Nature Human Behavior

SUPPLEMENTARY INFROMATION
Epigenome-wide association study of left-handedness for different tissues and ages

Veronika V. Odintsova, Matthew Suderman, Fiona A. Hagenbeek, Doretta Caramaschi, Jouke-Jan Hottenga, René Pool, BIOS consortium, Conor V. Dolan, Lannie Ligthart, Catharina E.M. van Beijsterveld, Gonneke Willemsen, Eco J.C. de Geus, Jeffrey J. Beck, Erik A. Ehli, Gabriel Cuellar-Partida, David M. Evans, Sara E. Medland, Caroline L. Relton, Dorret I. Boomsma, Jenny van Dongen

Content

Appendix 1. BIOS consortium authors

BIOS Consortium (Biobank-based Integrative Omics Study)

Management Team Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].
Cohort collection Dorret I. Boomsma[7], René Pool[7], Jenny van Dongen[7], Jouke J. Hottenga[7] (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Sasha Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).
Data Generation Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].
Data management and computational infrastructure Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Dasha V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].
Data Analysis Group Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Dasha V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter-Bram 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].
1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands
4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands
5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands
9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands
10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands
12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands
13. SURFsara, Amsterdam, the Netherlands
14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

Appendix 2. EWAS Model Equations

Primary and secondary analysis EWAS

The following models were fitted in each cohort in primary and secondary analysis. NTR: GEE. ALSPAC: linear regression.

Basic model:

NTR

For DNA methylation in peripheral blood in adults

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{Neu} \times Neu + \beta_{Eos} \times Eos + \beta_{Mono} \times Mono + \beta_{arrayrow} \times array\ row + \beta_{sampleplate2} \times sample\ plate\ 2 \ldots + \ldots \beta_{sampleplate\ N} \times sample\ plate\ N + \varepsilon$$

For DNA methylation in buccal cells in children

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{Epi} \times Epi + \beta_{NK} \times NK + \beta_{arrayrow} \times array\ row + \beta_{sampleplate2} \times sample\ plate\ 2 \ldots + \ldots \beta_{sampleplate\ N} \times sample\ plate\ N + \varepsilon$$

ALSPAC

For DNA methylation in peripheral blood in children and adults

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{B\ lym} \times B\ lym + \beta_{CD4T} \times CD4T + \beta_{CD8T} \times CD8T + \beta_{NK} \times NK + \beta_{NK} \times NK + \beta_{Mono} \times Mono + \beta_{Gran} \times Gran + \beta_{surrogate\ variable2} \times surrogate\ variable\ 2 + \ldots. \beta_{surrogate\ variableN} \times surrogate\ variable\ N + \varepsilon$$

For DNA methylation in cord blood

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{B\ lym} \times B\ lym + \beta_{CD4T} \times CD4T + \beta_{CD8T} \times CD8T + \beta_{NK} \times NK + \beta_{NK} \times NK + \beta_{Mono} \times Mono + \beta_{Gran} \times Gran + \beta_{nRBC} \times nRBC + \beta_{surrogate\ variable2} \times surrogate\ variable\ 2 + \ldots. \beta_{surrogate\ variableN} \times surrogate\ variable\ N + \varepsilon$$

Adjusted model:

NTR

For DNA methylation in peripheral blood in adults

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{BMI} \times BMI + \beta_{smoking} \times Smoking\ status + \beta_{Neu} \times Neu + \beta_{Eos} \times Eos + \beta_{Mono} \times Mono + \beta_{arrayrow} \times array\ row + \beta_{sampleplate2} \times sample\ plate\ 2 \ldots + \ldots \beta_{sampleplate\ N} \times sample\ plate\ N + \varepsilon$$

For DNA methylation in buccal cells in children

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{gestational\ age} \times gestational\ age + \beta_{birth\ weight} \times Birth\ Weight + \beta_{maternal\ smoking} \times Maternal\ Smoking + \beta_{Epi} \times Epi + \beta_{NK} \times NK + \beta_{arrayrow} \times array\ row + \beta_{sampleplate2} \times sample\ plate\ 2 \ldots + \ldots \beta_{sampleplate\ N} \times sample\ plate\ N + \varepsilon$$

ALSPAC

For DNA methylation in peripheral blood in adults (from 16 years old)

