Blood-based epigenome-wide analyses of cognitive abilities

Daniel L. McCartney[1], Robert F. Hillary[1], Daniel Trejo Banos[2,3], Danni A. Gadd[1], Rosie M. Walker[1,4], Cliff Nangle[1], Robin Flaig[1], Archie Campbell[1], Alison D. Murray[5], Sarah E. Harris[6], David J. Porteous[1], Elliot M Tucker-Drob[7,8], Andrew M. McIntosh[1,9], Kathryn L. Evans[1], Ian J. Deary[6], Simon R. Cox[6], Matthew R. Robinson[10,*], Riccardo E. Marioni[1,*]

[1] Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK.

[2] Department of Quantitative Biomedicine, University of Zurich, Zurich.

[3] Biomedical Informatics, University Hospital of Zurich, Zurich, Switzerland

[4] Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh BioQuarter, Edinburgh, EH16 4SB, UK

[5] Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, Scotland, UK

[6] Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK.

[7] Department of Psychology, University of Texas at Austin, TX, USA

[8] Population Research Center and Center on Aging and Population Sciences, University of Texas at Austin, TX, USA

[9] Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

[10] Institute of Science and Technology Austria, 3400, Klosterneuburg, Austria

*Corresponding author:

Names: Riccardo Marioni and Matthew Robinson

Contact Details: riccardo.marioni@ed.ac.uk and matthew.robinson@ist.ac.at

Abstract

We present a blood-based epigenome-wide association study and variance-components analysis of cognitive functions (n=9,162). Individual differences in DNA methylation (DNAm) accounted for up to 41.5% of the variance in cognitive functions; together, genetic and epigenetic markers accounted for up to 70.4% of the variance. A DNAm predictor accounted for 3.4% and 4.5% ($P \leq 9.9 \times 10^{-6}$) of the variance in general cognitive ability, independently of a polygenic score, in two external cohorts.

Manuscript

Blood-based markers of cognitive functioning might provide an accessible way to track neurodegeneration years prior to clinical manifestation of cognitive impairment and dementia. Blood-based DNA methylation can be used to generate predictors (with r>0.35) of lifestyle factors, such as smoking, alcohol consumption, and obesity [1] – factors that are linked with poorer cognitive function and an increased risk of dementia [2]. However, blood-based DNA methylation predictors of cognitive function itself, rather than its known correlates, may index a wider range of risk factors for neurodegeneration. Despite being peripheral to the Central Nervous System, blood is an easily accessible tissue, and its DNA methylation patterns may enable early diagnosis and provide mechanistic insights of early phases of disease progression. Although DNA methylation in brain tissue may provide more direct insights into the biology of neurodegeneration, [3, 4] acquiring in-vivo brain tissue samples is not feasible outside of extraordinary circumstances. Recent methodological advances [5, 6] have enabled the estimation of variance that DNA methylation can account for in complex traits. Therefore, we can now quantify how well blood-based DNA methylation predicts cognitive test outcomes.

There is modest evidence for associations between individual blood-based methylation sites and cognitive functioning; six CpG probes were identified as genome-wide significant in meta-analysis epigenome-wide association studies (EWASs) of seven cognitive traits [7]. That study was limited by heterogeneous cognitive outcomes across cohorts, which also varied in age and ethnicity (meta-analysis n ranging from 2,557 to 6,809). Large-scale single cohort studies with consistent cognitive phenotyping and DNA methylation typing and quality control are lacking. Here, we overcome these limitations by utilising phenotypic cognitive data and blood-based methylation data from a single large cohort of European ancestries.

Blood-based DNA methylation was assessed in 9,162 adult participants from the Generation Scotland cohort [8, 9]. Cognitive abilities were assessed across four domains: verbal declarative memory (Wechsler Logical Memory test); processing speed (Wechsler Digit Symbol test); executive function (phonemic verbal fluency); and verbal crystallised ability (using Mill Hill Vocabulary synonyms test) [10]. Scores from the first three tests were used to derive a general fluid cognitive function score (g_f); scores from all four tests were used to derive a general cognitive function score (g). The first unrotated component from a principal component analysis was extracted in both instances, accounting for 49% and 42% of the variance, respectively. Full details can be found in the **Online Methods**.

Variance components analyses and epigenome-wide association studies (EWAS) were conducted using BayesR+ software (**Online Methods**). The cognitive phenotypes were corrected for age, sex, BMI and epigenetic smoking score [11], and the DNA methylation data were corrected for batch, age, sex and epigenetic smoking.

A summary of cohort demographics and the phenotypes assessed is presented in **Supplementary Table 1 and Supplementary Figure 1**. The study cohort comprised 59% females and had a mean age of 49.8 years (SD 13.6; range 18 – 93).

Variance components analyses indicated that the proportion of variance of cognitive tests' scores accounted for by all DNA methylation probes were: 5.3% (95% Credible Interval [CI] = [0%, 21.6%]) for logical memory; 23.0% [12.1%, 37.8%] for verbal fluency; 24.2% [12.5%, 38.5%] for digit symbol; 45.8% [37.3%, 54.1%] for vocabulary; 29.2% [17.3%, 42.0%] for general fluid cognitive function; and 41.6% [31.0%, 53.0%] for general cognitive function (**Figure 1 and Supplementary Table 2**). A sensitivity analysis using a linear mixed model approach [5] with genetic and epigenetic relationship matrices yielded near identical DNAm variance components estimates; a sensitivity analysis using data from an unrelated subset (n=4,261) of the study cohort that was processed in a single methylation batch also showed similar estimates (**Supplementary Table 3**). The majority of the variance was accounted for by CpG sites with small effects; three prior mixture distributions were set with small (0.01%), medium (0.1%), and large (1%) variances for effect sizes (**Figure 1 and Supplementary Table 4**).

Genetic variation across common SNPs accounted for between 8% [0%, 34.4%] of the variance for logical memory to 42.9% [24.4%, 60.1%] for vocabulary; these estimates are in line with previous GREML estimates from the cohort [12]. The proportion of variance accounted for when combining the independent effects of genetics and DNAm increased to between 12.4% [0%, 41.9%] for logical memory, and 70.4% [59.9%, 78.6%] for vocabulary (**Figure 1; Supplementary Table 2**). Notably, the CpG contribution to the variance accounted for was largely independent of the genetic component. There was an absolute attenuation <7% (relative attenuation <17%) to the mean epigenetic effect size estimates of the variance components in the model that included genetics.

The EWAS models for the six cognitive traits identified three unique lead DNAm sites with a posterior inclusion probability (PIP) greater than 0.80. Each of these had a PIP >0.95 (probability of a false discovery <0.05) for at least one trait (**Supplementary Figure 2; Supplementary Table 5;** the entire output is available at https://gitlab.com/danielmccartney/ewas_of_cognitive_function). For the three lead CpG sites we queried the EWAS catalog (accessed 5 April 2021) for associations with other traits (**Supplementary Table 6**) [13]. They have been linked to age, metabolite levels, lipid levels, kidney cancer, and lung function. Of 28 CpGs identified in previous blood-based EWAS analyses of cognitive ability and Alzheimer's disease [4, 7, 14], 24 were available for lookup in the present dataset; Generation Scotland data was not included in any of these studies. There was no evidence for replication (maximum PIP of 0.05; **Supplementary Table 7**).

