

1 **DNAm scores for serum GDF15 and NT-proBNP levels associate with**
2 **a range of traits affecting the body and brain**

3 Danni A. Gadd¹, Hannah M. Smith¹, Donncha Mullin^{2,3}, Ola Chybowska¹, Robert F. Hillary¹, Dorien
4 M Kimenai⁴, Elena Bernabeu¹, Yipeng Cheng¹, Chloe Fawns-Ritchie^{1,5}, Archie Campbell¹, Danielle
5 Page⁶, Adele Taylor⁶, Janie Corley⁶, Maria Del C. Valdés-Hernández^{6,7,8}, Susana Muñoz
6 Maniega^{6,7,8}, Mark E. Bastin^{6,7,8}, Joanna M. Wardlaw^{6,7,8,9}, Rosie M. Walker^{8,10}, Kathryn L. Evans¹,
7 Andrew M. McIntosh², Caroline Hayward^{1,13}, Tom Russ^{2,3}, Sarah E. Harris⁶, Paul Welsh¹⁴, Naveed
8 Sattar¹⁴, Simon R. Cox⁶, Daniel L. McCartney¹, Riccardo E. Marioni^{1,†}

9 ¹ Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of
10 Edinburgh, Edinburgh, UK

11 ² Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

12 ³ Alzheimer Scotland Dementia Research Centre, University of Edinburgh, UK

13 ⁴ BHF Centre for Cardiovascular Science, University of Edinburgh, UK

14 ⁵ Division of Psychology, University of Dundee, Dundee, Scotland, UK

15 ⁶ Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, UK

16 ⁷ Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration,
17 Edinburgh, UK

18 ⁸ Centre for Clinical Brain Sciences, Edinburgh Imaging and UK Dementia Research Institute,
19 University of Edinburgh, UK

20 ⁹ Neurovascular Imaging Research Core, UCLA, Los Angeles, CA

21 ¹⁰ School of Psychology, University of Exeter, Perry Road, Exeter, UK

22 ¹¹ British Heart Foundation Data Science Centre, Health Data Research UK, London, UK

23 ¹² Health Data Research UK, London, UK

24 ¹³ Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of
25 Edinburgh, Edinburgh, UK

26 ¹⁴ School of Cardiovascular and Metabolic Health, BHF Glasgow Cardiovascular Research Centre,
27 University of Glasgow, 126 University Place, Glasgow, UK

28

29 [†] Corresponding author: Riccardo Marioni, riccardo.marioni@ed.ac.uk

30

31

32 **Abstract**

33 **Background:** Plasma growth differentiation factor 15 (GDF15) and N-terminal pro-B-type
34 natriuretic peptide (NT-proBNP) are cardiovascular biomarkers that associate with a range of
35 diseases. Epigenetic scores (EpiScores) for GDF15 and NT-proBNP may provide new routes for risk
36 stratification.

37 **Results:** In the Generation Scotland cohort ($N \geq 16,963$), GDF15 levels were associated with
38 incident dementia, ischaemic stroke and type 2 diabetes, whereas NT-proBNP levels were associated
39 with incident ischaemic heart disease, ischaemic stroke and type 2 diabetes (all $P_{FDR} < 0.05$).
40 Bayesian Epigenome-wide association studies (EWAS) identified 12 and 4 DNA methylation
41 (DNAm) CpG sites associated (Posterior Inclusion Probability [PIP] $> 95\%$) with levels of GDF15
42 and NT-proBNP, respectively. EpiScores for GDF15 and NT-proBNP that were trained in a subset of
43 the population. The GDF15 EpiScore replicated protein associations with incident dementia, type 2
44 diabetes and ischaemic stroke in the Generation Scotland test set (Hazard Ratios (HR) range 1.36 –
45 1.41, $P_{FDR} < 0.03$). The EpiScore for NT-proBNP replicated the protein association with type 2
46 diabetes, but failed to replicate an association with ischaemic stroke. EpiScores explained
47 comparable variance in protein levels across both the Generation Scotland test set and the external
48 LBC1936 test cohort (R^2 range of 5.7-12.2%). In LBC1936, both EpiScores were associated with
49 indicators of poorer brain health. Neither EpiScore was associated with incident dementia in the
50 LBC1936 population.

51 **Conclusions:** EpiScores for serum levels of GDF15 and Nt-proBNP associate with body and brain
52 health traits. These EpiScores are provided as potential tools for disease risk stratification.

53

54 **Keywords:** GDF15, NT-proBNP, epigenetic, DNA methylation, dementia, diabetes,
55 cardiovascular, stroke, risk stratification, brain

56 **Background**

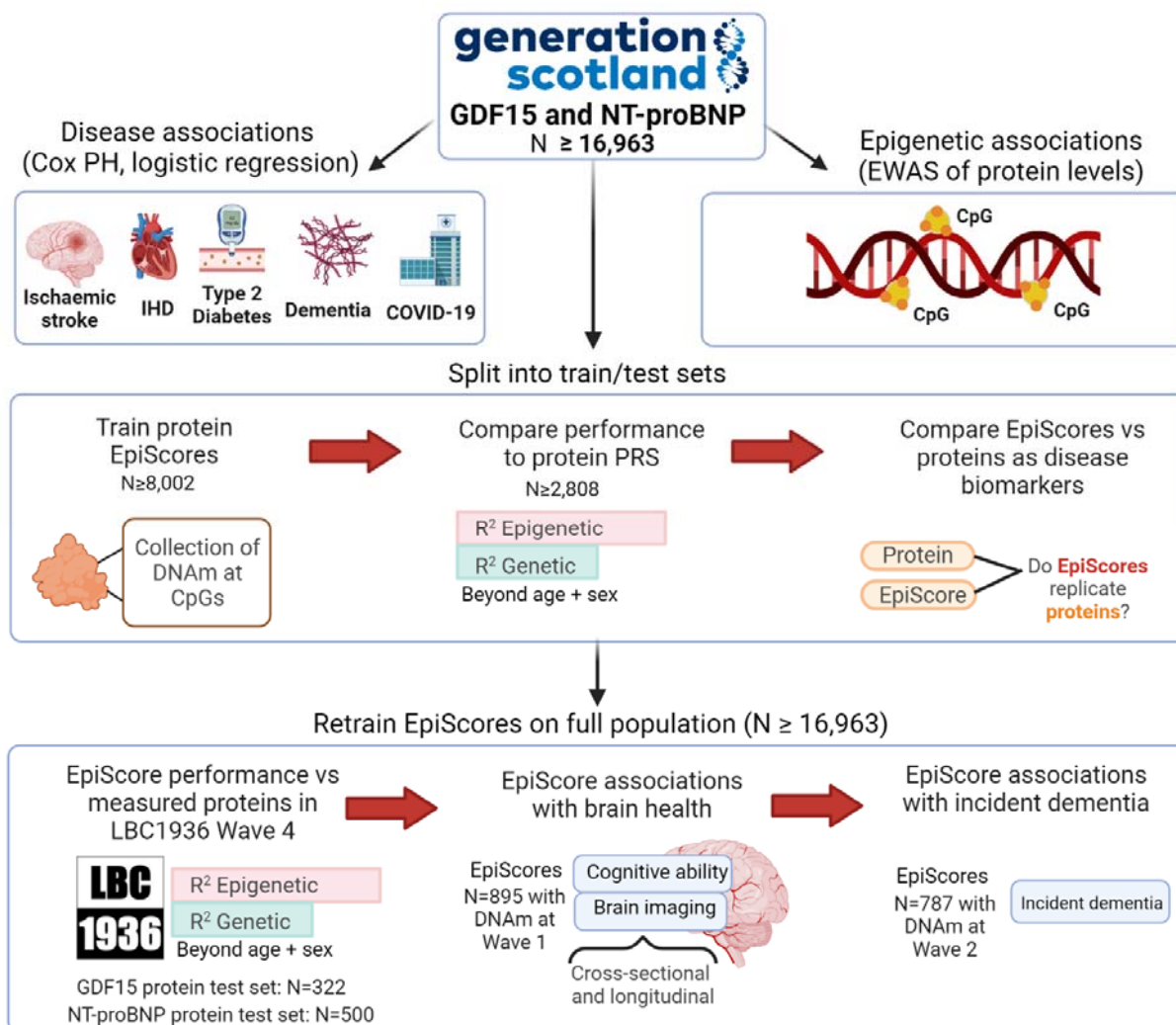
57 Delaying or preventing the onset of chronic diseases is a major challenge. Traditional risk factor
58 models provide a foundation to achieve this (1,2), but can be augmented by molecular-level data.
59 Plasma growth differentiation factor 15 (GDF15) and N-terminal pro-B-type natriuretic peptide
60 (NT-proBNP) are biomarker candidates that have yielded promising results as indicators of a range
61 of morbidities. GDF15 is associated with low-grade inflammation and age-related
62 immunosuppression (3). Higher levels of GDF15 have been found, through Mendelian
63 randomisation, to causally associate with increased risk of cardiometabolic stroke, atrial fibrillation,
64 coronary artery disease and myocardial infarction (4). Two recent proteome-wide studies that
65 assessed 1,301 (5) and 1,468 (6) proteins identified GDF15 as the top marker of multimorbidity. NT-
66 proBNP is a metabolite of pro B-type natriuretic peptide (BNP), which is a natriuretic and diuretic
67 hormone released by heart muscle in response to wall stretch (7). An inverse relationship between
68 the levels of NT-proBNP in blood and incident diabetes has been reported (8), whereas lower levels
69 of NT-proBNP have been associated with more favourable cardiovascular outcomes in randomised
70 control trials (9–11). Elevated levels of GDF15 and NT-proBNP in individuals diagnosed with
71 COVID-19 have been linked to more severe outcomes (12,13). Both protein markers have also been
72 found to associate with vascular brain injury, poorer neurocognitive performance and incident
73 dementias (14,15).

