term also yielded highly similar results (main effect correlation r=0.93). No significant protein* *APOE*-ε4 interaction has observed at a Bonferroni adjusted threshold of p-value < 1.7x10^{-5} (Supplementary Table 6).

The 109 proteome-wide significant proteins were found to be significantly enriched (q-value < 0.05) in several important inflammation and immune response biological pathways, such as interleukin and Tumour Necrosis Factor (TNF) signalling (Supplementary Table 7).

Restricting our PWAS on the dementia-free subset of UKB (31,692 participants; 349 delirium cases), reduced the number of proteins significantly associated with incident delirium to 75 (Figure 5b and Supplementary Table 8). Notably, the APOE plasma protein, although still had a negative effect size,

249 lost nominal significance and had reduced magnitude that in the non-stratified PWAS (OR [95% CI]:
250 0.97 [0.87 – 1.08]; p = 0.63).

251 **Protein selection and prediction models**

252 A machine learning framework was applied to further pin down which plasma proteins are robustly
253 associated with incident delirium. Our approach, based on LASSO regularised regression[37] and stability
254 selection[38], revealed 19 proteins (stability-selected proteins, Supplementary Table 9) consistently
255 selected as predictive of incident delirium in the training set. All of the stability-selected proteins
256 represent a subset of the top individually significant proteins identified through the proteome-wide
257 association analysis. The FGL1 protein was removed from subsequent analyses, as it had a non-
258 significant contribution to the re-fit prediction models and was dropped during stepwise regression.
259 The 18 remaining stability-selected proteins provide marginal prediction improvements of incident
260 delirium on an independent test set, compared to predictions based on demographic factors alone
261 (Figure 5 and Supplementary Tables 10-11). Specifically, adding the 18 stability-selected proteins to
262 the "basic" model that includes age, sex and BMI as predictors increased the AUC from 0.764 to 0.791,
263 but the increase was not significant (DeLong test p-value = 0.09, Supplementary Table 11). However,
264 adding proteins and *APOE*-ε4 status to the basic model showed a significant prediction improvement
265 (AUC from 0.764 to 0.794, p = 0.049, Supplementary Table 11). Finally, the model fit only with the
266 selected proteins performed worse that the basic model (AUC from 0.764 to 0.729, p = 0.21, Figure 5a
267 and Supplementary Table 11). The precision-recall performance showed similar pattern (Figure 5b),
268 with the full models having higher PR-AUC: 0.065 and 0.06 for the "*APOE*+proteomic+basic" and the
269 "proteomic+basic" models respectively, compared to 0.043 for the basic model.

270 We repeated our protein selection and prediction framework in the dementia-free sub-cohort of the
271 UKB proteomic set. 14 out of the 19 stability-selected proteins still had a significant effect on incident
272 delirium in the training set (Supplementary Table 9). However, only the AREG and MSLN proteins were
273 consistently selected in the dementia-free sub-cohort (selection frequency > 0.5). We still observed a
274 marginally better prediction of incident delirium on the test set using *APOE*-ε4 and the 14 stability-
275 selected proteins on top of demographic factors (AUC from 0.758 to 0.79, p = 0.13; PR-AUC from 0.033
276 to 0.036) (Supplementary Figure 8 and Supplementary Table 10), although now with a non-significant
277 AUC improvement.

278

279

280

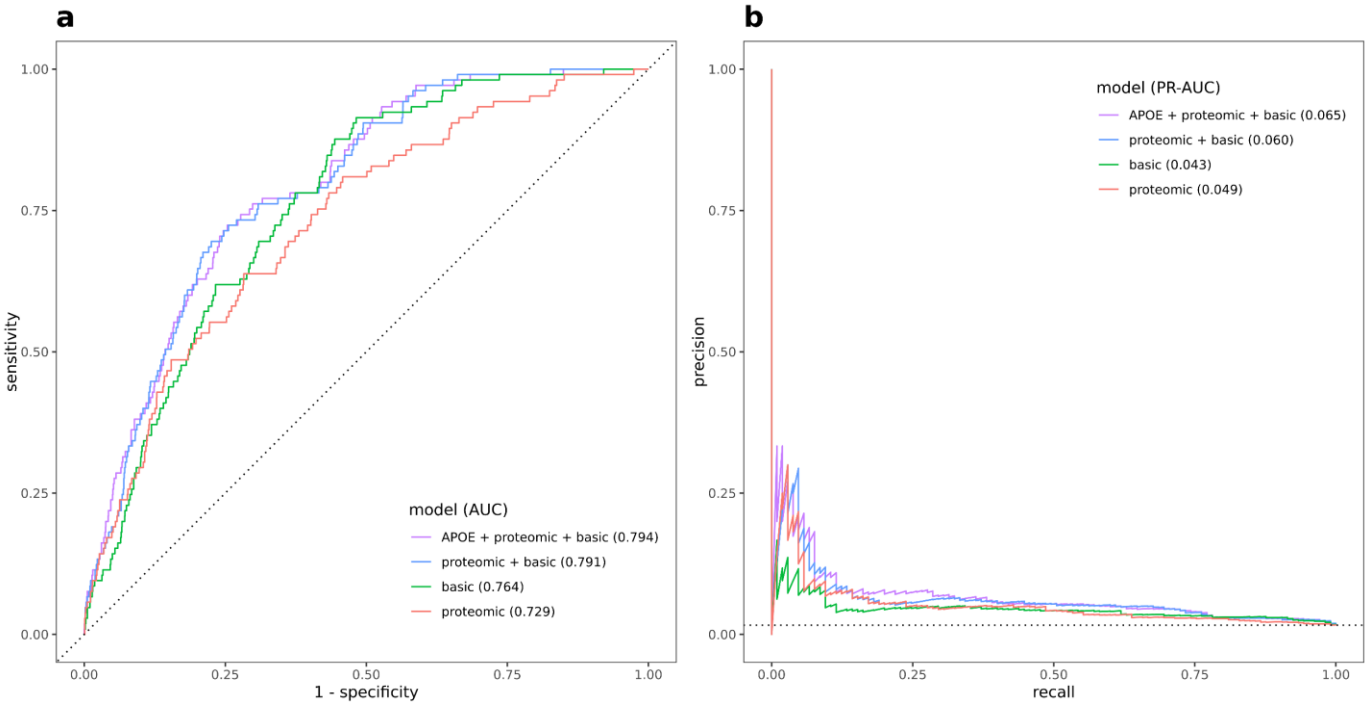


Figure 6: Proteomic predictive performance on test set. *Receiver Operating Characteristic (ROC) (**a**) and Precision-Recall (PR) curves (**b**) showing performance metrics of different models for predicting incident delirium on the held-out test set. The "proteomic" logistic regression model is fit with the 18 stability-selected proteins, the "basic" model is fit using only age at protein collection, sex and BMI and the "proteomic+basic" includes both. The "APOE+proteomic+basic" model additionally includes APOE-ε4 haplotype count as predictor. The dotted black lines indicate the performance of a randomly classifying model, that is the diagonal of the ROC curve (**a**) and the disease prevalence horizontal line in the PR curve (**b**). AUC: Area Under the Curve. ROC and PR plots and AUC calculations are made using the yardstick R package (v1.3.0).*

Mendelian randomisation

We performed mendelian randomisation (MR), to test for causal associations between 1,989 plasma proteins (exposures) and delirium (outcome). We found 53 proteins causally associated with delirium at a false discovery rate (FDR) q threshold < 0.05 using inverse variance weighted (IVW) or Wald ratio MR. (Figure 7 and Supplementary Table 12). Out of the 53 MR significant proteins, DSC2 was the only one with a significant association in our PWAS analysis of protein risk factors (Figure 5) and a consistent direction of effect. Using a less strict FDR q < 0.05 threshold in our PWAS, 7 additional

301 proteins were identified (ADAM8, APOE, DPP10, DSC2, LAYN, NT5C1A, PILRA and PVR) that were
302 significant in both the PWAS and MR. At a nominal PWAS p-value level ($p < 0.05$), three more proteins
303 (CCL25, PON3 and GGH) overlapped the significant MR results. Four of the above PWAS - MR
304 overlapping significant proteins successfully replicated in our FinnGen replication MR: GGH, PVR,
305 APOE and PILRA (Figure 7). All of the replicated delirium-associated proteins were also significant using
306 sensitivity MR methods (weighted median, maximum likelihood and MR-Egger; $p < 0.05$)
307 (Supplementary Table 13). The MR-Egger intercept test did not reveal evidence of horizontal
308 pleiotropy ($p > 0.05$). Genetic variant heterogeneity was suggested by the Cochran's Q test ($p < 0.05$)
309 for the genetic instruments used in the APOE and PVR MR (Supplementary Table 13). Repeating MR
310 for these two proteins using more strictly independent pQTLs ($r^2 < 0.001$), APOE did not show evidence
311 for heterogeneity (Q test $p = 0.37$), while its causal effect on delirium remained significant (IVW MR
312 beta (standard error) = -0.57 (0.03), $p = 1.13 \times 10^{-78}$). For PVR, heterogeneity remained high (Q test $p =$
313 0.03), and lost significance for its causal effect (IVW MR beta (SE) = 0.06 (0.06), $p = 0.34$), suggesting
314 that PVR's initially observed significant MR effect may be driven by pleiotropy and linkage
315 disequilibrium of its genetic instruments with the *APOE*-ε4 haplotype that is located in close proximity.
316 Using delirium in the dementia-free UKB GWAS as outcome, ADAM8, CCL25, PILRA, and LAYN lost their
317 MR causal effect significance (FDR $q > 0.05$). A summary of analyses supporting each protein's
318 association with delirium is presented in Table 1, with full summary results in Supplementary Data 1.