We applied the mean posterior effect sizes for all possible CpGs from the EWAS model of general cognitive function to CpGs in two independent studies (The Lothian Birth Cohort (LBC) 1936 and The Lothian Birth Cohort 1921, n=844 and n=427 with DNAm, cognitive scores, and cognitive polygenic score available, respectively – **Online Methods**). The DNAm predictor had a Pearson correlation of r=0.18 (LBC1936) and r=0.22 (LBC1921) with general cognitive function (**Supplementary Figure 3**). The incremental R^2 upon the addition of the DNAm predictor to a linear regression model adjusting for age and sex was 3.4% ($P=2.0x10^{-8}$) in LBC1936 and 4.5% ($P=9.9x10^{-6}$) in LBC1921. The corresponding R^2 for the polygenic score derived from a GWAS of 168,033 individuals in UK Biobank – a training sample over 18-times greater than the DNAm training sample – was 7.3% ($P<2x10^{-16}$) and 6.9% ($P=3.1x10^{-8}$), respectively; the additive incremental R^2 from the two omics-based predictors was 10.7% and 10.5% (**Figure 2**). There was a 0.66 SD (SE 0.15, $P=7.4x10^{-6}$) and 0.56 SD (SE 0.15, $P=2.0x10^{-4}$) difference in the mean cognitive score between the decile ($n_{per\ decile}$=84-85, LBC1936) and quintile ($n_{per\ quintile}$=85-86, LBC1921) with the highest versus lowest

DNAm cognitive score in linear regression models adjusting for age and sex. After including smoking, years of education, BMI, lung function, walking speed, grip strength, high blood pressure, alcohol consumption, a depression questionnaire score, and an index of social deprivation as covariates (**Online Methods**), the incremental R^2 for the epigenetic and polygenic scores dropped from 3.4% and 7.3% to 0.8% ($P=1.9x10^{-3}$) and 4.7% ($P=1.5x10^{-14}$) in LBC1936, respectively. The corresponding results for LBC1921 were a drop from 4.5% and 6.9% to 1.1% (P=0.010) and 2.4% ($P=1.3x10^{-4}$). A multivariate linear regression model of the epigenetic score (outcome) against the covariates showed modest and inconsistent associations across the two LBC studies (**Supplementary Table 8**). Based on a training sample size of 10,000 and assuming 100,000 CpGs affect the trait, with a DNAm variance components estimate of 41.6% for g, we would expect a DNAm prediction R^2 value of about 4.0% (following Formula 1 from [15]), which is very similar to the estimates obtained for LBC1936 (3.4%) and LBC1921 (4.5%). If the training sample increased to 20,000 or 100,000 then the expected R^2 should increase to around 8% and 29%, respectively.

This is the first variance components analysis of DNAm and cognitive function. We show substantial contributions to several of the tests in addition to the development of a novel epigenetic score with application to two independent test cohorts where it associated with cognitive ability. We used a large homogeneous discovery cohort with consistent cognitive testing and DNA collection across all participants. This will minimise the biases that are an inherent problem in heterogeneous and small EWAS meta-analyses. The estimation of posterior effect sizes using BayesR+ has been shown to be robust to various potential sources of heterogeneity, including family structure [16] and population structure based on genetic or epigenetic principal components [6]. Nonetheless, replication of our findings in cohorts of different ages and backgrounds will help to refine and generalise the estimates presented here.

Methylation-based predictors of cognitive function may improve longitudinal disease prediction and risk profiling of neurodegenerative health outcomes such as dementia. Here, the epigenetic variance components and prediction score were independent of genetic contributions. Unlike DNA differences which are largely fixed throughout life, DNAm differences may reflect environmental effects and phenotypic causation, directly through cognitive-related pathways or indirectly via related lifestyle and health outcomes. Given the overwhelming contribution of CpGs with small effects to our estimates, increasing EWAS sample sizes and employing powerful statistical methods like BayesR+ will clearly lead to both locus discovery and more accurate DNAm-based predictors of cognitive function.

Declarations

Ethics approval and consent to participate

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

Availability of Data and Material

According to the terms of consent for GS participants, access to data must be reviewed by the GS Access Committee. Applications should be made to access@generationscotland.org.

All code is available with open access at the following Gitlab repository:

https://gitlab.com/danielmccartney/ewas_of_cognitive_function

Competing Interests

R.E.M has received a speaker fee from Illumina and is an advisor to the Epigenetic Clock Development Foundation. A.M.M has previously received speaker fees from Janssen and Illumina and research funding from The Sackler Trust. All other authors declare no competing interests.

Acknowledgements

GS received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping and DNA methylation profiling of the GS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL; Reference 104036/Z/14/Z). The DNA methylation data assayed for Generation Scotland was partially funded by a 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Ref: 27404; awardee: Dr David M Howard) and by a JMAS SIM fellowship from the Royal College of Physicians of Edinburgh (Awardee: Dr Heather C Whalley). We thank the LBC1936 participants and team members who contributed to these studies. The LBC1936 is supported by Age UK (Disconnected Mind project, which supports S.E.H.), the Medical Research Council (G0701120, G1001245, MR/M013111/1, MR/R024065/1, which supports S.R.C.), and the University of Edinburgh. Methylation typing of LBC1936 was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of

Edinburgh, and The University of Queensland. Genotyping was funded by the Biotechnology and Biological Sciences Research Council (BB/F019394/1). R.F.H. and D.A.G. are supported by funding from the Wellcome Trust 4-year PhD in Translational Neuroscience–training the next generation of basic neuroscientists to embrace clinical research [108890/Z/15/Z]. E.M.T-D. was supported by National Institutes of Health (NIH) grants R01AG054628, R01MH120219, R01HD083613, P2CHD042849 and P30AG066614. SRC was also supported by a National Institutes of Health (NIH) research grant R01AG054628. D.L.Mc.C. and R.E.M. are supported by Alzheimer's Research UK major project grant ARUK-PG2017B−10.

Online Methods

The Generation Scotland Cohort

Details of the Generation Scotland: Scottish Family Health Study (GS) have been described in detail elsewhere [9, 10]. Briefly, GS comprises over 20,000 individuals comprehensively profiled for genetic, clinical, lifestyle, and sociodemographic data. A subset of 9,162 individuals from GS (aged 18 to 93 years, mean=49.7, SD=13.6) had genome-wide DNA methylation measured [17]. This subset was processed in two batches, hereafter referred to as "Set 1" and "Set 2".

Methylation preparation in Generation Scotland

Quality control was performed on Illumina HumanMethylationEPIC BeadChip DNA methylation data from blood samples of 5,200 related individuals from Set 1, and 4,588 genetically unrelated individuals from Set 2, also genetically unrelated to those in Set 1. Three Set 1 individuals who had answered "yes" to presence of all of 16 self-reported disease conditions in the study's health questionnaire were excluded from the analysis. Filtering for outliers, sex mismatches, non-blood samples, and poorly detected probes and samples was performed [17]. Further filtering was then carried out to remove CpGs with missing values, non-autosomal and non-CpG sites. Five individuals with a self-reported diagnosis of Alzheimer's disease were removed, along with samples with missing covariate information or cognitive variables. Following sample filtering, 9,162 complete cases remained comprising 4,901 Set 1 individuals and 4,261 Set 2 individuals.

Cognitive Phenotypes in Generation Scotland

Six cognitive phenotypes were assessed in this study: logical memory, digit symbol test score, verbal fluency, vocabulary, general cognitive ability, and general fluid cognitive ability. The logical memory phenotype (verbal declarative memory) was calculated from the Wechsler Logical Memory test, taking the sum of immediate and delayed recall of one oral story [18]. The digit symbol phenotype is often used as a measure of processing speed; calculated from the Wechsler Digit Symbol Substitution test in which participants must recode digits to symbols over a 120 second period [19]. The verbal fluency phenotype is often used as a measure of executive functioning and was derived from the phonemic verbal fluency test, using the letters C, F, and L, each for one minute [20]. Vocabulary was measured using the Mill Hill Vocabulary Scale, junior and senior synonyms combined [21]. General fluid cognitive ability (g_f) was calculated from the first unrotated principal component of logical memory, verbal fluency, and digit symbol tests. General cognitive ability (g) was derived from the first unrotated principal component from the same variables plus vocabulary.