74 DNAm-based epigenetic scores (EpiScores) for blood proteins have been found to serve as markers
75 of incident diseases (16) and augment clinically-used risk factors for risk stratification (17,18).
76 DNAm reflects the body's chronic response to low-grade inflammation, environmental and
77 biological exposures (19–21). A study that directly compared an EpiScore for C-Reactive protein
78 (CRP) to measured CRP found that the EpiScore had greater test-retest reliability over time point
79 measures (22). This suggests that EpiScores may be more stable indicators than measured proteins in
80 some instances. An EpiScore for GDF15 levels based on changes to DNAm at CpG sites across the

81 genome is one of seven protein EpiScores that contribute to GrimAge, a leading epigenetic predictor
82 of biological age acceleration, healthspan and lifespan (23,24). However, the performance of protein
83 EpiScores against within-sample protein measurements in relation to incident diseases has not been
84 comprehensively investigated. Additionally, EpiScores have typically been trained in samples of
85 restricted size, with training sets typically ranging from 775 to 2,356 individuals (25) (23).

86 Here, we assess the viability of EpiScores for serum GDF15 and NT-proBNP as markers of disease
87 outcomes and brain health (**Fig. 1**). Using GDF15 and NT-proBNP measures available in Generation
88 Scotland ($N \geq 16,963$), we first profile associations between GDF15 and NT-proBNP and four
89 incident diseases (type 2 diabetes, ischaemic heart disease, ischaemic stroke and dementia), in
90 addition to COVID-19 outcome severity. These diseases were chosen for the study as they have been
91 linked to GDF15 and NT-proBNP and were available through electronic health linkage. We next
92 map the epigenetic architectures of the two proteins, before training and testing protein EpiScores for
93 them in subsets of Generation Scotland. In the test set, direct biomarker comparisons between
94 measured proteins and the EpiScore equivalents are performed in relation to the four incident
95 diseases assessed in the full Generation Scotland sample initially. EpiScores are then retrained in the
96 full sample available and tested externally in the Lothian Birth Cohort 1936 (LBC1936), where
97 associations with brain health traits are also profiled cross-sectionally and longitudinally.

98



99
100 **Figure 1. Study design for this assessment of GDF15 and NT-proBNP EpiScores as**
101 **biomarkers.** Disease associations and epigenome-wide association studies (EWAS) for each protein
102 were first characterised in the full Generation Scotland sample. EpiScores for each protein were
103 initially trained and tested in subsets of the population. This allowed EpiScores to be compared with
104 measured proteins in associations with the four incident diseases profiled in the test set. EpiScores
105 were then retrained on the full sample and tested externally in the LBC1936 Wave 4 population,
106 which had measures of both proteins and DNAm available. EpiScores were projected into the larger
107 LBC1936 Wave 1 population (that has DNAm but no protein measures) and profiled in associations
108 with brain health traits, cross-sectionally and longitudinally. Consent for dementia linkage was
109 available from Wave 2 of the LBC1936; therefore, we also tested whether EpiScores were associated
110 with time-to-dementia. EpiScores were modelled with polygenic risk scores (PRS) for the proteins.
111 CpG: cytosine-phosphate-guanine. IHD: ischaemic heart disease.

112 **Results**

113 **Sample populations**

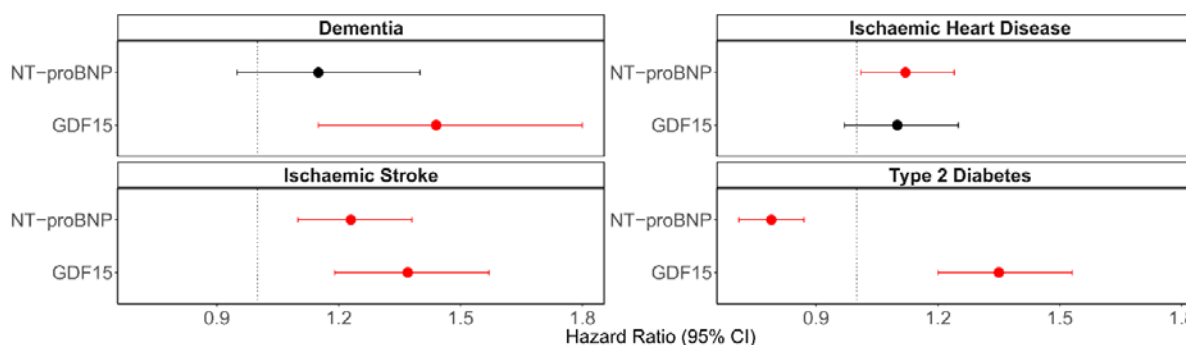
114 There were 18,413 Generation Scotland participants (59% female) that had DNAm measurements,
115 with a mean age of 48 years (SD 15), a minimum age of 17 years and maximum age of 98 years
116 (**Supplementary Table 1**) (26,27). Of these, 17,489 had GDF15 measurements and 16,963 had NT-
117 proBNP measurements. Subsets of this sample that were unrelated to one another were used to
118 initially train and test EpiScores for GDF15 ($N_{\text{train}} = 8,207$, $N_{\text{test}} = 2,954$) and NT-proBNP ($N_{\text{train}} =$
119 $8,002$, $N_{\text{test}} = 2,808$) (**Supplementary Table 1**). Measurements of serum GDF15 and NT-proBNP
120 levels were available at Wave 4 (mean age 79 years, with 0.6 SD) of the LBC1936 study. These
121 samples were used as external test sets for EpiScores trained on the full Generation Scotland sample.
122 Of 507 individuals at Wave 4, 322 had GDF15 measures (48% female) and 500 had NT-proBNP
123 measures (49% female). LBC1936 has successive Waves of measurements (Waves 1-5, collected at
124 mean ages of 70, 73, 76, 79 and 82 years old, with $SD < 1$ at each Wave) (28,29). EpiScores were
125 projected into Wave 1 (895 individuals with DNAm, but no protein measures) and evaluated in
126 relation to cross-sectional and longitudinal brain health traits. As consent to dementia linkage was
127 available from Wave 2, associations between EpiScores and time-to-dementia were also tested in
128 LBC1936.

129 **GDF15 and NT-proBNP disease associations**

130 Six associations (**Fig. 2**) were identified in Cox proportional hazards (PH) mixed effects models
131 between protein levels and incident diseases in Generation Scotland ($N \geq 16,963$). These associations
132 had False Discovery Rate (FDR) $P < 0.05$ in basic (age and sex adjusted) models and $P < 0.05$ in
133 fully-adjusted models (that further adjusted for smoking, alcohol intake, body mass index (BMI),
134 social deprivation and years of education) (**Supplementary Table 2**). Counts for cases, controls and
135 mean time-to-onset for cases are provided in **Supplementary Table 2**. In basic logistic regression
136 models, GDF15 was associated with subsequent hospitalisation due to COVID-19 (odds ratio (OR)

137 per SD = 2.0, 95% confidence interval (CI) = [1.2, 3.4], FDR P = 0.037), as opposed to having
138 COVID-19 without hospitalisation. An inverse association was identified between a one SD increase
139 in NT-proBNP levels and COVID-19 hospitalisation (OR = 0.59, 95% confidence interval (CI) =
140 [0.38, 0.93], FDR P = 0.046). No associations in relation to long-COVID as a binary outcome had
141 FDR P < 0.05. Full summary statistics are provided in **Supplementary Table 3**.

142



143

144 **Figure 2. Disease associations for GDF15 and NT-proBNP in Generation Scotland (N ≥ 16,963).**
145 Fully-adjusted hazard ratios from Cox PH mixed effects regression models between protein levels
146 and incident diseases are plotted with 95% confidence intervals. The six associations in red had FDR
147 P < 0.05 in basic and P < 0.05 in fully-adjusted models, whereas associations that had P > 0.05 are
148 shown in black. Hazard ratios are plotted per 1 SD increase in the rank-base inverse normalised
149 levels of each marker. Fully-adjusted models controlled for age, sex, relatedness and common health
150 and lifestyle factors (smoking, alcohol intake, BMI, social deprivation and years of education).

151 **GDF15 and NT-proBNP epigenetic associations**

152 In variance components analysis of GDF15 and NT-proBNP in Generation Scotland (N ≥ 16,963),
153 DNAm explained 36% of the variance in GDF15 levels (lower and upper confidence intervals [CIs]
154 = 32% to 39%) and 32% of the variance in NT-proBNP levels (lower and upper CIs = 27% to 36%).
155 In the EWAS, there were 12 and 4 associations (Bayesian Posterior Inclusion Probability [PIP] >
156 95%) between differential DNAm at 14 unique CpG sites and the levels of GDF15 and NT-proBNP,
157 respectively. The CpG sites cg03546163 (*FKBP5*) and cg13108341 (*DNAH9*) were associated with

158 both GDF15 and NT-proBNP. **Table 1** summarises the CpG sites, the biomarkers they associated
 159 with, the genes the CpGs are annotated to, and a selection of traits that DNAm at these CpGs have
 160 previously been associated with in EWAS studies. The full index of MRC IEU EWAS Catalogue
 161 associations (available as of August 2023) for these 14 CpG sites are available in **Supplementary**
 162 **Table 4**.