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{BMI} \times BMI + \beta_{smoking} \times Smoking\ status + \beta_{B\ lym} \times B\ lym + \beta_{CD4T} \times CD4T + \beta_{CD8T} \times CD8T + \beta_{NK} \times NK + \beta_{NK} \times NK + \beta_{Mono} \times Mono + \beta_{Gran} \times Gran + \beta_{surrogate\ variable2} \times surrogate\ variable\ 2 + \ldots. \beta_{surrogate\ variableN} \times surrogate\ variable\ N + \varepsilon$$

For DNA methylation in cord blood

$$CpGi = \alpha + \beta_{handedness} \times handedness + \beta_{sex} \times sex + \beta_{gestational\ age} \times gestational\ age + \beta_{birth\ weight} \times Birth\ Weight + \beta_{maternal\ smoking} \times Maternal\ Smoking + \beta_{B\ lym} \times B\ lym + \beta_{CD4T} \times CD4T + \beta_{CD8T} \times CD8T + \beta_{NK} \times$$

$NK + ß_{NK}\ x\ NK + ß_{Mono}\ x\ Mono + ß_{Gran}\ x\ Gran + ß_{nRBC}\ x\ nRBC + ß_{surrogate\ variable2}\ x\ surrogate\ variable\ 2 + \ldots. ß_{surrogate\ variableN}\ x\ surrogate\ variable\ N + \varepsilon$

For DNA methylation in peripheral blood in children

$CpGi = \alpha + ß_{handedness}\ x\ handedness + ß_{sex}\ x\ sex + ß_{age}\ x\ age + ß_{gestational\ age}\ x\ gestational\ age + ß_{birth\ weight}\ x\ Birth\ Weight + ß_{maternal\ smoking}\ x\ Maternal\ Smoking + ß_{B\ lym}\ x\ B\ lym + ß_{CD4T}\ x\ CD4T + ß_{CD8T}\ x\ CD8T + ß_{NK}\ x\ NK + ß_{Gran}\ x\ Gran + ß_{NK}\ x\ NK + ß_{Mono}\ x\ Mono + ß_{nRBC}\ x\ nRBC + ß_{surrogate\ variable2}\ x\ surrogate\ variable\ 2 + \ldots. ß_{surrogate\ variableN}\ x\ surrogate\ variable\ N + \varepsilon$

where CpGi = DNA methylation β-value at methylation site i, α = the intercept, *handedness* is coded as 0=left-handed, and 1=righthanded; sex is coded 0 for males and 1 for females, age = the age at DNA methylation measurement in years, BMI = body mass index, smoking = smoking status (0=no, 1=former smoking, 2=current smoking), maternal smoking (0=not smoked, 1=smoked), Mon = percentage of monocytes, Eos = percentage of eosinophils, Neu = percentage of neutrophils, Blym = percentage of B lymphocytes, CD4T = percentage of CD4 + T-lymphocytes, CD8T = percentage of CD8 + T-lymphocytes, NK = percentage/proportion of natural killer cells, Gran = percentage of granulocytes, nRBC=nucleated red blood cells, Epi = percentage/proportion of epithelial cells, arrow row = the row of the sample on the Illumina 450k (ranging from 1 to 6) or EPIC Beadchip (ranging from 1 to 8), sample plate = bisulfite plate (dummy-coding) in NTR, surrogate variable in ALSPAC (n=20), and ε is residual.

Secondary analysis

GWAS follow-up

Differences between t-statistics of CpGs located in a 1 Mb window of SNPs derived from the GWAS of handedness and all other CpGs were tested with a linear regression model:

$t = \alpha + ß_{handCpGs\_w1Mb}\ x\ handCpGs\_w1Mb + \varepsilon$

where t = t-statistic, α = the intercept, handCpGs_w1Mb = variable indicating if the CpG was located in a 1 Mb window of SNPs derived from the GWAS of handedness (0=no/1=yes)

Polygenic and methylation scores testing

We tested whether a methylation score (MS) adds predictive value for handedness over and above the polygenic score (PGS), and calculated the variance in handedness that is explained by the PGS and MS. We calculated the variance explained on the liability scale (Lee et al[1]). To this end we ran five logistic regression models with the R function glm:

Prediction by PGS:

Model 1: PGS and GWAS covariates*

$handedness = \alpha + ß_{PGS}\ x\ PGS + ß_{age}\ x\ age + GWAS\ covariates$

Model 2: genotype covariates

$handedness = \alpha + ß_{age}\ x\ age + GWAS\ covariates$

Prediction by PGS and MS:

Model 3: PGS, MS, and GWAS and EWAS covariates**

$handedness = \alpha + ß_{PGS}\ x\ PGS + ß_{age}\ x\ age + ß_{sex}\ x\ sex + ß_{PGS}\ x\ MS + GWAS\ covariates + EWAS\ covariates$

