

1 **Recurrent stroke prediction by applying a stroke polygenic risk score in the Japanese**
2 **population**

3

4 Naoki Kojima (MSc)¹, Masaru Koido (PhD)¹, Yunye He (PhD)¹, Yuka Shimmori (MSc)¹,
5 Tsuyoshi Hachiya (PhD)¹, BioBank Japan, Stéphanie Debette (MD, PhD)^{2,3}, Yoichiro
6 Kamatani (MD, PhD)¹

7

8 1. Laboratory of Complex Trait Genomics, Department of Computational Biology and
9 Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo,
10 Tokyo, Japan.

11 2. Bordeaux Population Health Research Center, University of Bordeaux, Inserm, UMR
12 1219, Bordeaux, France.

13 3. Department of Neurology, Institute for Neurodegenerative Diseases, CHU de Bordeaux,
14 Bordeaux, France.

15

16 **Short title:** Recurrent stroke prediction by PRS

17

18 **Corresponding author:**

19 Yoichiro Kamatani¹

20 Address: 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639 Japan

21 Email: kamatani.yoichiro@edu.k.u-tokyo.ac.jp

22

23 **Total word count of the manuscript (including Title Page, Abstract, Text, References,**
24 **Tables, and Figures Legends): 6306**

25

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

26

27 **Abstract**

28 **Background**

29 Recently, various polygenic risk score (PRS)-based methods were developed to improve
30 stroke prediction. However, current PRSs (including cross-ancestry PRS) poorly predict
31 recurrent stroke. Here, we aimed to determine whether the best PRS for Japanese individuals
32 can also predict stroke recurrence in this population by extensively comparing the methods
33 and maximizing the predictive performance for stroke onset.

34

35 **Methods**

36 We used data from the BioBank Japan (BBJ) 1st cohort (n=179,938) to derive and optimize
37 the PRSs using a 10-fold cross-validation. We integrated the optimized PRSs for multiple
38 traits, such as vascular risk factors and stroke subtypes to generate a single PRS using the
39 meta-scoring approach (metaGRS). We used an independent BBJ 2nd cohort (n=41,929) as a
40 test sample to evaluate the association of the metaGRS with stroke and recurrent stroke.

41

42 **Results**

43 We analyzed recurrent stroke cases (n=174) and non-recurrent stroke controls (n=1,153)
44 among subjects within the BBJ 2nd cohort. After adjusting for known risk factors, metaGRS
45 was associated with stroke recurrence (adjusted OR per SD 1.18 [95% CI: 1.00–1.39,
46 p=0.044]), although no significant correlation was observed with the published PRSs. We
47 administered three distinct tests to consider the potential index event bias; however, the
48 outcomes derived from these examinations did not provide any significant indication of the
49 influence of index event bias. The high metaGRS group without a history of hypertension
50 had a higher risk of stroke recurrence than that of the low metaGRS group (adjusted OR 2.24
51 [95% CI: 1.07–4.66, p=0.032]). However, this association was weak in the hypertension

52 group (adjusted OR 1.21 [95% CI: 0.69–2.13, p=0.50]).

53

54 **Conclusions**

55 The metaGRS developed in a Japanese cohort predicted stroke recurrence in an independent

56 cohort of patients. In particular, it predicted an increased risk of recurrence among stroke

57 patients without hypertension. These findings provide clues for additional genetic risk

58 stratification and help in developing personalized strategies for stroke recurrence prevention.

59

60 **Keywords:** recurrent stroke, stroke, polygenic risk score, LDpred2, risk factor, index event

61 bias, hypertension, metaGRS

62 **Non-standard abbreviations and acronyms**

63	PRS	polygenic risk score
64	P+T	pruning and thresholding
65	BBJ	BioBank Japan
66	BBJ1	BBJ 1 st cohort
67	BBJ2	BBJ 2 nd cohort
68	ToMMo	Tohoku Medical Megabank
69	AIS	all ischemic stroke
70	IPW	inverse probability weighting
71	LD	linkage disequilibrium
72	GWAS	genome-wide association study
73	PC	principal component
74	IPW	inverse probability weight
75	AUC	area under the curve
76	LAS	large artery stroke
77	SVS	small vessel stroke
78	CES	cardioembolic stroke
79	TIS	transient ischemic attack
80	HWE	Hardy-Weinberg equilibrium
81	WGS	whole genome sequencing
82	MI	myocardial infarction
83	SAP	stable angina pectoris
84	AP	unstable angina pectoris
85	AF	atrial fibrillation
86	DM	diabetes mellitus

87	SM	smoking
88	BMI	body mass index
89	HE	height
90	SBP	systolic blood pressure
91	DBP	diastolic blood pressure
92	TC	total cholesterol
93	TG	triglyceride
94	HDL	high-density lipoprotein
95	LDL	low-density lipoprotein
96		

97 Introduction

98

99 Stroke is a major cause of mortality in Japan, with 56,000 deaths reported in 2020.¹ The
100 conventional risk factors for stroke include hypertension, high waist-to-hip ratio, smoking,
101 cardiac causes, dyslipidemia, and diabetes mellitus.² In Japan, the stroke recurrence rate is up
102 to 30–50% during 5-10 years of follow-up after the first stroke.^{3,4} Accordingly, it will be
103 medically beneficial to stratify high-risk groups for recurrent stroke among those who have
104 experienced a stroke to potentially generate more intensive secondary prevention strategies
105 than current recommendations.

106 Genome-wide association studies (GWAS) have identified many disease-susceptibility
107 variants associated with complex traits.⁵ A polygenic risk score (PRS) is the weighted
108 summation of the individual genetic effects of these variants. Its weighting strategy varies
109 depending on the construction method; traditionally, only significant variants are used in
110 developing this score. The recently developed PRS methods involve non-significant variants
111 and updated effect weights and consider the linkage disequilibrium (LD) structure. The
112 development of PRS methods has helped stratify high-risk groups for complex traits,^{6–11}
113 including stroke.¹²

114 Polygenic risk scores developed using the 32 genome-wide significant ($p < 5.0 \times 10^{-8}$) variants
115 or 90 marginally associated ($p < 1.0 \times 10^{-5}$) variants (PRS₉₀) from the MEGASTROKE study¹²
116 are associated with stroke onset in subjects of European ancestry.^{12,13} The meta-scoring PRS
117 approach (metaGRS) includes 3.2 million variants by combining PRSs for stroke subtypes,
118 risk factors, and comorbidities by adjusting the effect weight via elastic-net logistic
119 regression; this approach has an improved predictive performance for stroke compared to
120 that of PRS₉₀.¹⁴ MetaGRS can predict stroke incidence independent of environmental factors
121 and could help motivate individuals with high genetic risk to make lifestyle changes for

122 stroke prevention (although not yet implemented in clinical practice outside a research
123 setting).¹⁵ The PRS shows reduced transferability between populations. Additionally, a PRS
124 developed using various variants derived from Japanese GWAS successfully predicted stroke
125 onset in the Japanese population.^{16,17} Most recently, the GIGASTROKE study proposed an
126 integrated PRS approach among PRSs derived from populations of multiple ancestries using
127 the metaGRS framework (iPGS), which showed a better predictive ability than the
128 MEGASTROKE European or East Asian PRS.¹⁸ However, the PRS did not successfully
129 predict stroke recurrence; for example, PRS₃₂ and iPGS did not significantly predict stroke
130 recurrence after adjusting for clinical comorbidities, with notably smaller effect sizes than for
131 non-recurrent stroke.^{12,18} Furthermore, the potential effect of index event (also known as
132 “collider”) bias that may distort the association of PRS was suspected.^{19,20}

133 The optimal method to improve the predictive accuracy of PRS depends on the population-
134 and trait-specific genetic architecture.^{21–30} Therefore, we compared different PRS methods
135 and determined whether the best PRS can predict the onset of recurrent stroke in a Japanese
136 population.

137

138

139 **Methods**

140 The workflow of this study is shown in Figure 1. This article follows the TRIPOD
141 (Transparent reporting of a multivariable prediction model for individual prognosis or
142 diagnosis) reporting guidelines.

143

144 **Study subjects and quality control**

145 BioBank Japan (BBJ) involves physicians diagnosing all ischemic stroke (AIS) cases at the
146 collaborating hospitals. BBJ was established in 2003 and recruited 267,000 patients from 12
147 medical institutions (66 hospitals) in two phases.^{31–33} The recruited patients had at least one
148 of the 51 primarily multifactorial (common) diseases, which accounted for 440,000 cases.
149 We used BBJ 1st cohort (BBJ1) data to derive PRSs and available independent BBJ 2nd
150 cohort (BBJ2) data to evaluate the performance of PRSs in predicting AIS and recurrent AIS.
151 Recurrent AIS information was unavailable for the BBJ1 data. In BBJ2, any AIS cases
152 (n=1,470), AIS-free controls (n=40,459), recurrent AIS cases (n=174), and non-recurrent AIS
153 controls (n=1,153) were available. The mean duration from the first episode of AIS onset to
154 recurrent AIS onset was 4.88 years. Detailed sample characteristics are listed in
155 Supplementary Methods and Table 1.

156 This study was approved by the ethics committee of the Institute of Medical Science, the
157 University of Tokyo, Japan. Quality control, pre-phasing, and genotype imputation were
158 conducted using PLINK(v2.0),^{34–37} Eagle (v2.4.1), and Minimac4 (v1.0.2), respectively. The
159 detailed processes are presented in the Supplementary Methods.

160

161 **Constructing PRSs for AIS**

162 Unbiased PRSs were obtained by applying a 10-fold cross-validation to select the model and
163 optimize the parameters.^{26,38} Briefly, BBJ1 samples were randomly split into ten equal-sized

164 subsamples. We retained one subsample for validation and the others for training. We
165 repeated this process 10 times, with each of the ten subsamples used exactly once for
166 validation. A GWAS was conducted on the training set in each iteration, adjusted for age,
167 sex, and the first 10 principal components (PCs) via Firth logistic regression using PLINK
168 (v.2.0).³⁴

169 We obtained the weights of variants for PRS from the GWAS summary statistics of the
170 training set using five PRS methods—P+T (PLINK (v.1.9)³⁵ for clumping), LDpred2,³⁹
171 Lassosum2,⁴⁰ (LDpred2 and Lassosum2 by bigsnpr package (v.1.7.2) in R (v.3.5.0)), PRS-CS
172 (v.1.0.0),⁴¹ and PRS-CSx (v.1.0.0).⁴² The PRS-CSx integrated BBJ1 with the European
173 stroke GWAS summary statistics (MEGASTROKE; the largest study available at this study
174 design)⁴³ by learning an optimal linear combination. We used combinations of parameters for
175 P+T (1,224 parameters), LDpred2 (126 parameters), Lassosum2 (200 parameters), PRS-CS
176 (9 parameters), and PRS-CSx (9 parameters), as described in the Supplementary Methods.

177 Subsequently, the PRSs for the validation sample were calculated using the weights obtained
178 from the training samples. The accuracy for predicting AIS cases was evaluated from
179 Nagelkerke's R^2 (simply " R^2 " from this point onwards)^{29,44} after adjusting for age, sex, and
180 the first 10 PCs. We calculated the mean R^2 over 10 cross-validation results for each method
181 with each parameter after a 10-fold cross-validation. We chose the method and parameters
182 that maximized incremental R^2 (PRS_{AIS}) among these PRSs.

183 We further integrated the PRS_{AIS} with PRSs of vascular risk factors, such as stroke subtypes
184 and comorbid diseases presence, using the elastic net framework to construct a metaGRS
185 using the glmnet package (v.4.1.3) in R (v.4.1.0). Nine binary traits and eight quantitative
186 traits of vascular risk factors reported in a previous study¹⁴ are described in the
187 Supplementary Methods. Binary traits were determined by conducting GWAS and attempting
188 to obtain unbiased weights using cross-validations. The effect weights from the derivation

189 sample every 17 traits were calculated using PRS-CS-auto since it did not require an
190 independent validation sample set for parameter optimization and performed well for various
191 traits.^{21,26,39,41,45} Subsequently, we used a validation sample to calculate the weight of AIS
192 and the 17-trait PRSs to predict AIS using elastic-net logistic regression. We conducted a 10-
193 fold cross-validation and used the mean weight for testing.

194 We used PLINK (v2.0)³⁴ to calculate the individual PRS by aggregating the effect estimates
195 multiplied by each imputed dosage into a single score per person.

196

197 **Risk factors**

198 The following seven risk factors that were previously utilized as covariates¹² were used as
199 covariates for testing: hypertension (systolic blood pressure>140 mmHg, diastolic blood
200 pressure>90 mmHg, or hypertension history), hyperlipidemia, diabetes (all types), smoking
201 (current smoker), vascular disease (myocardial infarction, peripheral artery disease, stable
202 angina pectoris, and unstable angina pectoris), congestive heart failure, and atrial fibrillation
203 (including atrial flutter). A sample was considered to have a risk factor status if it had that
204 status at enrollment or from historical records (Tables 1 and S1).

205

206 **Assessment of the association of PRS with AIS and AIS recurrence**

207 We used a single selected method with optimized parameters and calculated the metaGRS in
208 independent testing sample sets. We used a logistic regression model to assess the association
209 of the PRS using the two case-control settings for AIS (any-AIS versus AIS-free controls)
210 and AIS recurrence (recurrent AIS versus non-recurrent AIS). We also applied two other
211 combinations of case-controls: recurrent AIS versus AIS-free controls and non-recurrent AIS
212 versus AIS-free controls (Figure 1). Furthermore, we examined additional PRS contributions
213 of the seven risk factors to predictive accuracy and discriminative performance using the

214 values of R^2 and area under the curve (AUC), according to our previous studies.^{14,46}
215 Additionally, we evaluated the performance of the following PRSs derived from other
216 milestone studies for stroke prediction: 32 genome-wide significant variants (for any stroke,
217 ischemic stroke, or ischemic stroke subtypes) from the MEGASTROKE cross-ancestry study
218 of 524,354 individuals (PRS₃₂),⁴³ 89 genome-wide significant variants of 1,614,080 multi-
219 population individuals (PRS₈₉), and 6,010,730 variants of the East Asian PRS developed
220 from 9,809 individuals (iPGS_{EAS}),¹⁸ both from the GIGASTROKE study. The PRS
221 calculation process is described in the Supplementary Methods.

222

223 **Considering potential index event bias**

224 An index event bias may be induced when the samples are only selected from cases.⁴⁷ We
225 evaluated the extent to which the index event bias affected our results since we used case-
226 only samples in this study. The association of PRSs with recurrent AIS was evaluated after
227 adjusting for seven risk factors, in addition to age, sex, and the first 10 PCs. Adjusting for
228 such confounding bias will not be enough to eliminate bias for a recurrence association
229 study.⁴⁸ Therefore, we sought to mitigate a potential index event bias by applying three
230 distinct methodologies.^{48,49} First, we utilized linear and logistic regression models to assess
231 the relationships between metaGRS and covariates within the any-AIS case (n=1,327) and
232 AIS-free control groups (n=40,459). Initially, we did not adjust for age, sex, the seven risk
233 factors, or the first 10 PCs. Subsequently, we observed the distributions of covariate values
234 across the metaGRS quintiles. We performed statistical tests to detect heterogeneity in the
235 estimates between the prevalent case and control groups, following the methodology of a
236 prior study.^{19,50} Second, we refined our analysis by adjusting for associations between
237 metaGRS and AIS recurrence while considering the differential effects of covariates,
238 according to a previous method.¹⁹ Finally, we applied the inverse probability weighted (IPW)

239 approach⁴⁷ to comprehensively account for index event bias. Collectively, these analytical
240 approaches were adopted to enhance the validity of our findings.