CpG	PIP	Biomarker	CpG Gene	CpG trait associations (MRC-IEU EWAS Catalogue)
cg03546163*	0.98, 0.96	GDF15, NT-proBNP	<i>FKBP5</i>	Chronic kidney disease, fetal vs adult liver, body mass index, waist circumference, mortality, age, neurodegenerative disorders.
cg13108341*	1.00, 0.98	GDF15, NT-proBNP	<i>DNAH9</i>	Cancer treatment
cg00757033	1.00	NT-proBNP	<i>WDR51B</i>	Crohn's disease, inflammatory bowel disease, age
cg05412028	0.99	NT-proBNP	<i>ABCC4</i>	Age, ageing, primary Sjogrens syndrome
cg19693031	1.00	GDF15	<i>TXNIP</i>	Fetal vs adult liver, triglycerides, sex, hbA1c, alcohol consumption, blood pressures, hepatic fat, waist circumference, cholesterol measures, age.
cg02650017	1.00	GDF15	<i>PHOSPHO1</i>	Type 2 diabetes, primary Sjogrens syndrome, C-reactive protein, body mass index, serum high-density lipoprotein cholesterol, Crohn's disease, body mass index, coagulation factor VIII, eosinophilia, age.
cg06918740	1.00	GDF15	<i>N/A</i>	<i>N/A</i>
cg08900409	1.00	GDF15	<i>PGPEP1</i>	Age
cg25460262	1.00	GDF15	<i>GDF15</i>	Fetal vs adult liver
cg21088460	1.00	GDF15	<i>GDF15</i>	<i>N/A</i>
cg05575921	1.00	GDF15	<i>AHRR</i>	Extensive set of smoking-associated traits, lung function/cancer traits, body mass index, serum cotinine, C-reactive protein, IgG glycosylation measures, educational attainment, cognitive ability, statin use, urinary cadmium, mortality, post-traumatic stress disorder, age, acute myocardial infarction.
cg25410121	1.00	GDF15	<i>N/A</i>	<i>N/A</i>
cg15058033	0.97	GDF15	<i>PLXNB2</i>	<i>N/A</i>
cg16993186	0.97	GDF15	<i>CELF2</i>	<i>N/A</i>

163
 164 **Table 1. EWAS of GDF15 and NT-proBNP levels in Generation Scotland (N ≥ 16,963).**
 165 Posterior inclusion probabilities (PIPs) are provided for all CpG-protein associations (PIP > 0.95) in
 166 the BayesR EWAS. * Two CpGs were associated with both GDF15 and NT-proBNP. A selection of
 167 traits implicated in associations ($P < 3.8 \times 10^{-6}$, with $n > 100$) with the CpGs from the MRC-IEU EWAS
 168 Catalogue (as of August 2023) are shown. HbA1c: glycated haemoglobin. IgG: immunoglobulin G.

169 **EpiScores for GDF15 and NT-proBNP within Generation Scotland**

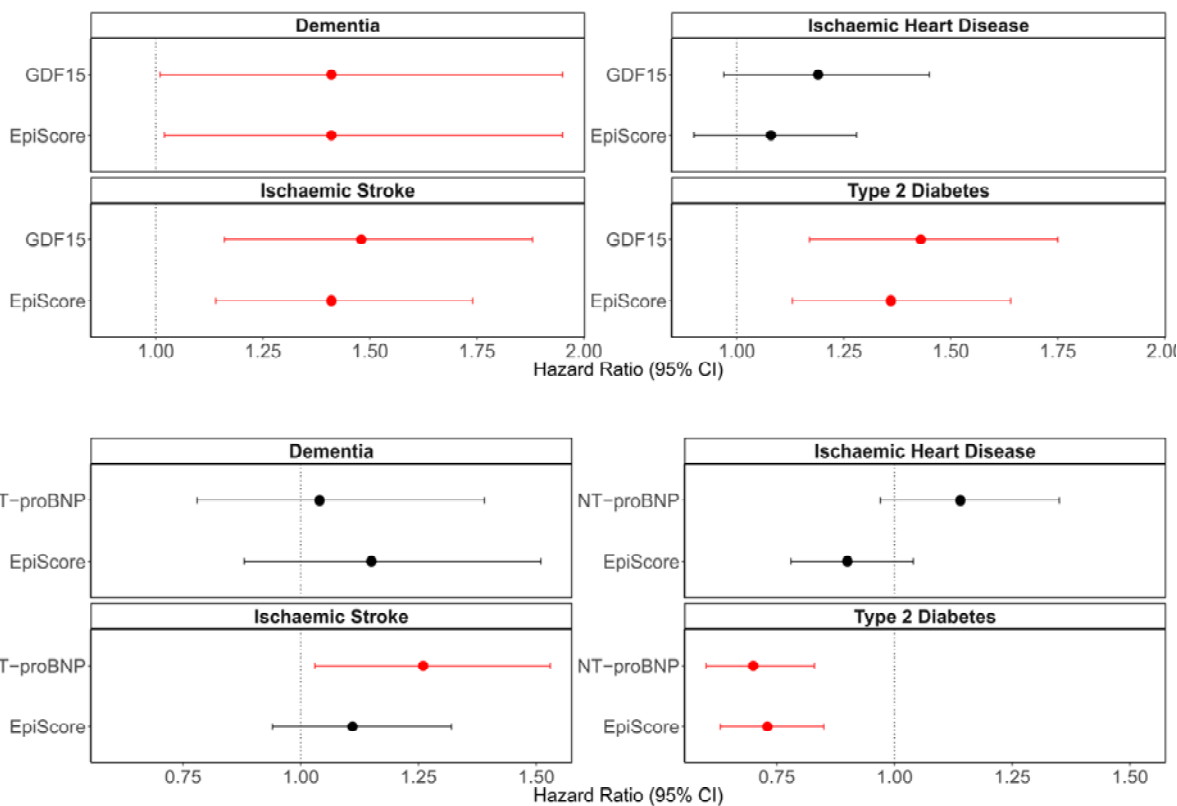
170 EpiScores for GDF15 and NT-proBNP were initially trained and tested in subsets of Generation
171 Scotland that were unrelated to one another. Predictor weights for EpiScores are available in
172 **Supplementary Table 5**. Performance in the test set was modelled through the incremental variance
173 (R^2) in protein levels that scores could explain beyond a null linear regression model that adjusted for
174 age and sex. The EpiScore for GDF15 trained using the full set of EPIC array probes had an R^2 of
175 12.2%, whereas the NT-proBNP EpiScore had an R^2 of 5.7%. Similar performance was observed
176 when comparing with EpiScores trained using sites available on the 450k array subset
177 (**Supplementary Fig. 1**). When modelling EpiScores and polygenic risk scores (PRS) derived for
178 each protein (**see Methods**), additive effects beyond the null model were observed for GDF15 (R^2 of
179 15.5%) and NT-proBNP (R^2 of 6.9%). A full summary of the results is provided in **Supplementary**
180 **Table 6**.

181 **EpiScore replication of protein biomarker associations within Generation Scotland**

182 The same Cox PH model structure (as shown in **Fig. 2**) was used to directly compare protein levels
183 and EpiScores in the Generation Scotland test. All protein-disease associations – except the
184 association between NT-proBNP and ischaemic stroke – were replicated by EpiScores in fully-
185 adjusted models (**Fig. 3, Supplementary Table 7**). Mean time-to-onset, counts for cases and
186 controls and full summary statistics are available in **Supplementary Table 7**. Mean attenuation in
187 the absolute log of the HR due to the additional adjustment for lifestyle factors beyond age and sex
188 was 6% for protein associations and 12% for EpiScore associations. Of the four protein EpiScore
189 associations identified in fully-adjusted models, three withstood further adjustment for estimated
190 immune cell proportions (attenuation in the absolute log of the HR ranging from 0 – 9%). The
191 association between the GDF15 EpiScore and dementia had $P = 0.064$, with 5% attenuation in the
192 absolute log of the HR. **Supplementary Table 7**). As only five instances of COVID-19

193 hospitalisation and nine instances of long-COVID were reported in the test population, we did not
194 conduct protein EpiScore and protein comparisons for these outcomes.

195



196

197 **Figure 3. Comparison of EpiScores versus measured protein equivalents in fully-adjusted**
198 **associations with incident diseases in the Generation Scotland test sample (N ≥ 2,808).** For each
199 disease, the protein-disease association is plotted, with the equivalent protein EpiScore-disease
200 association shown directly beneath it for comparison. Hazard ratios are plotted per 1 SD increase in
201 the rank-based inverse normalised levels of each marker. Nine associations (red) had FDR $P < 0.05$
202 in basic and $P < 0.05$ in fully-adjusted Cox proportional hazards mixed effects models in the test
203 samples. Fully-adjusted models adjusted for age, sex, relatedness and common lifestyle risk factors
204 (smoking, alcohol intake, BMI, social deprivation and years of education). Associations that were
205 non-significant ($P > 0.05$ in fully-adjusted models) are shown in black.

206

207 **EpiScore application to the LBC1936 external cohort**

208 EpiScores for each protein were then retrained in the entire Generation Scotland sample ($N_{\text{GDF15}} =$
209 17,489 and $N_{\text{NT-proBNP}} = 16,963$). Although we make predictor weights for EpiScores trained on the
210 full EPIC array probes and the subset of probes on the older 450k array available (**Supplementary**
211 **Table 8**), the LBC1936 external test cohort in this study measured DNAm using the 450k array.
212 Thus, the EpiScores trained on the 450k probe subset were projected into this population for external
213 validation.

214 In the LBC1936 test sample ($N_{\text{GDF15}} = 322$ and $N_{\text{NT-proBNP}} = 500$), incremental R^2 values of 8.9% for
215 GDF15 and 8.1% for NT-proBNP EpiScores were observed, beyond age and sex-adjusted linear
216 regression models (**Fig. 4a**). When a PRS for each protein was modelled together with the EpiScores,
217 incremental variance explained rose to 13.7% and 9.1% for GDF15 and NT-proBNP, respectively
218 (**Supplementary Table 9**). Finally, the GDF15 EpiScore generated previously by Lu *et al* as part of
219 the GrimAge biological age acceleration predictor (23) was projected into the Wave 4 GDF15 test
220 set (322 individuals) and evaluated. It explained 5.6% of the variance in GDF15 beyond age and sex,
221 as compared to the 8.9% observed modelling our updated GDF15 score. The two GDF15 EpiScores
222 had a Pearson correlation $r = 0.32$ in the test sample.