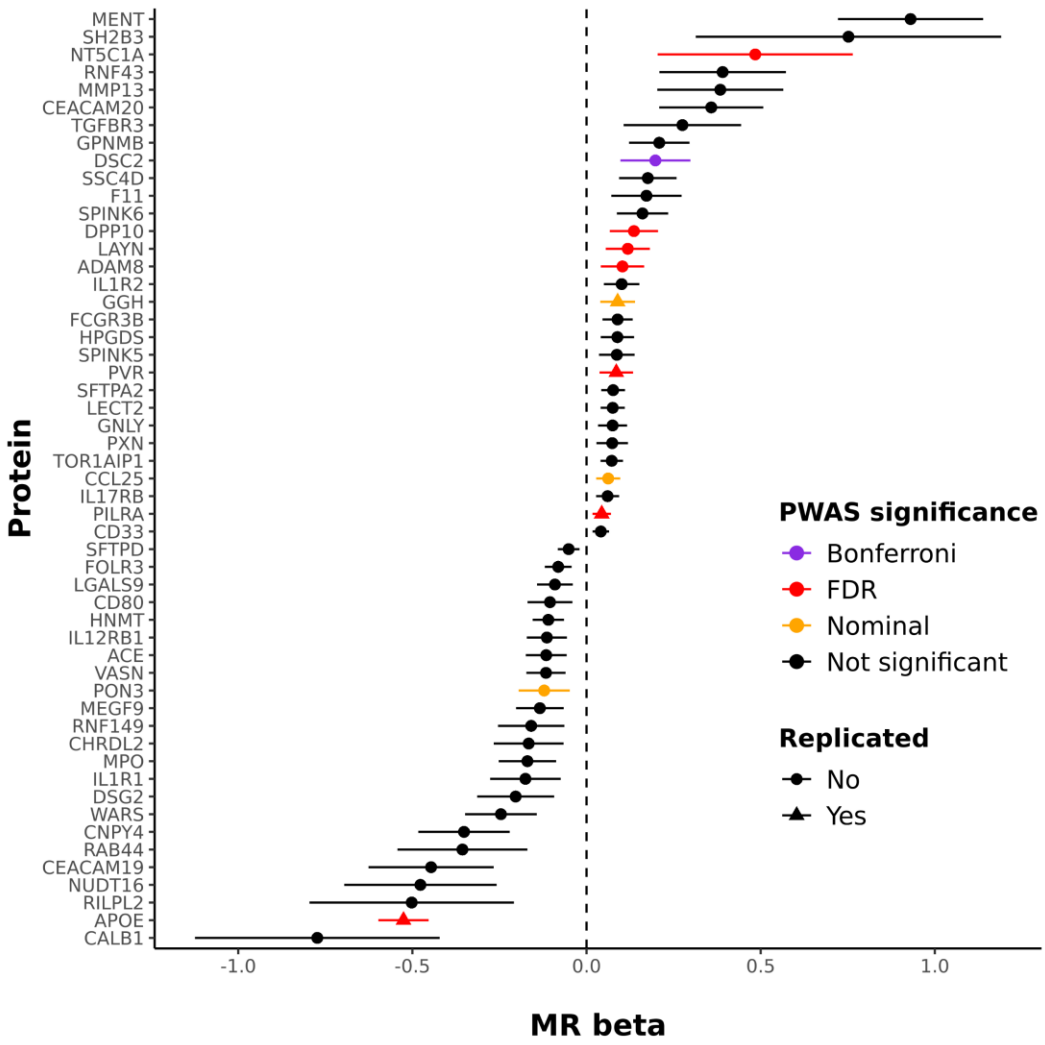


Figure 7: Mendelian randomisation results of plasma proteins on delirium. *MR associations were derived using the IVW or Wald ratio method and UKB delirium GWAS summary statistics. Only significant (FDR q < 0.05) MR associations are shown (IVW/Wald ratio betas and 95% confidence intervals). Replication MR was conducted using delirium GWAS summary statistics from Finngen. Colors are assigned to proteins based on their significance in the PWAS analysis: Bonferroni: $p < 1.7x10^{-5}$; FDR: q < 0.05; nominal: p < 0.05 and having consistent direction of effect. MR: mendelian randomisation; PWAS: proteome-wide association study; FDR: false discovery rate. IVW: inverse variance weighted.*

329 **Table 1: Overview of follow-up analyses on proteins significantly associated with delirium through PWAS and MR.**

Protein	Effect direction	PWAS significance	Discovery MR (UKB)							Replication MR IVW (FinnGen)	Co-localisation	Druggability Tier
			IVW	IVW dementia-free	Weighted median	Maximum Likelihood	MR-Egger	MR-Egger intercept test	Cochran's Q test			
PON3	-	Nominal	✓	✓	✓	✓	✓	✓	✓	✗	✗	Tier 1
ADAM8	+	FDR	✓	✗	✓	✓	✗	✓	✓	✗	✗	Tier 2
GGH	+	Nominal	✓	✓	✓	✓	✓	✓	✓	✓	✓	Tier 2
PVR [a]	+	FDR	✓	✓	✓	✓	✓	✓	✗	✓	✗	Tier 3A
APOE [b]	-	FDR	✓	✓	✓	✓	✓	✓	✗	✓	✓	Tier 3A
DPP10	+	FDR	✓	✓	✓	✓	✗	✓	✓	✗	✗	Tier 3B
CCL25	+	Nominal	✓	✗	✓	✓	✓	✓	✓	✗	✗	Tier 3B
DSC2	+	Bonferroni	✓	✓	✓	✓	✗	✓	✓	✗	✗	Not classified
NT5C1A	+	FDR	✓	✓	✓	✓	✗	✓	✓	✗	✗	Not classified
PILRA	+	FDR	✓	✗	✓	✓	✓	✓	✓	✓	✗	Not classified
LAYN	+	FDR	✓	✗	✗	✓	✗	✓	✓	✗	✗	Not classified

330

331 *Proteins that were supported by both the PWAS ($p < 0.05$) and MR via IVW or Wald ratio (FDR < 0.05) analyses in UKB and had consistent effect direction are*
332 *shown. ✓ (green) denotes significance ($p < 0.05$) in the MR sensitivity analyses (Weighted median, Maximum Likelihood and MR-Egger); non-significance in*
333 *the MR-Egger intercept and Cochran's Q tests ($p > 0.05$); and PP.H4 > 0.8 in the colocalisation analysis. ✓ (orange) denotes suggestive colocalisation: 0.8 >*
334 *PP.H4 > 0.5 ✗ denotes failure. Full summary statistics are available in Supplementary Data 1. PWAS: proteome-wide association study; MR: mendelian*
335 *randomisation; FDR: false discovery rate; IVW: inverse variance weighted; UKB: UK Biobank; pQTL: protein Quantitative Trait Loci*

336 *[a] PVR lost MR significance and showed high heterogeneity using strictly independent pQTL ($r^2 < 0.001$).*
337 *[b] APOE remained significant in MR and did not show heterogeneity evidence using strictly independent pQTL ($r^2 < 0.001$).*

338 **Colocalisation**

339 We further conducted colocalisation analysis to test whether the protein and delirium genetic effects
340 are exerted by the same rather than distinct genetic variants in each genomic region. 37 proteins
341 colocalised with delirium within the *APOE* genomic region (Supplementary Table 14 and
342 Supplementary Data 2) with high probability (posterior probability of a shared causal variant, PP.H4 >
343 0.9 as per coloc software notation[39]). 30 more proteins showed suggestive evidence of colocalisation
344 with delirium throughout the genome (PP.H4 > 0.5). It should be noted that colocalisation methods
345 are generally not well powered to detect shared casual genetic signals unless both traits shown strong
346 genetic support (i.e. GWAS $p < 5x10^{-8}$)[39,40]. Here, only the *APOE* region is strongly associated with
347 delirium, thus many colocalisation signals in other parts of the genome may be missed.

348 **Druggability assessment**

349 For delirium-associated proteins with support from both our PWAS and MR analyses, we assessed
350 their suitability to act as drug targets (Table 1). We used the "druggable genome" curated by Finan et
351 al (2017)[41], a resource which stratified human proteins on tiers based on the strength of their
352 druggability evidence (see Methods). Of the selected proteins, PON3 belonged in the druggability tier
353 1 (targets of approved drugs or clinical-phase drug candidates); ADAM8 and GGH belonged in tier 2
354 ("targets with known bioactive drug-like small-molecule binding partners"[41] or having high similarity
355 with approved drug targets); PVR, APOE, DPP10, CCL25 in tier 3 (secreted or extracellular proteins,
356 proteins with medium similarity to approved drug targets, and members of key druggable gene
357 families not included in tiers 1 or 2[41]). DSC2, NT5C1A, PILRA, LAYN did not have druggability
358 information.

359 **KGWAS**

360 Knowledge Graph GWAS (KGWAS) analysis of our GWAMA summary statistics revealed 4 disease-
361 critical genes within the *APOE* genomic region (*APOE*, *TOMM40*, *PVRL2*, *BCAM*), prioritized via
362 functional genomics knowledge graph[42]. All of the KGWAS-identified significant SNPs were also
363 significant in our GWAMA (Supplementary Table 15).

364

365 # Discussion

366 In this analysis we conducted the largest to our knowledge multi-ancestry genome-wide meta-analysis
367 on delirium. Genetic variants on the *APOE* gene on chromosome 19 were identified as significantly
368 associated with delirium, with the top variant hit, rs429358, showing population-specific association

369 patterns. The *APOE* gene encodes for APOE, a lipid transporter protein in the periphery and the brain.
370 APOE is a strong risk factor for Alzheimer Disease (AD), through its diverse roles in pathways such as
371 amyloid-β plaque deposition, neuroinflammation and dysregulation of lipid metabolism in the brain[43].

372 The role of *APOE* gene in delirium is currently unclear, with previous meta-analyses reporting no
373 association between *APOE* and delirium[12,23]. In UKB, a previous study on European participants found
374 an association between *APOE*-ε4 status and delirium (hazard ratio = 3.73 [2.68 - 5.21])[44]. It has been
375 suggested that interactions between *APOE*-ε4 and inflammation-related proteins can drive delirium
376 development[6]. To assess this hypothesis, we tested whether the interaction term between each
377 plasma protein level and *APOE*-ε4 status significantly associated with incident delirium
378 (Supplementary Table 5). No protein**APOE*-ε4 interaction reached significance adjusted for multiple
379 testing (p-value threshold = $1.7x10^{-5}$). However, the CEND1 protein exhibited an interaction with
380 *APOE*-ε4 marginally below threshold ($beta_{(interaction)}$ [95% CI] = 0.27 [0.10 – 0.42]; $p_{(interaction)} = 8x10^{-4}$).
381 CEND1 is a mitochondrial neural differentiation protein, expressed in the nervous system[45]. It has
382 previously been implicated in cognitive impairment in mice[45] and AD in human[46] brains. Additionally,
383 *APOE*-ε4 expression in astrocytes has been implicated in impaired mitochondrial function[47], although
384 to our knowledge CEND1 and APOE have not been linked in previous studies. CEND1 role in delirium
385 has also not been investigated so far.

386 Ancestry-dependent *APOE*-ε4 genetic effects on delirium have not been systematically assessed
387 previously. For AD, the risk conferred by *APOE*-ε4 varies by ancestral background, with African/African
388 Americans and Hispanics having less pronounced risk than white Europeans and Asians[48,49].
389 Additionally, higher *APOE*-ε4 expression levels have been observed in carries of European compared
390 to African ancestry[50]. Here, findings suggest a similar pattern, that is rs429358-C having a higher effect
391 in European, Finnish and south Asian populations than African and Hispanic/Admixed American.
392 Overall, sub-populations from the USA (AoU, MGI) have weaker *APOE* effects than those from the UK
393 or Finland (UKB, FinnGen). This might reflect the younger age of participants in USA-based studies
394 (Supplementary Table 1) or phenotypic differences in delirium diagnoses across the healthcare
395 systems of different countries. Age-dependent genetic effects have been previously described for
396 *APOE*-ε4 with regard to AD[48,51] and progression to mild cognitive impairment and AD[52], showing
397 increasing effects until an age of 70 – 75, with reduced effect on later ages[48,51,52].

398 It is also possible that underlying dementia is driving the strong association between *APOE* and
399 delirium observed here. 36% of UKB European delirium cases had a dementia diagnosis, compared to
400 1.4% in the control group (Supplementary Table 1). This is to be expected given the close relationship
401 of the two disorders[6], but it may hide delirium specific genetic effects and overemphasize the role of

402 *APOE*. To this end, *APOE* remained significant in our range of sensitivity analyses – GWAS adjusting for
403 all-cause dementia (Supplementary Figure 3), mediation analysis (Figure 4), dementia and AD
404 stratified GWAS (Supplementary Figure 4-5 and Supplementary Table 2) – although with weaker
405 effect. This result may suggest that *APOE* association with delirium is not entirely through its role in
406 dementia or AD.

407 Regarding our mediation analysis, it should be noted that unmeasured confounding could bias the
408 estimate of *APOE's* direct effect on delirium[36,53] (see Methods section). Unmeasured confounding
409 between *APOE* (exposure) and dementia (mediator) or delirium (outcome) could be introduced by
410 residual population stratification or linkage disequilibrium between *APOE*-ε4 and linked causal genetic
411 variants. It is also possible that *APOE*-induced pleiotropic genetic effects can confound the delirium-
412 dementia relationship, also introducing bias. Our results remained significant after adjusting for
413 possible confounders between dementia and delirium, i.e. shared risk factors such as cognitive
414 function, socioeconomic deprivation and comorbidity burden[3–5,54,55]. None of the additional
415 confounders had an independent odds ratio effect on both delirium and dementia above 1.54 each,
416 the minimum required to explain away the observed *APOE* direct effect on delirium. However, some
417 residual confounding may still remain, for example due to precipitating factors of delirium episodes
418 during hospitalisation, such as infections, medication status, sleep disruptions or severity of
419 presenting illness. These factors have been previously estimated to have a strong effect on delirium[3,4],
420 and may be also related to dementia[55]. Assessing the effect of those factors was outside the scope of
421 our study.