Model 4: with PGS, GWAS and EWAS covariates

$handedness = \alpha + ß_{PGS}\ x\ PGS + ß_{age}\ x\ age + ß_{sex}\ x\ sex + GWAS\ covariates + EWAS\ covariates$

Model 5: with MS, GWAS and EWAS covariates

$handedness = \alpha + ß_{PGS}\ x\ MS + ß_{age}\ x\ age + ß_{sex}\ x\ sex + GWAS\ covariates + EWAS\ covariates$

* GWAS covariates included dummy variables for platforms and 10 principal components based on genotype data in NTR and 10 principal components in ALSPAC.
** EWAS covariates included BMI, smoking (for adults), gestational age, birth weight, maternal smoking (for children), percentage/proportions of cells, bisulfite plate (dummy-coding) in NTR, surrogate variable in ALSPAC (n=20).

Model 3 was run for three MS with different sets of CpGs included by varying the p-value threshold (p-value $<1x10^{-1}$, $<1x10^{-3}$, $<1x10^{-5}$).

Calculation of R^2 (based on Lee et al[1])

Residual (homoskedastic) variance:

$$Res.Var = \pi^2 / 3$$

Explained variance whole model:

$$Ex.Var = var(\beta_1 * predictor_1 + \beta_2 * predictor_2 + \ldots + \beta N * predictorN)$$

Proportion of explained variance in total variance:

$$R^2 = Ex.Var / (Ex.Var + Res.Var)$$

where β_N=regression coefficient of the N^{th} predictor in the model.

Explained variance for PGS:

Exp.Var. PGS = Total Exp.Var Model 1 – Total Exp.Var.Model 2

Explained variance for PGS in the combined model:

Exp.Var. PGS = Total Exp.Var Model 3 – Total Exp.Var.Model 5

Explained variance for MS in the combined model:

Exp.Var. MS = Total Exp.Var Model 3 – Total Exp.Var.Model 4

Reference

1. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

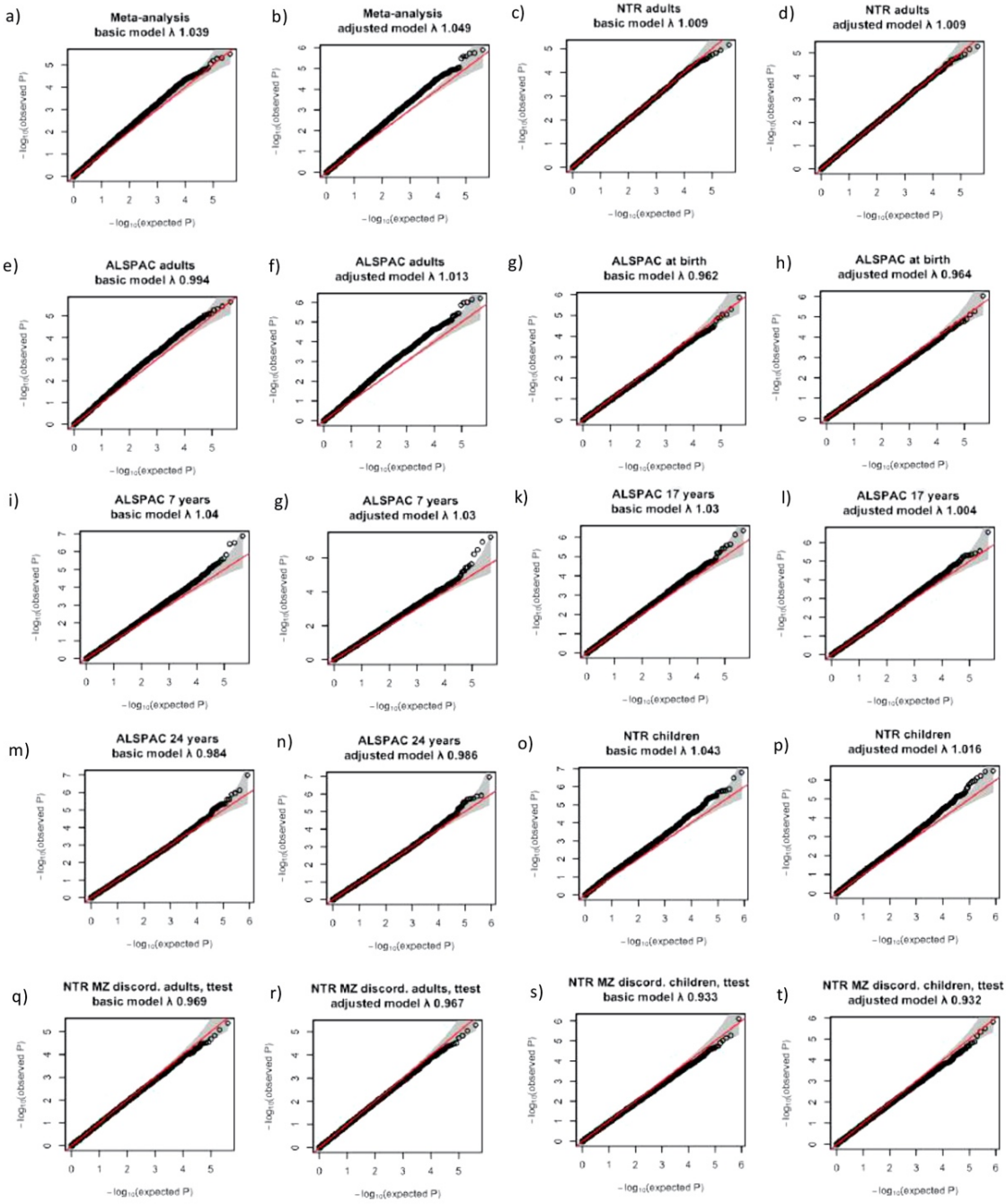


Figure S1. QQ plots of EWAS results on left-handedness

Quantile-quantile (QQ) plot from the EWAS of handedness in different analyses: a) Meta-analysis (NTR adults and ALSPAC adults) basic model; b) Meta-analysis (NTR adults and ALSPAC adults) adjusted model; c) NTR adults basic model; d) NTR adults adjusted model; e) ALSPAC adults basic model; f) ALSPAC adults adjusted model; g) ALSPAC at birth basic model; h) ALSPAC at birth adjusted model; i) ALSPAC 7 years old basic model; j) ALSPAC 7 years old adjusted model; k) ALSPAC 17 years old basic model; l) ALSPAC 17 years old adjusted model; m) ALSPAC 24 years old basic model; n) ALSPAC 24 years old adjusted model; o) NTR children basic model; p) NTR children adjusted model; q) MZ within-pair analysis, NTR adults basic model; r) MZ within-pair analysis NTR adults adjusted model; s) MZ within-pair analysis, NTR children basic model; t) MZ within-pair analysis, NTR children adjusted model.

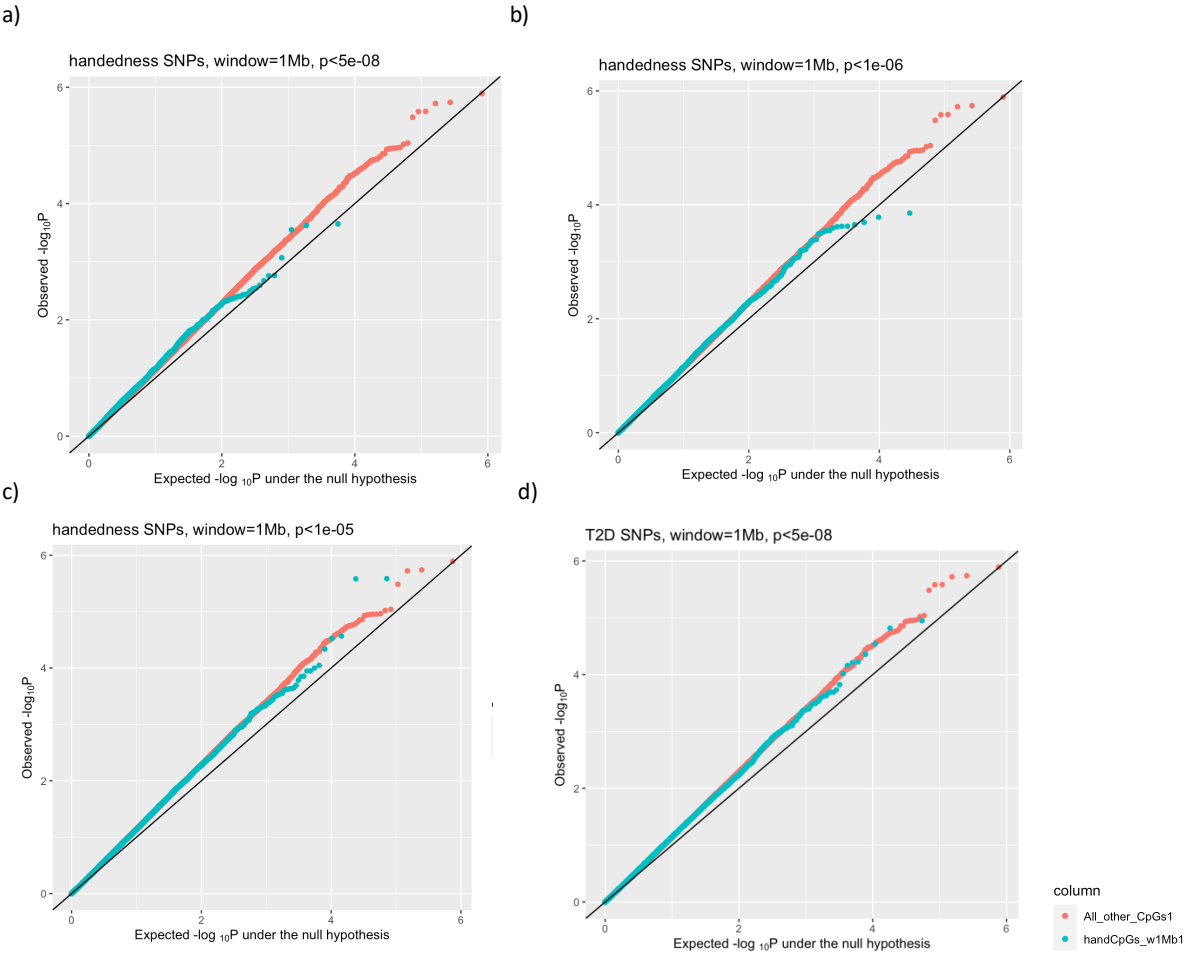
The observed p-values (y-axis) are plotted against the p-values expected under the null hypothesis (x-axis). The straight diagonal line denotes the pattern expected under the null hypothesis, with 95% confidence intervals indicated by the shaded grey line. λ = Bayesian estimate of inflation