Statistical Analysis

BayesR+ was also used for the EWAS and to estimate the genetic and DNA methylation-based heritability of the cognitive traits. BayesR+ is a software implemented in C++ for performing Bayesian penalised regression and Gaussian mixture-based variance partitioning on complex traits [6]. The joint and conditional effects of methylation sites (n=764,525) on cognitive traits were examined. Phenotypic and methylation data were scaled to mean zero and unit variance. The prior distribution comprised a series of Gaussian distributions which corresponded to effect sizes of different magnitudes (i.e. methylation sites with small, medium and large effect sizes), as well as a discrete spike at zero which allows for the

omission of probes with non-identifiable effects. The prior mixture variances were set to 0.0001, 0.001 and 0.01. Phenotypes were corrected for age, sex, BMI and epigenetic smoking score [11], and the DNA methylation data were corrected for batch, age, sex and epigenetic smoking. Adding a quadratic term for age in the correction of cognitive phenotypes made negligible differences to the trait residuals (r>0.98 between linear model residuals for each trait). To obtain estimates of variance accounted for in cognitive traits by methylation data and individual CpG associations with the cognitive test scores, Gibbs sampling was performed to sample over the posterior distribution conditional on the input data. The Gibbs algorithm consisted of 10,000 samples and 5,000 samples of burn-in after which a thinning of 5 samples was applied to reduce autocorrelation. The process was repeated over four chains, initializing a different random number seed for each chain. The last 250 iterations from each chain were combined for downstream analyses. For the EWAS, CpGs within 2.5kb and highly correlated (absolute Pearson correlation >0.5) with a lead CpG with posterior inclusion probability greater than 0.2 were grouped together. For each probe group, we calculated the proportion of iterations for which at least one probe was included in the model, yielding the group posterior inclusion probability. We then calculated the average (across the 1,000 iterations) sum of the squared regression coefficients for the probe group to give the contribution of the group to the total variance. Finally, we highlighted the lead CpG for the groups where the combined posterior inclusion probability was >0.80. Genetic effects at 560,797 SNPs (minor allele frequency > 1%; scaled to mean 0, variance 1) from the Illumina HumanOmniExpressExome-8 v1.0 Bead Chip or Illumina HumanOmniExpressExome-8 v1.2 Bead Chip were examined [22], setting prior mixture variances to 0.00001, 0.0001 and 0.001. For the genetic analysis, phenotypes were pre-corrected for age, sex and 20 genetic PCs. To estimate the additive and independent effects of DNAm and genetic data on complex traits, a

combined analysis was run – using the phenotype corrections specified for the EWAS model – setting prior mixture variances as above.

The mean posterior effect sizes for the EWAS model of general cognitive function, g, were used to generate an epigenetic predictor in two independent datasets, the Lothian Birth Cohort 1936 (LBC1936) and the Lothian Birth Cohort 1921 (LBC1921). The LBC are studies of cognitive ageing in older adults from the area around Edinburgh, Scotland [23, 24]. Briefly, participants were born in either 1921 or 1936 and completed the Scottish Mental Survey of 1932 or1947 at age 11. From age 70 (LBC1936) and age 79 (LBC1921), they were assessed triennially for a variety of health and lifestyle outcomes, with DNA collected at each visit. Here, we considered blood-based DNA methylation data from the age 70 (LBC1936) and age 79 (LBC1921) samples. Methylation was assessed on the Illumina 450k array – the predecessor of the EPIC array. Processing and quality control have been described previously [1, 25, 26]. This included steps to remove methylation samples and individuals with poor quality control measures, along with individuals who had mismatching genotypes or predicted sex information. This left a dataset of 381,846 CpGs (overlapping with those included in the Generation Scotland analyses) for 861 LBC1936 (436 LBC1921) individuals – 34 LBC1936 individuals were excluded due to DNAm being assessed as part of a separate analysis batch. Within each cohort, each CpG was scaled to mean 0, variance 1 with missing values mean imputed (i.e., set to 0) prior to multiplication by the mean CpG weights (for all available CpGs) and summation to give the epigenetic score. A polygenic score for cognitive ability was derived from Z-scores from all possible SNPs (GWAS P<1) in a UK Biobank GWAS of verbal numerical reasoning (n=168,033) [29] and applied to LBC1936 and LBC1921 genotype data using default settings in the PRSice software [30, 31].In LBC1936, general cognitive ability was defined as the first unrotated principal component (that accounted for 52% of the variance) from a PCA of six cognitive tests from the Wechsler

Adult Intelligence Scale-III UK (matrix reasoning, letter number sequencing, block design, symbol search, digit symbol, and digit span backward) [19] plus a test of vocabulary (National Adult Reading Test) [27]. A similar approach was taken in LBC1921, where we considered the first unrotated principal component (that also accounted for 51% of the variance) from a PCA of four cognitive tests: Raven's Standard Progressive Matrices [28], letter-number sequencing [19], digit symbol coding [19], and the National Adult Reading Test) [27]. The cognitive tests were completed at the same visit that blood was drawn for DNA profiling in both LBC studies. There were 844 LBC1936 (427 LBC1921) individuals with cognitive, epigenetic, and polygenic score data. Linear regression was used to test for an association between the predicted epigenetic score (predictor) and measured general cognitive ability (outcome), in models adjusting for age, sex, and the polygenic score. Additional models included common risk factors for cognitive decline, frailty, and dementia: body mass index (BMI in kg/m^2); year of education; self-reported smoking (current, ever, never); self-reported weekly units of alcohol; self-reported high blood pressure (yes/no); lung function (forced expiratory volume in one second) adjusted for age, sex, and height; time taken to walk six metres (seconds); socioeconomic deprivation (Scottish Index of Multiple Deprivation in LBC1936 and social grades based on highest reached occupation [32] in LBC1921); and depression (HADS-D total from the Hospital Anxiety and Depression questionnaire) [33, 34, 35]. BMI, six metre walk time, and units of alcohol were log transformed to reduce skew – a constant of one was added to the alcohol units prior to transforming. The listwise sample size in the fully-adjusted models was 825 and 412 individuals in LBC1936 and LBC1921, respectively. Linear regression model assumptions were visually inspected via regression diagnostic plots.

References:

1. McCartney, Daniel L., Robert F. Hillary, Anna J. Stevenson, Stuart J. Ritchie, Rosie M. Walker, Qian Zhang, Stewart W. Morris, et al. 2018. “Epigenetic Prediction of Complex Traits and Death.” Genome Biology. https://doi.org/10.1186/s13059-018-1514-1.
2. Livingston, Gill, Andrew Sommerlad, Vasiliki Orgeta, Sergi G. Costafreda, Jonathan Huntley, David Ames, Clive Ballard, et al. 2017. “Dementia Prevention, Intervention, and Care.” The Lancet. https://doi.org/10.1016/S0140-6736(17)31363-6.
3. Lunnon, Katie, Rebecca Smith, Eilis Hannon, Philip L. De Jager, Gyan Srivastava, Manuela Volta, Claire Troakes, et al. 2014. “Methylomic Profiling Implicates Cortical Deregulation of ANK1 in Alzheimer’s Disease.” Nature Neuroscience. https://doi.org/10.1038/nn.3782.
4. Lunnon, Katie, Rebecca Smith, Eilis Hannon, Philip De Jager, Gyan Srivastava, Manuela Volta, Claire Troakes, et al. 2014. “Cross-Tissue Methylomic Profiling Strongly Implicates a Role for Cortex-Specific Deregulation of ANK1 in Alzheimer’s Disease Neuropathology.” Nat Neurosci.
5. Zhang, Futao, Wenhan Chen, Zhihong Zhu, Qian Zhang, Marta F. Nabais, Ting Qi, Ian J. Deary, et al. 2019. “OSCA: A Tool for Omic-Data-Based Complex Trait Analysis.” Genome Biology. https://doi.org/10.1186/s13059-019-1718-z.
6. Trejo Banos, Daniel, Daniel L. McCartney, Marion Patxot, Lucas Anchieri, Thomas Battram, Colette Christiansen, Ricardo Costeira, et al. 2020. “Bayesian Reassessment of the Epigenetic Architecture of Complex Traits.” Nature Communications. https://doi.org/10.1038/s41467-020-16520-1.
7. Marioni, Riccardo E., Allan F. McRae, Jan Bressler, Elena Colicino, Eilis Hannon, Shuo Li, Diddier Prada, et al. 2018. “Meta-Analysis of Epigenome-Wide Association

Studies of Cognitive Abilities." Molecular Psychiatry. https://doi.org/10.1038/s41380-017-0008-y.