422 Adjusting for *APOE*-ε4 attenuated the genetic effects within the whole *APOE* region. This observation
423 suggests that the significance of the genetic variants in the close proximity is driven by linkage
424 disequilibrium with the *APOE*-ε4 haplotype, not secondary independent signals. Moreover, an intronic
425 variant on the *ADAM32* gene gained significance after adjusting for *APOE*-ε4. *ADAM32* belongs to the
426 ADAM family of metalloproteinases, some of which have been implicated in AD[56]. ADAM proteins are
427 involved diverse functions, including immunity related pathways[57].

428 Regarding our multi-trait analysis of delirium with AD (MTAG), five genetic loci were found to have a
429 significant effect on delirium, supported by replication in the AoU EUR cohort. Among the replicated
430 loci, several important AD risk genes[58] were detected: *BIN1*, *CLU*, *CR1*, *MS4A4A*, *TOMM40*. The
431 *MS4A4A* gene is expressed in macrophages and has been linked with AD[59], vascular dementia and
432 systemic lupus erythematosus[60]. The lead variant's minor allele on the *MS4A4A* gene, rs1582763, has
433 been previously associated with decreased risk of AD[59]. This variant's association with delirium has
434 not been reported before, but we also found a protective effect of the rs1582763 minor allele on

435 delirium. *BIN1* has been recently implicated in the regulation of calcium homeostasis in glutamatergic
436 neurons, and its expression in AD human brains is reduced comparted to healthy brains[61]. The
437 clusterin gene (*CLU*), also named Apolipoprotein J (*APOJ*) codes for a multifactorial protein, with
438 apparent role in neurodegenerative diseases[62]. *CLU*, much like *APOE* is thought to be involved in
439 amyloid-β plaque deposition in AD pathologies. With regard to delirium, protein expression of
440 apolipoproteins including CLU and APOE were previously found to be downregulated in the
441 cerebrospinal fluid (CSF) of delirium subjects compared to mild AD controls[63]. In our proteomic
442 analysis, CLU protein levels were also downregulated in incident delirium subjects' plasma
443 (Supplementary Table 3), but not significantly so (p-value > 0.05). This discrepancy may reflect
444 different CLU protein abundance between CSF and plasma tissues. The *CR1* gene, implicated in
445 complement activation, is believed to exert its role to AD pathogenesis through amyloid-β clearance,
446 neuroinflammation and tauopathy (the deposition of abnormal tau protein in the brain)[64]. To the best
447 of our knowledge, the role of *CR1* in delirium has not been investigated previously. In our study, CR1
448 plasma protein levels had a nominally significant association with incident delirium (p-value = 0.013).
449 *TOMM40* genetic variants have been associated with AD before[65]. Given *TOMM40*'s close proximity
450 and high linkage disequilibrium with the *APOE* gene, its role in AD has frequently been contested[65].
451 However, it has also been suggested that *TOMM40* independently affects AD risk through its role in
452 regulating protein transportation in the mitochondria[65,66]. *TOMM40* has not been previously
453 implicated in delirium. Overall, given the small effect sizes of the MTAG-identified genes for delirium
454 (Supplementary Table 2), and their prominent role in AD, it may be possible that their significance in
455 our analysis is mainly driven due to their role in AD. Nonetheless, our findings suggest novel genes
456 (e.g. *MS4A4A*, CR1, *TOMM40*) that may be of relevance to future delirium research and therapeutic
457 targets investigations.

458 Previous proteomic studies for delirium have, to our knowledge, focused on the protein landscape
459 shortly before or following delirium episodes[31]. Here, an important novelty of our proteomic study is
460 the large follow-up period between protein measurements and delirium incidence (median time to
461 first delirium event: 11.4 years, interquartile range (IQR) [9.5-12.7]), allowing identification of relevant
462 biomarkers and biological mechanisms at an early stage, such as several neurologically relevant and
463 immune system related proteins.

464 For example, we found plasma GFAP and NEFL levels – biomarkers of neuronal injury – to be
465 associated with incident delirium. Previous research paints a complex, possibly bidirectional
466 relationship between neuronal injury and delirium. Increased plasma and CSF NEFL levels, a marker of
467 neuroaxonal injury, have been observed in delirium patients pre- and postoperatively[6,8,9,67,68]. It is
468 unclear however whether this is due to underlying neurodegeneration (e.g. due to preclinical

469 dementia) or a feature of delirium[6]. In our proteomic analyses we observed significant and consistent
470 effect of NEFL on incident delirium in both the full and dementia-free populations. This may implicate
471 neuroaxonal injury as an underlying driver of delirium independently of dementia-related
472 neurodegeneration, even long before hospitalisation. Our findings, thus, add support for the role NEFL
473 as a useful predictive biomarker for delirium[6].

474 GFAP is an astrogliosis marker, that is an increase of astrocytes levels in the central nervous system in
475 response to injury. High GFAP levels have been observed previously in post-mortem brains of delirium
476 cases[69], CSF of persistent delirium[8], blood and CSF of delirium patients undergoing elective surgery[32,67].
477 In our analysis, GFAP, was only predictive of incident delirium in our full proteomic set (Supplementary
478 Table 4), not in the dementia-free population (Supplementary Tables 8-9). This may indicate that, in
479 the years preceding delirium, astrogliosis is only relevant for delirium in the context of underlying
480 dementia. GFAP may be more useful as a delirium biomarker in the short term, possibly in patients
481 with traumatic brain injury or underlying neurodegeneration.

482 Systemic inflammation markers have been observed among the delirium-associated proteins in our
483 study. For instance, C7, BTLA, FGL1 participate in the immune response, whereas LRG1 and LTA4H in
484 inflammatory processes. LRG1 has been implicated in brain injury after sepsis in mice[70], sepsis being
485 a main driver of delirium aetiology[3]. Interleukins (IL), identified through our enrichment analysis, play
486 an important role in regulating immune response and inflammation and have frequently been
487 implicated in delirium[71,72]. Indeed, IL-6 – found to have a strong delirium association in our proteome
488 wide association study – is among the most consistently identified biomarkers of delirium and markers
489 of postoperative cognitive decline[31,32]. All of the inflammatory related proteins were significant in our
490 dementia-free sensitivity analyses, suggesting a role for delirium partially independent of dementia
491 pathology.

492 Other notable proteins robustly associated with delirium in our proteomic analysis are BCAN,
493 SELENOP, AREG and MSLN. BCAN, a protein with a role in brain extracellular matrix formation, has
494 been observed to be downregulated in brains of post-infection delirium and in AD patiens[73]. Lower
495 plasma levels of SELENOP, an important selenium transporter in the brain has been associated with
496 worse global cognition and AD[74]. Here, to the best of our knowledge, is the first time that an
497 association between SELENOP and delirium has been reported. The associations of AREG and MSLN
498 with delirium were also novel findings. AREG is a member of epidermal growth factor protein family,
499 involved in many aspects of cell proliferation[75]. Possible mechanisms linking AREG to delirium could
500 include neuroinflammation and regulation of T cells in astrocytes in sepsis-associated delirium[76], type
501 2 inflammatory response and tolerance against pathogens[77], or tumor development[75]. MSLN is a

502 mesothelial cell surface glycoprotein, overexpressed in many cancers[78]. Delirium is common in cancer
503 patients, arising either as a complication of the disease or indirectly through medical and surgical
504 intervention for cancer treatment[79]. Further research on the overlapping pathophysiology between
505 delirium and cancer (e.g. shared biomarkers[80]) is important for better management of delirium
506 throughout cancer care trajectories[79].

507 Overall, our proteomic analysis results align with the proposed mechanisms of delirium
508 pathophysiology, that is brain vulnerability - indicated by brain injury marker proteins -, systemic and
509 nervous system inflammation being driving factors for delirium[3,6]. At the same time, our results could
510 inform future research in delirium prediction biomarkers, some of which are novel in delirium research
511 (e.g SELENOP, CEND1, AREG and MSLN).

512 Our findings have notable clinical and biological clinical implications. We found that *APOE*-ε4 is a
513 robust genetic risk factor, remaining significant after adjustment for presence of dementia. The
514 findings suggest that *APOE*-ε4 confers vulnerability to adverse acute brain changes induced by triggers
515 of delirium. Indeed, it is known that *APOE*-ε4 is associated with exaggerated neuroinflammatory
516 responses[81], altered blood brain barrier integrity[82,83], increased beta-amyloid accumulation following
517 various injuries[84], and increased white matter pathology[85]. *APOE*-ε4 therefore could contribute to
518 increased risk of delirium, and also cause or amplify neuropathological changes linked with dementia.
519 A key next step is to test whether *APOE*-ε4 carriers are disproportionately likely to develop new
520 dementia after a delirium episode, even in the absence of prior cognitive impairment.

521 Clinically, *APOE*-ε4 status has potential to assist in stratification for delirium prevention trials, in
522 prediction of dementia following delirium, and as a target for neuroprotective therapies during acute
523 illness especially if delirium is present. Animal studies targeting APOE have shown encouraging results
524 in terms of improving AD phenotypes[83]. Some promising therapeutic approaches include increasing
525 APOE levels and lipidation, APOE mimetics (small peptides mimicking APOE's receptor-binding
526 structure) and *APOE*-ε4 gene therapy[43,83]. However, therapeutic translation in human clinical trials is
527 still limited[83]. A currently unpublished phase 2 clinical trial (NCT03802396)[86] found that administration
528 of CN-105 – an APOE mimetic drug – lowered postoperative delirium incidence and severity compared
529 to administration of placebo, but not significantly. Further clinical research is needed in order to clarify
530 the therapeutic potential of APOE targeting therapies.

531 In terms of delirium treatment, we explored whether our identified proteins could be suitable drug
532 targets. For this, we triangulated our proteomic study findings with proteome-wide mendelian
533 randomisation (MR) and colocalisation analyses. Incorporating genetics in drug target identification,
534 e.g. through MR, is particularity attractive, as drugs with genetic support are more likely to be

535 successful in clinical trials[87,88] or can offer drug repurposing opportunities[89]. Our druggability
536 assessment found PON3 as an notable drug target for delirium. PON3 is a secreted high density
537 lipoprotein-associated enzyme[90], participating in the metabolism of the statin lovastatin, a licenced
538 lipid-lowering medication. The role of statins in delirium management is unclear. Statin usage has
539 been associated with improved delirium outcomes in randomised controlled trials and observational
540 studies, but overall evidence is inconsistence, with a recent meta-analysis finding no significant
541 protective effect and high heterogeneity[91]. In our analysis, increased PON3 plasma level was found to
542 have protective effect against delirium. This is consistent with high PON3 levels being protective
543 against atherosclerosis and cardiovascular disease[92], possibly through its antioxidant action[92,93]. It is
544 not clear, however, how PON3 levels can affect statin metabolism. It may be possible that PON3 can
545 affect lovastatin and other statins' efficacy via their involvement in the drug activation[90], which may
546 explain part of the high heterogeneity observed regarding statins' effect on delirium[91]. Our results
547 may, thus, suggest a reconsideration of a possible role for statins in improving delirium prevention or
548 treatment.