223 **EpiScore associations with brain health traits in LBC1936**

224 The EpiScores that were validated against protein measures in the LBC1936 Wave 4 external test set
225 were then projected into methylation measured at Wave 1 (a time point nine years prior), which has a
226 larger DNAm sample available but no protein measures. Structural equation models were then run to
227 characterise associations between the protein EpiScores and five brain health traits (cognitive ability
228 and four structural brain imaging measures). This allowed for EpiScore relationships with both cross-
229 sectional brain health (Wave 1, $N=895$ individuals with EpiScores, total model $N=1,091$) and
230 longitudinal change in brain health (Waves 1-5 for cognitive change and Waves 2-5 for brain

231 imaging changes) to be tested (five brain health traits x two EpiScores x cross-sectional and
232 longitudinal associations = 20 hypothesis tests).

233 Seven of the twenty basic model associations tested had FDR $P < 0.05$ (**Supplementary Table 10**).

234 All seven associations involved cross-sectional brain health phenotypes and had negative effect

235 estimates (standardised betas ranging from -0.05 to -0.19). Higher GDF15 and NT-proBNP

236 EpiScores were associated with lower general cognitive ability and lower brain volumes. None of the

237 ten slope associations assessing relationships between the EpiScores and longitudinal change in the

238 five brain health phenotypes were significant at FDR $P < 0.05$. In models that further adjusted for

239 additional health and lifestyle factors, five associations had $P < 0.05$ (**Fig.4b, Supplementary Table**

240 **10**). A one standard deviation increase in GDF15 EpiScore levels was associated with lower normal

241 appearing white matter volume (Beta = -0.07, SE = 0.02, $P = 2.2 \times 10^{-3}$), poorer general cognitive

242 ability (Beta = -0.09, SE = 0.04, $P = 9.1 \times 10^{-3}$) and lower total brain volume (Beta = -0.05, SE = 0.02,

243 $P = 3.5 \times 10^{-3}$). A one standard deviation increase in NT-proBNP EpiScore levels was associated with

244 lower normal appearing white matter volume (Beta = -0.05, SE = 0.02, $P = 0.02$) and lower total

245 brain volume (Beta = -0.03, SE = 0.01, $P = 0.03$). In a sensitivity analysis that further adjusted for

246 immune cell proportions, the two NT-proBNP associations were attenuated ($P > 0.07$) and the

247 associations between NT-proBNP and lower cognitive ability was found to be significant (Beta = -

248 0.10, SE = 0.04, $P = 6.0 \times 10^{-3}$). In the sensitivity analysis, the three GDF15 associations remained

249 significant ($P < 0.05$), with a mean attenuation of 11% in Beta effect magnitude (**Supplementary**

250 **Table 10**).

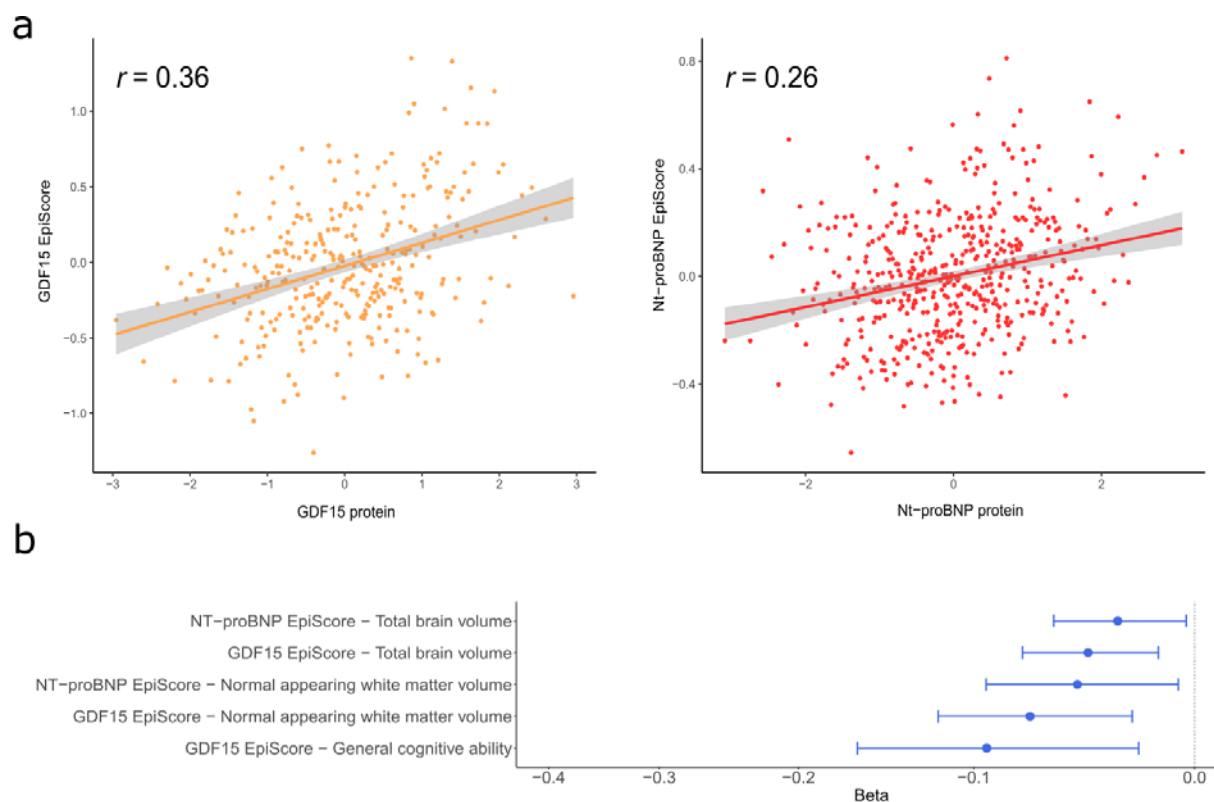
251 Individuals consented to share disease information from electronic health records from Wave 2 of the

252 study onwards. In Cox regression models that utilised Wave 2 as the baseline and modelled incident

253 dementia as the outcome ($N_{\text{cases}} = 108$, $N_{\text{controls}} = 672$, mean time-to-event for cases = 8.6 years [SD

254 3.42] and maximum follow-up of 14.3 years), no associations were identified for either EpiScore

255 (**Supplementary Table 11**).



256

257 **Figure 4. External assessment of the GDF15 and NT-proBNP EpiScores in LBC1936. a,**
258 Correlation plots between measured protein levels and GDF15 (orange) and NT-proBNP (red)
259 EpiScores in the LBC1936 Wave 4 external test set ($N_{\text{GDF}} = 322$, $N_{\text{NT-proBNP}} = 500$). Pearson
260 correlation coefficients are annotated in each instance. **b,** Standardised beta coefficients derived from
261 structural equation models (SEMs) between EpiScore levels at LBC1936 Wave 1 ($N=895$ with
262 DNAm, $N=1,091$ total) and cross-sectional measures of brain health traits that had FDR $P < 0.05$ in
263 basic (age and sex adjusted) models and $P < 0.05$ after adjustment for further lifestyle covariates. All
264 associations had a negative beta coefficient (blue). Twenty EpiScore-trait associations were tested in
265 total: 10 cross-sectionally and 10 assessing longitudinal change in brain traits.

266

267 Discussion

268 Here, biomarker-disease associations for GDF15 and NT-proBNP were first observed in Generation
269 Scotland, prior to developing EpiScores for these proteins. EpiScores replicated protein associations
270 with incident diseases in the Generation Scotland test sample. In the LBC1936 external test
271 population, the GDF15 and NT-proBNP EpiScores explained 9% and 8% of the variance in the

272 protein levels, respectively, with higher levels of the EpiScores associated with poorer brain health
273 cross-sectionally. EWAS of each protein highlighted 14 CpGs with differential DNAm.

274 This study provides EpiScores for GDF15 and NT-proBNP trained in the largest samples to date as
275 tools for health stratification. Despite the LBC1936 test set being older than the Generation Scotland
276 cohort (mean age of 79 *versus* 48 years), the EpiScores had R^2 values comparable to those observed
277 in the Generation Scotland test set. In the external LBC1936 test set, the GDF15 EpiScore had
278 improved performance (an additional R^2 of 3.3%) when compared to the GDF15 EpiScore derived
279 by Lu *et al* in 2019 as part of the GrimAge biological age acceleration predictor (23). This is likely
280 due to differences in the sample sizes used for training the two GDF15 scores (2,356 individuals as
281 compared to 17,489 individuals in our study). It may also be due to our training and testing
282 populations having more homogeneous ancestry (Scottish) than the populations used to train the
283 original GrimAge GDF15 score (mixed white European ancestry). No other EpiScores for either
284 GDF15 or NT-proBNP exist in the literature to our knowledge; these EpiScores can therefore be
285 utilised as new tools for risk stratification and can be projected into any cohort with Illumina-based
286 DNAm. We provide EpiScore weights trained on both the 450k and EPIC arrays for use in future
287 research.

288 Generation Scotland is one of the world's largest single cohort resources with DNAm, protein
289 measures, and extant data linkage to electronic health records. This allowed for direct comparisons
290 between protein and EpiScore measures in the context of incident disease analyses, which has only
291 recently been possible owing to the expansion of the cohort's epigenetic resource. As DNAm may
292 record chronic exposure to a range of environmental risk factors (30) and biological processes such
293 as inflammation (20,31), EpiScores may be reflective of a range of biological pathways that occur
294 upstream of disease diagnoses. Given that GDF15 and Nt-pro-BNP are promising biomarkers for a
295 range of diseases, our EpiScores are well-positioned candidates with many potential use-cases. The
296 results of inclusion of the PRS for proteins in incremental variance models suggested that EpiScore

297 signals were largely independent of genetic architectures on the proteins, as additive improvements
298 in incremental variance observed when PRS and EpiScores were modelled together. This is in
299 concordance with previous studies that found additive epi/genetic heritability estimates for plasma
300 protein levels (32,33). While we have previously regressed out protein quantitative trait loci (pQTLs)
301 from proteins prior to training EpiScores (25), there is an argument that EpiScores capturing a
302 combination of genetic and epigenetic signatures may enhance the disease-predictive signal
303 available. Both approaches are likely viable for the creation of new biomarkers.