8. McCartney, Daniel L, Futao Zhang, Robert F Hillary, Qian Zhang, Anna J Stevenson, Rosie M Walker, Mairead L Bermingham, et al. 2019. "An Epigenome-Wide Association Study of Sex-Specific Chronological Ageing." Genome Medicine 12 (1): 1. https://doi.org/10.1186/s13073-019-0693-z.
9. Smith, Blair H., Archie Campbell, Pamela Linksted, Bridie Fitzpatrick, Cathy Jackson, Shona M. Kerr, Ian J. Deary, et al. 2013. "Cohort Profile: Generation Scotland: Scottish Family Health Study (GS: SFHS). The Study, Its Participants and Their Potential for Genetic Research on Health and Illness." International Journal of Epidemiology. https://doi.org/10.1093/ije/dys084.
10. Smith, Blair H., Harry Campbell, Douglas Blackwood, John Connell, Mike Connor, Ian J. Deary, Anna F. Dominiczak, et al. 2006. "Generation Scotland: The Scottish Family Health Study; a New Resource for Researching Genes and Heritability." BMC Medical Genetics 7 (1): 74. https://doi.org/10.1186/1471-2350-7-74.
11. Bollepalli, Sailalitha, Tellervo Korhonen, Jaakko Kaprio, Simon Anders, and Miina Ollikainen. 2019. "EpiSmokEr: A Robust Classifier to Determine Smoking Status from DNA Methylation Data." Epigenomics 11 (13): 1469–86. https://doi.org/10.2217/epi-2019-0206.
12. Marioni, Riccardo E., Gail Davies, Caroline Hayward, Dave Liewald, Shona M. Kerr, Archie Campbell, Michelle Luciano, et al. 2014. "Molecular Genetic Contributions to Socioeconomic Status and Intelligence." Intelligence 44 (1): 26–32. https://doi.org/10.1016/j.intell.2014.02.006.
13. http://www.ewascatalog.org/

14. Roubroeks, Janou A.Y., Adam R. Smith, Rebecca G. Smith, Ehsan Pishva, Zina Ibrahim, Martina Sattlecker, Eilis J. Hannon, et al. 2020. "An Epigenome-Wide Association Study of Alzheimer's Disease Blood Highlights Robust DNA Hypermethylation in the HOXB6 Gene." Neurobiology of Aging 95 (November): 26–45. https://doi.org/10.1016/j.neurobiolaging.2020.06.023.
15. Daetwyler, Hans D., Beatriz Villanueva, and John A. Woolliams. 2008. "Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach." Edited by Michael Nicholas Weedon. PLoS ONE 3 (10): e3395. https://doi.org/10.1371/journal.pone.0003395.
16. Patxot, Marion, Daniel Trejo Banos, Athanasios Kousathanas, Etienne J Orliac, Sven E Ojavee, Gerhard Moser, Julia Sidorenko, et al. 2020. "Probabilistic Inference of the Genetic Architecture of Functional Enrichment of Complex Traits." MedRxiv, September, 2020.09.04.20188433. https://doi.org/10.1101/2020.09.04.20188433.
17. Madden, Rebecca A., Daniel L. McCartney, Rosie M. Walker, Robert F. Hillary, Mairead L. Bermingham, Konrad Rawlik, Stewart W. Morris, et al. 2020. "Birth Weight Associations with DNA Methylation Differences in an Adult Population." Epigenetics. https://doi.org/10.1080/15592294.2020.1827713.
18. Wechsler, D. 1997. "WMS-III Wechsler Memory Scale." Psychological Corporation.
19. Wechsler, D. 1997. WAIS--III: Wechsler Adult Intelligence Scale (3rd Ed.) Administration and Scoring Manual. The Psychological Corporation, San Antonio, TX.
20. Lezak, Muriel Deutsch. 1995. Neuropsychological Assessment (3rd Ed.). Neuropsychological Assessment (3rd Ed.).
21. Raven, J C, J Raven, and J H Court. 1988. The Mill Hill Vocabulary Scale. Manual for Raven's Progressive Matrices and Vocabulary Scales. Raven JC, Raven J, Court

JH. The Mill Hill vocabulary scale. Manual for Raven's progressive matrices and vocabulary scales. 1988.

22. Howard, David M., Lynsey S. Hall, Jonathan D. Hafferty, Yanni Zeng, Mark J. Adams, Toni Kim Clarke, David J. Porteous, et al. 2017. "Genome-Wide Haplotype-Based Association Analysis of Major Depressive Disorder in Generation Scotland and UK Biobank." *Translational Psychiatry* 7 (11): 1263. https://doi.org/10.1038/s41398-017-0010-9.
23. Deary, Ian J., Alan J. Gow, Alison Pattie, and John M. Starr. 2012. "Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936." International Journal of Epidemiology 41 (6): 1576–84. https://doi.org/10.1093/ije/dyr197.
24. Taylor, Adele M, Alison Pattie, and Ian J Deary. 2018. "Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936." International Journal of Epidemiology 47 (4): 1042-1042r. https://doi.org/10.1093/ije/dyy022.
25. Shah, Sonia, Allan F. McRae, Riccardo E. Marioni, Sarah E. Harris, Jude Gibson, Anjali K. Henders, Paul Redmond, et al. 2014. "Genetic and Environmental Exposures Constrain Epigenetic Drift over the Human Life Course." Genome Research 24 (11): 1725–33. https://doi.org/10.1101/gr.176933.114.
26. Marioni, Riccardo E., Sonia Shah, Allan F. McRae, Brian H. Chen, Elena Colicino, Sarah E. Harris, Jude Gibson, et al. 2015. "DNA Methylation Age of Blood Predicts All-Cause Mortality in Later Life." Genome Biology 16 (1). https://doi.org/10.1186/s13059-015-0584-6.
27. Nelson, Hazel E. & Willison, Jonathan (1991). The National Adult Reading Test (NART). Windsor: NFER-Nelson

28. Raven, J C, J E Raven, J H Court, and Oxford Psychologists Press. 1972. "Manual for Raven's Progressive Matrices and Vocabulary Scales." Oxford: Oxford Psychologists Press.

29. Davies, Gail, Max Lam, Sarah E. Harris, Joey W. Trampush, Michelle Luciano, W. David Hill, Saskia P. Hagenaars, et al. 2018. "Study of 300,486 Individuals Identifies 148 Independent Genetic Loci Influencing General Cognitive Function." Nature Communications 9 (1): 1–16. https://doi.org/10.1038/s41467-018-04362-x.

30. Euesden, Jack, Cathryn M. Lewis, and Paul F. O'Reilly. 2015. "PRSice: Polygenic Risk Score Software." Bioinformatics 31 (9): 1466–68. https://doi.org/10.1093/bioinformatics/btu848.

31. Ritchie, Stuart J., W. David Hill, Riccardo E. Marioni, Gail Davies, Saskia P. Hagenaars, Sarah E. Harris, Simon R. Cox, et al. 2020. "Polygenic Predictors of Age-Related Decline in Cognitive Ability." Molecular Psychiatry 25 (10): 2584–98. https://doi.org/10.1038/s41380-019-0372-x.

32. General Register Office. Census 1951: classification of occupations. London, UK: Her Majesty's Stationary Office, 1956.