241

242 **Association of metaGRS and AIS recurrence in patients with/without hypertension**

243 Logistic regression was conducted in subgroups with and without hypertension and metaGRS
244 tertiles among recurrent AIS cases (with hypertension n=107, without hypertension n=67)
245 and non-recurrent AIS controls (with hypertension n=731, without hypertension n=422) to
246 assess the relationship between the PRS and the risk of AIS recurrence. The low metaGRS
247 tertile was set as the reference group and adjusted for age, sex, and the first 10 PCs.

248

249 **Statistical analysis**

250 The mean with standard deviation (SD) or proportion of factors was reported for the baseline
251 characteristics of testing samples. The incremental value (R^2 or AUC) was estimated from
252 the differences between patients with and without PRSs of the fitted values of age, sex, first
253 10 PCs, and seven risk factors^{48,57,58} and calculated as the 95% confidence interval. The
254 pROC package (v.1.18.0) in R was used to determine the discriminative ability of the AUC.
255 The IPW package (v.1.2) in R was used for the IPW approach. R (v. 3.5.0) was used to
256 perform logistic regression to calculate R^2 , Pearson's correlation coefficient, and linear
257 regression. All statistical tests were two-sided. The significance level was set at $p = 0.05$.

258

259

260 Results

261

262 Derivation of effect weight

263 The imputed genotype data of 17,621 AIS cases and 162,317 controls without an AIS
264 diagnosis were used for 9,622,629 autosomal variants after implementing quality control of
265 the BBJ1 dataset (Tables S2-S4).

266 We conducted a 10-fold cross-validation to adjust the parameters and select the best PRS
267 associated with AIS. We performed GWAS 10 times on 90% of the randomly selected BBJ1
268 dataset (training data). We successfully detected previously reported^{43,51} signals in each
269 dataset ($p < 5 \times 10^{-8}$), including *SH3PXD2A*, *CCDC63* (eight times), *CUX2*, and *LINC02356*
270 (every time) (Figure S1, Table S5).

271 We confirmed some expected characteristics of each PRS method (such as low accuracy)
272 using only genome-wide significant variants (Tables S6–10 and Supplementary Notes). The
273 mean incremental R^2 values of each scoring method with the best-performed parameters
274 were 0.0038 (95% CI: 0.0030–0.0046), 0.00443 (95% CI: 0.0035–0.0054), 0.0039 (95% CI:
275 0.0030–0.0048), 0.00441 (95% CI: 0.0036–0.0053), and 0.0037 (95% CI: 0.0031–0.0042)
276 for P+T, LDpred2, Lassosum2, PRS-CS, and PRS-CSx, respectively (Table 2). We chose
277 LDpred2 with the parameter set of ρ -value = 0.0056, a heritability-value = $1.0 \times h^2_{LDSC}$,
278 where h^2_{LDSC} is the heritability estimate from the constrained LD score regression⁵², and a
279 no-sparse model for subsequent analyses, since it showed the best mean incremental R^2 value
280 among the five methods.

281 We observed an average number of nonzero weights for 8.4 traits after computing the
282 metaGRS via elastic net regularization 10 times (10-fold). The metaGRS weight of AIS was
283 highest (mean=0.123, SD=0.026), followed by diastolic blood pressure (mean=0.039,
284 SD=0.039), atrial fibrillation (mean=0.023, SD=0.024), and myocardial infarction

285 (mean=0.018, SD=0.025) (Figure S2 and Table S11). Only the triglyceride weights were zero
286 at all 10 measurements among the 18 traits included in the metaGRS calculation. The number
287 of variants used for metaGRS was 1,014,026; a total of 1,011,847 variants (99.8%) remained
288 after matching with the BBJ2 dataset.

289

290 **Association of metaGRS with AIS cases and recurrent AIS**

291 We used the imputed genotype data of 1,470 AIS cases and 40,459 controls without a
292 diagnosis of AIS for 59,387,070 variants from the BBJ2 dataset to test the association of
293 metaGRS with AIS and AIS recurrence. The AIS case-only sample of the BBJ2 was used to
294 analyze AIS recurrence. Table 1 presents the characteristics of the test samples.

295 MetaGRS was associated with AIS diagnosis after adjusting for age, sex, first 10 PCs, and
296 seven risk factors (adjusted OR, 1.21 [95% CI: 1.15–1.27, $p=2.89\times 10^{-12}$]), as previously
297 reported.^{14,18} MetaGRS was also associated with AIS recurrence compared with recurrence-
298 free AIS (adjusted OR 1.18 [95% CI: 1.00–1.39, $p=0.044$]; Table 3 and Figure S3). MetaGRS
299 showed stronger association when comparing recurrent AIS with AIS-free controls (adjusted
300 OR 1.37 [95% CI: 1.18–1.59, $p=5.35\times 10^{-5}$]; Table S12).

301 The contribution of the metaGRS and traditional risk factors showed an AIS prediction
302 accuracy with an R^2 value of 0.06 and an AUC of 0.689 after constructing the baseline model
303 using age, sex, the first 10 PCs, and seven risk factors. The incremental AUCs were 0.0087
304 and 0.0123 for AIS and AIS recurrence, respectively when metaGRS was added to the
305 baseline model (Table 3). In our dataset, clinical risk factors (including hypertension) were
306 related to AIS diagnosis but were insignificantly associated with AIS recurrence (Table S1).

307 We assessed the prediction performance of previously developed PRSs for AIS and AIS
308 recurrence in our dataset. After matching with the BBJ2 dataset (Supplementary Methods),
309 27, 84, and 5,756,652 variants remained in PRS₃₂, PRS₈₉, and iPGS, respectively. We

310 confirmed their association with AIS diagnosis; adjusted ORs were 1.11 [95% CI: 1.06–1.17,
311 $p=4.23\times 10^{-5}$], 1.08 [95% CI: 1.03–1.14, $p=2.96\times 10^{-3}$], and 1.26 [95% CI: 1.20–1.33,
312 $p=1.24\times 10^{-17}$] for PRS₃₂, PRS₈₉, and iPGS, respectively (Table 3, Figure S3); however, a
313 significant association was not observed between PRSs and AIS recurrence (p-values of 0.41,
314 0.054, and 0.37, respectively; Table 3 and Figure S3). Our Japanese optimized metaGRS was
315 the only PRS significantly associated with AIS recurrence in this study.

316

317 **Analyzing for potential index event bias**

318 We observed the values of covariates at the AIS-free control group and any-AIS case group
319 in each quintile. We did not find any significant heterogeneous relationships between the
320 covariates and the metaGRS in terms of regression estimates in the prevalent case and
321 control samples ($p>0.05$, Table S13).

322 We used three different variable models—i) age and sex; ii) age, sex, and seven risk factors;
323 and iii) age, sex, the first 10 PCs, and seven risk factors—to determine the association
324 between metaGRS and recurrent AIS; none of these confounders significantly influenced our
325 results (Figure S4).

326 We compared the association results of IPW adjusted (accounting for index event bias) with
327 those of non-adjusted IPW (accounting for confounding bias). The results remained almost
328 unchanged, but the 95% confidence intervals overlapped (Figure S5). A comparison of the
329 three distinct models did not indicate an effect of index event bias.

330

331 **Association of metaGRS and AIS recurrence in patients with/without hypertension**

332 We divided the test sample into subgroups according to the presence or absence of a history
333 of hypertension and evaluated the risk effect of the metaGRS tertile. The high metaGRS
334 group without a history of hypertension showed a higher risk effect for AIS recurrence

335 compared to the low metaGRS group (OR of the high metaGRS group compared to that of
336 the low metaGRS group was 2.24 [95% CI: 1.07–4.66, $p=0.032$], Figure 2 and Table S14).
337 However, no significant association was observed between the metaGRS and AIS recurrence
338 in the group with a history of hypertension (the OR of the high metaGRS group compared to
339 that of the low metaGRS group was 1.21 [95% CI: 0.69–2.13, $p=0.50$] (Figure 2 and Table
340 S14).

341

342

343 Discussion

344 We successfully examined the association between recurrent AIS and our best model PRS
345 (metaGRS using LDpred2); the adjusted OR was 1.18 for each unit of SD increase in PRS.
346 Our metaGRS showed stronger (adjusted OR per SD=1.37) association when comparing
347 recurrent AIS with AIS-free controls. Furthermore, a high PRS was associated with AIS
348 recurrence particularly in groups without a history of hypertension (OR of the top vs. bottom
349 metaGRS tertile=2.24). These results are consistent with the result of a previous study
350 wherein the stroke prediction accuracy of the PRS was high in the group with low
351 CHA2DS2-VASc scores.¹² These results indicate the utility of the PRS in developing more
352 precise strategies to prevent AIS recurrence in individuals with a high PRS who do not have
353 high profiles based on clinical risk factors.

354 We attempted to mitigate potential index event bias since our purpose was to specifically
355 determine the efficacy of PRS among AIS patients. It is difficult to predict and provide an
356 accurate assessment of recurrent AIS based on genetic predisposition owing to the possible
357 effect of index event bias leading to a distorted association in studies on recurrent
358 stroke.^{19,20,53} This study found no evidence of heterogeneous associations between covariates
359 and the metaGRS; we did not find any evidence of a solid collider bias of known variables.
360 By applying IPW, we confirmed that our results support the association between metaGRS
361 and recurrent AIS.

362 There are three putative reasons our metaGRS could predict AIS recurrence. First, the
363 metaGRS algorithm combines the genetic profiles of related traits and slightly improves the
364 performance, reaching the level of significance. Second, the performances of PRS-CS
365 (incremental $R^2=0.00441$) and LDpred2 (incremental $R^2=0.00443$) in our validation analysis
366 were better than those of other traditional PRS methods, such as P+T (incremental
367 $R^2=0.0038$). This demonstrated the importance of using shrinkage estimation methods that

368 consider LD to predict AIS and AIS recurrence. Third, we restricted to use only single
369 matched ancestry throughout.

370 Nevertheless, our study had several limitations. First, the sample size for recurrent AIS needs
371 to be increased (n=174 at testing), even in the largest hospital-based biobank in Japan.
372 Compared to our metaGRS, iPGS constructed in GIGASTROKE showed a stronger
373 association for AIS and weaker association for AIS recurrence. Although potential
374 discrepancies exist, both PRSs (metaGRS and iPGS) exhibit the same direction of effects and
375 have overlapping confidence intervals (Table 3, Figure S3). Second, despite using as many
376 covariates (age, sex, the first 10 PCs, and seven risk factors (hypertension, hyperlipidemia,
377 diabetes mellitus, smoking, vascular disease, congestive heart failure, and atrial fibrillation))
378 as possible based on a previous study,¹² other confounders might have affected our results.
379 Finally, there may have been an index event bias that was not fully detected by each method
380 that we implemented; however this risk was minimized using multiple approaches. Further
381 studies using different sample sets (including other ancestry groups) are warranted to confirm
382 the prediction of recurrent stroke using the PRS.

383

384 In conclusion, our study indicated that PRS can be applied to predict AIS recurrence in
385 addition to traditional clinical risk factors. This shows the potential utility of PRS in
386 population-based screening and in the clinical setting. Overall, our results indicate that
387 stratifying high-risk groups for recurrent stroke among those who have experienced a stroke
388 could be medically beneficial and help in developing personalized strategies for recurrence
389 prevention. Our results suggest that it might be particularly useful in patients with AIS
390 without hypertension, although this requires confirmation in independent datasets.

391

392 **Acknowledgments**

393 We want to acknowledge all the participants and investigators of BioBank Japan.
394 Supercomputing resources were provided by the Human Genome Center, Institute of Medical
395 Science, and the University of Tokyo (<http://sc.hgc.jp/shirokane.html>). The ToMMo
396 summary statistics were derived from jMorp ([https://jmorp.megabank.tohoku.ac.jp/gwas-](https://jmorp.megabank.tohoku.ac.jp/gwas-studies/TGA000007)
397 [studies/TGA000007](https://jmorp.megabank.tohoku.ac.jp/gwas-studies/TGA000007)). The MEGASTROKE project received funding from sources specified
398 at <https://www.megastroke.org/acknowledgements.html>

399

400 **Data availability**

401 The weights of metaGRS derived in this study will be publicly available after acceptance.
402 Genotype datasets were deposited in the National Bioscience Database Center Human
403 Database (BBJ1, Research ID: hum0014; BBJ2, Research ID: hum0311).

404

405 **Source of Funding**

406 This research was supported by the Ministry of Education, Culture, Sports, Sciences and
407 Technology (MEXT) of the Japanese government and the Japan Agency for Medical
408 Research and Development (AMED) under grant nos. JP18km0605001/ JP23tm0624002 (the
409 BioBank Japan project), JP223fa627011 (Y.K.), and JP23tm0524003 (Y.K.).