304 The higher proportion of variance explained by the GDF15 EpiScore as compared to the NT-proBNP
305 EpiScore suggests that GDF15 was better-characterised by DNAm differences across the genome.
306 This may be due to its association with chronic inflammation, as we have observed particularly
307 strong DNAm signatures associated with inflammatory proteins in previous work (25,34). A stronger
308 DNAm signature was also observed for GDF15 in our EWAS analyses. To our knowledge, this
309 represents the first EWAS of NT-proBNP. The only other EWAS of GDF15 levels was performed by
310 us previously, using GDF15 measures from the SomaLogic assay (34), where we identified no
311 associations for GDF15 passing Bonferroni correction. The improved power to detect associations in
312 the present study (17,489 rather than 774 individuals) may have facilitated identification of
313 associations in the present study. There were two CpG sites associated with both GDF15 and NT-
314 proBNP (cg03546163 in *FKBP5* and cg13108341 in *DNAH9*), which suggests a partially-shared
315 DNAm signature across the proteins. FK506-binding protein 5 (FKBP5) is implicated in cellular
316 stress response (35). One previous study found cg03546163 to be differentially methylated in 107
317 individuals with type 2 diabetes that went onto develop end stage renal disease, versus 253 controls
318 who did not (36).

319 The lack of associations with incident ischaemic heart disease in the Generation Scotland test set
320 may be due to limited sample size, as an association between protein NT-proBNP and ischaemic
321 heart disease was observed in the full Generation Scotland sample. Additionally, the GDF15

322 EpiScore association with incident dementia observed in Generation Scotland did not replicate in
323 LBC1936. This may be due to differences in the way the phenotypes were defined across LBC1936
324 (consensus committee) versus Generation Scotland (Read and ICD codes only), or different cohort
325 sampling frames and recruitment strategies.

326 Our findings support previous work identifying associations between GDF15 and Nt-proBNP protein
327 levels and severe COVID-19 outcomes in hospitalised individuals (12,13). Although very few
328 hospitalisation cases were available (n=28), both proteins (sampled a mean of 11 years prior to
329 COVID-19 diagnoses) associated with subsequent hospitalisation due to COVID-19. GDF15 is likely
330 to be elevated in individuals with multiple morbidities that may contribute towards greater risk of
331 hospitalisation due to viral illnesses. Diabetes has been associated with increased risk of
332 hospitalisation and adverse outcomes in COVID-19 (37,38). Both proteins (and equivalent
333 EpiScores) should be investigated in populations that have DNAm quantified nearer to, or at
334 COVID-19 diagnoses to further resolve these signals.

335 This study has several limitations. First, whereas it was advantageous to have common ancestry
336 across both (Scottish) training and validation datasets, future studies should test EpiScores across
337 larger populations that include non-European ancestries and larger age ranges as data become
338 available. Second, emerging evidence has quantified differences in genetic associations with the
339 measurements of the same proteins across panels that use antibody-based versus aptamer-based
340 quantification methods (39). A particular example of interest from this study was GDF15 levels,
341 which was highlighted as a protein that may have different conformational shapes (isoforms) that are
342 targeted by the two assay methods (39). While it is likely that increased training sample size led to
343 improved performance of our GDF15 score versus the GrimAge GDF15 score in the LBC1936 test
344 set, it is possible that technical or biological variability across protein assays may also underlie
345 differences in performance of scores. EpiScores trained on protein measurements from different
346 panels should be therefore be compared further when data become available. Similarly, differences

347 in the protein assay method across the previous EWAS of GDF15 (aptamer-based) that we ran and
348 the present study (immunochemiluminescence) may also introduce variability and EWAS across
349 multiple protein panels should be compared when samples are available.

350 **Conclusions**

351 In conclusion, EpiScores for blood-based GDF15 and NT-proBNP levels are generated in this study
352 and have been found to be useful indicators of disease risk stratification, with disease-specific use-
353 cases. The EpiScores can be derived in any population with Illumina-based DNAm measurements
354 and may be integrated into epigenetic screening panels in future to better-identify high-risk
355 individuals.

356 **Methods**

357

358 **Generation Scotland**

359 Generation Scotland is a population-based cohort study that includes ~8,000 families from across
360 Scotland (26,27). Study recruitment of 23,960 participants occurred between 2006 and 2011, while
361 participants were aged between 18 and 99 years. In addition to completing health and lifestyle
362 questionnaires, participants donated blood samples for biomarker and omics measurement. Details
363 on DNAm quality control in Generation Scotland are provided in **Supplementary Information**. The
364 quality-controlled DNAm dataset comprised a total of 18,413 individuals with 760,838 CpG sites
365 available on the EPIC array. GDF15 and NT-proBNP measurement details are provided in
366 **Supplementary Information**. There were 18,413 individuals with GDF15 measures
367 (**Supplementary Table 1**) and these were subset to 17,489 individuals that had DNAm (mean
368 1038.7 pg/mL [SD 928]). There were 17,863 individuals with NT-proBNP measures, with 16,963
369 that had DNAm available (mean 94.6 pg/mL [SD 211.2]). Electronic health records via data linkage
370 to GP records (Read 2 codes) and hospital records (10th revision of the International Classification of

371 Diseases codes [ICD-10 codes]) were assessed prospectively from the time of blood draw. Incident
372 data for all-cause dementia, type 2 diabetes, ischaemic stroke and ischaemic heart disease were
373 considered with censoring date October 2020. Dementia cases were defined as per a previous review
374 of dementia linkage codes (40), whereas code lists for all other diseases are available in
375 **Supplementary Tables 12-14**. Prevalent cases (ascertained via retrospective linkage or self-report
376 from a baseline questionnaire) were excluded from each disease trait, leaving only incident
377 diagnoses. Dementia analyses were limited to cases/controls with age of event/censoring ≥ 65 years.
378 Type 1 and juvenile diabetes cases were treated as control observations in the type 2 diabetes
379 analyses. Death was treated as a censoring event.

380 **Lothian Birth Cohort 1936**

381 The Lothian Birth Cohort 1936 (LBC1936) is a longitudinal study of ageing of people residing in
382 Edinburgh and surrounding areas in Scotland (N = 1,091) (28,29). Individuals were born in 1936 and
383 completed an intelligence test when they were 11 years old. They were later recruited to the cohort at
384 a mean age of 70 years old and have been followed up triennially for a series of cognitive, clinical,
385 social and physical measurements in five Waves (mean ages 70, 73, 76, 79 and 82 – all with SD
386 below 1 for measures at each Wave). Blood samples were taken and used to derive protein,
387 epigenetic and genetic measurements. Sample measurement details for the DNAm measures
388 available in LBC1936 are provided in **Supplementary Information**. DNAm is available at the four
389 successive waves of the study (N=895, 787, 619 and 507 in Waves 1, 2, 3 and 4, respectively). Both
390 GDF15 (N=322) and NT-proBNP (N=500) serum levels were measured at Wave 4 of the study
391 (mean age 79 years, SD 0.6) and were used to externally test EpiScore performance. From Wave 2 of
392 the LBC1936, individuals consented for linkage to health records for research. Dementia cases were
393 defined by a consensus committee that completed decisions in August 2022 (41). Potential cases
394 were identified through a combination of electronic health record linkage, death certificate data and
395 clinician visits to individuals that were suspected of having cognitive impairments or dementia. Of

396 the 865 individuals who had provided linkage consent at Wave 2, 118 were confirmed as having
397 dementia.

398 **Epigenome-wide association studies in Generation Scotland**

399 The xBayesR+ software implements penalised Bayesian regression on complex traits (30). The
400 BayesR method has been found to outperform linear and mixed model approaches and implicitly
401 adjusts for probe correlations, data structure (such as relatedness) and unobserved confounders
402 (30,32). Prior mixture variances for the methylation data (760,838 CpG sites) were set to 0.001, 0.01
403 and 0.1 and epigenome-wide associations studies (EWAS) were run for GDF15 (N=17,489) and NT-
404 proBNP (N=16,963) levels in Generation Scotland. Protein measurements were transformed by rank-
405 based inverse normalisation, regressed onto age, sex and 20 genetic principal components and scaled
406 to have a mean of 0 and variance of 1. DNAm measurements in beta format were regressed onto age,
407 sex and processing batch and scaled to have a mean of 0 and variance of 1. Houseman immune cell
408 estimates were included as fixed effect covariates (42). Effect size estimates were obtained through
409 Gibbs sampling over the posterior distribution, conditional on input data. The Gibbs protocol had
410 10,000 samples, with 5,000 samples of burn-in followed by a thinning of 5 samples to reduce
411 autocorrelation. Methylation probes that had a posterior inclusion probability of $\geq 95\%$ were
412 deemed to be significant for each protein.

413 **EpiScore development**

414 Elastic net penalised regression was used to train EpiScores for GDF15 and Nt-pro-BNP levels. As
415 Generation Scotland has extensive phenotyping and extant linkage to primary care and hospital
416 records, EpiScores were first trained and tested in subsets of the full sample that were unrelated to
417 one another to facilitate direct comparisons between EpiScore and protein levels in associations with
418 incident diseases. EpiScores were then retrained on the full Generation Scotland sample and tested in
419 LBC1936 – an external cohort. For both analyses, DNAm beta values were considered with missing

420 CpG measurements mean imputed. To generate alternative versions of the EpiScores that can be
421 applied to existing cohort studies with older Illumina array data (450k array), CpGs were filtered to
422 the intersection of the 450k and EPIC array sites. A total of 760,838 EPIC array probes and 390,461
423 450k probes were available. CpG measurements were scaled to have a mean of 0 and variance of 1,
424 prior to training. Protein measurements in training samples were transformed by rank-based inverse
425 normalisation, regressed onto age, sex and 20 genetic principal components and scaled to have a
426 mean of 0 and variance of 1. Penalised regression models were performed using Big Lasso (Version
427 1.5.1) in R (Version 4.0) (43). GDF15 and NT-proBNP protein levels were the outcomes. An elastic
428 net penalty was specified ($\alpha=0.5$). In the within-Generation Scotland analyses 10-fold cross
429 validation was applied to select the lambda value that minimised the mean prediction error, whereas
430 20-fold cross validation was applied when training EpiScores in the full Generation Scotland sample.