33. Zigmond, A. S., and R. P. Snaith. 1983. "The Hospital Anxiety and Depression Scale." *Acta Psychiatrica Scandinavica* 67 (6): 361–70. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x.

34. https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/?utm_source=redirect&utm_medium=shorturl&utm_campaign=simd

35. Deary, Ian J., Alan J. Gow, Michelle D. Taylor, Janie Corley, Caroline Brett, Valerie Wilson, Harry Campbell, et al. 2007. "The Lothian Birth Cohort 1936: A Study to Examine Influences on Cognitive Ageing from Age 11 to Age 70 and Beyond." BMC Geriatrics 7 (1): 28. https://doi.org/10.1186/1471-2318-7-28.

Figure 1: Proportion of variance accounted for that is attributed to small, medium and large effect DNA methylation (DNAm) probes for each trait (upper figure), and proportion of variance accounted for by DNAm probes (green column), genetic single nucleotide polymorphism (SNP) variants (orange column) and both DNAm and SNP data (purple column) for six cognitive traits. Error bars represent 95% credible intervals.

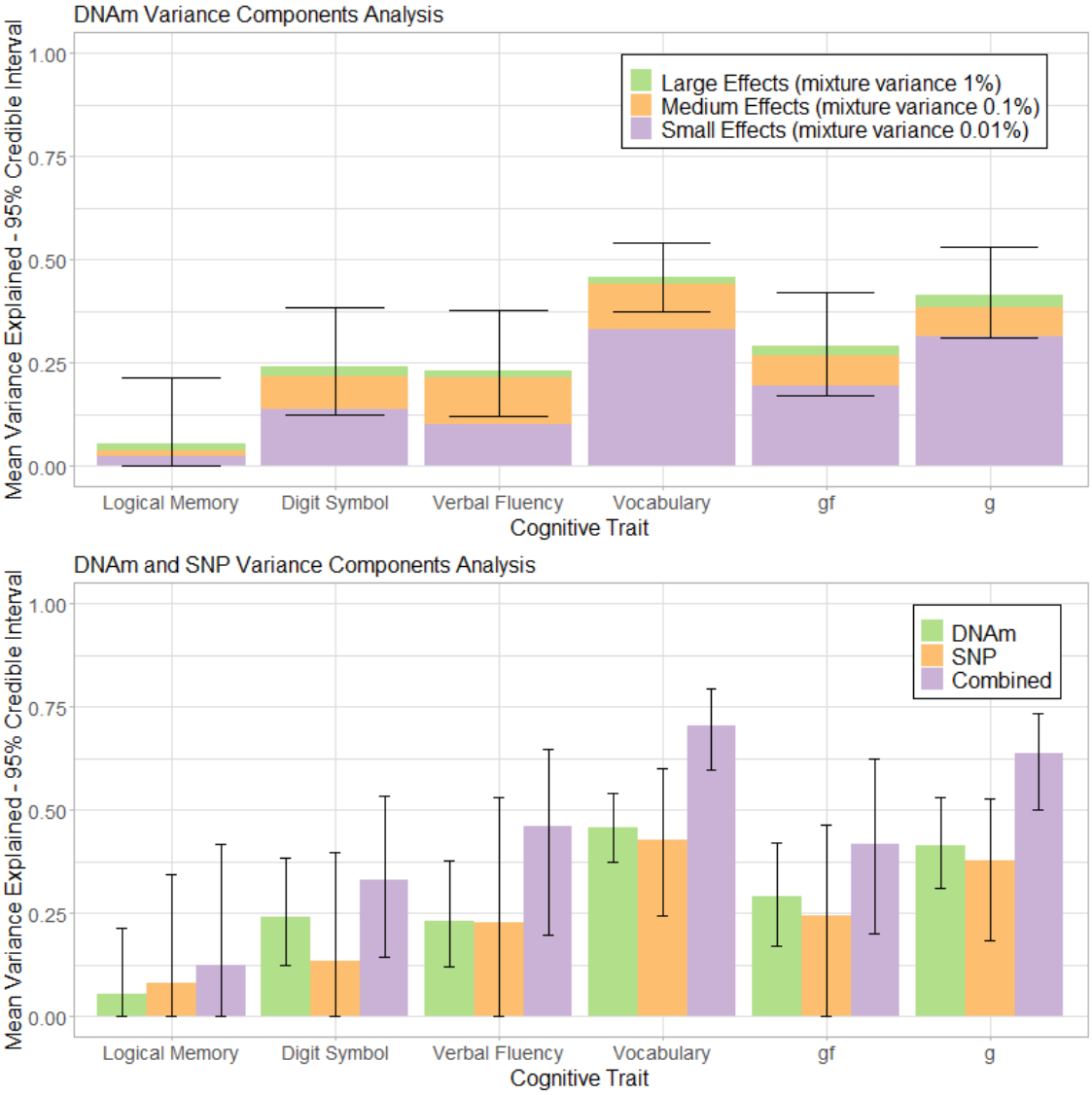
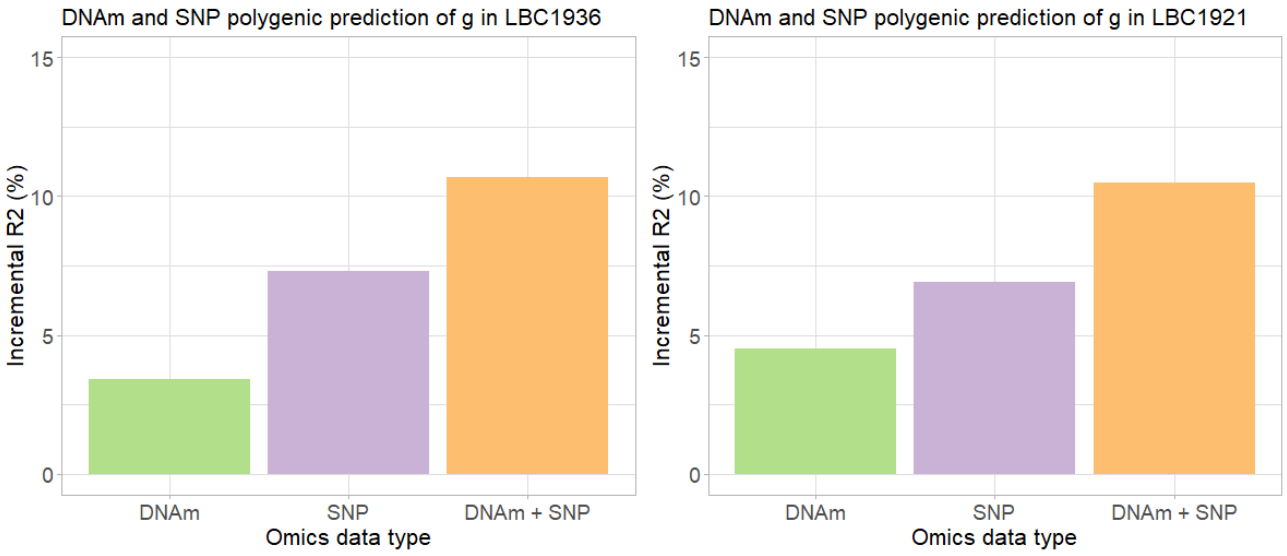


Figure 2: Genetic and Epigenetic prediction of general cognitive ability in the Lothian Birth Cohort 1936 (LBC1936) and the Lothian Birth Cohort 1921 (LBC1921). DNAm = DNA methylation (epigenetic predictor derived from BayesR+ weights); SNP = Single Nucleotide Polymorphism (polygenic predictor derived from GWAS weights). Incremental R^2 is the difference in R^2 between a linear regression model adjusting for age and sex plus the omics-data derived score/s. The weights used to derive the DNAm score were derived from Generation Scotland (N=9,162) whereas those used to derive the polygenic (SNP) score were derived from UK Biobank (N=168,033).