410

411 **Disclosures**

412 Y.K. holds stock of StaGen Co, Ltd.

413

414 **Supplementary Material**

415 Supplementary Methods

416 Supplementary Notes

417 Tables S1–S15

418 Figures S1–S5

419

420 References

421

- 422 1. Ministry of Health, L. and W. Vital Statistics of Japan,
423 <https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei20/index.html>. (2020).
- 424 2. O'Donnell, M. J. *et al.* Global and regional effects of potentially modifiable risk
425 factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control
426 study. *The Lancet* **388**, 761–775 (2016).
- 427 3. Hata, J. *et al.* Ten year recurrence after first ever stroke in a Japanese community: The
428 Hisayama study. *J Neurol Neurosurg Psychiatry* **76**, 368–372 (2005).
- 429 4. Takashima, N. *et al.* Long-term survival after stroke in 1.4 million Japanese
430 population: Shiga stroke and heart attack registry. *J Stroke* **22**, 336–344 (2020).
- 431 5. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide
432 association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896–D901 (2017).
- 433 6. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of
434 polygenic risk scores. *Nat Rev Genet* **19**, 581–590 (2018).
- 435 7. Khera, A. v. *et al.* Whole-Genome Sequencing to Characterize Monogenic and
436 Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial
437 Infarction. *Circulation* **139**, 1593–1602 (2019).
- 438 8. Konuma, T. & Okada, Y. Statistical genetics and polygenic risk score for precision
439 medicine. *Inflamm Regen* **41**, (2021).
- 440 9. Levin, M. G. & Rader, D. J. Polygenic Risk Scores and Coronary Artery Disease.
441 *Circulation* 637–640 (2020) doi:10.1161/CIRCULATIONAHA.119.044770.
- 442 10. Plagnol, V. Polygenic score development in the era of large-scale biobanks. *Cell*
443 *Genomics* **2**, 100088 (2022).
- 444 11. Khera, A. v. *et al.* Genome-wide polygenic scores for common diseases identify
445 individuals with risk equivalent to monogenic mutations. *Nature Genetics* vol. 50
446 1219–1224 Preprint at <https://doi.org/10.1038/s41588-018-0183-z> (2018).
- 447 12. Marston NA *et al.* Clinical Application of A Novel Genetic Risk Score Predicts
448 Ischemic Stroke in Patients With Cardiometabolic Disease. *Circulation*. 2021 Feb
449 2;143(5):470-478. doi: 10.1161/CIRCULATIONAHA.120.051927. Epub 2020 Nov 13.
450 PMID: 33185476; PMCID: PMC7856243. (2020).
- 451 13. Rutten-Jacobs, L. C. A. *et al.* Genetic risk, incident stroke, and the benefits of adhering
452 to a healthy lifestyle: Cohort study of 306 473 UK Biobank participants. *BMJ (Online)*
453 **363**, 1–8 (2018).
- 454 14. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to
455 clinical risk factors for ischaemic stroke. *Nat Commun* **10**, 1–10 (2019).
- 456 15. Thomas, E. A. *et al.* Polygenic Risk, Midlife Life's Simple 7, and Lifetime Risk of
457 Stroke. *J Am Heart Assoc* **11**, (2022).
- 458 16. Hachiya, T. *et al.* Genetic Predisposition to Ischemic Stroke: A Polygenic Risk Score.
459 *Stroke* **48**, 253–258 (2017).
- 460 17. Hachiya, T. *et al.* Genome-wide polygenic score and the risk of ischemic stroke in a
461 prospective cohort: The Hisayama study. *Stroke* 759–765 (2020)

- 462 doi:10.1161/STROKEAHA.119.027520.
- 463 18. Mishra, A. *et al.* Stroke genetics informs drug discovery and risk prediction across
464 ancestries. *Nature* (2022) doi:10.1038/s41586-022-05165-3.
- 465 19. Howe, L. J. *et al.* Polygenic risk scores for coronary artery disease and subsequent
466 event risk amongst established cases. *Hum Mol Genet* **29**, 1388–1395 (2020).
- 467 20. Dahabreh, I. J. & Kent, D. M. *Index Event Bias as an Explanation for the Paradoxes*
468 *of Recurrence Risk Research*. <https://jamanetwork.com/>.
- 469 21. Ni, G. *et al.* A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders
470 Applied Across Multiple Cohorts. *Biol Psychiatry* **90**, 611–620 (2021).
- 471 22. Hahn, G. *et al.* A Smoothed Version of the Lasso Penalty for Fitting Integrated
472 Risk Models Using Summary Statistics or Individual-Level Data. *Genes (Basel)* **13**,
473 (2022).
- 474 23. Osterman, M. D., Kinzy, T. G. & Bailey, J. N. C. Polygenic Risk Scores. *Curr Protoc*
475 **1**, (2021).
- 476 24. Zhao, Z., Song, J., Wang, T. & Lu, Q. Polygenic risk scores: effect estimation and
477 model optimization. *Quantitative Biology* **0**, 0 (2021).
- 478 25. Ma, Y. & Zhou, X. Genetic prediction of complex traits with polygenic scores: a
479 statistical review. *Trends in Genetics* **37**, 995–1011 (2021).
- 480 26. Pain, O. *et al.* Evaluation of polygenic prediction methodology within a reference-
481 standardized framework. *PLoS Genet* **17**, 1–22 (2021).
- 482 27. Zhang, Q., Privé, F., Vilhjálmsón, B. & Speed, D. Improved genetic prediction of
483 complex traits from individual-level data or summary statistics. *Nat Commun* **12**, 1–9
484 (2021).
- 485 28. Yang, S. & Zhou, X. PGS-server: accuracy, robustness and transferability of polygenic
486 score methods for biobank scale studies. *Brief Bioinform* **23**, 1–19 (2022).
- 487 29. Wang, Y., Tsuo, K., Kanai, M., Neale, B. M. & Martin, A. R. Challenges and
488 opportunities for developing more generalizable polygenic risk scores. 293–320
489 (2022).
- 490 30. Kulm, S., Marderstein, A., Mezey, J. & Elemento, O. A systematic framework for
491 assessing the clinical impact of polygenic risk scores. *medRxiv preprint* doi:
492 <https://doi.org/10.1101/2020.04.06.20055574> (2021)
493 doi:10.1101/2020.04.06.20055574.
- 494 31. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J*
495 *Epidemiol* **27**, S2–S8 (2017).
- 496 32. Hirata, M. *et al.* Cross-sectional analysis of BioBank Japan clinical data: A large
497 cohort of 200,000 patients with 47 common diseases. *J Epidemiol* **27**, S9–S21 (2017).
- 498 33. Hirata, M. *et al.* Overview of BioBank Japan follow-up data in 32 diseases. *J*
499 *Epidemiol* **27**, S22–S28 (2017).
- 500 34. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and
501 richer datasets. *Gigascience* **4**, 1–16 (2015).
- 502 35. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-
503 based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
- 504 36. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality
505 control and statistical analysis. *Int J Methods Psychiatr Res* **27**, 1–10 (2018).
- 506 37. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic
507 risk score analyses. *Nat Protoc* **15**, 2759–2772 (2020).
- 508 38. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses
509 identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* **52**,
510 1169–1177 (2020).
- 511 39. Privé, F., Arbel, J. & Vilhjálmsón, B. J. LDpred2: Better, faster, stronger.

- 512 *Bioinformatics* **36**, 5424–5431 (2020).
- 513 40. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for
514 misspecifications in GWAS summary statistics and polygenic scores. *Human Genetics*
515 *and Genomics Advances* 100136 (2022) doi:10.1016/j.xhgg.2022.100136.
- 516 41. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via
517 Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1–10 (2019).
- 518 42. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat*
519 *Genet* (2022) doi:10.1038/s41588-022-01054-7.
- 520 43. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects
521 identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524–537
522 (2018).
- 523 44. Nagelkerke N. J. D. A Note on a General Definition of the Coefficient of
524 Determination. *Biometrika* **78**, 691–692 (1991).
- 525 45. Wang, Y. *et al.* Global Biobank analyses provide lessons for developing polygenic risk
526 scores across diverse cohorts. *Cell Genomics* **3**, (2023).
- 527 46. Hindy, G. *et al.* Genome-wide polygenic score, clinical risk factors, and long-term
528 trajectories of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2738–2746
529 (2020) doi:10.1161/ATVBAHA.120.314856.
- 530 47. Mitchell, R. E. *et al.* Strategies to investigate and mitigate collider bias in genetic and
531 Mendelian randomisation studies of disease progression. *PLoS Genetics* vol. 19
532 Preprint at <https://doi.org/10.1371/journal.pgen.1010596> (2023).
- 533 48. Sep, S. J., Van Kuijk, S. M. & Smits, L. J. Index event bias: Problems with eliminating
534 the paradox. *Journal of Stroke and Cerebrovascular Diseases* vol. 23 2464 Preprint at
535 <https://doi.org/10.1016/j.jstrokecerebrovasdis.2014.06.025> (2014).
- 536 49. Levine, D. A. *et al.* Smoking and mortality in stroke survivors: Can we eliminate the
537 paradox? *Journal of Stroke and Cerebrovascular Diseases* **23**, 1282–1290 (2014).
- 538 50. Altman, D. G. & Bland, J. M. Interaction revisited: The difference between two
539 estimates. *BMJ* **326**, 219 (2003).
- 540 51. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population
541 identifies novel susceptibility loci across different diseases. *Nat Genet* **52**, 669–679
542 (2020).
- 543 52. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from
544 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
- 545 53. Cho, S. M. J. *et al.* Genetic, sociodemographic, lifestyle, and clinical risk factors of
546 recurrent coronary artery disease events: a population-based cohort study. *Eur Heart J*
547 (2023) doi:10.1093/eurheartj/ehad380.
- 548 54. Weale, E. M. Quality control for genome-wide association studies. *Methods Mol Biol*
549 (2010) doi:10.1007/978-1-60327-367-1_19.
- 550 55. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants
551 in the Japanese population. *Nat Commun* **10**, (2019).
- 552 56. Privé, F., Vilhjálmsson, B. J., Aschard, H. & Blum, M. G. B. Making the Most of
553 Clumping and Thresholding for Polygenic Scores. *Am J Hum Genet* **105**, 1213–1221
554 (2019).
- 555 57. Chagnon, M., O’Loughlin, J., Engert, J. C., Karp, I. & Sylvestre, M. P. Missing single
556 nucleotide polymorphisms in Genetic Risk Scores: A simulation study. *PLoS One* **13**,
557 1–14 (2018).
- 558 58. Goldstein, B. A., Yang, L., Salfati, E. & Assimes, T. L. Contemporary Considerations
559 for Constructing a Genetic Risk Score: An Empirical Approach. *Genet Epidemiol* **39**,
560 439–445 (2015).
- 561 59. Wang, Y. *et al.* Global Biobank analyses provide lessons for developing polygenic risk

- 562 scores across diverse cohorts. *Cell Genomics* **3**, (2023).
- 563 60. Tadaka, S. *et al.* jMorp updates in 2020: Large enhancement of multi-omics data
564 resources on the general Japanese population. *Nucleic Acids Res* **49**, D536–D544
565 (2021).
- 566 61. Restuadi, R. *et al.* Polygenic risk score analysis for amyotrophic lateral sclerosis
567 leveraging cognitive performance, educational attainment and schizophrenia.
568 *European Journal of Human Genetics* (2021) doi:10.1038/s41431-021-00885-y.
- 569 62. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics
570 using MTAG. *Nat Genet* **50**, 229–237 (2018).
- 571 63. Maier, R. M. *et al.* Improving genetic prediction by leveraging genetic correlations
572 among human diseases and traits. *Nat Commun* **9**, 1–17 (2018).
- 573 64. Ho, W. K. *et al.* European polygenic risk score for prediction of breast cancer shows
574 similar performance in Asian women. *Nat Commun* **11**, 1–11 (2020).
- 575 65. Amariuta, T. *et al.* Improving the trans-ancestry portability of polygenic risk scores by
576 prioritizing variants in predicted cell-type-specific regulatory elements. *Nat Genet* **52**,
577 1346–1354 (2020).
- 578
- 579

580 **Tables**

581

582 **Table 1. Characteristic of the testing sample (BBJ2)**

583

	Any-AIS versus AIS-free controls				Recurrent AIS versus non-recurrent AIS			
	N sample (%, SD)	OR	95% CI	p- value	N sample (%, SD)	OR	95% CI	p- value
Total participants	41,929 (100%)	-	-	-	1,327 (100%)	-	-	-
AIS / AIS recurrent	1,470 (3.5%)	-	-	-	174 (13.1%)	-	-	-
Age (SD)*	70.0 (SD=12.6)	1.18	[1.06– 1.31]	0.002	64.1 (SD=11.5)	0.83	[0.59– 1.16]	0.256
Female sex	19,407 (46.3%)	0.48	[0.43– 0.54]	<0.001	383 (28.9%)	0.63	[0.42– 0.93]	0.019
Hypertension	17,710 (42.2%)	2.27	[2.04– 2.53]	<0.001	838 (63.1%)	0.92	[0.66– 1.3]	0.673
Hyperlipidemia	11,604 (27.7%)	2.17	[1.95– 2.41]	<0.001	603 (45.4%)	0.92	[0.66– 1.29]	0.625
Diabetes	3,622 (8.6%)	1.16	[0.97– 1.39]	0.089	125 (9.4%)	0.97	[0.52– 1.7]	1.000
Smoking**	21,570 (51.9%)	1.43	[1.28– 1.59]	<0.001	795 (60.5%)	1.28	[0.90– 1.82]	0.179
Vascular disease	3,136 (7.5%)	1.31	[1.08– 1.56]	0.005	122 (9.2%)	1.08	[0.59– 1.87]	0.778
Heart failure	1,329 (3.2%)	1.27	[0.95– 1.66]	0.095	48 (3.6%)	1.79	[0.78– 3.75]	0.124
Atrial fibrillation	1,004 (2.4%)	2.22	[1.71– 2.84]	<0.001	65 (4.9%)	1.07	[0.46– 2.23]	0.850

584

585 All risk factor characteristics were derived from history and not at the time of registration.

586 Odds ratio (OR), 95% CI (95% confidence intervals, and p-values were calculated using

587 logistic regression for onset and AIS recurrence (unadjusted for other factors). *Age at AIS

588 case-control was that at recruitment, while age at recurrent AIS case-control was that at first
589 incidence. The numbers indicate the median threshold age. ** The total number of missing
590 values of smoking was 389 for all case-control and 12 for recurrent AIS case-control
591 samples. Abbreviations: BBJ2, BioBank 2nd cohort; AIS, all ischemic stroke; SD = standard
592 deviation
593

594 **Table 2. Polygenic risk score performance at validation**

595

Method	Best parameters	Mean number of variants	Best mean incremental Nagelkerke R ²	Standard deviation	95% confidence interval
P+T	Clumping R ² =0.95, Clumping kb=526, Imputation R ² =0.8, p-value threshold=1	3,144,737	0.0038	0.0012	0.0030– 0.0046
LDpred2	ρ value=0.0056, heritability value x 1.0, no sparse	898,456	0.00443	0.0013	0.0035– 0.0054
Lassosum2	S=0.9, lambda=0.00388	282,520	0.0039	0.0013	0.0030– 0.0048
PRS-CS	Phi=1.00E-04	985,439	0.00441	0.0012	0.0036– 0.0053
PRS-CSx	Phi=1.00E-05	1,016,745	0.0037	0.0008	0.0031– 0.0042

596

597

598 **Table 3. Polygenic risk score performance at testing**

599

PRS Method	Association tests	OR per SD [95% CI]	p-value	Incremental AUC	Incremental Nagelkerke R ²
MetaGRS	AIS	1.21[1.15–1.27]	2.89E-12	0.0087	0.0044
	Recurrence	1.18[1.00–1.39]	0.044	0.0123	0.0057
MEGASTROKE 27 SNVs	AIS	1.11[1.06–1.17]	4.23E-05	0.0033	0.0015
	Recurrence	1.07[0.91–1.25]	0.41	0.0032	0.0009
GIGASTROKE 84 SNVs	AIS	1.08[1.03–1.14]	2.96E-03	0.0022	0.0008
	Recurrence	1.17[1.00–1.38]	0.054	0.0125	0.0052
GIGASTROKE iPGS	AIS	1.26[1.20–1.33]	1.24E-17	0.0130	0.0066
	Recurrence	1.08[0.91–1.27]	0.37	0.0044	0.0011