431 A summary of the individuals with protein measurements available that were used to train and test
432 EpiScores in the initial, within-Generation Scotland analyses is provided in **Supplementary Fig. 2**.
433 Briefly, individuals that were part of the same family as disease cases in the test sample were
434 excluded from the training sample. In the test subset of Generation Scotland, control individuals that
435 were related to those in the training sample were excluded. A total of 8,207 individuals with GDF15
436 and 8,002 individuals with NT-proBNP measurements were therefore used to train EpiScores, while
437 2,954 individuals with GDF15 and 2,808 individuals with NT-proBNP measurements comprised the
438 test samples. When retraining the EpiScores on the full Generation Scotland sample, there were
439 17,489 and 16,963 individuals available for the GDF15 and NT-proBNP EpiScores, respectively.
440 **Supplementary Fig. 3** summarises the training and testing samples used, which included 500
441 individuals with NT-proBNP and 322 individuals with GDF15 measures in the external LBC1936
442 test set.

443 **EpiScore testing**

444 To test EpiScores, the additional variance in protein levels that the EpiScores explained over a null
445 model was quantified by running the following models:

446 Model 1: protein ~ age + sex

447 Model 2: protein ~ age + sex + protein EpiScore

448 The incremental variance (R^2) in protein levels explained due to the protein EpiScore was calculated
449 by subtracting the R^2 derived from model 1 from that in model 2. In these models, scaled, rank-based
450 inverse normalised protein levels were used for testing. Pearson correlation coefficients were also
451 calculated between GDF15 and NT-proBNP levels and their respective EpiScores in the test set and
452 plotted. Protein EpiScores were tested using the described approach in both the Generation Scotland
453 test subset ($N \geq 2,808$) and the individuals in Wave 4 of the LBC1936 external cohort that had
454 measures of the proteins available ($N_{\text{GDF}} = 322$, $N_{\text{NT-proBNP}} = 500$). To assess the incremental variance
455 that could be attributed to genetic architectures of the proteins, polygenic risk scores (PRS) for the
456 proteins were calculated using genome-wide association study (GWAS) summary statistics generated
457 in the Generation Scotland population via BOLT-LMM (44) (see **Supplementary Information**). A
458 summary of sentinel protein quantitative trait loci (pQTLs) identified by conditional and joint
459 analyses (COJO) via Genome-wide Complex Trait Analysis (GCTA) software (45) for the GWAS
460 results is available in **Supplementary Table 15**. PRS were derived using PRSice software (46). The
461 PRS utilised pQTLs that had $P < 5 \times 10^{-8}$, with clumping (parameters: $R^2 = 0.25$, distance = 250kb, $p1$
462 = 1). PRS were modelled in incremental variance assessments singularly and additively with the
463 EpiScores in the test sets in relation to measured proteins.

464 **Cox proportional hazards analyses in Generation Scotland**

465 Cox proportional hazards mixed effects regression models were used to assess the relationship
466 between measured levels of GDF15 ($N=17,489$) and NT-proBNP ($N=16,693$) levels in the baseline
467 Generation Scotland sample and four incident morbidities. The same model structure was also used

468 in the test subset of the Generation Scotland sample where proteins and EpiScores were available for
469 direct comparisons. All models were run using coxme (47) (Version 2.2-16) with a kinship matrix
470 accounting for relatedness. Cases included those diagnosed after baseline who had died, in addition
471 to those who received a diagnosis and remained alive. Controls were censored if disease free at time
472 of death, or at the end of the follow-up period. Date of censoring was set to October 2020, which was
473 the latest date of the GP data linkage information. Protein levels were rank-based inverse normalised
474 and scaled to have a mean of 0 and variance of 1 prior to analyses. Basic models were run adjusting
475 for age and sex. Fully-adjusted models further controlled for alcohol consumption (units consumed
476 in the previous week); social deprivation (assessed by the Scottish Index of Multiple Deprivation
477 (48)); body mass index (kilograms/height in metres squared); educational attainment (an 11-category
478 ordinal variable) and a DNAm-based score for smoking status (49). Each of these covariates was
479 sampled at baseline.

480 An FDR multiple testing correction $P < 0.05$ was applied to basic model associations across all
481 diseases tested. Basic associations were considered to be significant if they had FDR $P < 0.05$.
482 Associations in fully-adjusted models were considered to be significant if they had unadjusted $P <$
483 0.05 . Proportional hazards assumptions were checked through Schoenfeld residuals (global test and a
484 test for the protein variable) using the coxph and cox.zph functions from the survival package (50)
485 (Version 3.2-7). For each association failing to meet the assumption (Schoenfeld residuals $P < 0.05$),
486 a sensitivity analysis was run across yearly follow-up intervals. There were minimal differences in
487 hazard ratios between follow-up periods that did not violate the assumption and those that did. All
488 associations were therefore retained.

489 **COVID-19 analyses in Generation Scotland**

490 Associations between measured levels of GDF15 and NT-proBNP and subsequent long-COVID
491 (derived through CovidLife study survey 3 questionnaire (51)) or COVID-19 hospitalisation (derived

492 through hospital linkage) were tested in the full Generation Scotland population. The preparation of
493 the two binary outcome variables (long-COVID or hospitalisation due to COVID-19) is detailed in
494 **Supplementary Information**. Logistic regression models with either hospitalisation (28 of 491
495 possible individuals) or long-COVID status (87 of 269 possible individuals) were run, with
496 standardised (measured) proteins as the independent variables. Controls were defined as individuals
497 that had COVID-19 but did not experience hospitalisation or long-COVID. Sex and age at COVID
498 testing were adjusted for in the models. The latter was defined as the age at positive COVID-19 test
499 or 1st January 2021 if COVID-19 test data were not available.

500 **EpiScore associations with brain health traits in LBC1936**

501 As longitudinal cognitive testing and brain morphology measures are available in LBC1936,
502 structural equation models (SEM) were used to examine the relationship between each EpiScore and
503 brain health traits (cross-sectionally and longitudinally). Outcomes included: general cognitive
504 ability (*g*), total brain volume, normal-appearing white matter volume, global grey matter volume
505 and white matter hyperintensity volume. Cognitive test data were available at all measurement
506 Waves (mean ages 70, 73, 76, 79 and 82) and brain magnetic resonance imaging (MRI) data were
507 available from the second Wave (mean ages 73, 76, 79 and 82). Information on how the SEM
508 analyses were constructed, with information on the number of individuals with cognitive and brain
509 imaging measures at each Wave is included in **Supplementary Information**. Basic models were run
510 with adjustment for age and sex, whereas fully-adjusted models included further covariates: DNAm-
511 derived immune cell proportion estimates, DNAm-derived smoking score (49), self-reported alcohol
512 consumption, BMI and the Scottish Index of Multiple Deprivation (52). Intercept (cross-sectional
513 associations) and slope (longitudinal change) coefficients were extracted. A total of 1,091 individuals
514 were modelled as part of the SEM analyses, with 895 individuals that had EpiScore measures
515 available at Wave 1.

516 Individuals consented to share disease information from electronic health records from Wave 2 of the
517 study onwards. Cox PH models were run to test associations between Wave 2 GDF15 and NT-
518 proBNP EpiScores and incident dementia diagnoses after Wave 2 baseline, with basic adjustments
519 for age and sex. The test population included 780 individuals who had dementia ascertainment and
520 EpiScore information available at Wave 2, with 108 of these individuals having received a dementia
521 diagnosis post-baseline (mean time-to-event 8.6 years [SD 3.4]). For the 108 incident cases, time-to-
522 event was calculated using age at diagnosis. For controls who had died age at death was used for
523 censoring, whereas age at the date of the dementia consensus meeting decision was taken forward for
524 controls that remained alive.

525 **List of abbreviations**

526 BMI – Body mass index
527 CI – Confidence interval
528 COJO – Conditional and joint analyses
529 COVID-19 – Coronavirus disease 2019
530 CpG – Cytosine-phosphate-guanine
531 DNAm – DNA methylation
532 EpiScore – Epigenetic score
533 EWAS – Epigenome-wide association study
534 FDR – False discovery rate
535 GCTA – Genome-wide Complex Trait Analysis
536 GDF15 – Plasma growth differentiation factor 15
537 GWAS – Genome-wide association study
538 HbA1c – Glycated haemoglobin
539 HR – Hazard ratio
540 ICD – International Classification of Diseases codes
541 IHD – Ischaemic heart disease
542 IgG – Immunoglobulin G

543 MRI – Magnetic resonance imaging
544 NT-proBNP – N-terminal pro-B-type natriuretic peptide
545 LBC1936 – The Lothian Birth Cohort of 1936
546 OR – odds ratio
547 PH – Proportional hazards
548 PIP – Posterior inclusion probability
549 pQTL – Protein quantitative trait loci
550 PRS – Polygenic risk score
551 SD – Standard deviation
552 SE – Standard error
553 SEM – Structural equation model

554

555 **Declarations**

556 **Ethics approval and consent to participate**

557 All components of Generation Scotland received ethical approval from the NHS Tayside Committee
558 on Medical Research Ethics (REC Reference Number: 05/S1401/89). Generation Scotland has also
559 been granted Research Tissue Bank status by the East of Scotland Research Ethics Service (REC
560 Reference Number: 20-ES-0021), providing generic ethical approval for a wide range of uses within
561 medical research.