549 Additionally, we identified proteins previously suggested as therapeutic targets for e.g. cancers
550 (ADAM8[94], GGH[95], PILRA[96]) and AD (PILRA[97]). This may offer promising opportunities for investigating
551 new delirium treatment targets. However, results should be interpreted with caution, as liberal p-
552 value thresholds – up to $p < 0.05$ – were used to indicate proteomic support for our druggability
553 assessment, and some proteins (e.g. PON3 and ADAM8) were not replicated in our replication MR,
554 although the same effect direction was still observed (Table 1 and Supplementary Data 1).

555 The main strength of our analysis is the large-scale investigation of delirium genetic and proteomic
556 risk factors. Both in terms of sample size, follow-up time before protein measurement and delirium
557 episodes, and number of genetic variants / proteins tested, this is to our knowledge the largest study
558 on the molecular background of delirium risk conducted so far. On the other hand, some of the
559 limitations of the study include the underdiagnosis of delirium in hospital health records from which
560 the phenotype was mostly derived[98]. This misclassification could introduce noise to the results,
561 limiting discovery of genetic and protein effects. Also, the small sample sizes of the non-European sub-
562 populations hinders the identification of risk factors specific to them. Moreover, not the full spectrum
563 of the human proteome is captured in the assayed plasma proteins, potentially missing proteins
564 important for delirium biology and prognosis. Finally, although the use of plasma proteins as predictive
565 biomarkers is of great significance, proteomic profiles of delirium-relevant tissues, such as brain or
566 CSF would be invaluable.

567 In conclusion, our results point out to an oligogenic genetic architecture for delirium, with the *APOE*
568 locus identified as a strong, potentially population specific genetic risk factor, independently of
569 dementia. However, further replication in larger non-European cohorts is required. Our plasma
570 proteome analysis supports previous findings and discovers putatively novel proteins implicated to
571 delirium, some of which suggested for therapeutic applications. Taken together, genetic and
572 proteomic risk factors suggest a shared aetiology between delirium and dementias, possibly
573 contributing to a better understanding of delirium's complex biological origin and the discovery of
574 clinically relevant biomarkers.

575

576 # Methods

577 **Study populations**

578 The project utilises biomedical data from ancestrally diverse large-scale cohorts. Included cohorts
579 were either (a) databases containing individual-level genomic measurements linked to healthcare
580 records, or (b) previously published summary results from genomic studies on delirium phenotypes.
581 Contributing individual-level cohorts include the UK Biobank[99,100] (UKB) and the All of Us Research
582 Program[101] (AoU). Summary results have been obtained from ancestrally Finnish (FinnGen[102]; $n_{cases} =$
583 $3{,}371$; $n_{controls} = 388{,}560$) and European participants (Michigan Genomics Initiative cohort
584 (MGI)[103]; $n_{cases} = 160$; $n_{controls} = 44{,}654$).

585 The UKB is a population-based prospective study, containing a rich set of genetic and phenotypic data
586 for approximately 500,000 participants living across the United Kingdom. Participants, aged 40 to 69
587 years old at recruitment between 2006 – 2010, have been linked to their annually updated electronic
588 health records, allowing longitudinal investigation of healthcare outcomes. Similarly, AoU includes
589 genomic data and healthcare outcomes for approximately 245,000 individuals from diverse
590 populations in the USA.

591 Age distributions for UKB and AoU participants, along with delirium and all-cause dementia prevalence
592 by age group, can be found in Supplementary Figure 1. Age distributions for the FinnGen and MGI
593 cohort can be found in their respective original publications, that is Kurki et al. (2023; Supplementary
594 Methods)[102] and Zawistowski et al. (2023)[103] respectively. Age distribution of delirium first events in
595 FinnGen can be found here: https://r10.risteys.finregistry.fi/endpoints/F5_DELIRIUM.

596 **Delirium phenotype**

597 The analysis focused on delirium episodes that were not triggered by substance intoxication or
598 withdrawal[2]. For convenience, such delirium episodes will hereby be referred as simply delirium.
599 Delirium cases were defined as individuals with one or more delirium-corresponding codes in their
600 electronic health records (EHR), that is hospital inpatient, death register or primary care data. The
601 relevant codes were: "F05" (delirium, not induced by alcohol and other psychoactive substances) for
602 International Classification of Diseases, 10th version (ICD-10)[104] and "293.0" (Acute confusional state)
603 for ICD-9[105]. Read v2 and v3 codes[106] for primary care data mapping to delirium were obtained from a
604 previously defined list by Kuan et al (2019)[107], published in the HDRUK Phenotype Library
605 (https://phenotypes.healthdatagateway.org/).

606 **Discovery of genetic risk factors**

607 A Genome-Wide Association Study (GWAS) framework was implemented in order to identify genetic
608 variants associated with delirium in UKB's and AoU's ancestrally distinct sub-populations. For this
609 purpose, the REGENIE software (version 3.2.2) was used, which carries out a logistic regression
610 analysis between a disease phenotype and each genetic variant, accounting for covariates, population
611 structure and relatedness of participants[108].

612 In UKB, the set of imputed genotypes was used[99] (Data-Field 22828), filtered to include variants with
613 > 5 minor alleles in cases and controls, imputation score > 0.5, missingness rate < 3% and deviation
614 from Hardy-Weinberg Equilibrium (HWE) with p-value < 10^{-6}. Individuals were filtered to include those
615 with missingness rate < 5%, no mismatch between reported and genetically inferred sex (Data-Field
616 22001), no sex chromosome aneuploidy (Data-Field 22019), no excessive heterozygosity (Data-Field
617 22027) and no more than ten 3rd degree relatives (Data-Field 22021). The covariates considered for
618 the UKB GWAS included: age, sex, genotyping batch (Data-Field 22000) and the first 20 pre-computed
619 genomic principal components (data-field 22009). Here, age was defined as age at first delirium
620 occurrence for cases and age at last data freeze (31 October 2022) or age at death for controls. GWAS
621 were conducted separately for sub-populations of white British ancestry (EUR; $n_{cases} = 7{,}176$;
622 $n_{controls} = 385{,}097$), African (AFR; black / black British; $n_{cases} = 115$; $n_{controls} = 7{,}480$) and south
623 Asian (SAS; $n_{cases} = 107$; $n_{controls} = 9{,}356$) ethnic backgrounds. Summary statistics from the UKB
624 GWAS were converted from GRCh37 to GRCh38 genomic coordinates using the LiftOver software[109].
625 In total the analysis covered approximately 22.5, 13.6 and 8.8 million genetic variants in EUR, AFR, SAS
626 ancestries respectively.

627 In AoU, short read whole genome sequencing genotypes were used[101] for conducting GWAS on
628 European (EUR; $n_{cases} = 652$; $n_{controls} = 119{,}817$), African/African American (AFR; $n_{cases} = 233$;
629 $n_{controls} = 52{,}300$), and admixed American/Hispanic (AMR; $n_{cases} = 117$; $n_{controls} = 39{,}977$) sub-

630 populations. The same GWAS framework as described for UKB was followed, with the exception of
631 not including genotyping batch and principal components 11-20 as covariates, as they were not
632 available in the AoU datasets. Sub-populations with a low number of delirium cases (<20) were
633 excluded from the analysis. Those consisted of east Asian ethnic background in UKB and east Asian
634 and middle Eastern genetic ancestries in AoU.

635 In order to increase power to detect genetic associations, our GWAS summary statistics and previously
636 published GWAS results were combined into a multi-ancestry genome-wide meta-analysis. The METAL
637 software (version 2020-05-05)[110] was used to conduct a fixed effects inverse-variance meta-analysis
638 on the set of 24,951,029 variants that were present in at least two studies. The genomic control
639 correction method was applied in METAL to allow for multi-ancestry analysis. In total, up to 1,059,130
640 individuals ($n_{cases} = 11{,}931$; $n_{controls} = 1{,}047{,}199$) were included in the meta-analysis. Genome-
641 wide significance was considered at a $5x10^{-8}$ p-value threshold. Significantly associated variants at the
642 multi-ancestry meta-analysis were inspected for consistency at each contributing sub-cohort.
643 Descriptive statistics for each sub-cohort are presented in Supplementary Table 1.

644 Wherever reported, Odds Ratios (OR) were calculated as $OR = e^{\beta}$, where β is the logistic regression
645 coefficient. 95% Confidence Intervals (CI) for the ORs were calculated as $OR_{95\%\,CI} = e^{\beta} \pm 1.96 *$
646 $SE_{\beta} * e^{\beta}$.

647 **Sensitivity analyses**

648 A series of sensitivity analyses were conducted in addition to our main genetic analysis, to:

649 - identify genetic variants associated with delirium independently of *APOE*. For this purpose,
650 we performed a GWAS on the UKB EUR cohort conditional on the *APOE*-ε4 haplotype count
651 (0, 1 or 2) (Supplementary Figure 2). We inferred *APOE*-ε4 haplotypes for each participant
652 based on their rs429358 and rs7412 genotypes, as described in previous studies[35].
653 - Identify genetic signals for delirium independently of dementia. To this end, we conducted in
654 the UKB EUR cohort (a) delirium GWAS conditional on dementia (dementia-adjusted), by
655 including all-cause dementia status as covariate (Supplementary Figure 3); (b) delirium GWAS
656 on sub-cohorts stratified by all-cause dementia status (i.e. whether or not individuals have
657 developed all-cause dementia at any time point) (Supplementary Figure 4); and similarly (c)
658 delirium GWAS on sub-cohorts stratified by Alzheimer disease (AD) status (Supplementary
659 Figure 5).

660 • better distinguish genetic signals in true delirium cases and controls in the elderly, by
661 conducting a delirium GWAS on UKB EUR participants aged 60 years or older (Supplementary
662 Figure 6).

663 For all sensitivity analyses, the same GWAS framework as in our main analysis of genetic risk factors
664 was followed. All sensitivity analyses were repeated in the AoU EUR cohort for replication (significance
665 p-value threshold < 0.05). Sample sizes for each analysis and results for rs429358, the top delirium-
666 associated genetic variant from our main genetic risk factor analysis, are presented in Supplementary
667 Table 2.

668 For the analyses that utilised dementia or AD status in UKB, the algorithmically defined all-cause
669 outcomes were used (Data-Field 42018), provided by UKB. Briefly, disease outcomes were constructed
670 based on self-reported medical conditions reported at the baseline assessment, ICD-9/10 codes in
671 hospital diagnoses and procedures, and ICD-10 codes in Death Register records. The full list of ICD
672 codes for all-cause dementia and AD is included in Supplementary Table 16 (as reported in UKBiobank
673 Recourse 460, https://biobank.ctsu.ox.ac.uk/ukb/refer.cgi?id=460). All-cause dementia and AD
674 diagnoses codes for the AoU cohort are reported in Supplementary Table 17.

675 **Multi-trait analysis between delirium and Alzheimer disease**

676 Given the close inter-relationship between delirium and AD[6], we wanted to quantify the shared
677 genetic architecture between the two traits. For this purpose, we used the high-definition likelihood
678 (HDL) software (v1.4.0)[111] to estimate the SNP heritability (h^2_{SNP}) of delirium and the genetic
679 correlation (r_g) between delirium and AD using GWAS summary statistics. SNP heritability, the
680 proportion of phenotypic variance explained by the additive SNP genetic factors[112], was estimated in
681 the observed scale, and converted to the liability scale using the equation: $h^2_{SNP,l} =$
682 $h^2_{SNP}\frac{K(1-K)}{z^2}\frac{K(1-K)}{P(1-P)}$, where K=0.015 is the population prevalence of delirium[4], P=0.018 is delirium
683 prevalence in UKB, and z is standard normal probability density function's height at the quantile
684 corresponding to cumulative probability K. The equation is derived from Lee et al (2011)[113]. Genetic
685 correlation is a measure of the variance shared by two traits due to genetic causes[112]. For AD, we used
686 the largest to-date AD GWAS meta-analysis, Bellenguez et al (2022)[58]. The AD GWAS was conducted
687 on 487,511 European individuals (Stage I: 39,106 clinically diagnosed AD cases and 46,828 proxy AD
688 cases) and 21 million variants. AD summary statistics were obtained from European Bioinformatics
689 Institute GWAS Catalog (https://www.ebi.ac.uk/gwas/) under accession no. GCST90027158. Delirium
690 summary statistics were taken from our UKB EUR GWAS sub-cohort, to better match the ancestry of
691 the AD GWAS summary statistics and the HDL pre-computed European-ancestry reference panel.