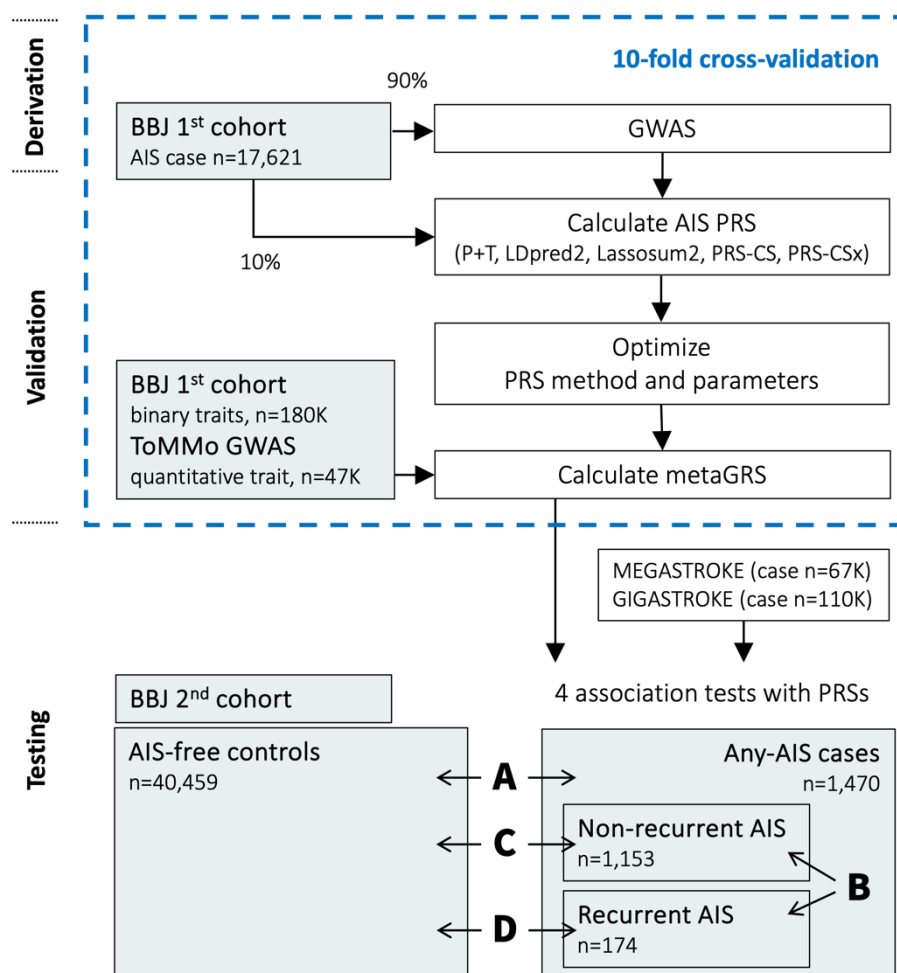
600

601 Polygenic risk score performance was evaluated using an independent testing set for AIS and
602 recurrent AIS. We showed two main association tests; AIS (any-AIS cases vs. AIS-free
603 controls) and AIS recurrence (recurrent AIS vs. non-recurrent AIS). Incremental AUC and R²
604 are the differences in the values when fitting with/without PRS, along with age, sex, the first
605 10 PCs, and seven risk factors. Abbreviations: AIS, all ischemic strokes; OR, odds ratio;
606 AUC, area under the curve; PC, principal components; PRS, polygenic risk score; SD =
607 standard deviation

608

609 **Figure Legends**

610 **Figure 1. Workflow**



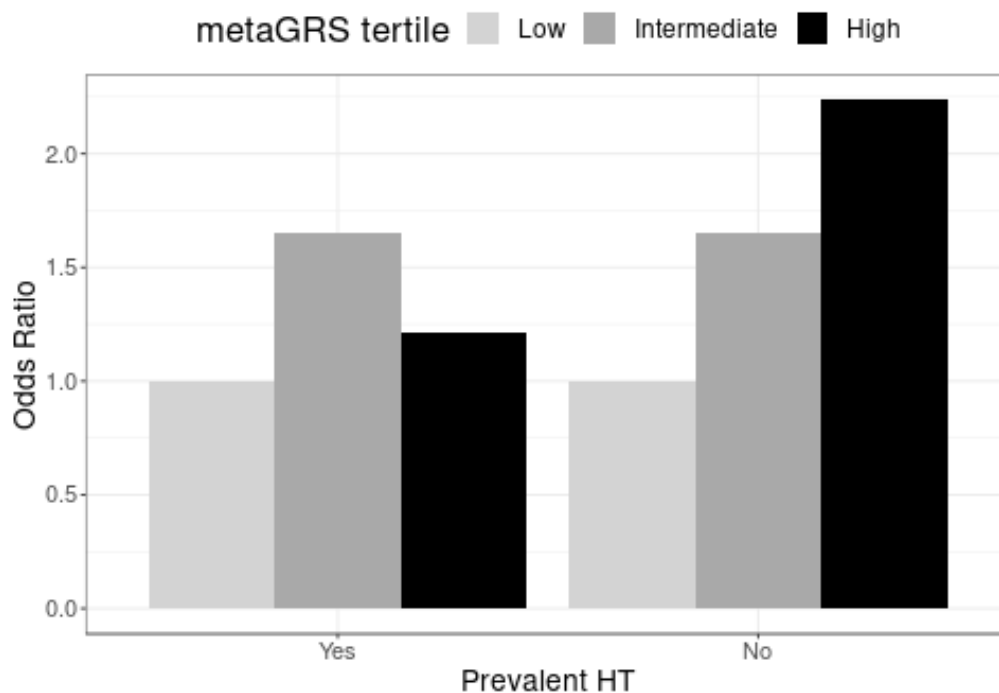
611

612 MEGASTROKE AIS summary statistics of European (EUR) studies were only used for
 613 PRS-CSx. 1000 Genomes Project super population samples (EAS or EUR) were used for the
 614 LD reference panel. Abbreviations: GWAS = genome-wide association study, PRS =
 615 polygenic risk score, P+T = pruning and thresholding, OR = odds ratio, AIS = all ischemic
 616 stroke, ToMMo = Tohoku Medical Megabank; LD, linkage disequilibrium.

617 We used a logistic regression model to assess the association of the PRS using the two case-
 618 control settings for AIS (**A**: any-AIS vs. AIS-free controls) and AIS recurrence (**B**: recurrent
 619 AIS vs. non-recurrent AIS). We also applied two other combinations of case-controls (**C**:
 620 recurrent versus AIS-free controls and **D**: non-recurrent versus AIS-free controls).

621

622 **Figure 2. Odds ratio of metaGRS tertiles with/without a history of hypertension**



623

624

625 Association of AIS recurrence and meta-GRS tertiles with or without a history of
626 hypertension (HT) in the testing sample, with reference to the low metaGRS tertiles.

627

628

629

630

631

632

633

634

635

SUPPLEMENTARY MATERIALS

636

637

638

Recurrent stroke prediction by applying a stroke polygenic risk score

639

in the Japanese population

640

641

642

643 **Supplementary Methods**

644

645 **Samples**

646 BioBank Japan (BBJ) collected DNA, serum, and medical records (clinical information) with
647 consent from patients. The BBJ 1st cohort (BBJ1) dataset contained all ischemic stroke (AIS,
648 n=17,621) cases, including large artery stroke (LAS, n=981), small vessel stroke (SVS,
649 n=3,108), and cardioembolic stroke (CES, n=608) cases. The patients without AIS were
650 included as controls (n=162,317).

651 Testing involved the use of data from part of the BBJ 2nd cohort (BBJ2) dataset, which
652 contains information about AIS cases (n=1,470), including LAS (n=268), SVS (n=508), CES
653 (n=122), and transient ischemic attack (TIA, n=105). All patients without AIS were included
654 as controls (n=40,459). Among these cases, recurrent ischemic stroke (n=187) was used as a
655 case of recurrent AIS, which included LAS (n=40), SVS (n=57), CES (n=11), and transient
656 ischemic attack (TIA, n=20). Samples with information on the first onset date and follow-up
657 of under 30 days were excluded; AIS cases with recurrence (n=174) were set as the case
658 group, and AIS cases without recurrence (n=1,153) remained in the control group in the
659 testing sample.

660

661 **Quality control and imputation process of BBJ1 data**

662 We removed variants with call rates <0.99, samples with call rates <0.98, non-East Asian
663 samples, and sex-discordant samples. We used 939 samples whose genotypes were analyzed
664 using whole-genome sequencing (WGS); we added an additional quality control based on the
665 concordance rate between the genotyping array and WGS. We excluded variants with
666 concordance rates <99.5% or non-reference discordance rates $\geq 0.5\%$ and Hardy-Weinberg
667 equilibrium (HWE) ($p < 1e-6$). The 10 principal components (PCs) were calculated by

668 applying additional quality control (QC) to the BBJ1 genotyped data. We removed samples
669 from close relatives (King's cutoff >0.0884), and 24 long LD regions,⁵⁴ including the MHC
670 region (chr6, position 25,000,000-35,000,000), and $MAF < 0.01$. We pruned (PLINK2
671 parameters: `--indep-pairwise 200 50 0.20`) and finally used 92,231–92,303 variants to
672 conduct projection and calculate the first 10 PCs depending on a 10-fold sample set of the
673 90% BBJ1 dataset.

674 Subsequently, the datasets were phased (Eagle v2.4.1) and imputed (Minimac4 v1.0.2) using
675 the developed panel.⁵⁵ We further conducted quality control to remove variants with minor
676 allele counts < 10 , close relatives (King cutoff >0.0884), and imputation r-square < 0.8 .

677

678 **Quality control and imputation process for BBJ2 data**

679 We removed samples with no age/sex information, sex discrepancy, call rates < 0.98 ,
680 heterozygosity rates with $SD > 4$ or < -4 , from duplicate or twins ($\pi\text{-hat} \geq 0.75$), and from
681 non-East Asian subjects. We then removed variants with a call rate < 0.99 , duplicate SNPs,
682 heterozygous count < 5 , HWE ($p < 1e-6$), and allele frequency discrepancies (gap from 1000
683 genomes EAS > 0.16). A total of 41,929 samples and 525,239 variants were analyzed.

684 We applied additional quality control to the BBJ2 genotyped data to calculate the 10 PCs. We
685 removed variants in 24 long LD regions,⁵⁴ pruned them (PLINK2 parameters: `--indep-`
686 `pairwise 200 50 0.05`), extracted close relatives (King cut-off >0.0884), and used 69,068
687 variants to calculate the first 10 PCs.

688 We removed variants of the imputation r-squared < 0.3 after phasing (Eagle v2.4.1) and
689 imputing (Minimac4 v1.0.2) the developed panel.⁵⁵ We used a lower r-square threshold for
690 BBJ2 to reduce the number of variants unmatched with BBJ1.

691

692 **Polygenic risk score parameters at the derivation**

693 Pruning and threshold used a total of 1,224 parameter combinations of three stricter
694 imputation r-squared score thresholds {0.8, 0.9, and 0.95}, four base sizes of the clumping
695 window {50, 100, 200, and 500}, six squared correlations of clumping {0.01, 0.05, 0.2, 0.5,
696 0.8, and 0.95}, and 17 p-value threshold {1e-8, 3e-8, 1e-7, 3e-7, 1e-6, 3e-6, 1e-5, 3e-5, 1e-4,
697 3e-4, 1e-3, 3e-3, 0.01, 0.03, 0.1, 0.3, and 1}. We used clumping windows and divided the
698 base size by the squared correlation of clumping.⁵⁶⁻⁵⁹ The following default parameters were
699 used for LDpred2: three heritability {0.7, 1, and 1.4} $\times h^2_{LDSC}$, where h^2_{LDSC} is the heritability
700 estimate from the constrained LD score regression,⁵² 21 ρ estimates {equally spaced on a log
701 scale between 1e-5 and 1}, and sparse or not. The following default parameters were used for
702 Lassosum2: 10 values of s {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0} and 20 values of
703 λ {equally spaced on a log scale between 0.1 and 0.001}. Regarding PRS-CS and PRS-CSx,
704 we used slightly more extended parameters: phi {1e-7, 1e-6, 1e-5, 1e-4, 1e-3, 0.01, 0.1, 1,
705 and auto}, than the default parameters {1e-4, 1e-3, 0.01, 0.1, 1, and auto}, because the
706 optimized parameter in the initial trials using our dataset was the smallest (phi=1e-4) in the
707 default range. Other parameters were set to default values (a=1, b=0.5).^{41,42}

708

709 **Linkage disequilibrium (LD) reference**

710 We used the EAS superpopulation of the 1000 Genomes panel (n=504) as a LD reference for
711 P+T clumping. For Ldpred2, Lassosum2, PRS-CS, and PRS-CSx models, we restricted our
712 use of external LD reference panels to the HapMap 3 variants, which were also constructed
713 from 1000 Genomes EAS. HapMap 3 variant restriction resulted in 898,456, 898,456,
714 985,440, and 1,076,835 variants of Ldpred2, Lassosum2, PRS-CS, and PRS-CSx,
715 respectively. To use PRS-CSx for the EUR, we used the EUR superpopulation of the 1000
716 Genomes panel and its HapMap3 variants (1,016,745 variants).

717

718 **Genome wide association study summary statistics for metaGRS construction**

719 MetaGRS was developed following a previous study¹⁴ and included nine binary and eight
720 quantitative traits. The nine binary traits were SVS, LAS, CES, myocardial infarction (MI),
721 stable angina pectoris (SAP), unstable angina pectoris (AP), atrial fibrillation (AF), diabetes
722 (DM), and ever smoking (SM) from the BBJ1 dataset. Subsequently, we conducted GWAS.
723 The number of cases and controls is listed in Table S15. Summary statistics were obtained
724 from the jMorp website (<https://jmorp.megabank.tohoku.ac.jp>) and were used for eight
725 quantitative traits—body mass index (BMI), height (HE), systolic blood pressure (SBP),
726 diastolic blood pressure (DBP), total cholesterol (TC), triglyceride (TG), high-density
727 lipoprotein (HDL), and low-density lipoprotein (LDL) from the Tohoku Medical Megabank
728 Project (ToMMo; 22,033 to 47,056 samples).⁶⁰

729

730 **Polygenic risk score calculation from other milestone studies**

731 We obtained effect weights from the PGS catalog (PGS000665 and PGS002725 of
732 <https://www.pgscatalog.org> for PRS₃₂ and iPGS_{EAS}, respectively) and for PRS₈₉ from the
733 supplementary tables.¹⁸ We used variants that matched with the BBJ2 dataset (imputation r-
734 squared > 0.3). The unmatched variants in PRS₃₂ and PRS₈₉ included proxy variants that
735 showed the highest r-squared values with the index variants only from an r-squared value
736 greater than 0.3. The r-squared values were calculated using the plink --r2 command³⁵ with
737 the 1000 Genome EAS as a reference.