562 LBC1936 ethical approval was obtained from the Multicentre Research Ethics Committee for
563 Scotland (Wave 1, MREC/01/0/56), the Lothian Research Ethics Committee (Wave 1,
564 LREC/2003/2/29), and the Scotland A Research Ethics Committee (Waves 2-5, 07/MRE00/58). All
565 participants provided written informed consent.

566 **Consent for publication**

567 Not applicable.

568 **Availability of data and materials**

569 The source datasets from the cohorts that were analysed during the current study are not publicly
570 available due to them containing information that could compromise participant consent and
571 confidentiality. Data can be obtained from the data owners. Instructions for accessing Generation
572 Scotland data can be found here: <https://www.ed.ac.uk/generation-scotland/for-researchers/access>;
573 the 'GS Access Request Form' can be downloaded from this site. Completed request forms must be
574 sent to access@generationscotland.org to be approved by the Generation Scotland Access
575 Committee. According to the terms of consent for Generation Scotland participants, access to data
576 must be reviewed by the Generation Scotland Access Committee. Applications should be made to
577 access@generationscotland.org.

578 LBC1936 data are available on request from the Lothian Birth Cohort Study, University of
579 Edinburgh <https://www.ed.ac.uk/lothian-birth-cohorts/data-access-collaboration>.

580 All correspondence and material requests should be sent to Riccardo Marioni at
581 riccardo.marioni@ed.ac.uk. All R code used in analyses is provided at:
582 <https://github.com/DanniGadd/EpiScores-GDF15-NT-proBNP>. Latent Curve Growth Model
583 (LCGM) LBC cognitive code that was adapted for these analyses is also available at:
584 <https://www.ed.ac.uk/lothian-birth-cohorts/summary-data-resources>.

585 **Competing interests**

586 R.E.M is an advisor to the Epigenetic Clock Development Foundation, and has received consultant
587 fees from Optima partners. A.M.M has previously received speaker's fees from Illumina and Janssen
588 and research grant funding from The Sackler Trust. R.F.H. has received consultant fees from
589 Illumina and Optima partners. D.A.G. has received consultant fees from and is currently employed in
590 part-time capacity by Optima partners. D.L.M. is currently employed in part-time capacity by

591 Optima partners. PW reports grant income from Roche Diagnostics in relation to and outside of the
592 submitted work, as well as grant income from AstraZeneca, Boehringer Ingelheim, and Novartis,
593 outside the submitted work and speaker fees from Novo Nordisk, and Raisio outside the submitted
594 work. NS has consulted for Afimmune, Amgen, AstraZeneca, Boehringer Ingelheim, Eli Lilly,
595 Hanmi Pharmaceuticals, Merck Sharp & Dohme, Novartis, Novo Nordisk, Pfizer, and Sanofi; and
596 received grant support paid to his University from AstraZeneca, Boehringer Ingelheim, Novartis, and
597 Roche Diagnostics outside the submitted work. All other authors declare no competing interest.

598 **Funding**

599 **This research was funded in whole, or in part, by the Wellcome Trust [104036/Z/14/Z,**
600 **108890/Z/15/Z, 221890/Z/20/Z, 218493/Z/19/Z]. For the purpose of open access, the author has**
601 **applied a CC BY public copyright licence to any Author Accepted Manuscript version arising**
602 **from this submission** D.A.G. is supported by funding from the Wellcome Trust 4 year PhD in
603 Translational Neuroscience: training the next generation of basic neuroscientists to embrace clinical
604 research [108890/Z/15/Z]. R.F.H is supported through a MRC IEU Short-term Fellowship. E.B. and
605 R.E.M are supported by Alzheimer's Society major project grant AS-PG-19b-010. Hannah Smith is a
606 student on the Translational Neuroscience PhD Programme and is funded by Wellcome grant
607 [218493/Z/19/Z].

608 Generation Scotland received core support from the Chief Scientist Office of the Scottish
609 Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006).
610 Genotyping and DNAm profiling of the GS samples was carried out by the Genetics Core
611 Laboratory at the Edinburgh Clinical Research Facility, Edinburgh, Scotland and was funded by the
612 Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award
613 STratifying Resilience and Depression Longitudinally (STRADL; Reference 104036/Z/14/Z). The
614 DNAm data assayed for Generation Scotland was partially funded by a 2018 NARSAD Young

615 Investigator Grant from the Brain & Behavior Research Foundation (Ref: 27404; awardee: Dr David
616 M Howard) and by a JMAS SIM fellowship from the Royal College of Physicians of Edinburgh
617 (Awardee: Dr Heather C Whalley). Recruitment to the CovidLife study was facilitated by SHARE -
618 the Scottish Health Research Register and Biobank. Roche Diagnostics supported this study through
619 provision of free reagents and a grant for measurement of NT-proBNP and GDF-15 in Generation
620 Scotland.

621 LBC1936 is supported by the Biotechnology and Biological Sciences Research Council, and the
622 Economic and Social Research Council [BB/W008793/1] (which supports S.E.H.), Age UK
623 (Disconnected Mind project), the Milton Damerel Trust, and the University of Edinburgh. S.R.C. is
624 supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal
625 Society (221890/Z/20/Z). Methylation typing was supported by Centre for Cognitive Ageing and
626 Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic
627 Support Fund, The University of Edinburgh, and The University of Queensland.

628 J.M.W., is funded by the UK Dementia Research Institute Ltd which is funded by the Medical
629 Research Council, Alzheimer's Society and Alzheimer's Research UK Institute (award no. UKDRI –
630 Edin002, DRIEdi17/18, and MRC MC_PC_17113). D.M.K. is supported by a British Heart
631 Foundation Intermediate Basic Science Research Fellowship (FS/IBSRF/23/25161).

632

633 **Authors' contributions**

634 D.A.G., and R.E.M. were responsible for the conception and design of the study. D.A.G. carried out
635 all data analyses. D.A.G. and R.E.M. drafted the article. O.C., H.S., E.B., Y.C., R.F.H., D.M.,
636 C.F.R., and D.L.Mc. contributed to methodology. A.C., R.F., S.W.M., R.M.W., K.L.E., S.E.H., D.P.,
637 A.M.M., C.H., T.R., R.M.W., P.W. and N.S. contributed to the data collection and preparation. D.P.,
638 A.T., J.C. and S.R.C. collected and processed LBC1936 data. M.E.B., J.M.W., M.D.C.V.H., and

639 S.M.M., derived brain imaging variables in LBC1936. R.E.M. supervised the project. All authors
640 read and approved the manuscript.

641 **Acknowledgements**

642 We acknowledge the participants of the Generation Scotland and Lothian Birth Cohort studies for
643 their involvement in the data that have made this work possible. The authors thank all LBC1936 and
644 Generation Scotland study research team members who have contributed, and continue to contribute,
645 to ongoing studies.

646 **References**

- 647 1. Hageman SHJ, McKay AJ, Ueda P, Gunn LH, Jernberg T, Hagström E, et al. Estimation of recurrent
648 atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the
649 updated SMART2 algorithm. *Eur Heart J*. 2022 May 7;43(18):1715–27.
- 650 2. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms:
651 new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021 Jul
652 1;42(25):2439–54.
- 653 3. Pence BD. Growth Differentiation Factor-15 in Immunity and Aging. *Front Aging*. 2022 Feb 9;0:8.
- 654 4. Wang Z, Yang F, Ma M, Bao Q, Shen J, Ye F, et al. The impact of growth differentiation factor 15 on the
655 risk of cardiovascular diseases: two-sample Mendelian randomization study. *BMC Cardiovasc Disord*.
656 2020 Dec 1;20(1):1–7.
- 657 5. Tanaka T, Basisty N, Fantoni G, Candia J, Moore AZ, Biancotto A, et al. Plasma proteomic biomarker
658 signature of age predicts health and life span. *eLife*. 2020 Oct 1;9:1–24.
- 659 6. Gadd DA, Hillary RF, Kuncheva Z, Mangelis T, Admanit R, Gagnon J, et al. Blood protein levels predict
660 leading incident diseases and mortality in UK Biobank [Internet]. *medRxiv*; 2023 [cited 2023 May 22]. p.
661 2023.05.01.23288879. Available from:
662 <https://www.medrxiv.org/content/10.1101/2023.05.01.23288879v1>
- 663 7. Shrivastava A, Haase T, Zeller T, Schulte C. Biomarkers for Heart Failure Prognosis: Proteins, Genetic
664 Scores and Non-coding RNAs. *Front Cardiovasc Med*. 2020 Nov 23;7:252.
- 665 8. Birukov A, Eichelmann F, Kuxhaus O, Polemiti E, Fritsche A, Wirth J, et al. Opposing Associations of NT-
666 proBNP With Risks of Diabetes and Diabetes-Related Complications. *Diabetes Care*. 2020 Dec
667 1;43(12):2930–7.
- 668 9. Noveanu M, Bredthardt T, Potocki M, Reichlin T, Twerenbold R, Uthoff H, et al. Direct comparison of
669 serial B-type natriuretic peptide and NT-proBNP levels for prediction of short- and long-term outcome in
670 acute decompensated heart failure. *Crit Care*. 2011 Jan 5;15(1):1–15.