692 We also applied a joint analysis of summary statistics between delirium and AD. Such approach
693 increases statistical power to detect genetic associations for each trait[34]. The multi-trait analysis of
694 GWAS (MTAG) software (v1.0.8) was used for this purpose[34], jointly analysing summary statistics from
695 our delirium meta-analysis – excluding the AoU sets – and the Bellenguez et al[58] AD summary statistics.
696 The MTAG analysis for the discovery of delirium genetic risk variants was conducted on 9,883,704
697 SNPs that overlapped across the two disorders, filtered for minor allele frequency >= 0.01 and sample
698 size N >= (2/3) * 90th percentile for each trait. MTAG results are trait-specific summary statistics (i.e.,
699 effect estimates, standard errors and p-values), interpreted similarly with single-trait GWAS results.
700 The genome-wide significance threshold was defined as a p-value = $5x10^{-8}$. Independent lead variants
701 were defined as the most significant variants within a ±500kb region, using the GWASLab python
702 package (version 3.4.46)[114].

703 The AoU EUR set was held out for replication of the MTAG lead hits. A multi-trait analysis with AD was
704 conducted on the replication set as described above. Lead variants were considered replicated if they
705 had a Bonferroni adjusted p-value < (0.05 / number of lead variants) and same direction of effect
706 across the discovery and replication set.

707 **Mediation Analysis**

708 We employed a mediation analysis framework in order to quantify to what extend the effect of *APOE*
709 gene on delirium is mediated through underlying dementia status. Mediation analysis enables the
710 decomposition of an exposure's total causal effect on an outcome, here *APOE* and delirium
711 respectively, into direct and indirect effect through a hypothesised mediator variable (here
712 dementia)[53]. We used the R package 'medflex'[53] (v0.6-10) to calculate the so-called natural direct and
713 natural indirect effects of the number of *APOE*-ε4 haplotypes (0,1,2) on delirium, mediated through
714 dementia. When exponentiated, natural direct and indirect effect estimates can be interpreted as
715 odds ratios. Medflex's weighting-based approach was used, to estimate the unobserved potential
716 outcomes. For the natural effect model's parameters, bootstrap-based (n=1,000) standard errors and
717 Wald-type two-sided p-values were obtained. Sex and age were used as covariates.

718 Mediation analysis, as implemented under the counterfactual framework in 'medflex', relies on the
719 assumption that, after adjusting for baseline covariates, there is no unmeasured confounding
720 between (i) exposure and outcome, (ii) exposure and mediator, and (iii) mediator and outcome[53]. It is
721 also assumed that (iv) there are no confounders of the mediator-outcome relationship that are
722 themselves induced by the exposure[53]. We assessed the robustness of our mediation analysis to
723 violations of unmeasured confounding assumptions, using the E-value sensitivity analysis method[36,115].
724 An E-value, calculated as $E - value = RR + \sqrt{RR * (RR - 1)}$, where $RR \approx \mathrm{OR}, \text{for } \mathrm{OR} > 1$ is the

725 observed direct or indirect effect risk ratio[115], can be interpreted as the minimum unmeasured
726 confounder effect size required to completely explain away the observed effect[36].

727 We also conducted a sensitivity mediation analysis additionally adjusted for measurements of
728 cognitive function, socioeconomic deprivation and chronic disease burden, to assess the robustness
729 of our analysis to common dementia and delirium risk factors[4]. General cognitive ability was calculated
730 as a composite measure – i.e. the first unrotated principal component – of four UKB cognitive tests:
731 fluid intelligence (Data-Field 20016), numeric memory (Data-Field 4282), reaction time (Data-Field
732 20023, log transformed) and visual memory (Data-Field 399, log1p transformed), as described and
733 validated in previous studies[116,117]. When cognitive tests were unavailable at baseline, measurements
734 of the same tests at repeat assessments or online follow-up were used if available. The Townsend
735 deprivation index immediately before joining UKB (Data-Field: 22189, decile transformed) was used
736 as a measure of socioeconomic deprivation[118]. The Charlson Comorbidity Index (CCI) was calculated
737 as an estimate of chronic disease burden[119]. CCI is a weighted sum of 12 chronic comorbid diseases,
738 here calculated at the time of delirium episode. Diseases were extracted from self-reported medical
739 conditions and cancers (Data-Fields: 20002 and 20001), and ICD-9 and ICD-10 disease codes[120] from
740 linked healthcare records. Weights were obtained from Quan et al[119]. This sensitivity analysis was
741 conducted on a subset of n = 141,864 individuals, due to missing cognitive function data.

742 **Proteomic study population**

743 Plasma proteome data were available in UKB for a subset of 53,075 participants. Protein
744 measurements of 2,923 unique plasma proteins[121] were derived from blood samples taken during
745 randomly selected participants' initial UKB assessment visit between 2006-2010. Proteins were
746 measured using the antibody-based Olink Explore 3072 proximity extension assay. Proteome data
747 have been previously undergone extensive quality control[121]. As additional filtering in the present
748 analysis, European participants from batches 0 to 6 were extracted, as they have been reported to be
749 highly representative of the UKB European population[121]. Moreover, protein measurements with
750 >20% missing data were removed and the remaining proteins were mean-imputed, inverse-rank
751 normalised and standardised (mean zero standard deviation 1) to ensure homogeneity across the
752 proteins. Delirium incident cases were defined as the participants whose first reported delirium
753 episode was > 1 years after baseline, that is, the date of blood sample collection at the first UKB
754 assessment visit (Data-Field 53-0.0). Delirium data were available for up to 16 years of follow-up after
755 baseline (censoring date: 31 October 2022). The final population consisted of 32,652 European
756 participants and 2,919 plasma proteins. 541 participants had an incident delirium diagnosis within the
757 follow-up period (cases). Of the remaining 32,111 participants (controls), 29,440 reached the end of

758 follow-up without a delirium diagnosis, and 2,671 died before the end of follow-up. Follow-up
759 information was available for all participants.

760 **Discovery of protein risk factors**

761 To explore the relationships between baseline protein levels and incident delirium, we performed a
762 proteome-wide association study (PWAS) on the UKB proteomic study population. Multivariable
763 logistic regression models were fit between each protein as predictor and incident delirium status as
764 outcome. In total, 2,919 models were fit, equal to the number of proteins. The models were adjusted
765 for sex, Body Mass Index (BMI) and age at baseline. Associations were deemed significant at a
766 Bonferroni adjusted p-value threshold: p-value < 1.7×10^{-5}. For sensitivity analyses, we further adjusted
767 protein models for *APOE*-ε4 haplotype status – zero, one or two copies of the haplotype – and for
768 interaction between each protein and *APOE*-ε4.

769 We performed a pathway enrichment analysis on the set of Bonferroni-adjusted significant proteins
770 emerging from the protein-wide association analysis. We used the Enrichr[122] web-based tool and
771 pathway annotations based on the Reactome[123], Molecular Signatures Database (MSigDB)[124] and
772 Kyoto Encyclopedia of Genes and Genomes (KEGG)[125] databases. The full set of Olink proteins was
773 used as background genes. Fisher's exacts tests were implemented to assess whether the identified
774 proteins significantly overlap with the proteins in any of the pathways. Q-values were obtained by
775 adjusting p-values for multiple testing using the Benjamini-Hochberg method.

776 **Protein selection and prediction models**

777 To investigate whether baseline plasma proteome can improve prediction of incident delirium, a
778 supervised machine learning approach was implemented. The LASSO (Least Absolute Shrinkage and
779 Selection Operator) method[37] was utilised for the selection of important proteins and to avoid
780 overfitting given the high multicollinearity of proteomics data. For this analysis, the full set was
781 randomly split into a training (80%; $n_{cases} = 436$; $n_{controls} = 25{,}733$) and test (20%; $n_{cases} =$
782 105; $n_{controls} = 6{,}378$) set. A LASSO model for binary outcomes was implemented in the training set
783 using the glmnet R package (version 4.1.8)[126]. Here, the whole set of 2,919 proteins adjusted for
784 demographic covariates: age, sex and BMI were used as predictors of incident delirium. In brief, the
785 coefficients penalty parameter lambda was tuned using a 10-fold Cross-Validation (CV) framework for
786 100 lambdas between 10^{-6} and 0.07. The model with lambda.1se was chosen as the most
787 parsimonious, giving the strictest model such that cross-validated error is within one standard error
788 of the minimum[126]. To increase the robustness of the LASSO protein selection, a stability selection[38]
789 approach was additionally applied. For each of 100 random subsampling iterations of the training set,
790 including all 436 delirium cases and an equal number of 436 randomly selected controls, a LASSO

model as described above was fit using the penalty factor tuned in the full training set. The proteins that were selected on at least half of the subsampling iterations were chosen as robustly selected (stability-selected proteins).

Four logistic regression models with incident delirium as outcome were subsequently re-fit in the training set: (a) using only demographic covariates (age, sex and BMI) as predictors (basic model); (b) using only the stability selected proteins as predictors (proteomic model); (c) using both demographic covariates and the stability selected proteins as predictors (proteomic + basic model); and (d) using demographics, stability selected proteins and *APOE*-ε4 haplotype status as predictors (*APOE* + proteomic + basic model). For the models that included proteins as predictors (b,c and d), stepwise regression models were fit, starting with all the stability-selected proteins and removing predictors until no AIC improvement was observed.

The performance of the models was evaluated in the held-out test set. Receiver operating characteristic (ROC) curves, Area Under the ROC Curve (AUC) and Precision – Recall curves and AUC (PR-AUC) estimates were used to compare the predictive performance of the three models in the test set. Precision – Recall metrics were chosen as they are more sensitive to binary outcome imbalance[127], as is the situation here. Two-sided Delong tests were used to compare whether AUCs were significantly different between each model pair[128] (ΔAUC p-value < 0.05).

pQTL analysis

We obtained protein quantitative trait loci (pQTL), that is genetic variants affecting plasma protein levels, from the UKB proteomic study population. A GWAS was conducted for each of the 2,923 inverse-rank normalised and standardised protein traits. Autosomal and X chromosome genetic variants were derived from UKB's called genotypes (Data-Field 22418), filtered to include variants with missingness < 1%, HWE deviation p-value > 10^{-15}, minor allele frequency > 1%. The proteomic study population was further filtered to include unrelated, up to 2nd degree, individuals, with missingness rate < 10%. The GWAS on each protein, were conducted using REGENIE (version 3.2.2)[108] on 30,272 individuals and ~540,000 genetic variants. Age, age^2, sex, sex*age, and sex*age^2, batch, UKB assessment centre, UKB genetic array, time between blood sampling and measurement, and the first 20 genetic principal components were used as covariates.

2,464 proteins were significantly associated with at least one genetic variant ($p < 5x10^{-8}$). Per protein, independent pQTLs were obtained through linkage disequilibrium (LD)-based clumping, including variants with LD $r^2 < 0.2$ within ±250 kb windows of the top associated pQTLs. Genotype filtering and LD-clumping was performed using PLINK (v1.90b)[129].