Supplementary Table 1: Generation Scotland cohort summary

	N	Mean	SD
Age (years)	9162	49.8	13.6
Digit Symbol	9162	71.2	16.5
Verbal Fluency	9162	40.0	11.5
Vocabulary	9162	30.5	4.5
Logical Memory	9162	31.1	7.9
Body Mass Index (kg/m^2)	9162	26.9	5.2
EpiSmoker*	9162	2.09	4.5
	N	%	
Sex			
Male	3752	41.0	
Female	5410	59.0	

* A plot of the epigenetic smoker (EpiSmoker) variable against self-reported smoking (current, former, never) is provided in **Supplementary Figure 1**.

Supplementary Table 2: Mean variance accounted for by the effects of DNA methylation (DNAm) and genome-wide DNA single nucleotide polymorphisms (GWAS) alone, DNAm data conditioned on GWAS data (DNAmAdjGWAS), and the additive effects of DNAm and GWAS data for six cognitive traits.

	DNAm			**GWAS**			**DNAmAdjGWAS**			**DNAm+GWAS**		
	Mean	**2.5%ile**	**97.5%ile**	**Mean**	**2.5%ile**	**97.5%ile**	**Mean**	**2.5%ile**	**97.5%ile**	**Mean**	**2.5%ile**	**97.5%ile**
Digit Symbol	0.242	0.125	0.385	0.134	0.000	0.396	0.220	0.115	0.338	0.330	0.145	0.533
Logical Memory	0.053	0.000	0.216	0.081	0.000	0.344	0.045	0.000	0.196	0.124	0.000	0.419
Verbal Fluency	0.230	0.121	0.378	0.227	0.000	0.532	0.192	0.078	0.390	0.461	0.198	0.648
Vocabulary	0.458	0.373	0.541	0.429	0.244	0.601	0.415	0.294	0.533	0.704	0.599	0.796
g_f	0.292	0.173	0.420	0.243	0.000	0.466	0.253	0.151	0.373	0.418	0.200	0.626
g	0.416	0.310	0.530	0.379	0.183	0.529	0.350	0.248	0.467	0.638	0.500	0.735

Supplementary Table 3: Comparison of epigenetic variance components estimates between BayesR+ and a linear mixed model approach, OSCA. OSCA[1] analyses included genetic and epigenetics relationship matrices on all 9,162 participants; OSCA[2] analyses included an epigenetics relationship matrix on a genetically unrelated subset of 4,261 participants who were processed in a single methylation batch ("Set 2" – **Online Methods**).

	DNAm variance component estimate		
	BayesR+ (mean)	**OSCA[1] (mean, SE)**	**OSCA[2] (mean, SE)**
Logical Memory	0.053	0.062 (0.03)	0.142 (0.08)
Digit Symbol	0.242	0.257 (0.04)	0.231 (0.07)
Verbal Fluency	0.230	0.254 (0.04)	0.168 (0.07)
Vocabulary	0.458	0.484 (0.03)	0.604 (0.06)
g_f	0.292	0.313 (0.04)	0.417 (0.08)
g	0.416	0.434 (0.03)	0.584 (0.07)

Supplementary Table 4: Contribution of mixtures with small, medium, and large effect sizes (variances of 0.01%, 0.1%, and 1%, respectively) to the mean variance accounted for by the effects of DNA methylation (DNAm) for six cognitive traits.

	DNAm effect sizes			
	Small	**Medium**	**Large**	**Total**
Logical Memory	0.025	0.014	0.014	0.053
Digit Symbol	0.138	0.079	0.025	0.242
Verbal Fluency	0.101	0.114	0.016	0.230
Vocabulary	0.330	0.112	0.016	0.458
g_f	0.196	0.073	0.023	0.292
g	0.313	0.073	0.030	0.416

Supplementary Table 5: List of blood-based CpG sites that are with strong evidence (PIP>0.8) for association with of cognitive function in Generation Scotland. PIP: Posterior Inclusion Probability.

For the EWAS, CpGs within 2.5kb and highly correlated (absolute Pearson correlation >0.5) with a lead CpG with posterior inclusion probability greater than 0.2 were grouped together. For each probe group (with number of CpGs – **nCpG**), we calculated the proportion of iterations for which at least one probe was included in the model, yielding the group posterior inclusion probability (**Group PIP**). We then calculated the sum of the squared regression coefficients at each iteration for the probe group to give the contribution of the group to the total variance (**Group R^2**). Finally, we highlighted the lead CpG for the groups.

Cognitive Phenotype	Group PIP	Group R^2	Lead CpG	Nearest Gene(s)	nCpG	Lead PIP	Lead Median Beta	Lead Beta2.5	Lead Beta97.5
Digit Symbol	1.000	0.009	cg13108341	*DNAH9*	1	1.000	-0.092	-0.124	-0.059
g	1.013	0.005	cg09984392	*SQLE*	4	0.965	0.072	0.000	0.104
g	1.000	0.009	cg13108341	*DNAH9*	1	1.000	-0.092	-0.124	-0.054
g	0.952	0.005	cg05325763	*CPT1A*	5	0.870	-0.070	-0.108	0.000
g_f	0.998	0.007	cg13108341	*DNAH9*	1	0.998	-0.081	-0.114	-0.043
g_f	0.846	0.003	cg09984392	*SQLE*	4	0.714	0.048	0.000	0.087
Verbal Fluency	0.879	0.004	cg13108341	*DNAH9*	1	0.879	-0.062	-0.096	0.000
Vocabulary	1.041	0.006	cg05325763	*CPT1A*	5	0.962	-0.075	-0.111	0.000

Supplementary Table 6: Epigenome-wide association study (EWAS) catalog lookup of three CpGs with strong (PIP>0.95) associations with cognitive traits. Lookup is restricted to associations at $P<3.6x10^{-8}$. * denotes multiple models in study, where summary statistics are presented for the model with the most significant association.

CpG	Max. PIP	Cognitive Traits	First Author	Study ID	EWAS Trait	Beta	SE	P
cg24503796	0.97	g, g_f	Mulder RH	33450751	Age*	0.002	$9.3x10^{-5}$	0
cg24503796	0.97	g, g_f	Battram T	30602389	Tissue type	-0.168	0.005	$6.3x10^{-90}$
cg24503796	0.97	g, g_f	Wozniak MB	23526956	Clear cell renal carcinoma	NA	NA	$3.8x10^{-27}$
cg24503796	0.97	g, g_f	Petersen A-K	24014485	M15964_arabitol	34.500	0.219	$2.0x10^{-12}$
cg24503796	0.97	g, g_f	Gomez-Alonso MDC	33413638	Docosahexaenoic acid	1.690	0.26	$4.1x10^{-11}$
cg24503796	0.97	g, g_f	Petersen A-K	24014485	1-linoleoylglycerophosphoethanolamine	4.210	0.109	$7.0x10^{-11}$
cg24503796	0.97	g, g_f	Battram T	23177740	Age	0.001	$1.1x10^{-4}$	$3.6x10^{-10}$
cg13108341	1.00	Digit Symbol, g	Bermingham M	30935889	Forced Vital Capacity	-0.032	NA	$2.6x10^{-8}$
cg09984392	0.97	g	Mulder RH	33450751	Age*	0.0031	0.00011	0
cg09984392	0.97	g	Battram T	30602389	Tissue	-0.0694	0.005	$1.2x10^{-28}$
cg09984392	0.97	g	Petersen A-K	24014485	1-linoleoylglycerophosphoethanolamine*	4.26	0.11	$6.50x10^{-11}$
cg09984392	0.97	g	Singmann P	26500701	Sex	NA	NA	$1.52x10^{-8}$

Supplementary Table 7: Lookup of CpGs associated with cognitive abilities and Alzheimer's disease. For each CpG, we report the cognitive phenotype with the maximum PIP in the current study. PIP: Posterior Inclusion Probability.