738

739 **Supplementary Notes**

740

741 **Characteristics of the PRS methods during validation**

742 The following characteristics of each method were observed while testing several
743 parameters: the best mean incremental R^2 of 0.0038 was observed for the P + T method when
744 we liberalized the p-value threshold to 1 and clumped $r^2 > 0.8$, whereas it was below 0.001
745 when we used a low p-value threshold ($< 10^{-8}$, Table S5). For LDpred2, ρ values higher than
746 0.001 showed good performance (mean incremental R^2 was 0.0038 for ρ values ≥ 0.001
747 compared to 0.0016 for ρ values < 0.001), while heritability estimates and sparse parameters
748 made less difference (Table S6). For Lassosum2, the closer the value of parameter “s” is to 1,
749 the higher the prediction accuracy, but it wasn’t the case at exactly 1. Larger lambda
750 parameters correlate with greater accuracy (mean incremental R^2 was between 0.0025 to
751 0.0039). However, the accuracy sharply decreases (mean incremental $R^2 < 0.001$) if the value
752 is too small (< 0.01). The prediction performance was maximized when the parameters were
753 $s=0.9$ and $\lambda=0.0038$ (Table S7). High and low phi values resulted in low performance
754 for PRS-CS (mean incremental R^2 was 0.0041 for $\phi=10^{-3}$, 10^{-4} , and 10^{-5} compared to
755 0.0029 for other phi parameters; Table S8). Meanwhile, PRS-CSx was relatively consistent
756 regardless of the phi values (mean incremental R^2 was 0.0032 for $\phi=10^{-3}$, 10^{-4} , and 10^{-5}
757 compared to 0.0028 for other phi parameters; Table S9).

758 The performance of the PRS-CS ($R^2=0.00441$) was comparable to that of LDpred2
759 ($R^2=0.00443$) in the validation analysis. Low validation predictability (incremental $R^2 <$
760 0.001) was observed when we restricted the p-value threshold in the P+T method to genome-
761 wide significance ($p < 5 \times 10^{-8}$). PRS-CSx improves cross-population polygenic prediction by
762 integrating GWAS summary statistics from other populations;^{61–65} however, we could not
763 reproduce this result in our current study using PRS-CSx. Our results demonstrate the

764 importance of using Bayesian methods of high-dimensional techniques in variable selection
765 and shrinkage estimation considering LD (such as LDpred2 and PRS-CS) to predict AIS and
766 recurrent AIS.

767

768

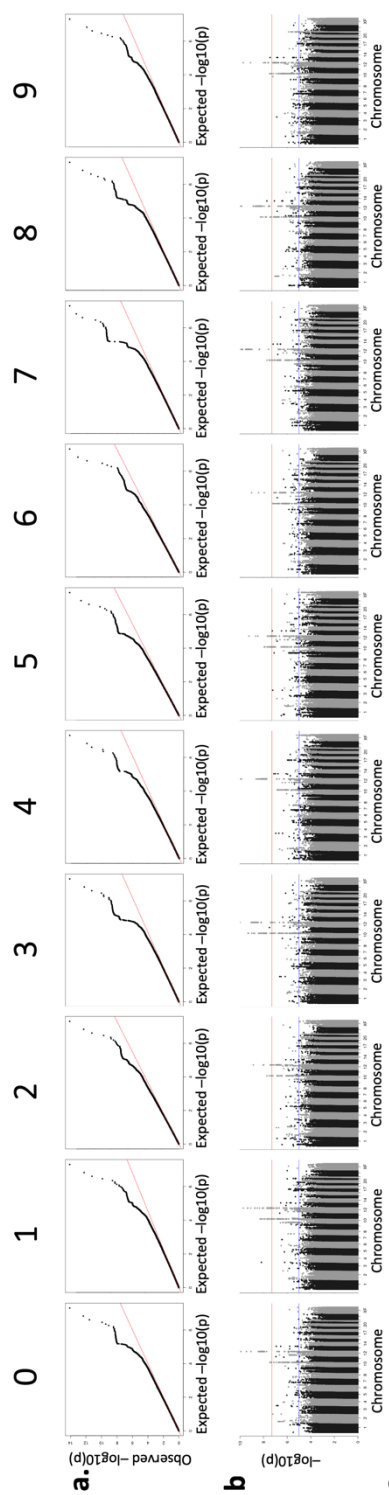
769 **Supplementary Figures**

770

771 **Figure S1. Genome-wide association study (GWAS) in derivation samples**

We conducted stroke GWAS using 90% of BBJ 1st cohort data as the derivation sample, using Firth logistic regression and PLINK (v.2.0) for each 10-fold cross-validation identification. a. Quantile-Quantile plot, b. Manhattan plot, and c. genome-wide significant ($p < 5e-8$) loci of each fold. Abbreviations: Chr = chromosome.

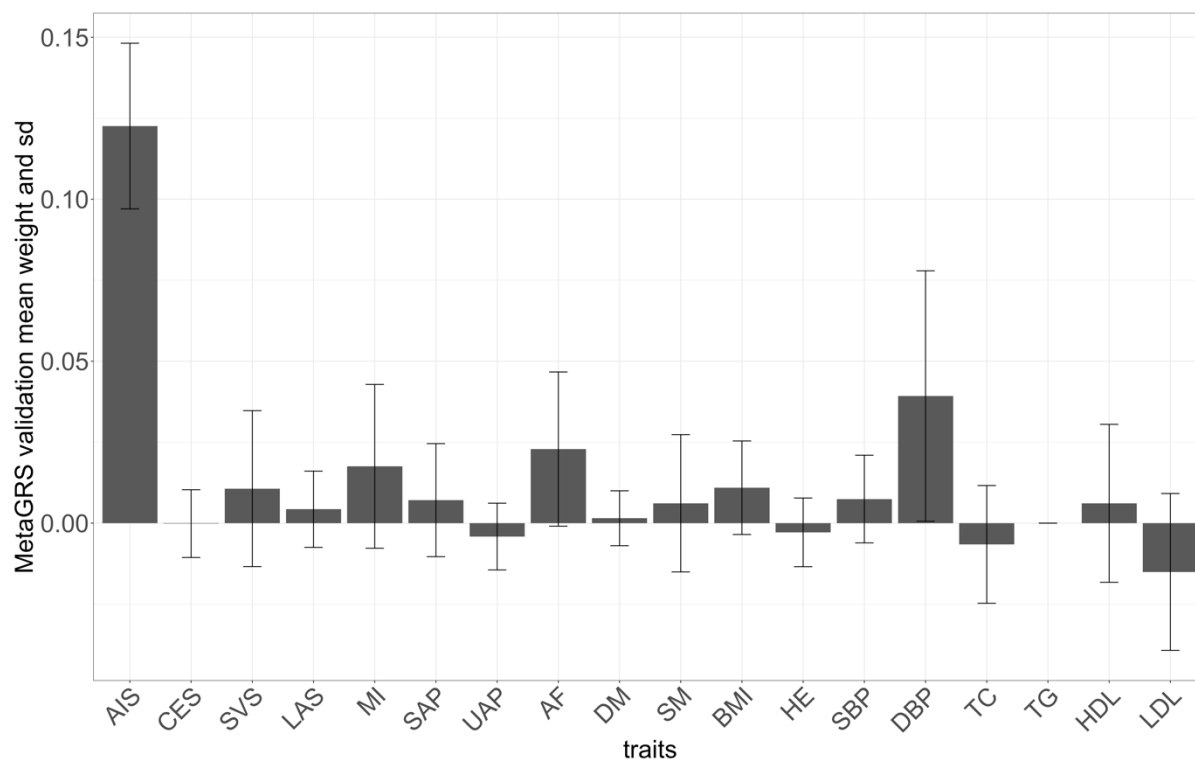
10-fold cross-validation identification



Chr.3																			
Chr.10	SH3PYXD2A																		
Chr.11																			
Chr.12	CGDC63 MYL2 CUX2 LINC02356																		
Chr.13																			

772

773



774

775

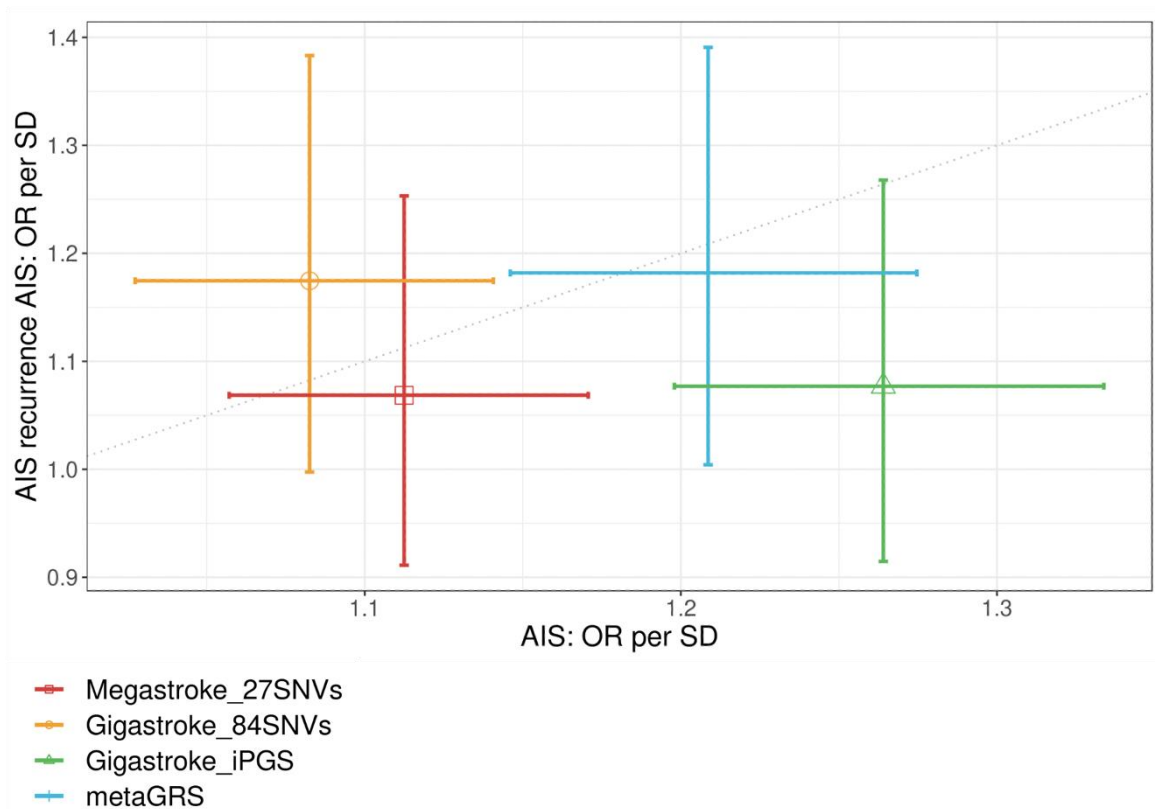
776 **Figure S2. Elastic net weight**

777 The mean weight of 10-fold elastic net regression as determined by “glmnet” in the
778 validation sample. The X-axis shows the AS and the 17 binary and quantitative traits. Error
779 bars represent standard deviations.

780 Abbreviations: AIS, all ischemic stroke; SVS, small vessel stroke; LAS, large artery stroke;
781 CES, cardioembolic stroke; MI, myocardial infarction; SAP, stable angina pectoris; UAP,
782 unstable angina pectoris; AF, atrial fibrillation; DM, diabetes; SM, smoking; BMI, body mass
783 index; HE, height; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total
784 cholesterol; TG, triglyceride; HCL, high-density lipoprotein; LDL, low-density lipoprotein;
785 SD, standard deviation.

786

787



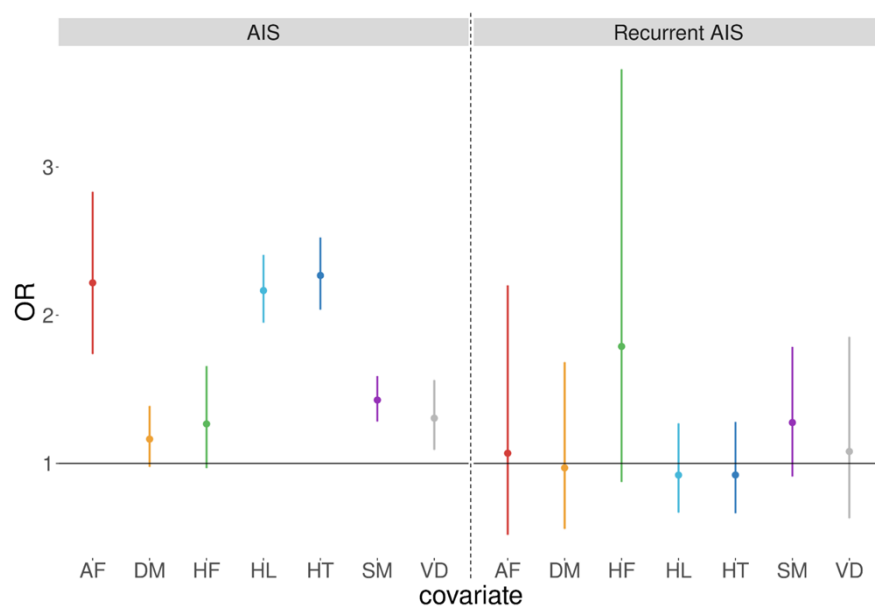
788

789

790 **Figure S3. Odds ratio per SD to predict AIS and AIS recurrence**

791 Odds ratio per standard deviation by metaGRS and three publicly available PRS in the
792 independent test set of AIS (case=1, 470, control=40,459) and AIS recurrence case-control
793 set (case=174, control=1,153). Age, sex, the first 10 PCs, and seven risk factors were
794 adjusted. Error bars represent 95% confidence intervals. The dotted line represents the point
795 at which the ORs per SD of AIS and AIS recurrence were equal. Abbreviations: OR, odds
796 ratio; AIS, all ischemic stroke; PRS, polygenic risk score; PC, principal component; SD,
797 standard deviation.