823 The independent pQTLs derived here were used as genetic instruments for protein levels in the
824 subsequent mendelian randomisation and colocalisation analyses.

825 **Mendelian randomisation**

826 Mendelian randomisation (MR) is a method used to assess potential causal influence of a modifiable
827 exposure on an outcome, often disease risk. MR analyses are based on the use of genetic variants as
828 instrumental variables (IV). IVs are variables associated with an exposure but not with the outcome of
829 interest through any other pathway[130,131]. Three assumptions are required for MR to be valid: (1) IVs
830 are significantly associated with the exposure; (2) there are no confounders of the IVs and the
831 outcome; and (3) IVs do not affect the outcome other than through the exposure (no pleiotropy)[130].

832 We assessed whether the identified protein risk factors (exposures) showed causal effects on delirium
833 (outcome), using a two-sample MR approach[132]. Genetic instruments for protein exposure traits
834 comprised of the per-protein independent pQTLs, as derived from our pQTL analysis in UKB (p-value
835 $< 5x10^{-8}$, LD $r^2 < 0.2$). To further ensure no weak instrument bias in our genetic instrument pQTLs -
836 relevant to MR assumption (1) - we also excluded pQTLs with an F-statistic < 10 in their per-protein
837 exposure GWAS. F-statistics were approximated as $F = \frac{\widehat{\gamma_j^2}}{\sigma_j^2}$ for each pQTL j, where $\hat{\gamma}$ is the pQTL-
838 exposure effect size and σ its respective standard error[133]. To minimise chances of including genetic
839 instruments with pleiotropic effects, we included only (a) pQTLs associated with less than 5 proteins
840 (p-value $< 5x10^{-8}$), and (b) cis-pQTLs, that is pQTLs in close proximity with their associated protein's
841 encoding gene. cis-pQTLs were defined as being located within ±1Mb (10^6 base pairs) of the protein-
842 encoding gene's coding region. Protein-encoding genes' start and end positions were derived as
843 reported in the metadata from Sun et al[121]. Genetic associations of the pQTLs with delirium were
844 derived from our UKB-EUR GWAS excluding the individuals in the proteomic set, to avoid sample
845 overlap in our two-sample setting ($n_{cases} = 6{,}650; n_{controls} = 354{,}251$). The same procedure as our
846 full UKB-EUR GWAS was followed, as described in the section 'Discovery of genetic risk factors'.

847 For our main MR analysis, 1989 proteins were tested in total. The inverse variance weighted (IVW)
848 method was used[134] to assess causal estimates between 1,738 proteins with > 1 cis-pQTLs and
849 delirium. For 251 proteins with 1 cis-pQTLs the Wald ratio method was used. We further employed
850 the weighted median (WM), maximum likelihood and MR-Egger sensitivity MR methods[135], to test for
851 consistency of estimates with our main analysis. The MR-Egger intercept and the Cochran's Q tests
852 were used to assess for the presence of horizontal pleiotropy and heterogeneity in the IVs
853 respectively[136–138]. MR associations were considered significant at an FDR q < 0.05 in our main IVW or
854 Wald ratio MR analysis. For significant proteins, IVW or Wald ratio MR were repeated using delirium

855 in the dementia-free UKB sub-cohort as outcome GWAS, again using an FDR q threshold of 0.05.
856 Finally, the FinnGen delirium GWAS was used as a replication MR outcome set, following the same MR
857 framework as described above. A significance threshold of p-value < 0.05 was used in the replication
858 MR. Full MR results are reported in Supplementary Table 12 and 13. All MR analyses were conducted
859 using the TwoSampleMR R package (version 0.5.11)[135]. A full list of the cis-pQTLs used, together with
860 their association statistics with each exposure and delirium is available in Supplementary Data 3. A
861 STROBE-MR checklist[131] for the conducted MR analyses can be found in the Supplementary Table 18.

862 **Colocalisation**

863 We tested whether genetic signals for delirium and each of the plasma proteins colocalise, i.e. the two
864 traits are affected by the same, rather than distinct genetic variants in a specific genomic region[40]. For
865 each of the 2,456 proteins with at least 1 pQTL, we analysed genomic regions around lead pQTLs with
866 boundaries defined by recombination hotspot locations, based on a previously published map[139]
867 (https://bitbucket.org/nygcresearch/ldetect-data/src/master/EUR). We used the coloc R package
868 (v5.2.3)[39], specifically the Approximate Bayes Factor colocalisation method (coloc.abf), which assumes
869 at most one causal variant per trait per region[39]. Using per-trait GWAS results and the default prior
870 probabilities of $\mathrm{p1} = \mathrm{p2} = 10^{-4}$ for each variant being the causal for delirium or the analysed protein
871 respectively, and $p12 = 10^{-5}$ for each variant being the causal for both, coloc.abf calculates per-
872 variant posterior probabilities that the two traits have a common causal genetic variant in a genomic
873 region (PP.H4 as per coloc notation). Proteins and delirium were considered to share a causal genetic
874 variant when PP.H4 > 0.9. Suggestive colocalisation was considered for protein-delirium pairs with
875 PP.H4 > 0.5.

876 **Druggability assessment**

877 We further investigated whether selected delirium-associated proteins are suitable to act as potential
878 drug targets. The selected proteins included those identified as significant in both our protein risk
879 factor (p < 0.05) and MR analyses (FDR q < 0.05), as well as having a consistent direction of effect in
880 the two analyses. We utilised the "druggable genome" resource developed by Finan et al (2017)[41]. In
881 this resource 4,479 genes encoding drugged or druggable proteins were systematically identified, and
882 further stratified into 3 tiers based on druggability evidence strength; as described by Finan et al: "Tier
883 1 (1427 genes) included efficacy targets of approved small molecules and biotherapeutic drugs as well
884 as clinical-phase drug candidates. Tier 2 was composed of 682 genes encoding targets with known
885 bioactive drug-like small-molecule binding partners as well as those with ≥50% identity (over ≥75% of
886 the sequence) with approved drug targets. Tier 3 contained 2370 genes encoding secreted or
887 extracellular proteins, proteins with more distant similarity to approved drug targets, and members

888 of key druggable gene families not already included in tier 1 or 2."[41] We therefore tested whether our
889 selected proteins overlap the with druggable genes' encoding proteins. For proteins in Tier 1
890 druggability we searched the Drug Bank (https://www.drugbank.ca/) and https://clinicaltrials.gov/
891 databases for additional information on drug indications and clinical trial status respectively. The
892 druggable genome can be found in the Supplementary Table 1 of the original Finan et al publication[41].

893 **KGWAS**

894 We implemented Knowledge Graph GWAS (KGWAS), a novel geometric deep learning approach that
895 integrates GWAS summary statistics with functional genomics data (variant and gene-level
896 annotations and interactions), to improve power in GWAS association testing[42]. Using functional
897 genomics knowledge graphs, KGWAS estimates variants' prior relevance to disease and recomputes
898 the original GWAS associations in a p-value weighting framework. We used our delirium GWAMA
899 summary statistics and KGWAS's default fast mode, which uses Enformer embedding for variant
900 feature and ESM embedding for gene features. We further used MAGMA[140], as integrated in KGWAS,
901 for gene prioritisation of KGWAS significant associations. We used a p-value threshold of 0.05 / 16,637
902 = $3x10^{-6}$ for gene-based p-values estimated by MAGMA, 16,637 being the number of genes tested.

903

904 # Acknowledgments

905 This work used the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/).

906 This research has been conducted using the UK Biobank Resource projects 788 and 44986. This work
907 uses data provided by patients and collected by the NHS as part of their care and support.

908 UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as
909 a Research Tissue Bank (RTB) approval.

910 We gratefully acknowledge *All of Us* participants for their contributions, without whom this research
911 would not have been possible. We also thank the National Institutes of Health's *All of Us* Research
912 Programme for making available the participant data examined in this study.

913 We want to acknowledge the participants and investigators of the *FinnGen* study.

914 The authors acknowledge the Michigan Genomics Initiative participants, Precision Health at the
915 University of Michigan, the University of Michigan Medical School Central Biorepository, and the
916 University of Michigan Advanced Genomics Core for providing data and specimen storage,
917 management, processing, and distribution services, and the Center for Statistical Genetics in the

Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of the research reported in this publication.

This research was funded by the *Legal & General Group* (research grant to establish the independent Advanced Care Research Centre at University of Edinburgh). The funder had no role in conduct of the study, interpretation or the decision to submit for publication. The views expressed are those of the authors and not necessarily those of Legal & General.

For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Data availability

Details for accessing individual-level data can be found here:

- for UK Biobank https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access
- for All of Us Research Programme https://www.researchallofus.org/register

Details on obtaining delirium GWAS summary statistics used in this work can be found here:

- for FinnGen R10 release https://www.finngen.fi/en/access_results
- for MGI freeze 3 https://precisionhealth.umich.edu/our-research/michigangenomics

GWAS summary statistics generated in this work will be made publicly available via University of Edinburgh's Datashare service [DOI to be released upon publication]. Summary statistics for the proteomic, mendelian randomisation and colocalisation analyses can be found in the Supplementary Tables file and Supplementary Data 2 and 3.

Code availability

All software used in the present study is publicly available. The code used for all analyses in the study can be found on GitHub: https://github.com/VasiliosRaptis/deliriumGen.

References

1. Anand, A. & MacLullich, Alasdair. M. J. Delirium in older adults. *Medicine* **49**, 26–31 (2021).

945 2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*.
946 (American Psychiatric Association, 2013). doi:10.1176/appi.books.9780890425596.

947 3. Wilson, J. E. *et al.* Delirium. *Nat Rev Dis Primers* **6**, 90 (2020).

948 4. Inouye, S. K., Westendorp, R. G. & Saczynski, J. S. Delirium in elderly people. *The Lancet* **383**,
949 911–922 (2014).

950 5. Gordon, E. H., Ward, D. D., Xiong, H., Berkovsky, S. & Hubbard, R. E. Delirium and incident
951 dementia in hospital patients in New South Wales, Australia: retrospective cohort study. *BMJ*
952 e077634 (2024) doi:10.1136/bmj-2023-077634.

953 6. Fong, T. G. & Inouye, S. K. The inter-relationship between delirium and dementia: the
954 importance of delirium prevention. *Nat Rev Neurol* **18**, 579–596 (2022).

955 7. Watne, L. O. *et al.* Cerebrospinal fluid quinolinic acid is strongly associated with delirium and
956 mortality in hip-fracture patients. *Journal of Clinical Investigation* **133**, e163472 (2023).

957 8. Tsui, A. *et al.* Persistent delirium is associated with cerebrospinal fluid markers of neuronal
958 injury. *Brain Communications* **6**, fcae319 (2024).

959 9. Krogseth, M. *et al.* Delirium, neurofilament light chain, and progressive cognitive
960 impairment: analysis of a prospective Norwegian population-based cohort. *The Lancet Healthy*
961 *Longevity* **4**, e399–e408 (2023).

962 10. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer disease. *Nat Rev Immunol* (2024)
963 doi:10.1038/s41577-024-01104-7.

964 11. Leslie, D. L. & Inouye, S. K. The Importance of Delirium: Economic and Societal Costs. *J*
965 *American Geriatrics Society* **59**, (2011).

966 12. Sepulveda, E. *et al.* The complex interaction of genetics and delirium: a systematic review
967 and meta-analysis. *Eur Arch Psychiatry Clin Neurosci* **271**, 929–939 (2021).