CpG	Maximum PIP	Cognitive Phenotype	Study
cg13344806	0.018	Vocabulary	Lunnon et al.
cg05667256	0.017	g	Lunnon et al.
cg03048889	0.017	Verbal Fluency	Lunnon et al.
cg17074958	0.016	g	Lunnon et al.
cg05242218	0.015	Vocabulary	Lunnon et al.
cg14067233	0.014	g	Lunnon et al.
cg07401045	0.011	g	Lunnon et al.
cg01986016	0.019	g_f	Roubroeks et al.
cg10906729	0.017	g	Roubroeks et al.
cg05888755	0.017	g	Roubroeks et al.
cg00072689	0.016	g_f	Roubroeks et al.
cg20471691	0.016	Vocabulary	Roubroeks et al.
cg15435170	0.015	g_f	Roubroeks et al.
cg18684142	0.013	Logical Memory	Roubroeks et al.
cg17179862	0.011	Digit Symbol	Roubroeks et al.
cg19828220	0.011	g	Roubroeks et al.
cg03803541	0.01	Verbal Fluency	Roubroeks et al.
cg16848873	0.009	Vocabulary	Roubroeks et al.
cg18181703	0.047	Vocabulary	Marioni et al.
cg04513006	0.024	g	Marioni et al.
cg03366574	0.017	Vocabulary	Marioni et al.
cg16201957	0.016	g	Marioni et al.
cg21450381	0.016	g_f	Marioni et al.
cg12507869	0.012	g	Marioni et al.

Supplementary Table 8: Multivariate linear regression of epigenetic cognitive scores on ageing phenotypes in the Lothian Birth Cohort 1936 (n=825) and the Lothian Birth Cohort 1921 (n=412). The outcome and all continuous variables were standardised to mean 0, variance 1.

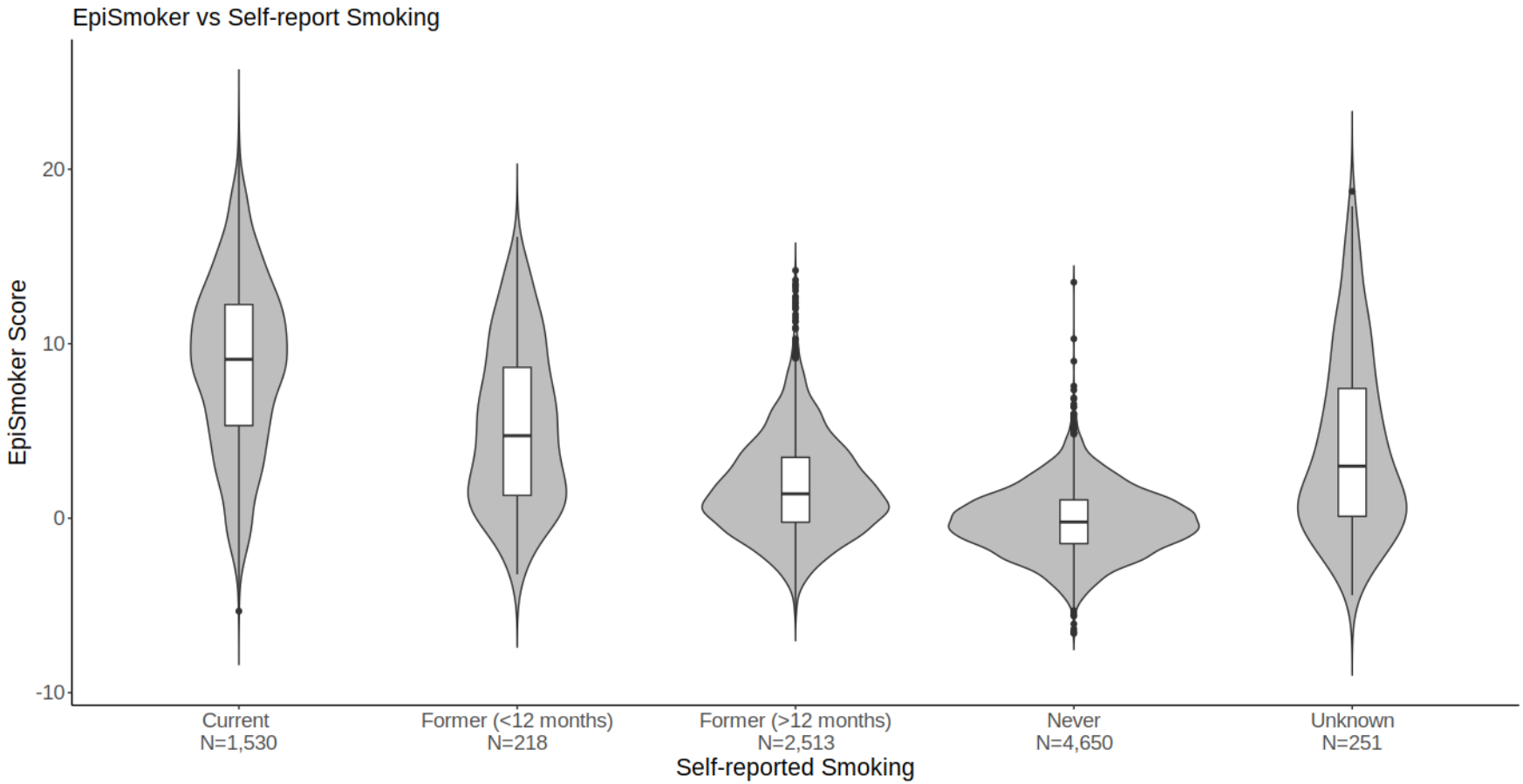
	LBC1936			**LBC1921**		
	Beta	**SE**	**P**	**Beta**	**SE**	**P**
Intercept	-0.02	0.08	0.78	-0.06	0.11	0.58
Age	0.06	0.04	0.086	-0.03	0.05	0.52
Sex (Male)	0.07	0.11	0.53	-0.22	0.16	0.16
Years of education	0.06	0.04	0.10	0.03	0.06	0.62
BMI[a]	-0.08	0.04	0.042	-0.05	0.05	0.35
Ex-smoker (vs Never)	-0.03	0.08	0.73	0.25	0.10	0.015
Current smoker (vs Never)	-0.06	0.12	0.64	0.17	0.21	0.42
Alcohol units per week[a]	0.08	0.04	0.023	0.07	0.05	0.15
High Blood Pressure (Yes)	0.01	0.07	0.94	0.02	0.10	0.83
HADS Depression Score	-0.08	0.04	0.034	-0.03	0.05	0.57
Deprivation[b]	0.08	0.04	0.049	-0.07	0.06	0.20
Walking speed[a]	0.00	0.04	0.99	-0.07	0.06	0.21
Grip strength	0.09	0.06	0.11	0.19	0.08	0.018
FEV[c]	0.01	0.04	0.88	0.05	0.05	0.29
Multiple R^2 (age + sex)	0.015			0.012		
Multiple R^2 (all covariates)	0.064			0.087		