798



799

800

801 **Figure S4. Association between seven comorbidities and AIS/AIS recurrence**

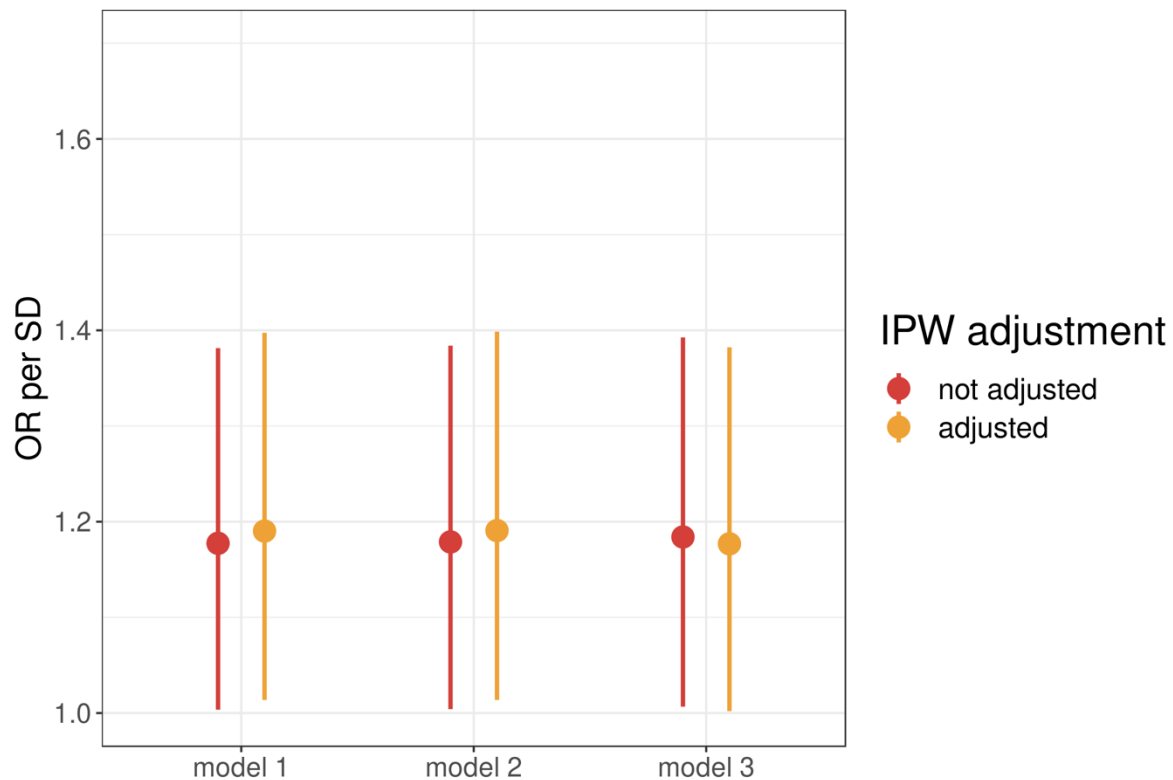
802 The Y-axis shows the odds ratio per standard deviation (SD) and 95% confidence interval
803 (CI) of the seven risk factors (six clinical comorbidities and smoking) in the BBJ 2nd cohort.

804 In our dataset, we used HT defined as SBP>140 mmHg, DBP>90 mmHg, or hypertension
805 history, DM inclusive of type 1 diabetes and other diabetes, SM as a current smoker at the
806 time of registration, VD representing myocardial infarction, arteriosclerosis obliterans, stable
807 angina pectoris, unstable angina pectoris, and AF inclusive of atrial flutter. We used the status
808 at the time of registration and history of these comorbidities.

809 Abbreviations: HT, hypertension; HL, hyperlipidemia; DM, diabetes; SM, smoking; VD,
810 vascular disease; HF, congestive heart failure; AF, atrial fibrillation.

811

812



813

814 **Figure S5. Association between metaGRS and AIS recurrence in patients with or**
815 **without IPW adjustment**

816 The y-axis shows the odds ratio per SD and the 95% confidence interval for predicting
817 recurrent AIS. Model 1: age and sex; model 2: age, sex, and seven risk factors; model 3: age,
818 sex, the first 10 principal components (PCs), and seven risk factors. The color represents
819 whether the inverse probability weight (IPW) is adjusted. We applied logistic regression
820 using variables as covariates when the IPW was not adjusted. We replaced 14 missing data
821 on the smoking status to mean values.

822

823 **Supplementary Tables**

824 In separate spreadsheets S1-15

825

826

827

828 **Consortium authors**

829

830 **BioBank Japan**

831 Koichi Matsuda^{1,2}, Yuji Yamanashi³, Yoichi Furukawa⁴, Takayuki Morisaki⁵, Yoshinori
832 Murakami⁶, Yoichiro Kamatani^{2,7}, Kaori Muto⁸, Akiko Nagai⁸, Wataru Obara⁹, Ken Yamaji¹⁰,
833 Kazuhisa Takahashi¹¹, Satoshi Asai^{12,13}, Yasuo Takahashi¹³, Takao Suzuki¹⁴, Nobuaki
834 Sinozaki¹⁴, Hiroki Yamaguchi¹⁵, Shiro Minami¹⁶, Shigeo Murayama¹⁷, Kozo Yoshimori¹⁸,
835 Satoshi Nagayama¹⁹, Daisuke Obata²⁰, Masahiko Higashiyama²¹, Akihide Masumoto²²,
836 Yukihiro Koretsune²³

837

838 ¹Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science,
839 The University of Tokyo, Tokyo, Japan.

840 ²Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The
841 University of Tokyo, Tokyo, Japan.

842 ³Division of Genetics, The Institute of Medical Science, The University of Tokyo, Tokyo,
843 Japan.

844 ⁴Division of Clinical Genome Research, Institute of Medical Science, The University of
845 Tokyo, Tokyo, Japan.

846 ⁵Division of Molecular Pathology IMSUT Hospital, Department of Internal Medicine Project
847 Division of Genomic Medicine and Disease Prevention The Institute of Medical Science The
848 University of Tokyo, Tokyo, Japan.

849 ⁶Department of Cancer Biology, Institute of Medical Science, The University of Tokyo,
850 Tokyo, Japan.

851 ⁷Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The
852 University of Tokyo, Tokyo, Japan.

853 ⁸Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo,
854 Japan.

855 ⁹Department of Urology, Iwate Medical University, Iwate, Japan.

856 ¹⁰Department of Internal Medicine and Rheumatology, Juntendo University Graduate School
857 of Medicine, Tokyo, Japan.

858 ¹¹Department of Respiratory Medicine, Juntendo University Graduate School of Medicine,
859 Tokyo, Japan.

860 ¹²Division of Pharmacology, Department of Biomedical Science, Nihon University School of
861 Medicine, Tokyo, Japan.

862 ¹³Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center,
863 Nihon University. School of Medicine, Tokyo, Japan.

864 ¹⁴Tokushukai Group, Tokyo, Japan.

865 ¹⁵Hiroki Yamaguchi. Department of Hematology, Nippon Medical School, Tokyo, Japan.

866 ¹⁶Department of Bioregulation, Nippon Medical School, Kawasaki, Japan.

867 ¹⁷Shigeo Murayama. Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology,
868 Tokyo, Japan.

869 ¹⁸Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan.

870 ¹⁹The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo,
871 Japan.

872 ²⁰Center for Clinical Research and Advanced Medicine, Shiga University of Medical
873 Science, Shiga, Japan.

874 ²¹Masahiko Higashiyama. Department of General Thoracic Surgery, Osaka International
875 Cancer Institute, Osaka, Japan.

876 ²²Iizuka Hospital, Fukuoka, Japan.

877 ²³National Hospital Organization Osaka National Hospital, Osaka, Japan.

878

879

880

881 **MEGASTROKE CONSORTIUM**

882

883 Rainer Malik¹, Ganesh Chauhan², Matthew Traylor³, Muralidharan
884 Sargurupremraj^{4,5}, Yukinori Okada^{6,7,8}, Aniket Mishra^{4,5}, Loes Rutten-Jacobs³,
885 Anne-Katrin Giese⁹, Sander W van der Laan¹⁰, Solveig Gretarsdottir¹¹, Christopher
886 D Anderson^{12,13,14}, Michael Chong¹⁵, Hieab HH Adams^{16,17}, Tetsuro Ago¹⁸, Peter
887 Almgren¹⁹, Philippe Amouyel^{20,21}, Hakan Ay^{13,22}, Traci M Bartz²³, Oscar R
888 Benavente²⁴, Steve Bevan²⁵, Giorgio B Boncoraglio²⁶, Robert D Brown, Jr.²⁷,
889 Adam S Butterworth^{28,29}, Caty Carrera^{30,31}, Cara L Carty^{32,33}, Daniel I
890 Chasman^{34,35}, Wei-Min Chen³⁶, John W Cole³⁷, Adolfo Correa³⁸, Ioana
891 Cotlarciuc³⁹, Carlos Cruchaga^{40,41}, John Danesh^{28,42,43,44}, Paul IW de Bakker^{45,46},
892 Anita L DeStefano^{47,48}, Marcel den Hoed⁴⁹, Qing Duan⁵⁰, Stefan T Engelter^{51,52},
893 Guido J Falcone^{53,54}, Rebecca F Gottesman⁵⁵, Raji P Grewal⁵⁶, Vilmundur
894 Gudnason^{57,58}, Stefan Gustafsson⁵⁹, Jeffrey Haessler⁶⁰, Tamara B Harris⁶¹, Ahamad
895 Hassan⁶², Aki S Havulinna^{63,64}, Susan R Heckbert⁶⁵, Elizabeth G Holliday^{66,67},
896 George Howard⁶⁸, Fang-Chi Hsu⁶⁹, Hyacinth I Hyacinth⁷⁰, M Arfan Ikram¹⁶, Erik
897 Ingelsson^{71,72}, Marguerite R Irvin⁷³, Xueqiu Jian⁷⁴, Jordi Jiménez-Conde⁷⁵, Julie A
898 Johnson^{76,77}, J Wouter Jukema⁷⁸, Masahiro Kanai^{6,7,79}, Keith L Keene^{80,81}, Brett M
899 Kissela⁸², Dawn O Kleindorfer⁸², Charles Kooperberg⁶⁰, Michiaki Kubo⁸³, Leslie A
900 Lange⁸⁴, Carl D Langefeld⁸⁵, Claudia Langenberg⁸⁶, Lenore J Launer⁸⁷, Jin-Moo
901 Lee⁸⁸, Robin Lemmens^{89,90}, Didier Leys⁹¹, Cathryn M Lewis^{92,93}, Wei-Yu Lin^{28,94},
902 Arne G Lindgren^{95,96}, Erik Lorentzen⁹⁷, Patrik K Magnusson⁹⁸, Jane Maguire⁹⁹,
903 Ani Manichaikul³⁶, Patrick F McArdle¹⁰⁰, James F Meschia¹⁰¹, Braxton D
904 Mitchell^{100,102}, Thomas H Mosley^{103,104}, Michael A Nalls^{105,106}, Toshiharu
905 Ninomiya¹⁰⁷, Martin J O'Donnell^{15,108}, Bruce M Psaty^{109,110,111,112}, Sara L Pulit^{45,113},
906 Kristiina Rannikmäe^{114,115}, Alexander P Reiner^{65,116}, Kathryn M Rexrode¹¹⁷,
907 Kenneth Rice¹¹⁸, Stephen S Rich³⁶, Paul M Ridker^{34,35}, Natalia S Rost^{9,13}, Peter M
908 Rothwell¹¹⁹, Jerome I Rotter^{120,121}, Tatjana Rundek¹²², Ralph L Sacco¹²², Saori
909 Sakaue^{7,123}, Michele M Sale¹²⁴, Veikko Salomaa⁶³, Bishwa R Sapkota¹²⁵, Reinhold
910 Schmidt¹²⁶, Carsten O Schmidt¹²⁷, Ulf Schminke¹²⁸, Pankaj Sharma³⁹, Agnieszka
911 Slowik¹²⁹, Cathie LM Sudlow^{114,115}, Christian Tanislav¹³⁰, Turgut Tatlisumak^{131,132},
912 Kent D Taylor^{120,121}, Vincent NS Thijs^{133,134}, Gudmar Thorleifsson¹¹, Unnur
913 Thorsteinsdottir¹¹, Steffen Tiedt¹, Stella Trompet¹³⁵, Christophe Tzourio^{5,136,137},
914 Cornelia M van Duijn^{138,139}, Matthew Walters¹⁴⁰, Nicholas J Wareham⁸⁶, Sylvia
915 Wassertheil-Smoller¹⁴¹, James G Wilson¹⁴², Kerri L Wiggins¹⁰⁹, Qiong Yang⁴⁷,
916 Salim Yusuf¹⁵, Najaf Amin¹⁶, Hugo S Aparicio^{48,185}, Donna K Arnett¹⁸⁶, John
917 Attia¹⁸⁷, Alexa S Beiser^{47,48}, Claudine Berr¹⁸⁸, Julie E Buring^{34,35}, Mariana
918 Bustamante¹⁸⁹, Valeria Caso¹⁹⁰, Yu-Ching Cheng¹⁹¹, Seung Hoan Choi^{48,192},
919 Ayesha Chowhan^{48,185}, Natalia Cullell³¹, Jean-François Dartigues^{193,194}, Hossein
920 Delavaran^{95,96}, Pilar Delgado¹⁹⁵, Marcus Dörr^{196,197}, Gunnar Engström¹⁹, Ian
921 Ford¹⁹⁸, Wander S Gurpreet¹⁹⁹, Anders Hamsten^{200,201}, Laura Heitsch²⁰², Atsushi
922 Hozawa²⁰³, Laura Ibanez²⁰⁴, Andreea Ilinca^{95,96}, Martin Ingelsson²⁰⁵, Motoki
923 Iwasaki²⁰⁶, Rebecca D Jackson²⁰⁷, Katarina Jood²⁰⁸, Pekka Jousilahti⁶³, Sara
924 Kaffashian^{4,5}, Lalit Kalra²⁰⁹, Masahiro Kamouchi²¹⁰, Takanari Kitazono²¹¹, Olafur
925 Kjartansson²¹², Manja Kloss²¹³, Peter J Koudstaal²¹⁴, Jerzy Krupinski²¹⁵, Daniel L
926 Labovitz²¹⁶, Cathy C Laurie¹¹⁸, Christopher R Levi²¹⁷, Linxin Li²¹⁸, Lars Lind²¹⁹,

927 Cecilia M Lindgren^{220,221}, Vasileios Lioutas^{48,222}, Yong Mei Liu²²³, Oscar L
928 Lopez²²⁴, Hirata Makoto²²⁵, Nicolas Martinez-Majander¹⁷², Koichi Matsuda²²⁵,
929 Naoko Minegishi²⁰³, Joan Montaner²²⁶, Andrew P Morris^{227,228}, Elena Muiño³¹,
930 Martina Müller-Nurasyid^{229,230,231}, Bo Norrving^{95,96}, Soichi Ogishima²⁰³, Eugenio
931 A Parati²³², Leema Reddy Peddareddygari⁵⁶, Nancy L Pedersen^{98,233}, Joanna
932 Pera¹²⁹, Markus Perola^{63,234}, Alessandro Pezzini²³⁵, Silvana Pileggi²³⁶, Raquel
933 Rabionet²³⁷, Iolanda Riba-Llena³⁰, Marta Ribasés²³⁸, Jose R Romero^{48,185}, Jaime
934 Roquer^{239,240}, Anthony G Rudd^{241,242}, Antti-Pekka Sarin^{243,244}, Ralhan Sarju¹⁹⁹,
935 Chloe Sarnowski^{47,48}, Makoto Sasaki²⁴⁵, Claudia L Satizabal^{48,185}, Mamoru
936 Satoh²⁴⁵, Naveed Sattar²⁴⁶, Norie Sawada²⁰⁶, Gerli Sibolt¹⁷², Ásgeir Sigurdsson²⁴⁷,
937 Albert Smith²⁴⁸, Kenji Sobue²⁴⁵, Carolina Soriano-Tárraga²⁴⁰, Tara Stanne²⁴⁹, O
938 Colin Stine²⁵⁰, David J Stott²⁵¹, Konstantin Strauch^{229,252}, Takako Takai²⁰³, Hideo
939 Tanaka^{253,254}, Kozo Tanno²⁴⁵, Alexander Teumer²⁵⁵, Liisa Tomppo¹⁷², Nuria P
940 Torres-Aguila³¹, Emmanuel Touze^{256,257}, Shoichiro Tsugane²⁰⁶, Andre G
941 Uitterlinden²⁵⁸, Einar M Valdimarsson²⁵⁹, Sven J van der Lee¹⁶, Henry Völzke²⁵⁵,
942 Kenji Wakai²⁵³, David Weir²⁶⁰, Stephen R Williams²⁶¹, Charles DA Wolfe^{241,242},
943 Quenna Wong¹¹⁸, Huichun Xu¹⁹¹, Taiki Yamaji²⁰⁶, Dharambir K Sanghera^{125,169,170},
944 Olle Melander¹⁹, Christina Jern¹⁷¹, Daniel Strbian^{172,173}, Israel Fernandez-
945 Cadenas^{31,30}, W T Longstreth, Jr^{65,174}, Arndt Rolfs¹⁷⁵, Jun Hata¹⁰⁷, Daniel Woo⁸²,
946 Jonathan Rosand^{12,13,14}, Guillaume Pare¹⁵, Jemma C Hopewell¹⁷⁶, Danish
947 Saleheen¹⁷⁷, Kari Stefansson^{11,178}, Bradford B Worrall¹⁷⁹, Steven J Kittner³⁷,
948 Sudha Seshadri^{48,180}, Myriam Fornage^{74,181}, Hugh S Markus³, Joanna MM
949 Howson²⁸, Yoichiro Kamatani^{6,182}, Stephanie Debette^{4,5}, Martin Dichgans^{1,183,184}