968 13. Massimo, L. *et al.* Genetic and environmental factors associated with delirium severity in
969 older adults with dementia. *Int J Geriat Psychiatry* **32**, 574–581 (2017).

970 14. Van Munster, B. C. *et al.* Genetic polymorphisms in the DRD2, DRD3, and SLC6A3 gene in
971 elderly patients with delirium. *American J of Med Genetics Pt B* **153B**, 38–45 (2010).

972 15. Ely, W. E. *et al.* Apolipoprotein E4 polymorphism as a genetic predisposition to delirium in
973 critically ill patients: *Critical Care Medicine* **35**, 112–117 (2007).

974 16. Kim, J. *et al.* Antipsychotics and dopamine transporter gene polymorphisms in delirium
975 patients. *Psychiatry Clin Neurosci* **59**, 183–188 (2005).

976 17. Van Munster, B. C., Korevaar, J. C., De Rooij, S. E., Levi, M. & Zwinderman, A. H. The
977 association between delirium and the apolipoprotein E ε4 allele in the elderly. *Psychiatric Genetics*
978 **17**, 261–266 (2007).

979 18. Van Munster, B. C., Zwinderman, A. H. & De Rooij, S. E. Genetic Variations in the Interleukin-
980 6 and Interleukin-8 Genes and the Interleukin-6 Receptor Gene in Delirium. *Rejuvenation Research*
981 **14**, 425–428 (2011).

982 19. Leung, J. M. *et al.* Apolipoprotein E e4 Allele Increases the Risk of Early Postoperative
983 Delirium in Older Patients Undergoing Noncardiac Surgery. *Anesthesiology* **107**, 406–411 (2007).

984 20. Cunningham, E. L. *et al.* Observational cohort study examining apolipoprotein E status and
985 preoperative neuropsychological performance as predictors of post-operative delirium in an older
986 elective arthroplasty population. *Age and Ageing* **46**, 779–786 (2017).

987 21. Van Munster, B. C., Korevaar, J. C., Zwinderman, A. H., Leeflang, M. M. & De Rooij, S. E. J. A.
988 The Association Between Delirium and the Apolipoprotein E Epsilon 4 Allele: New Study Results and
989 a Meta-Analysis. *The American Journal of Geriatric Psychiatry* **17**, 856–862 (2009).

990 22. Terrelonge, M. *et al.* KIBRA, MTNR1B, and FKBP5 genotypes are associated with decreased
991 odds of incident delirium in elderly post-surgical patients. *Sci Rep* **12**, 556 (2022).

992 23. Adamis, D., Meagher, D., Williams, J., Mulligan, O. & McCarthy, G. A systematic review and
993 meta-analysis of the association between the apolipoprotein E genotype and delirium. *Psychiatric*
994 *Genetics* **26**, 53–59 (2016).

995 24. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The*
996 *American Journal of Human Genetics* **101**, 5–22 (2017).

997 25. Blauwendraat, C., Nalls, M. A. & Singleton, A. B. The genetic architecture of Parkinson's
998 disease. *The Lancet Neurology* **19**, 170–178 (2020).

999 26. Orme, T., Guerreiro, R. & Bras, J. The Genetics of Dementia with Lewy Bodies: Current
1000 Understanding and Future Directions. *Curr Neurol Neurosci Rep* **18**, 67 (2018).

1001 27. Andrews, S. J. *et al.* The complex genetic architecture of Alzheimer's disease: novel insights
1002 and future directions. *eBioMedicine* **90**, 104511 (2023).

1003 28. McCoy, T. H., Hart, K., Pellegrini, A. & Perlis, R. H. Genome-wide association identifies a
1004 novel locus for delirium risk. *Neurobiology of Aging* **68**, 160.e9-160.e14 (2018).

1005 29. Heinrich, M. *et al.* Association between genetic variants of the cholinergic system and
1006 postoperative delirium and cognitive dysfunction in elderly patients. *BMC Med Genomics* **14**, 248
1007 (2021).

1008 30. Westphal, S. *et al.* Genome-wide association study of myocardial infarction, atrial fibrillation,
1009 acute stroke, acute kidney injury and delirium after cardiac surgery – a sub-analysis of the RIPHeart-
1010 Study. *BMC Cardiovasc Disord* **19**, 26 (2019).

1011 31. Wiredu, K., Aduse-Poku, E., Shaefi, S. & Gerber, S. A. Proteomics for the Discovery of Clinical
1012 Delirium Biomarkers: A Systematic Review of Major Studies. *Anesthesia & Analgesia* **136**, 422–432
1013 (2023).

1014 32. Mosharaf, M. P., Alam, K., Gow, J. & Mahumud, R. A. Cytokines and inflammatory
1015 biomarkers and their association with post-operative delirium: a meta-analysis and systematic
1016 review. *Sci Rep* **15**, 7830 (2025).

1017 33. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
1018 **562**, 203–209 (2018).

1019 34. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using
1020 MTAG. *Nat Genet* **50**, 229–237 (2018).

1021 35. Lumsden, A. L., Mulugeta, A., Zhou, A. & Hyppönen, E. Apolipoprotein E (APOE) genotype-
1022 associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK
1023 Biobank. *eBioMedicine* **59**, 102954 (2020).

1024 36. Smith, L. H. & VanderWeele, T. J. Mediational E-values: Approximate Sensitivity Analysis for
1025 Unmeasured Mediator-Outcome Confounding. *Epidemiology* **30**, 835–837 (2019).

1026 37. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal*
1027 *Statistical Society Series B: Statistical Methodology* **58**, 267–288 (1996).

1028 38. Meinshausen, N. & Bühlmann, P. Stability Selection. *Journal of the Royal Statistical Society*
1029 *Series B: Statistical Methodology* **72**, 417–473 (2010).

1030 39. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
1031 Association Studies Using Summary Statistics. *PLoS Genet* **10**, e1004383 (2014).

1032 40. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization:
1033 Review and comparison of approaches. *The American Journal of Human Genetics* **109**, 767–782
1034 (2022).

1035 41. Finan, C. *et al.* The druggable genome and support for target identification and validation in
1036 drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).

1037 42. Huang, K. *et al.* Small-cohort GWAS discovery with AI over massive functional genomics
1038 knowledge graph. Preprint at https://doi.org/10.1101/2024.12.03.24318375 (2024).

1039 43. Raulin, A.-C. *et al.* ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies.
1040 *Mol Neurodegeneration* **17**, 72 (2022).

1041 44. Bowman, K. *et al.* Vitamin D levels and risk of delirium: A mendelian randomization study in
1042 the UK Biobank. *Neurology* **92**, (2019).

1043 45. Xie, W. *et al.* CEND1 deficiency induces mitochondrial dysfunction and cognitive impairment
1044 in Alzheimer's disease. *Cell Death Differ* **29**, 2417–2428 (2022).

1045 46. Manavalan, A. *et al.* Brain site-specific proteome changes in aging-related dementia. *Exp Mol*
1046 *Med* **45**, e39–e39 (2013).

1047 47. Schmukler, E. *et al.* Altered mitochondrial dynamics and function in APOE4-expressing
1048 astrocytes. *Cell Death Dis* **11**, 578 (2020).

1049 48. Farrer, L. A. Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E
1050 Genotype and Alzheimer Disease: A Meta-analysis. *JAMA* **278**, 1349 (1997).

1051 49. Belloy, M. E. *et al. APOE* Genotype and Alzheimer Disease Risk Across Age, Sex, and
1052 Population Ancestry. *JAMA Neurol* **80**, 1284 (2023).

1053 50. Griswold, A. J. *et al.* Increased *APOE* ε4 expression is associated with the difference in
1054 Alzheimer's disease risk from diverse ancestral backgrounds. *Alzheimer's & Dementia* **17**, 1179–1188
1055 (2021).

1056 51. Bellou, E. *et al.* Age-dependent effect of APOE and polygenic component on Alzheimer's
1057 disease. *Neurobiology of Aging* **93**, 69–77 (2020).

1058 52. Bonham, L. W. *et al.* Age-dependent effects of *APOE* ε4 in preclinical Alzheimer's disease.
1059 *Ann Clin Transl Neurol* **3**, 668–677 (2016).

1060 53. Steen, J., Loeys, T., Moerkerke, B. & Vansteelandt, S. **medflex** : An *R* Package for Flexible
1061 Mediation Analysis using Natural Effect Models. *J. Stat. Soft.* **76**, (2017).

1062 54. Leighton, S. P. *et al.* Delirium and the risk of developing dementia: a cohort study of 12 949
1063 patients. *J Neurol Neurosurg Psychiatry* **93**, 822–827 (2022).

1064 55. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2024 report of the Lancet
1065 standing Commission. *The Lancet* **404**, 572–628 (2024).

1066 56. Edwards, D., Handsley, M. & Pennington, C. The ADAM metalloproteinases. *Molecular*
1067 *Aspects of Medicine* **29**, 258–289 (2008).

1068 57. Lambrecht, B. N., Vanderkerken, M. & Hammad, H. The emerging role of ADAM
1069 metalloproteinases in immunity. *Nat Rev Immunol* **18**, 745–758 (2018).

1070 58. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related
1071 dementias. *Nat Genet* **54**, 412–436 (2022).

1072 59. Mattiola, I. *et al.* The macrophage tetraspan MS4A4A enhances dectin-1-dependent NK cell–
1073 mediated resistance to metastasis. *Nat Immunol* **20**, 1012–1022 (2019).

1074 60. Chen, J., Zhao, X., Huang, C. & Lin, J. Novel insights into molecular signatures and pathogenic
1075 cell populations shared by systemic lupus erythematosus and vascular dementia. *Funct Integr*
1076 *Genomics* **23**, 337 (2023).

1077 61. Saha, O. *et al.* The Alzheimer's disease risk gene BIN1 regulates activity-dependent gene
1078 expression in human-induced glutamatergic neurons. *Mol Psychiatry* (2024) doi:10.1038/s41380-
1079 024-02502-y.

1080 62. Foster, E. M., Dangla-Valls, A., Lovestone, S., Ribe, E. M. & Buckley, N. J. Clusterin in
1081 Alzheimer's Disease: Mechanisms, Genetics, and Lessons From Other Pathologies. *Front. Neurosci.*
1082 **13**, 164 (2019).

1083 63. Poljak, A. *et al.* Quantitative proteomics of delirium cerebrospinal fluid. *Transl Psychiatry* **4**,
1084 e477–e477 (2014).

1085 64. Zhu, X.-C. *et al.* CR1 in Alzheimer's Disease. *Mol Neurobiol* **51**, 753–765 (2015).

1086 65. Roses, A. *et al.* Understanding the genetics of *APOE* and *TOMM40* and role of mitochondrial
1087 structure and function in clinical pharmacology of Alzheimer's disease. *Alzheimer's & Dementia*
1088 **12**, 687–694 (2016).

1089 66. Zhu, Z. *et al.* TOMM40 and APOE variants synergistically increase the risk of Alzheimer's
1090 disease in a Chinese population. *Aging Clin Exp Res* **33**, 1667–1675 (2021).

1091 67. Fong, T. G. *et al.* Biomarkers of neurodegeneration and neural injury as potential predictors
1092 for delirium. *Int J Geriat Psychiatry* **39**, e6044 (2024).

1093 68. Fong, T. G. *et al.* Association of Plasma Neurofilament Light with Postoperative Delirium.
1094 *Annals of Neurology* **88**, 984–994 (2020).

1095 69. Van Munster, B. C. *et al.* Neuroinflammation in Delirium: A Postmortem Case-Control Study.
1096 *Rejuvenation Research* **14**, 615–622 (2011).