[a] Natural log transformed – constant of one added to units alcohol per week

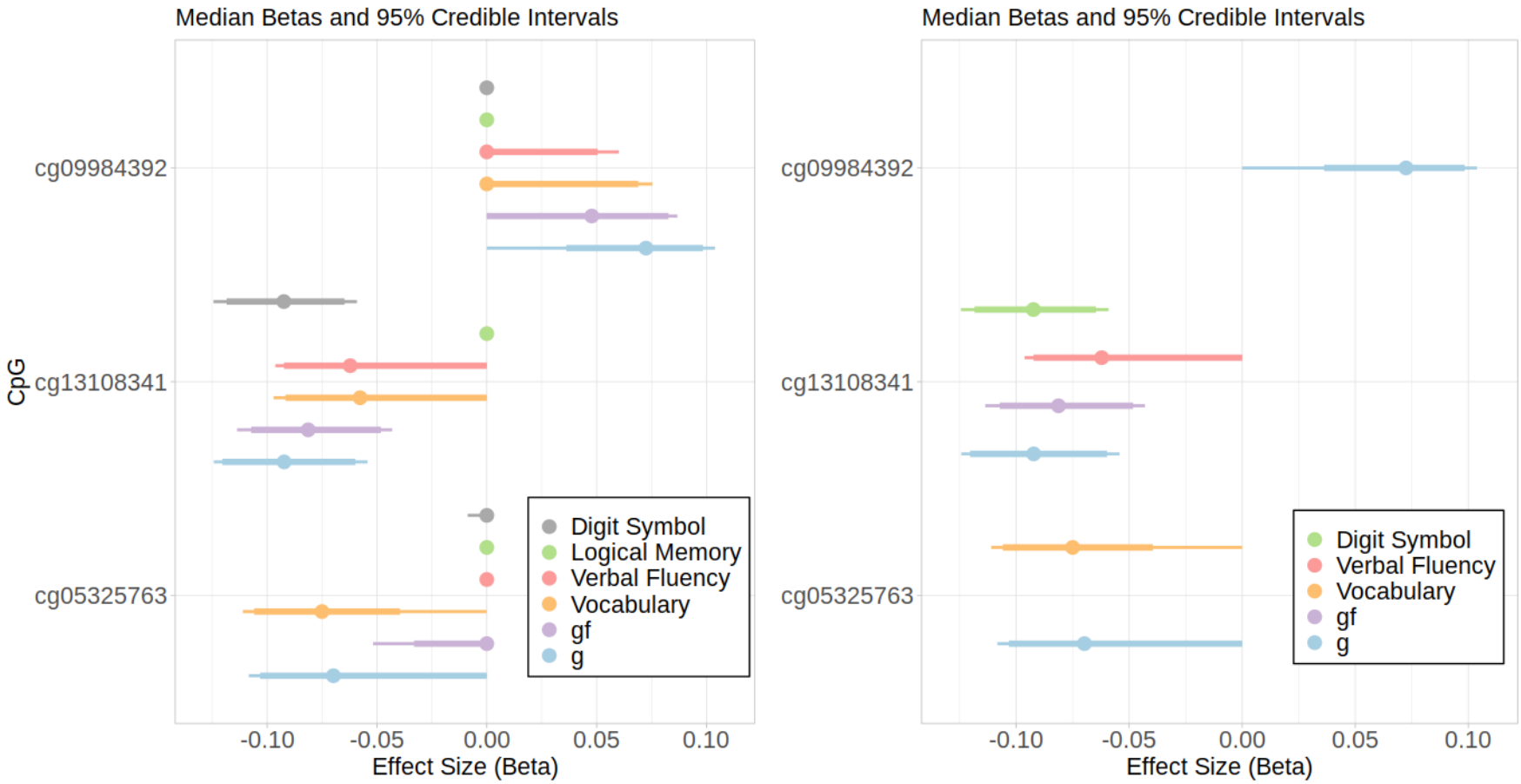
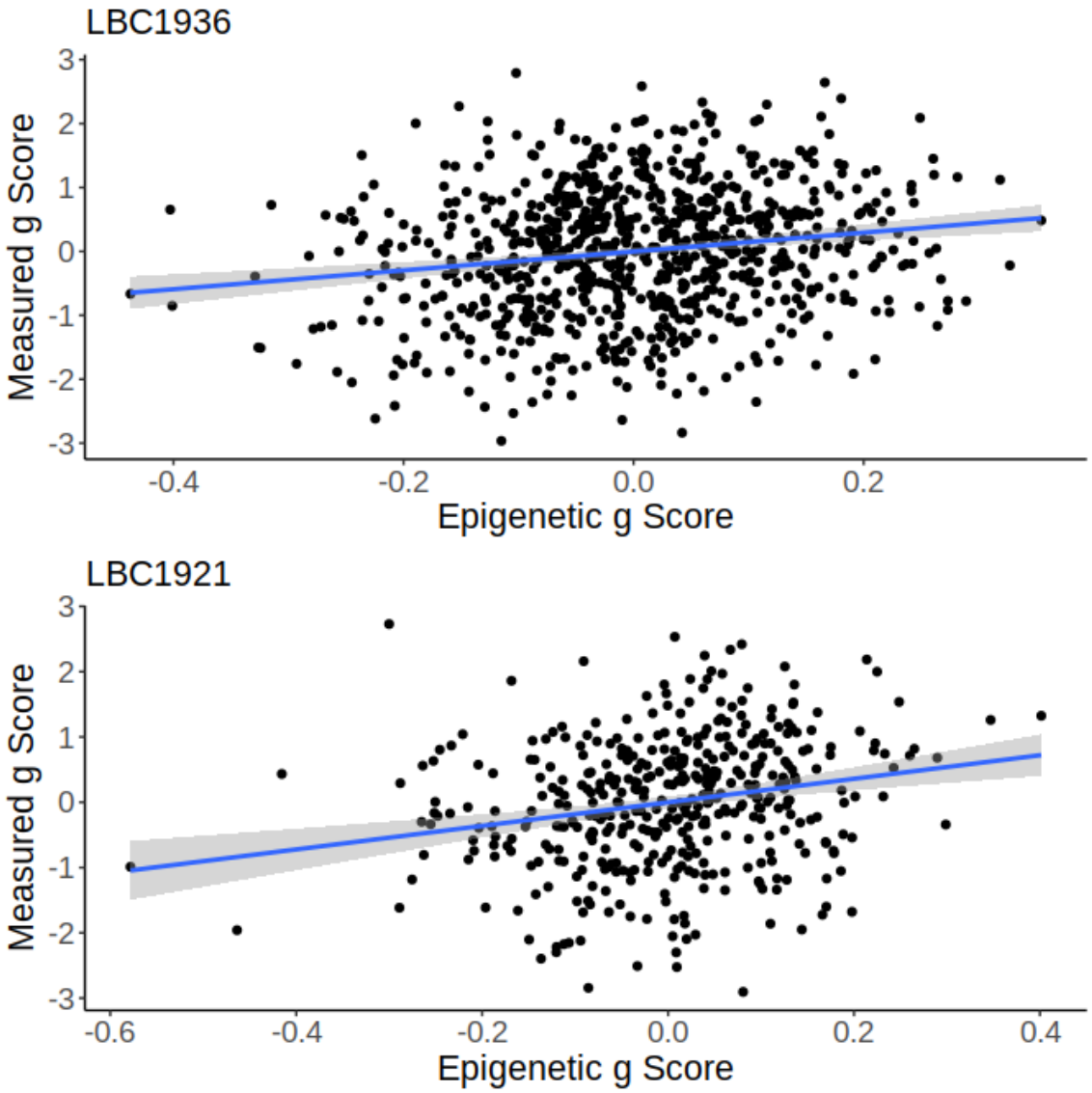
[b] Scottish Index of Multiple Deprivation in LBC1936 – higher scores indicate less deprived areas as measured across seven domains. Highest occupation-based social class coding in LBC1921 – higher scores indicate a lower social class grouping.

[c] Residuals from a regression of forced expiratory volume in one second (FEV_1) on age, sex, and height

Supplementary Figure 1: Plot of the epigenetic smoking variable (EpiSmoker) against self-reported smoking status (current, ever, never) in Generation Scotland (n=9,162).



Supplementary Figure 2: Median effects observed for DNA methylation probes with posterior inclusion probabilities (PIPs) > 0.8 across four cognitive tests and two composite measures (no probes identified for logical memory). Thick horizontal lines represent the 5th and 95th percentiles; thin lines represent 2.5th and 97.5th percentiles. Left panel displays summaries for all traits, right panel displays summaries for traits where PIP > 0.8.



Supplementary Figure 3. Scatter Plot of Epigenetic g Score by Measured g Score in the Lothian Birth Cohort 1936 (LBC1936) and the Lothian Birth Cohort 1921 (LBC1921).