950
951 ¹Institute for Stroke and Dementia Research (ISD), University Hospital, LMU
952 Munich, Munich, Germany

953 ²Centre for Brain Research, Indian Institute of Science, Bangalore, India

954 ³Stroke Research Group, Division of Clinical Neurosciences, University of
955 Cambridge, UK

956 ⁴INSERM U1219 Bordeaux Population Health Research Center, Bordeaux, France

957 ⁵University of Bordeaux, Bordeaux, France

958 ⁶Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical
959 Sciences, Yokohama, Japan

960 ⁷Department of Statistical Genetics, Osaka University Graduate School of
961 Medicine, Osaka, Japan

962 ⁸Laboratory of Statistical Immunology, Immunology Frontier Research Center
963 (WPI-IFReC), Osaka University, Suita, Japan.

964 ⁹Department of Neurology, Massachusetts General Hospital, Harvard Medical
965 School, Boston, MA, USA

966 ¹⁰Laboratory of Experimental Cardiology, Division of Heart and Lungs, University
967 Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands

968 ¹¹deCODE genetics/AMGEN inc, Reykjavik, Iceland

969 ¹²Center for Genomic Medicine, Massachusetts General Hospital (MGH), Boston,
970 MA, USA

971 ¹³J. Philip Kistler Stroke Research Center, Department of Neurology, MGH,
972 Boston, MA, USA

973 ¹⁴Program in Medical and Population Genetics, Broad Institute, Cambridge, MA,
974 USA

975 ¹⁵Population Health Research Institute, McMaster University, Hamilton, Canada

- 976 ¹⁶Department of Epidemiology, Erasmus University Medical Center, Rotterdam,
977 Netherlands
- 978 ¹⁷Department of Radiology and Nuclear Medicine, Erasmus University Medical
979 Center, Rotterdam, Netherlands
- 980 ¹⁸Department of Medicine and Clinical Science, Graduate School of Medical
981 Sciences, Kyushu University, Fukuoka, Japan
- 982 ¹⁹Department of Clinical Sciences, Lund University, Malmö, Sweden
- 983 ²⁰Univ. Lille, Inserm, Institut Pasteur de Lille, LabEx DISTALZ-UMR1167, Risk
984 factors and molecular determinants of aging-related diseases, F-59000 Lille,
985 France
- 986 ²¹Centre Hosp. Univ Lille, Epidemiology and Public Health Department, F-59000
987 Lille, France
- 988 ²²AA Martinos Center for Biomedical Imaging, Department of Radiology,
989 Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
- 990 ²³Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine,
991 University of Washington, Seattle, WA, USA
- 992 ²⁴Division of Neurology, Faculty of Medicine, Brain Research Center, University
993 of British Columbia, Vancouver, Canada
- 994 ²⁵School of Life Science, University of Lincoln, Lincoln, UK
- 995 ²⁶Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto
996 Neurologico "Carlo Besta", Milano, Italy
- 997 ²⁷Department of Neurology, Mayo Clinic Rochester, Rochester, MN, USA
- 998 ²⁸MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and
999 Primary Care, University of Cambridge, Cambridge, UK
- 1000 ²⁹The National Institute for Health Research Blood and Transplant Research Unit
1001 in Donor Health and Genomics, University of Cambridge, UK
- 1002 ³⁰Neurovascular Research Laboratory, Vall d'Hebron Institut of Research,
1003 Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall
1004 d'Hebrón Hospital, Barcelona, Spain
- 1005 ³¹Stroke Pharmacogenomics and Genetics, Fundacio Docència i Recerca
1006 MutuaTerrassa, Terrassa, Spain
- 1007 ³²Children's Research Institute, Children's National Medical Center, Washington,
1008 DC, USA
- 1009 ³³Center for Translational Science, George Washington University, Washington,
1010 DC, USA
- 1011 ³⁴Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA,
1012 USA
- 1013 ³⁵Harvard Medical School, Boston, MA, USA
- 1014 ³⁶Center for Public Health Genomics, Department of Public Health Sciences,
1015 University of Virginia, Charlottesville, VA, USA
- 1016 ³⁷Department of Neurology, University of Maryland School of Medicine and
1017 Baltimore VAMC, Baltimore, MD, USA
- 1018 ³⁸Departments of Medicine, Pediatrics and Population Health Science, University
1019 of Mississippi Medical Center, Jackson, MS, USA
- 1020 ³⁹Institute of Cardiovascular Research, Royal Holloway University of London,
1021 UK & Ashford and St Peters Hospital, Surrey UK
- 1022 ⁴⁰Department of Psychiatry, The Hope Center Program on Protein Aggregation and
1023 Neurodegeneration (HPAN), Washington University, School of Medicine, St. Louis,
1024 MO, USA

1025 ⁴¹Department of Developmental Biology, Washington University School of
1026 Medicine, St. Louis, MO, USA
1027 ⁴²NIHR Blood and Transplant Research Unit in Donor Health and Genomics,
1028 Department of Public Health and Primary Care, University of Cambridge,
1029 Cambridge, UK
1030 ⁴³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,
1031 Hinxton, Cambridge, UK
1032 ⁴⁴British Heart Foundation, Cambridge Centre of Excellence, Department of
1033 Medicine, University of Cambridge, Cambridge, UK
1034 ⁴⁵Department of Medical Genetics, University Medical Center Utrecht, Utrecht,
1035 Netherlands
1036 ⁴⁶Department of Epidemiology, Julius Center for Health Sciences and Primary
1037 Care, University Medical Center Utrecht, Utrecht, Netherlands
1038 ⁴⁷Boston University School of Public Health, Boston, MA, USA
1039 ⁴⁸Framingham Heart Study, Framingham, MA, USA
1040 ⁴⁹Department of Immunology, Genetics and Pathology and Science for Life
1041 Laboratory, Uppsala University, Uppsala, Sweden
1042 ⁵⁰Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
1043 ⁵¹Department of Neurology and Stroke Center, Basel University Hospital,
1044 Switzerland
1045 ⁵²Neurorehabilitation Unit, University and University Center for Medicine of
1046 Aging and Rehabilitation Basel, Felix Platter Hospital, Basel, Switzerland
1047 ⁵³Department of Neurology, Yale University School of Medicine, New Haven, CT,
1048 USA
1049 ⁵⁴Program in Medical and Population Genetics, The Broad Institute of Harvard and
1050 MIT, Cambridge, MA, USA
1051 ⁵⁵Department of Neurology, Johns Hopkins University School of Medicine,
1052 Baltimore, MD, USA
1053 ⁵⁶Neuroscience Institute, SF Medical Center, Trenton, NJ, USA
1054 ⁵⁷Icelandic Heart Association Research Institute, Kopavogur, Iceland
1055 ⁵⁸University of Iceland, Faculty of Medicine, Reykjavik, Iceland
1056 ⁵⁹Department of Medical Sciences, Molecular Epidemiology and Science for Life
1057 Laboratory, Uppsala University, Uppsala, Sweden
1058 ⁶⁰Division of Public Health Sciences, Fred Hutchinson Cancer Research Center,
1059 Seattle, WA, USA
1060 ⁶¹Laboratory of Epidemiology and Population Science, National Institute on Aging,
1061 National Institutes of Health, Bethesda, MD, USA
1062 ⁶²Department of Neurology, Leeds General Infirmary, Leeds Teaching Hospitals
1063 NHS Trust, Leeds, UK
1064 ⁶³National Institute for Health and Welfare, Helsinki, Finland
1065 ⁶⁴FIMM - Institute for Molecular Medicine Finland, Helsinki, Finland
1066 ⁶⁵Department of Epidemiology, University of Washington, Seattle, WA, USA
1067 ⁶⁶Public Health Stream, Hunter Medical Research Institute, New Lambton,
1068 Australia
1069 ⁶⁷Faculty of Health and Medicine, University of Newcastle, Newcastle, Australia
1070 ⁶⁸School of Public Health, University of Alabama at Birmingham, Birmingham, AL,
1071 USA
1072 ⁶⁹Department of Biostatistical Sciences, Wake Forest School of Medicine,
1073 Winston-Salem, NC, USA

1074 ⁷⁰Aflac Cancer and Blood Disorder Center, Department of Pediatrics, Emory
1075 University School of Medicine, Atlanta, GA, USA
1076 ⁷¹Department of Medicine, Division of Cardiovascular Medicine, Stanford
1077 University School of Medicine, CA, USA
1078 ⁷²Department of Medical Sciences, Molecular Epidemiology and Science for Life
1079 Laboratory, Uppsala University, Uppsala, Sweden
1080 ⁷³Epidemiology, School of Public Health, University of Alabama at Birmingham,
1081 USA
1082 ⁷⁴Brown Foundation Institute of Molecular Medicine, University of Texas Health
1083 Science Center at Houston, Houston, TX, USA
1084 ⁷⁵Neurovascular Research Group (NEUVAS), Neurology Department, Institut
1085 Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona,
1086 Barcelona, Spain
1087 ⁷⁶Department of Pharmacotherapy and Translational Research and Center for
1088 Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL,
1089 USA
1090 ⁷⁷Division of Cardiovascular Medicine, College of Medicine, University of Florida,
1091 Gainesville, FL, USA
1092 ⁷⁸Department of Cardiology, Leiden University Medical Center, Leiden, the
1093 Netherlands
1094 ⁷⁹Program in Bioinformatics and Integrative Genomics, Harvard Medical School,
1095 Boston, MA, USA
1096 ⁸⁰Department of Biology, East Carolina University, Greenville, NC, USA
1097 ⁸¹Center for Health Disparities, East Carolina University, Greenville, NC, USA
1098 ⁸²University of Cincinnati College of Medicine, Cincinnati, OH, USA
1099 ⁸³RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
1100 ⁸⁴Department of Medicine, University of Colorado Denver, Anschutz Medical
1101 Campus, Aurora, CO, USA
1102 ⁸⁵Center for Public Health Genomics and Department of Biostatistical Sciences,
1103 Wake Forest School of Medicine, Winston-Salem, NC, USA
1104 ⁸⁶MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine,
1105 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK
1106 ⁸⁷Intramural Research Program, National Institute on Aging, National Institutes of
1107 Health, Bethesda, MD, USA
1108 ⁸⁸Department of Neurology, Radiology, and Biomedical Engineering, Washington
1109 University School of Medicine, St. Louis, MO, USA
1110 ⁸⁹KU Leuven – University of Leuven, Department of
1111 Neurosciences, Experimental Neurology, Leuven, Belgium
1112 ⁹⁰VIB Center for Brain & Disease Research, University Hospitals Leuven,
1113 Department of Neurology, Leuven, Belgium
1114 ⁹¹Univ.-Lille, INSERM U 1171. CHU Lille. Lille, France
1115 ⁹²Department of Medical and Molecular Genetics, King's College London, London,
1116 UK
1117 ⁹³SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's
1118 College London, London, UK
1119 ⁹⁴Northern Institute for Cancer Research, Paul O'Gorman Building, Newcastle
1120 University, Newcastle, UK
1121 ⁹⁵Department of Clinical Sciences Lund, Neurology, Lund University, Lund,
1122 Sweden

- 1123 ⁹⁶Department of Neurology and Rehabilitation Medicine, Skåne University
1124 Hospital, Lund, Sweden
- 1125 ⁹⁷Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden
- 1126 ⁹⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet,
1127 Stockholm, Sweden
- 1128 ⁹⁹University of Technology Sydney, Faculty of Health, Ultimo, Australia
- 1129 ¹⁰⁰Department of Medicine, University of Maryland School of Medicine, MD, USA
- 1130 ¹⁰¹Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
- 1131 ¹⁰²Geriatrics Research and Education Clinical Center, Baltimore Veterans
1132 Administration Medical Center, Baltimore, MD, USA
- 1133 ¹⁰³Division of Geriatrics, School of Medicine, University of Mississippi Medical
1134 Center, Jackson, MS, USA
- 1135 ¹⁰⁴Memory Impairment and Neurodegenerative Dementia Center, University of
1136 Mississippi Medical Center, Jackson, MS, USA
- 1137 ¹⁰⁵Laboratory of Neurogenetics, National Institute on Aging, National Institutes of
1138 Health, Bethesda, MD, USA
- 1139 ¹⁰⁶Data Tecnica International, Glen Echo MD, USA
- 1140 ¹⁰⁷Department of Epidemiology and Public Health, Graduate School of Medical
1141 Sciences, Kyushu University, Fukuoka, Japan
- 1142 ¹⁰⁸Clinical Research Facility, Department of Medicine, NUI Galway, Galway,
1143 Ireland
- 1144 ¹⁰⁹Cardiovascular Health Research Unit, Department of Medicine, University of
1145 Washington, Seattle, WA, USA
- 1146 ¹¹⁰Department of Epidemiology, University of Washington, Seattle, WA, USA
- 1147 ¹¹¹Department of Health Services, University of Washington, Seattle, WA, USA
- 1148 ¹¹²Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
- 1149 ¹¹³Brain Center Rudolf Magnus, Department of Neurology, University Medical
1150 Center Utrecht, Utrecht, The Netherlands
- 1151 ¹¹⁴Usher Institute of Population Health Sciences and Informatics, University of
1152 Edinburgh, Edinburgh, UK
- 1153 ¹¹⁵Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
- 1154 ¹¹⁶Fred Hutchinson Cancer Research Center, University of Washington, Seattle,
1155 WA, USA
- 1156 ¹¹⁷Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
- 1157 ¹¹⁸Department of Biostatistics, University of Washington, Seattle, WA, USA
- 1158 ¹¹⁹Nuffield Department of Clinical Neurosciences, University of Oxford, UK
- 1159 ¹²⁰Institute for Translational Genomics and Population Sciences, Los Angeles
1160 Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA,
1161 USA
- 1162 ¹²¹Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA
1163 Medical Center, Torrance, CA, USA
- 1164 ¹²²Department of Neurology, Miller School of Medicine, University of Miami,
1165 Miami, FL, USA
- 1166 ¹²³Department of Allergy and Rheumatology, Graduate School of Medicine, the
1167 University of Tokyo, Tokyo, Japan
- 1168 ¹²⁴Center for Public Health Genomics, University of Virginia, Charlottesville, VA,
1169 USA
- 1170 ¹²⁵Department of Pediatrics, College of Medicine, University of Oklahoma Health
1171 Sciences Center, Oklahoma City, OK, USA
- 1172 ¹²⁶Department of Neurology, Medical University of Graz, Graz, Austria