1097 70. Miao, Y., Wang, M., Cai, X., Zhu, Q. & Mao, L. Leucine rich alpha-2-glycoprotein 1 (Lrg1)
1098 silencing protects against sepsis-mediated brain injury by inhibiting transforming growth factor
1099 beta1 (TGFβ1)/SMAD signaling pathway. *Bioengineered* **13**, 7316–7327 (2022).

1100 71. Oren, R. L. *et al.* Age-dependent differences and similarities in the plasma proteomic
1101 signature of postoperative delirium. *Sci Rep* **13**, 7431 (2023).

1102 72. Vasunilashorn, S. M., Dillon, S. T., Marcantonio, E. R. & Libermann, T. A. Application of
1103 Multiple Omics to Understand Postoperative Delirium Pathophysiology in Humans. *Gerontology* **69**,
1104 1369–1384 (2023).

1105 73. Peters Van Ton, A. M., Verbeek, M. M., Alkema, W., Pickkers, P. & Abdo, W. F.
1106 Downregulation of synapse-associated protein expression and loss of homeostatic microglial control
1107 in cerebrospinal fluid of infectious patients with delirium and patients with Alzheimer's disease.
1108 *Brain, Behavior, and Immunity* **89**, 656–667 (2020).

1109 74. Jujić, A. *et al.* Low Levels of Selenoprotein P Are Associated With Cognitive Impairment in
1110 Patients Hospitalized for Heart Failure. *Journal of Cardiac Failure* S1071916424000393 (2024)
1111 doi:10.1016/j.cardfail.2024.01.010.

1112 75. Berasain, C. & Avila, M. A. Amphiregulin. *Seminars in Cell & Developmental Biology* **28**, 31–
1113 41 (2014).

1114 76. Hong, Y. *et al.* Sepsis-associated encephalopathy: From pathophysiology to clinical
1115 management. *International Immunopharmacology* **124**, 110800 (2023).

1116 77. Zaiss, D. M. W., Gause, W. C., Osborne, L. C. & Artis, D. Emerging Functions of Amphiregulin
1117 in Orchestrating Immunity, Inflammation, and Tissue Repair. *Immunity* **42**, 216–226 (2015).

1118 78. Hassan, R. *et al.* Mesothelin-targeting T cell receptor fusion construct cell therapy in
1119 refractory solid tumors: phase 1/2 trial interim results. *Nat Med* **29**, 2099–2109 (2023).

1120 79. Lawlor, P. G. & Bush, S. H. Delirium in patients with cancer: assessment, impact, mechanisms
1121 and management. *Nat Rev Clin Oncol* **12**, 77–92 (2015).

1122 80. Amgarth-Duff, I., Hosie, A., Caplan, G. & Agar, M. A systematic review of the overlap of fluid
1123 biomarkers in delirium and advanced cancer-related syndromes. *BMC Psychiatry* **20**, 182 (2020).

1124 81. Tzioras, M., Davies, C., Newman, A., Jackson, R. & Spires-Jones, T. Invited Review: APOE at
1125 the interface of inflammation, neurodegeneration and pathological protein spread in Alzheimer's
1126 disease. *Neuropathology Appl Neurobio* **45**, 327–346 (2019).

1127 82. Reas, E. T. *et al. APOE* ε4-related blood–brain barrier breakdown is associated with
1128 microstructural abnormalities. *Alzheimer's & Dementia* **20**, 8615–8624 (2024).

1129 83. Serrano-Pozo, A., Das, S. & Hyman, B. T. APOE and Alzheimer's disease: advances in genetics,
1130 pathophysiology, and therapeutic approaches. *The Lancet Neurology* **20**, 68–80 (2021).

1131 84. Yang, S.-T. *et al.* Accumulation of amyloid in cognitive impairment after mild traumatic brain
1132 injury. *Journal of the Neurological Sciences* **349**, 99–104 (2015).

1133 85. Koizumi, K. *et al.* Apoε4 disrupts neurovascular regulation and undermines white matter
1134 integrity and cognitive function. *Nat Commun* **9**, 3816 (2018).

1135 86. VanDusen, K. W. *et al.* The MARBLE Study Protocol: Modulating ApoE Signaling to Reduce
1136 Brain Inflammation, DeLirium, and PostopErative Cognitive Dysfunction. *Journal of Alzheimer's*
1137 *Disease* **75**, 1319–1328 (2020).

1138 87. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications.
1139 *Nat Genet* **47**, 856–860 (2015).

1140 88. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic
1141 evidence on clinical success. *Nature* **629**, 624–629 (2024).

1142 89. Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug
1143 repurposing. *Nat Rev Genet* **22**, 658–671 (2021).

1144 90. Draganov, D. I., Stetson, P. L., Watson, C. E., Billecke, S. S. & La Du, B. N. Rabbit Serum
1145 Paraoxonase 3 (PON3) Is a High Density Lipoprotein-associated Lactonase and Protects Low Density
1146 Lipoprotein against Oxidation. *Journal of Biological Chemistry* **275**, 33435–33442 (2000).

1147 91. Chang, Y.-H. *et al.* Statin Use and Delirium Risk: An Updated Systematic Review and Meta-
1148 Analysis. *American Journal of Therapeutics* **30**, e326–e335 (2023).

1149 92. Camps, J., Marsillach, J. & Joven, J. The paraoxonases: role in human diseases and
1150 methodological difficulties in measurement. *Critical Reviews in Clinical Laboratory Sciences* **46**, 83–
1151 106 (2009).

1152 93. Grdic Rajkovic, M., Rumora, L. & Barisic, K. The paraoxonase 1, 2 and 3 in humans. *Biochem*
1153 *Med* 122–130 (2011) doi:10.11613/BM.2011.020.

1154 94. Schlomann, U. *et al.* ADAM8 as a drug target in pancreatic cancer. *Nat Commun* **6**, 6175
1155 (2015).

1156 95. Shubbar, E. *et al.* High levels of γ-glutamyl hydrolase (GGH) are associated with poor
1157 prognosis and unfavorable clinical outcomes in invasive breast cancer. *BMC Cancer* **13**, 47 (2013).

1158 96. Li, Q., Yang, Z., He, X. & Yang, X. Comprehensive analysis of PILRA's association with the
1159 prognosis, tumor immune infiltration, and immunotherapy in pan-cancer. *Sci Rep* **13**, 14334 (2023).

1160 97. Monroe, K. *et al.* PILRA regulates microglial neuroinflammation and lipid metabolism as a
1161 candidate therapeutic target for Alzheimer's disease. Preprint at https://doi.org/10.21203/rs.3.rs-
1162 3954863/v1 (2024).

1163 98. Ibitoye, T. *et al.* Delirium is under-reported in discharge summaries and in hospital
1164 administrative systems: a systematic review. *Delirium* (2023) doi:10.56392/001c.74541.

1165 99. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
1166 **562**, 203–209 (2018).

1167 100. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide
1168 Range of Complex Diseases of Middle and Old Age. *PLoS Med* **12**, e1001779 (2015).

1169 101. The All of Us Research Program Genomics Investigators *et al.* Genomic data in the All of Us
1170 Research Program. *Nature* **627**, 340–346 (2024).

1171 102. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
1172 population. *Nature* **613**, 508–518 (2023).

1173 103. Zawistowski, M. *et al.* The Michigan Genomics Initiative: A biobank linking genotypes and
1174 electronic clinical records in Michigan Medicine patients. *Cell Genomics* **3**, 100257 (2023).

1175 104. World Health Organization. ICD-10 : international statistical classification of diseases and
1176 related health problems : tenth revision. *ICD-10* (2004).

1177 105. World Health Organization. International classification of diseases : [9th] ninth revision,
1178 basic tabulation list with alphabetic index. *ICD-9 : basic tabulation list with alphabetic index* (1978).

1179 106. Stuart-Buttle, C. D., Read, J. D., Sanderson, H. F. & Sutton, Y. M. A language of health in
1180 action: Read Codes, classifications and groupings. *Proc AMIA Annu Fall Symp* 75–79 (1996).

1181 107. Kuan, V. *et al.* A chronological map of 308 physical and mental health conditions from 4
1182 million individuals in the English National Health Service. *The Lancet Digital Health* **1**, e63–e77
1183 (2019).

1184 108. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and
1185 binary traits. *Nat Genet* **53**, 1097–1103 (2021).

1186 109. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron:
1187 Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci.*
1188 *U.S.A.* **100**, 11484–11489 (2003).

1189 110. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide
1190 association scans. *Bioinformatics* **26**, 2190–2191 (2010).

1191 111. Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations
1192 across human complex traits. *Nat Genet* **52**, 859–864 (2020).

1193 112. Falconer, D. S. & Mackay, T. *Introduction to Quantitative Genetics*. (Pearson, Prentice Hall,
1194 Harlow, 2009).

1195 113. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating Missing Heritability for
1196 Disease from Genome-wide Association Studies. *The American Journal of Human Genetics* **88**, 294–
1197 305 (2011).

1198 114. He, Y., Koido, M., Shimmori, Y. & Kamatani, Y. GWASLab: a Python package for processing
1199 and visualizing GWAS summary statistics. Preprint at https://doi.org/10.51094/jxiv.370 (2023).

1200 115. VanderWeele, T. J. & Ding, P. Sensitivity Analysis in Observational Research: Introducing the
1201 E-Value. *Ann Intern Med* **167**, 268–274 (2017).

1202 116. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416
1203 Participants and Longitudinal Stability in 20,346 Participants. *PLoS ONE* **11**, e0154222 (2016).

1204 117. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLoS*
1205 *ONE* **15**, e0231627 (2020).

1206 118. Townsend, P. Deprivation. *J. Soc. Pol.* **16**, 125–146 (1987).

1207 119. Quan, H. *et al.* Updating and Validating the Charlson Comorbidity Index and Score for Risk
1208 Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *American Journal of*
1209 *Epidemiology* **173**, 676–682 (2011).

1210 120. Quan, H. *et al.* Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10
1211 Administrative Data. *Medical Care* **43**, 1130–1139 (2005).

1212 121. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank.
1213 *Nature* **622**, 329–338 (2023).

1214 122. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis
1215 tool. *BMC Bioinformatics* **14**, 128 (2013).

1216 123. Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research* **52**,
1217 D672–D678 (2024).

1218 124. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell*
1219 *Systems* **1**, 417–425 (2015).

1220 125. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**,
1221 27–30 (2000).

1222 126. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models
1223 via Coordinate Descent. *J. Stat. Soft.* **33**, (2010).

1224 127. Lever, J., Krzywinski, M. & Altman, N. Classification evaluation. *Nat Methods* **13**, 603–604
1225 (2016).

1226 128. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more
1227 correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–
1228 845 (1988).

1229 129. Purcell, S. M. & Chang, C. C. PLINK v1.9. www.cog-genomics.org/plink/1.9/.

1230 130. Sanderson, E. *et al.* Mendelian randomization. *Nat Rev Methods Primers* **2**, 6 (2022).

1231 131. Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in
1232 Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**, 1614 (2021).

1233 132. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian
1234 randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible
1235 technique. *International Journal of Epidemiology* **45**, 1717–1726 (2016).

1236 133. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian
1237 randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int. J. Epidemiol.*
1238 dyw220 (2016) doi:10.1093/ije/dyw220.

1239 134. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With
1240 Multiple Genetic Variants Using Summarized Data. *Genetic Epidemiology* **37**, 658–665 (2013).

1241 135. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the
1242 human phenome. *eLife* **7**, e34408 (2018).

136. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512–525 (2015).

137. Greco M, F. D., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* **34**, 2926–2940 (2015).

138. Bowden, J. *et al.* Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *International Journal of Epidemiology* **48**, 728–742 (2019).

139. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).

140. De Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).