Table S13. GWAS follow-up results

SNP fraction	N SNPs	N CpGs located within 1Mb window	Effect	Bootstrap SE	Bootstrap p
LH $p<5x10^{-08}$	420	2,784	0.027	0.013	0.040
LH $p<1x10^{-06}$	2,625	14,631	0.011	0.006	0.048
LH $p<1x10^{-05}$	3,464	35,975	0.003	0.004	0.472
T2D $p<5x10^{-08}$	2,392	27,229	0.005	0.004	0.265

LH, left-handedness; T2D, type 2 diabetes.; Effect, ß (regression coefficient); Bootstrap SE, standard error computed with bootstraps; Bootstrap p, p-value computed with bootstrap SE.
Note: Results of linear regression of absolute z-scores of CpGs on a variable (yes/no) indicating if CpGs were located within 1Mb from SNPs associated with the trait. The analysis was performed using EWAS summary statistics from the meta-analysis. GWAS summary statistics were obtained for left-handedness from Cuellar-Partida et al. (2020) and for type 2 diabetes from Watanabe et al, 2019 (available at GWAS atlas https://atlas.ctglab.nl/traitDB/3686; 41204_E11_logistic.EUR.sumstats.MACfilt.txt).

Figure S2a-d. QQ plots of p-values of CpGs located within 1Mb window from GWAS SNPs associated with left-handedness and type 2 diabetes



a) QQ-plot of p-values of CpGs located within 1Mb window from handedness GWAS SNPs associated with handedness at $p<5x10^{-08}$; b) QQ-plot of p-values of CpGs located within 1Mb window from handedness GWAS SNPs associated with handedness at $p<1x10^{-06}$; c) QQ-plot of p-values of CpGs located within 1Mb window from handedness GWAS SNPs associated with handedness at $p<1x10^{-05}$; d) QQ-plot of p-values of CpGs located within 1Mb window from T2D GWAS SNPs associated with handedness at $p<5x10^{-08}$