1173 ¹²⁷University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF,
1174 Greifswald, Germany
1175 ¹²⁸University Medicine Greifswald, Department of Neurology, Greifswald,
1176 Germany
1177 ¹²⁹Department of Neurology, Jagiellonian University, Krakow, Poland
1178 ¹³⁰Department of Neurology, Justus Liebig University, Giessen, Germany
1179 ¹³¹Department of Clinical Neurosciences/Neurology, Institute of Neuroscience and
1180 Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg,
1181 Sweden
1182 ¹³²Sahlgrenska University Hospital, Gothenburg, Sweden
1183 ¹³³Stroke Division, Florey Institute of Neuroscience and Mental Health, University
1184 of Melbourne, Heidelberg, Australia
1185 ¹³⁴Austin Health, Department of Neurology, Heidelberg, Australia
1186 ¹³⁵Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden
1187 University Medical Center, Leiden, the Netherlands
1188 ¹³⁶INSERM U1219, Bordeaux, France
1189 ¹³⁷Department of Public Health, Bordeaux University Hospital, Bordeaux, France
1190 ¹³⁸Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University
1191 Medical Center Rotterdam, Netherlands
1192 ¹³⁹Center for Medical Systems Biology, Leiden, Netherlands
1193 ¹⁴⁰School of Medicine, Dentistry and Nursing at the University of Glasgow,
1194 Glasgow, UK
1195 ¹⁴¹Department of Epidemiology and Population Health, Albert Einstein College of
1196 Medicine, NY, USA
1197 ¹⁴²Department of Physiology and Biophysics, University of Mississippi Medical
1198 Center, Jackson, MS, USA
1199 ¹⁴³A full list of members and affiliations appears in the Supplementary Note
1200 ¹⁴⁴Department of Human Genetics, McGill University, Montreal, Canada
1201 ¹⁴⁵Department of Pathophysiology, Institute of Biomedicine and Translation
1202 Medicine, University of Tartu, Tartu, Estonia
1203 ¹⁴⁶Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia
1204 ¹⁴⁷Clinical Gene Networks AB, Stockholm, Sweden
1205 ¹⁴⁸Department of Genetics and Genomic Sciences, The Icahn Institute for
1206 Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, New
1207 York, NY, USA
1208 ¹⁴⁹Department of Pathophysiology, Institute of Biomedicine and Translation
1209 Medicine, University of Tartu, Biomeedikum, Tartu, Estonia
1210 ¹⁵⁰Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska
1211 Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden.
1212 ¹⁵¹Clinical Gene Networks AB, Stockholm, Sweden
1213 ¹⁵²Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team
1214 Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France
1215 ¹⁵³ICAN Institute for Cardiometabolism and Nutrition, Paris, France
1216 ¹⁵⁴Department of Biomedical Engineering, University of Virginia, Charlottesville,
1217 VA, USA
1218 ¹⁵⁵Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
1219 ¹⁵⁶Seattle Epidemiologic Research and Information Center, VA Office of Research
1220 and Development, Seattle, WA, USA
1221 ¹⁵⁷Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA,
1222 USA

- 1223 ¹⁵⁸Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust,
1224 Gjettem, Norway
- 1225 ¹⁵⁹Saw Swee Hock School of Public Health, National University of Singapore and
1226 National University Health System, Singapore
- 1227 ¹⁶⁰National Heart and Lung Institute, Imperial College London, London, UK
- 1228 ¹⁶¹Department of Gene Diagnostics and Therapeutics, Research Institute, National
1229 Center for Global Health and Medicine, Tokyo, Japan
- 1230 ¹⁶²Department of Epidemiology, Tulane University School of Public Health and
1231 Tropical Medicine, New Orleans, LA, USA
- 1232 ¹⁶³Department of Cardiology, University Medical Center Groningen, University of
1233 Groningen, Netherlands
- 1234 ¹⁶⁴MRC-PHE Centre for Environment and Health, School of Public Health,
1235 Department of Epidemiology and Biostatistics, Imperial College London, London,
1236 UK
- 1237 ¹⁶⁵Department of Epidemiology and Biostatistics, Imperial College London,
1238 London, UK
- 1239 ¹⁶⁶Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK
- 1240 ¹⁶⁷National Heart, Lung and Blood Research Institute, Division of Intramural
1241 Research, Population Sciences Branch, Framingham, MA, USA
- 1242 ¹⁶⁸A full list of members and affiliations appears at the end of the manuscript
- 1243 ¹⁶⁹Department of Pharmaceutical Sciences, College of Pharmacy, University of
1244 Oklahoma Health Sciences Center, Oklahoma City, OK, USA
- 1245 ¹⁷⁰Oklahoma Center for Neuroscience, Oklahoma City, OK, USA
- 1246 ¹⁷¹Department of Pathology and Genetics, Institute of Biomedicine, The
1247 Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
- 1248 ¹⁷²Department of Neurology, Helsinki University Hospital, Helsinki, Finland
- 1249 ¹⁷³Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland
- 1250 ¹⁷⁴Department of Neurology, University of Washington, Seattle, WA, USA
- 1251 ¹⁷⁵Albrecht Kossel Institute, University Clinic of Rostock, Rostock, Germany
- 1252 ¹⁷⁶Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield
1253 Department of Population Health, University of Oxford, Oxford, UK
- 1254 ¹⁷⁷Department of Genetics, Perelman School of Medicine, University of
1255 Pennsylvania, PA, USA
- 1256 ¹⁷⁸Faculty of Medicine, University of Iceland, Reykjavik, Iceland
- 1257 ¹⁷⁹Departments of Neurology and Public Health Sciences, University of Virginia
1258 School of Medicine, Charlottesville, VA, USA
- 1259 ¹⁸⁰Department of Neurology, Boston University School of Medicine, Boston, MA,
1260 USA
- 1261 ¹⁸¹Human Genetics Center, University of Texas Health Science Center at Houston,
1262 Houston, TX, USA
- 1263 ¹⁸²Center for Genomic Medicine, Kyoto University Graduate School of Medicine,
1264 Kyoto, Japan
- 1265 ¹⁸³Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
- 1266 ¹⁸⁴German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
- 1267 ¹⁸⁵Boston University School of Medicine, Boston, MA, USA
- 1268 ¹⁸⁶University of Kentucky College of Public Health, Lexington, KY, USA
- 1269 ¹⁸⁷University of Newcastle and Hunter Medical Research Institute, New Lambton,
1270 Australia
- 1271 ¹⁸⁸Univ. Montpellier, Inserm, U1061, Montpellier, France
- 1272 ¹⁸⁹Centre for Research in Environmental Epidemiology, Barcelona, Spain

1273 ¹⁹⁰Department of Neurology, Università degli Studi di Perugia, Umbria, Italy
1274 ¹⁹¹Department of Medicine, University of Maryland School of Medicine, Baltimore,
1275 MD, USA
1276 ¹⁹²Broad Institute, Cambridge, MA, USA
1277 ¹⁹³Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR
1278 1219, Bordeaux, France
1279 ¹⁹⁴Bordeaux University Hospital, Department of Neurology, Memory Clinic,
1280 Bordeaux, France
1281 ¹⁹⁵Neurovascular Research Laboratory. Vall d'Hebron Institut of Research,
1282 Neurology and Medicine Departments-Universitat Autònoma de Barcelona. Vall
1283 d'Hebrón Hospital, Barcelona, Spain
1284 ¹⁹⁶University Medicine Greifswald, Department of Internal Medicine B,
1285 Greifswald, Germany
1286 ¹⁹⁷DZHK, Greifswald, Germany
1287 ¹⁹⁸Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK
1288 ¹⁹⁹Hero DMC Heart Institute, Dayanand Medical College & Hospital, Ludhiana,
1289 India
1290 ²⁰⁰Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska
1291 Institutet, Stockholm, Sweden
1292 ²⁰¹Karolinska Institutet, Stockholm, Sweden
1293 ²⁰²Division of Emergency Medicine, and Department of Neurology, Washington
1294 University School of Medicine, St. Louis, MO, USA
1295 ²⁰³Tohoku Medical Megabank Organization, Sendai, Japan
1296 ²⁰⁴Department of Psychiatry, Washington University School of Medicine, St. Louis,
1297 MO, USA
1298 ²⁰⁵Department of Public Health and Caring Sciences / Geriatrics, Uppsala
1299 University, Uppsala, Sweden
1300 ²⁰⁶Epidemiology and Prevention Group, Center for Public Health Sciences,
1301 National Cancer Center, Tokyo, Japan
1302 ²⁰⁷Department of Internal Medicine and the Center for Clinical and Translational
1303 Science, The Ohio State University, Columbus, OH, USA
1304 ²⁰⁸Institute of Neuroscience and Physiology, the Sahlgrenska Academy at
1305 University of Gothenburg, Goteborg, Sweden
1306 ²⁰⁹Department of Basic and Clinical Neurosciences, King's College London,
1307 London, UK
1308 ²¹⁰Department of Health Care Administration and Management, Graduate School
1309 of Medical Sciences, Kyushu University, Japan
1310 ²¹¹Department of Medicine and Clinical Science, Graduate School of Medical
1311 Sciences, Kyushu University, Japan
1312 ²¹²Landspítali National University Hospital, Departments of Neurology &
1313 Radiology, Reykjavik, Iceland
1314 ²¹³Department of Neurology, Heidelberg University Hospital, Germany
1315 ²¹⁴Department of Neurology, Erasmus University Medical Center
1316 ²¹⁵Hospital Universitari Mutua Terrassa, Terrassa (Barcelona), Spain
1317 ²¹⁶Albert Einstein College of Medicine, Montefiore Medical Center, New York,
1318 NY, USA
1319 ²¹⁷John Hunter Hospital, Hunter Medical Research Institute and University of
1320 Newcastle, Newcastle, NSW, Australia
1321 ²¹⁸Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical
1322 Neurosciences, University of Oxford, UK

1323 ²¹⁹Department of Medical Sciences, Uppsala University, Uppsala, Sweden
1324 ²²⁰Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human
1325 Genetics, University of Oxford, Oxford, UK
1326 ²²¹The Wellcome Trust Centre for Human Genetics, Oxford, UK
1327 ²²²Beth Israel Deaconess Medical Center, Boston, MA, USA
1328 ²²³Wake Forest School of Medicine, Wake Forest, NC, USA
1329 ²²⁴Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA
1330 ²²⁵BioBank Japan, Laboratory of Clinical Sequencing, Department of
1331 Computational biology and medical Sciences, Graduate school of Frontier Sciences,
1332 The University of Tokyo, Tokyo, Japan
1333 ²²⁶Neurovascular Research Laboratory, Vall d'Hebron Institut of Research,
1334 Neurology and Medicine Departments-Universitat Autònoma de Barcelona. Vall
1335 d'Hebrón Hospital, Barcelona, Spain
1336 ²²⁷Department of Biostatistics, University of Liverpool, Liverpool, UK
1337 ²²⁸Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
1338 ²²⁹Institute of Genetic Epidemiology, Helmholtz Zentrum München - German
1339 Research Center for Environmental Health, Neuherberg, Germany
1340 ²³⁰Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany
1341 ²³¹DZHK (German Centre for Cardiovascular Research), partner site Munich Heart
1342 Alliance, Munich, Germany
1343 ²³²Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto
1344 Neurologico "Carlo Besta", Milano, Italy
1345 ²³³Karolinska Institutet, MEB, Stockholm, Sweden
1346 ²³⁴University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, Estonia
1347 ²³⁵Department of Clinical and Experimental Sciences, Neurology Clinic,
1348 University of Brescia, Italy
1349 ²³⁶Translational Genomics Unit, Department of Oncology, IRCCS Istituto di
1350 Ricerche Farmacologiche Mario Negri, Milano, Italy
1351 ²³⁷Department of Genetics, Microbiology and Statistics, University of Barcelona,
1352 Barcelona, Spain
1353 ²³⁸Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions,
1354 Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona,
1355 Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona,
1356 Spain
1357 ²³⁹Department of Neurology, IMIM-Hospital del Mar, and Universitat Autònoma
1358 de Barcelona, Spain
1359 ²⁴⁰IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
1360 ²⁴¹National Institute for Health Research Comprehensive Biomedical Research
1361 Centre, Guy's & St. Thomas' NHS Foundation Trust and King's College London,
1362 London, UK
1363 ²⁴²Division of Health and Social Care Research, King's College London, London,
1364 UK
1365 ²⁴³FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland
1366 ²⁴⁴THL-National Institute for Health and Welfare, Helsinki, Finland
1367 ²⁴⁵Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate,
1368 Japan
1369 ²⁴⁶BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow,
1370 UK
1371 ²⁴⁷deCODE Genetics/Amgen, Inc., Reykjavik, Iceland
1372 ²⁴⁸Icelandic Heart Association, Reykjavik, Iceland

1373 ²⁴⁹Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg,
1374 Goteborg, Sweden
1375 ²⁵⁰Department of Epidemiology, University of Maryland School of Medicine,
1376 Baltimore, MD, USA
1377 ²⁵¹Institute of Cardiovascular and Medical Sciences, Faculty of Medicine,
1378 University of Glasgow, Glasgow, UK
1379 ²⁵²Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich,
1380 Germany
1381 ²⁵³Division of Epidemiology and Prevention, Aichi Cancer Center Research
1382 Institute, Nagoya, Japan
1383 ²⁵⁴Department of Epidemiology, Nagoya University Graduate School of Medicine,
1384 Nagoya, Japan
1385 ²⁵⁵University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF,
1386 Greifswald, Germany
1387 ²⁵⁶Department of Neurology, Caen University Hospital, Caen, France
1388 ²⁵⁷University of Caen Normandy, Caen, France
1389 ²⁵⁸Department of Internal Medicine, Erasmus University Medical Center,
1390 Rotterdam, Netherlands
1391 ²⁵⁹Landspítali University Hospital, Reykjavik, Iceland
1392 ²⁶⁰Survey Research Center, University of Michigan, Ann Arbor, MI, USA
1393 ²⁶¹University of Virginia Department of Neurology, Charlottesville, VA, USA
1394
1395