- 671 10. Myhre PL, Vaduganathan M, Claggett B, Packer M, Desai AS, Rouleau JL, et al. B-Type Natriuretic Peptide
672 During Treatment With Sacubitril/Valsartan: The PARADIGM-HF Trial. *J Am Coll Cardiol*. 2019 Mar
673 26;73(11):1264–72.
- 674 11. Zile MR, Claggett BL, Prescott MF, McMurray JJV, Packer M, Rouleau JL, et al. Prognostic Implications of
675 Changes in N-Terminal Pro-B-Type Natriuretic Peptide in Patients With Heart Failure. *J Am Coll Cardiol*.
676 2016 Dec 6;68(22):2425–36.
- 677 12. Myhre PL, Prebensen C, Strand H, Røysland R, Jonassen CM, Rangberg A, et al. Growth Differentiation
678 Factor 15 Provides Prognostic Information Superior to Established Cardiovascular and Inflammatory
679 Biomarkers in Unselected Patients Hospitalized With COVID-19. *Circulation*. 2020 Dec 12;142(22):2128.
- 680 13. Gao L, Jiang D, Wen XS, Cheng XC, Sun M, He B, et al. Prognostic value of NT-proBNP in patients with
681 severe COVID-19. *Respir Res*. 2020 Apr 15;21(1):1–7.
- 682 14. McGrath ER, Himali JJ, Levy D, Conner SC, Decarli C, Pase MP, et al. Growth Differentiation Factor 15 and
683 NT-proBNP as Blood-Based Markers of Vascular Brain Injury and Dementia. *J Am Heart Assoc [Internet]*.
684 2020 Oct 6 [cited 2021 Nov 24];9(19). Available from: <https://pubmed.ncbi.nlm.nih.gov/32921207/>
- 685 15. Walker KA, Chen J, Zhang J, Fornage M, Yang Y, Zhou L, et al. Large-scale plasma proteomic analysis
686 identifies proteins and pathways associated with dementia risk. *Nat Aging*. 2021 May 14;1(5):473–89.
- 687 16. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Nangle C, et al. Epigenetic scores for the
688 circulating proteome as tools for disease prediction. *bioRxiv*. 2021 Jul 7;8(5):2020.12.01.404681.
- 689 17. Cheng Y, Gadd DA, Gieger C, Monterrubio-Gómez K, Zhang Y, Berta I, et al. DNA Methylation scores
690 augment 10-year risk prediction of diabetes. *medRxiv*. 2021 Nov 21;2021.11.19.21266469.
- 691 18. Chybowska AD, Gadd DA, Cheng Y, Bernabeu E, Campbell A, Walker RM, et al. Augmenting clinical risk
692 prediction of cardiovascular disease through protein and epigenetic biomarkers [Internet]. *medRxiv*;
693 2022 [cited 2023 Jan 5]. p. 2022.10.21.22281355. Available from:
694 <https://www.medrxiv.org/content/10.1101/2022.10.21.22281355v1>
- 695 19. Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, et al. Creating and validating
696 a DNA methylation-based proxy for interleukin-6. [cited 2021 Mar 9]; Available from:
697 <https://academic.oup.com/biomedgerontology/advance-article/doi/10.1093/gerona/glab046/6141415>
- 698 20. Zaghlool SB, Kühnel B, Elhadad MA, Kader S, Halama A, Thareja G, et al. Epigenetics meets proteomics in
699 an epigenome-wide association study with circulating blood plasma protein traits. *Nat Commun*. 2020
700 Dec 1;11(1):15.
- 701 21. Conole ELS, Stevenson AJ, Maniega SM, Harris SE, Green C, Valdés Hernández MDC, et al. DNA
702 Methylation and Protein Markers of Chronic Inflammation and Their Associations With Brain and
703 Cognitive Aging. *Neurology*. 2021 Dec 7;97(23):e2340–52.
- 704 22. Stevenson AJ, McCartney DL, Hillary RF, Campbell A, Morris SW, Bermingham ML, et al. Characterisation
705 of an inflammation-related epigenetic score and its association with cognitive ability. *Clin Epigenetics*.
706 2020 Jul 27;12:113.
- 707 23. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts
708 lifespan and healthspan. *Aging*. 2019 Jan 1;11(2):303–27.

- 709 24. Hillary RF, Stevenson AJ, McCartney DL, Campbell A, Walker RM, Howard DM, et al. Epigenetic clocks
710 predict prevalence and incidence of leading causes of death and disease burden. *Clin Epigenet.*
711 2020;12(115).
- 712 25. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Cheng Y, et al. Epigenetic scores for the
713 circulating proteome as tools for disease prediction. *eLife.* 2022 Jan 1;11.
- 714 26. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: Generation
715 scotland: Scottish family health study (GS: SFHS). The study, its participants and their potential for
716 genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689–700.
- 717 27. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: The
718 Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet.*
719 2006;7:74.
- 720 28. Taylor AM, Pattie A, Deary IJ. Cohort profile update: The Lothian birth cohorts of 1921 and 1936. *Int J*
721 *Epidemiol.* 2018;47(4):1042r.
- 722 29. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: The lothian birth cohorts of 1921 and 1936. *Int J*
723 *Epidemiol.* 2012;41(6):1576–84.
- 724 30. Trejo Banos D, McCartney DL, Patxot M, Anchieri L, Battram T, Christiansen C, et al. Bayesian
725 reassessment of the epigenetic architecture of complex traits. *Nat Commun.* 2020;11.
- 726 31. Gadd DA, Hillary RF, McCartney DL, Shi L, Stolicy A, Robertson N, et al. Integrated methylome and
727 phenome study of the circulating proteome reveals markers pertinent to brain health. *Zenodo*
728 [Internet]. 2022 Jul 6 [cited 2022 Jul 6]; Available from: <https://zenodo.org/record/6801458>
- 729 32. Hillary RF, Trejo-Banos D, Kousathanas A, McCartney DL, Harris SE, Stevenson AJ, et al. Multi-method
730 genome- And epigenome-wide studies of inflammatory protein levels in healthy older adults. *Genome*
731 *Med.* 2020 Jul 8;12:60.
- 732 33. Hillary RF, McCartney DL, Harris SE, Stevenson AJ, Seeboth A, Zhang Q, et al. Genome and epigenome
733 wide studies of neurological protein biomarkers in the Lothian Birth Cohort 1936. *Nat Commun.* 2019
734 Dec 1;10(1):3160.
- 735 34. Gadd DA, Hillary RF, McCartney DL, Shi L, Stolicy A, Robertson NA, et al. Integrated methylome and
736 phenome study of the circulating proteome reveals markers pertinent to brain health. *Nat Commun.*
737 2022 Aug 9;13(1):4670.
- 738 35. Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene–Stress–Epigenetic Regulation of FKBP5: Clinical
739 and Translational Implications. *Neuropsychopharmacology.* 2016 Jan;41(1):261–74.
- 740 36. Smyth LJ, Kilner J, Nair V, Liu H, Brennan E, Kerr K, et al. Assessment of differentially methylated loci in
741 individuals with end-stage kidney disease attributed to diabetic kidney disease: an exploratory study.
742 *Clin Epigenetics.* 2021 May 1;13(1):99.
- 743 37. Rawshani A, Kjölhede EA, Rawshani A, Sattar N, Eeg-Olofsson K, Adiels M, et al. Severe COVID-19 in
744 people with type 1 and type 2 diabetes in Sweden: A nationwide retrospective cohort study. *Lancet Reg*
745 *Health – Eur* [Internet]. 2021 May 1 [cited 2023 Aug 20];4. Available from:
746 [https://www.thelancet.com/journals/lanep/article/PIIS2666-7762\(21\)00082-X/fulltext](https://www.thelancet.com/journals/lanep/article/PIIS2666-7762(21)00082-X/fulltext)

- 747 38. Ortega E, Corcoy R, Gratacòs M, Claramunt FXC, Mata-Cases M, Puig-Treserra R, et al. Risk factors for
748 severe outcomes in people with diabetes hospitalised for COVID-19: a cross-sectional database study.
749 *BMJ Open*. 2021 Jul 1;11(7):e051237.
- 750 39. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights
751 into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun* 2021 121. 2021
752 Nov 24;12(1):1–13.
- 753 40. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al. Identifying dementia
754 outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur*
755 *J Epidemiol*. 2019 Jun 15;34(6):557–65.
- 756 41. Mullin DS, Stirland LE, Buchanan E, Convery CA, Cox SR, Deary IJ, et al. Identifying dementia using
757 medical data linkage in a longitudinal cohort study: Lothian Birth Cohort 1936. *BMC Psychiatry*. 2023
758 May 1;23:303.
- 759 42. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA
760 methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- 761 43. R. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical
762 Computing, Vienna, Austria. URL <https://www.R-project.org/>. Accessed April 2021. 2020.
- 763 44. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed model association for biobank-scale data sets.
764 *Nat Genet*. 2018 Jul;50(7):906–8.
- 765 45. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-
766 SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat*
767 *Genet*. 2012 Apr;44(4):369–75.
- 768 46. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015 May
769 1;31(9):1466–8.
- 770 47. Therneau TM. *coxme: Mixed Effects Cox Models*. R package version 2.2-16. [https://CRAN.R-](https://CRAN.R-project.org/package=coxme)
771 [project.org/package=coxme](https://CRAN.R-project.org/package=coxme). Accessed April 2021. 2020;
- 772 48. GovScot. Scottish Government. The Scottish Index of Multiple Deprivation (SIMD); 1-20. (2016).
773 Available from: <http://www.gov.scot/Resource/0050/00504809.pdf>. Accessed April 2021. 2016.
- 774 49. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: A robust classifier to determine
775 smoking status from DNA methylation data. *Epigenomics*. 2019;11(13):1469–86.
- 776 50. Therneau TM. *A Package for Survival Analysis in R*. R package version 3.2-7, [https://CRAN.R-](https://CRAN.R-project.org/package=survival)
777 [project.org/package=survival](https://CRAN.R-project.org/package=survival). Accessed April 2021. 2020;
- 778 51. Fawns-Ritchie C, Altschul DM, Campbell A, Huggins C, Nangle C, Dawson R, et al. CovidLife: a resource to
779 understand mental health, well-being and behaviour during the COVID-19 pandemic in the UK.
780 *Wellcome Open Res* 2021 6176. 2021 Jul 7;6:176.
- 781 52. GovScot Scottish Government. Scottish Index of Multiple Deprivation 2006: Technical Report. Available
782 at: <https://www.gov.scot/publications/scottish-index-multiple-deprivation-2006-technical-report/>.
783 Accessed August 2023.

784