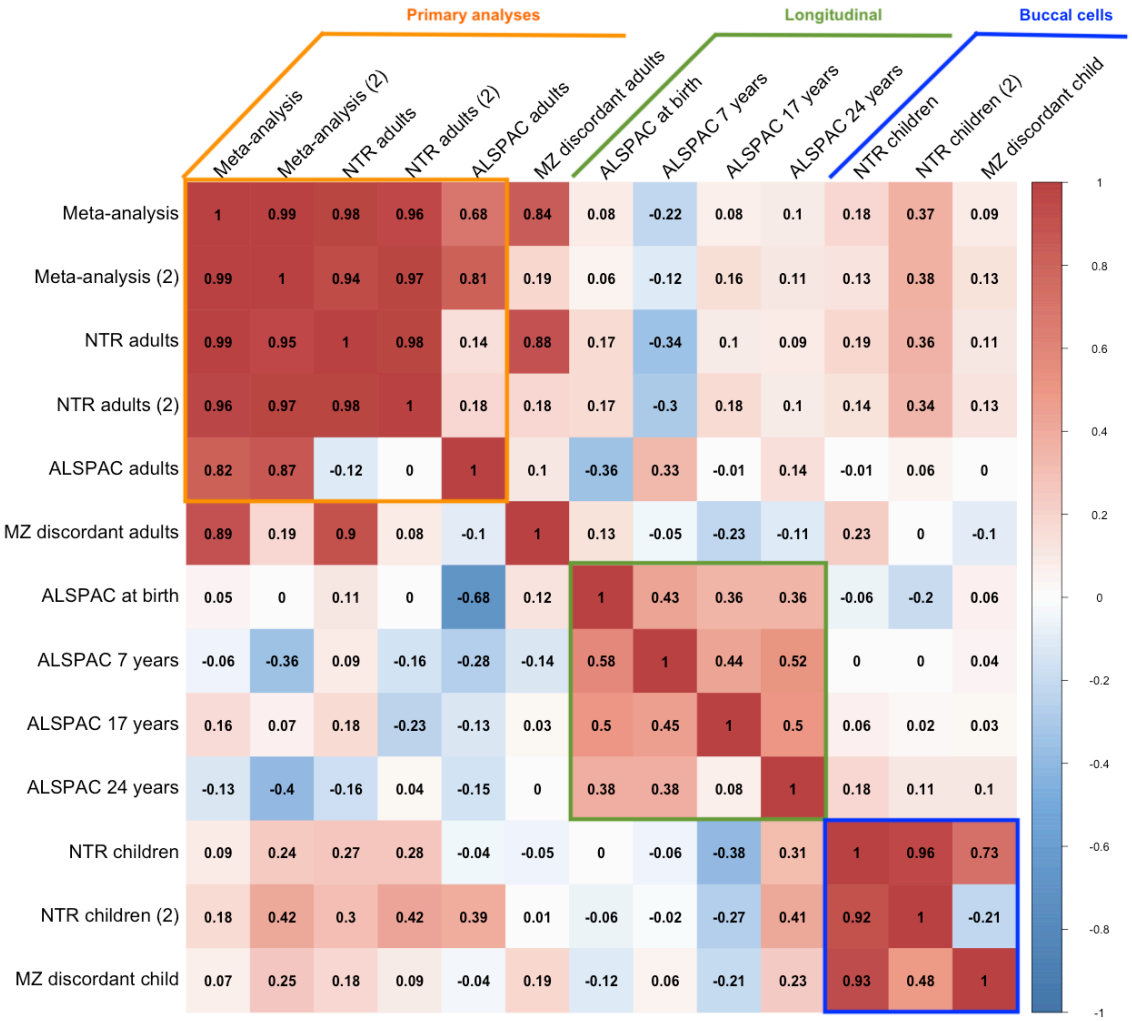


Figure S3. Correlations among the effect sizes of top 100 CpGs across analyses

Note: Orange frame = primary analysis. Green frame = longitudinal EWAS in ALSPAC at 4 points (at birth, 7 years, 17 years, 24 years). Blue frame = EWAS in NTR children (buccal cells). MZ discordant twins = MZ discordant twin within-pair EWAS. (2) = analyses without MZ discordant twins.

Correlation matrix is based on the 379,924 methylation sites available in all analyses (present on the EPIC array and 450k array). The lower triangle contains the correlations between effect estimates of the top 100 CpGs of models listed on the horizontal axis with the effect sizes of the same CpGs for the models listed on the vertical axis, and the upper triangle vice versa.

a) b)

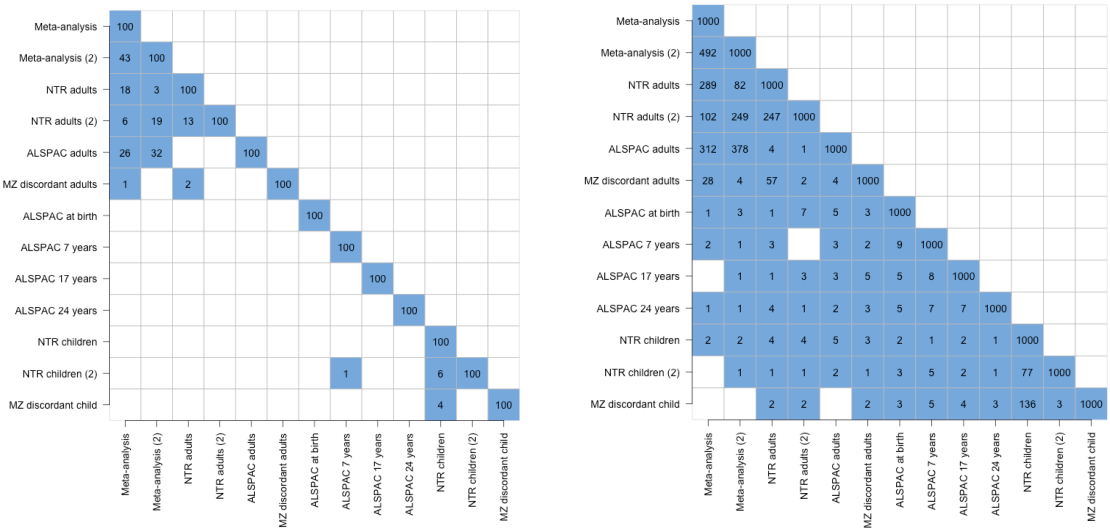
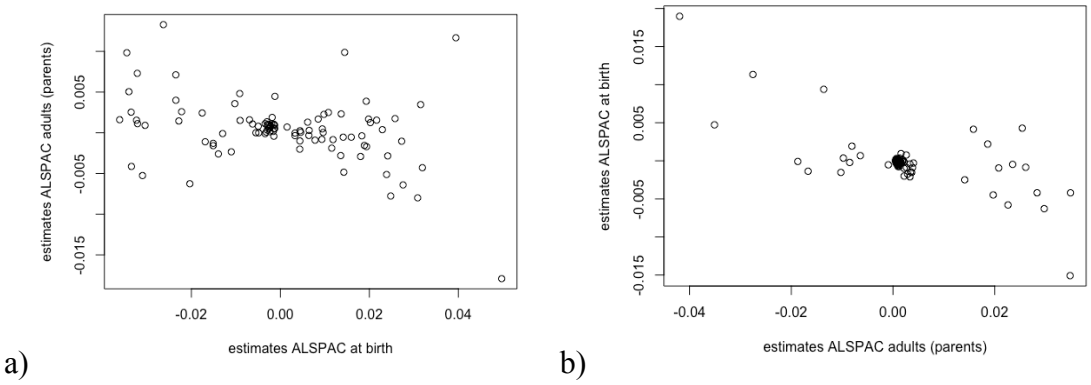


Figure S4. Overlaps in top CpGs across analyses

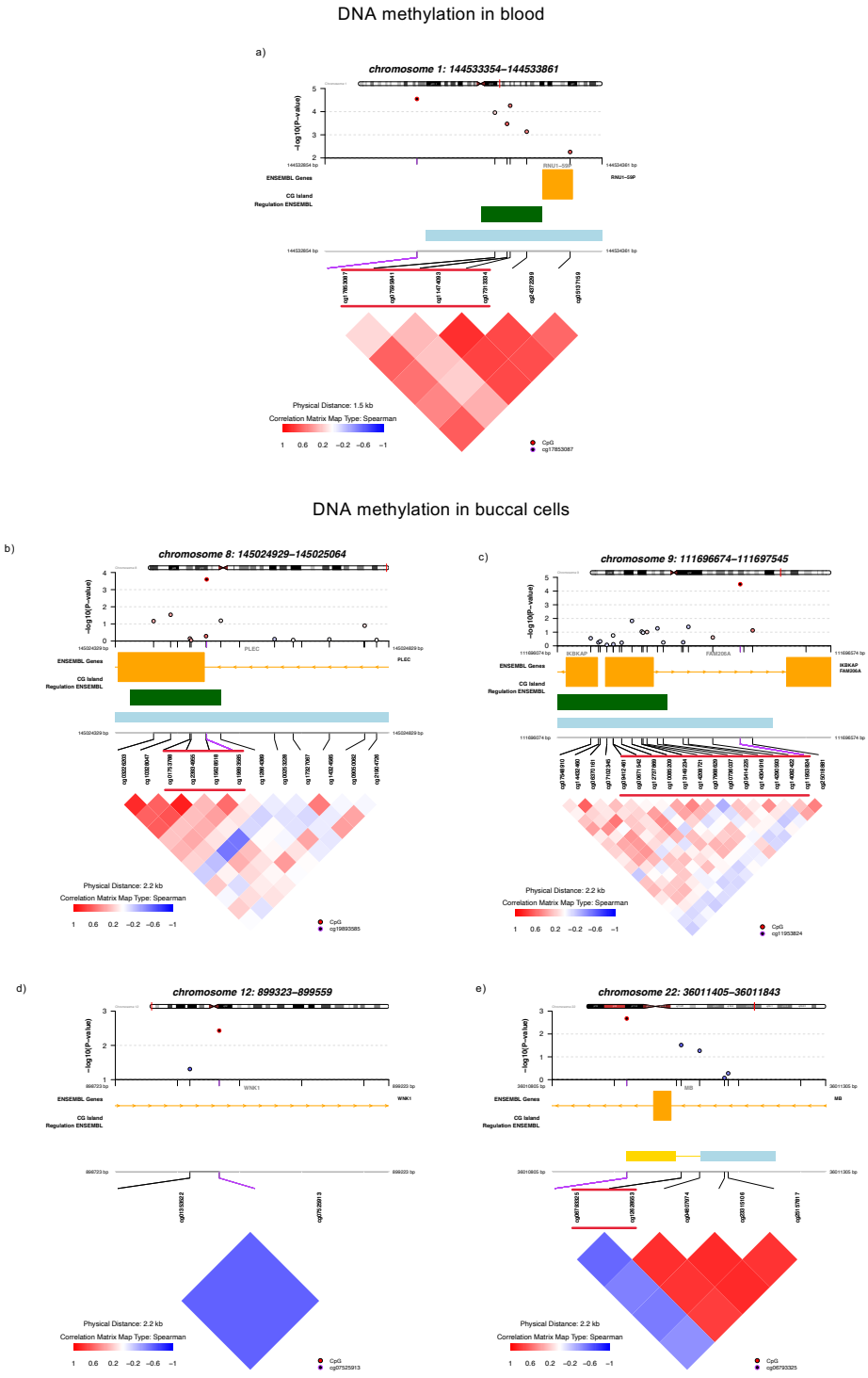
On the base of preselected list of 379,924 methylation sites available in all analyses. MZ discordant twins = MZ discordant twin within-pair EWAS.(2)= analyses without MZ discordant twins. a) top 100 CpGs; b) top 1000 CpGs.



a) b)

Figure S5. Scatterplot of effect sizes in ALPAC adults and ALSPAC at birth

a) top 100 CpGs from ALSPAC at birth EWAS, b) top 100 CpGs from ALSPAC adults (parents) EWAS
Note: Scatterplots were done to check if an outlier was the cause for unexpected negative correlation between top 100 estimates with lowest p-value $r^{ALSPACadults- ALSPACatbirth}$ =-0.680; p= 7.2 x 10^{-15}.



Figures S6a-i. Differentially methylated regions associated with left-handedness

CpGs from DMRs are indicated with red lines. The top panel of each plot shows the EWAS p-values for all CpGs in the window, with the most strongly associated CpG highlighted. The middle panel shows the genomic coordinates (genome build GRCh37/hg19) and the functional annotation of the region: the ENSEMBL Genes track shows the genes in the genomic region (orange); the CpG Island track shows the location of CpG islands (green); the Regulation ENSEMBL track shows regulatory regions (blue). The bottom panel shows the Spearman correlation between methylation levels of CpGs in the window.

DNA methylation in blood: a) DMR at chromosome 1, EWAS, basic model, ALSPAC adults (blood)
DNA methylation in buccal cells: d) DMR at chromosome 8, EWAS, adjusted model, NTR children (buccal cells); e) DMR at chromosome 9, EWAS, adjusted model, NTR children (buccal cells); f) DMR at chromosome 12, EWAS, adjusted model, NTR children (buccal cells); g) DMR at chromosome 22, EWAS, adjusted model, NTR children (buccal cells).

Table S28. Polygenic and methylation scores general description

	NTR adults	NTR children 9 years	ALSPAC mothers	ALSPAC children 7 years
PGS 0.5 fraction	6,452,863		6,779,197	
MS $p < 1x10^{-1}$	66,712	48,125	43,628	46,698
MS $p < 1x10^{-3}$	1536	5,041	413	546
MS $p < 1x10^{-5}$	14	12	2	7

PGS=polygenic score; MS=methylation score

Figure S7. Polygenic and methylation scores histograms

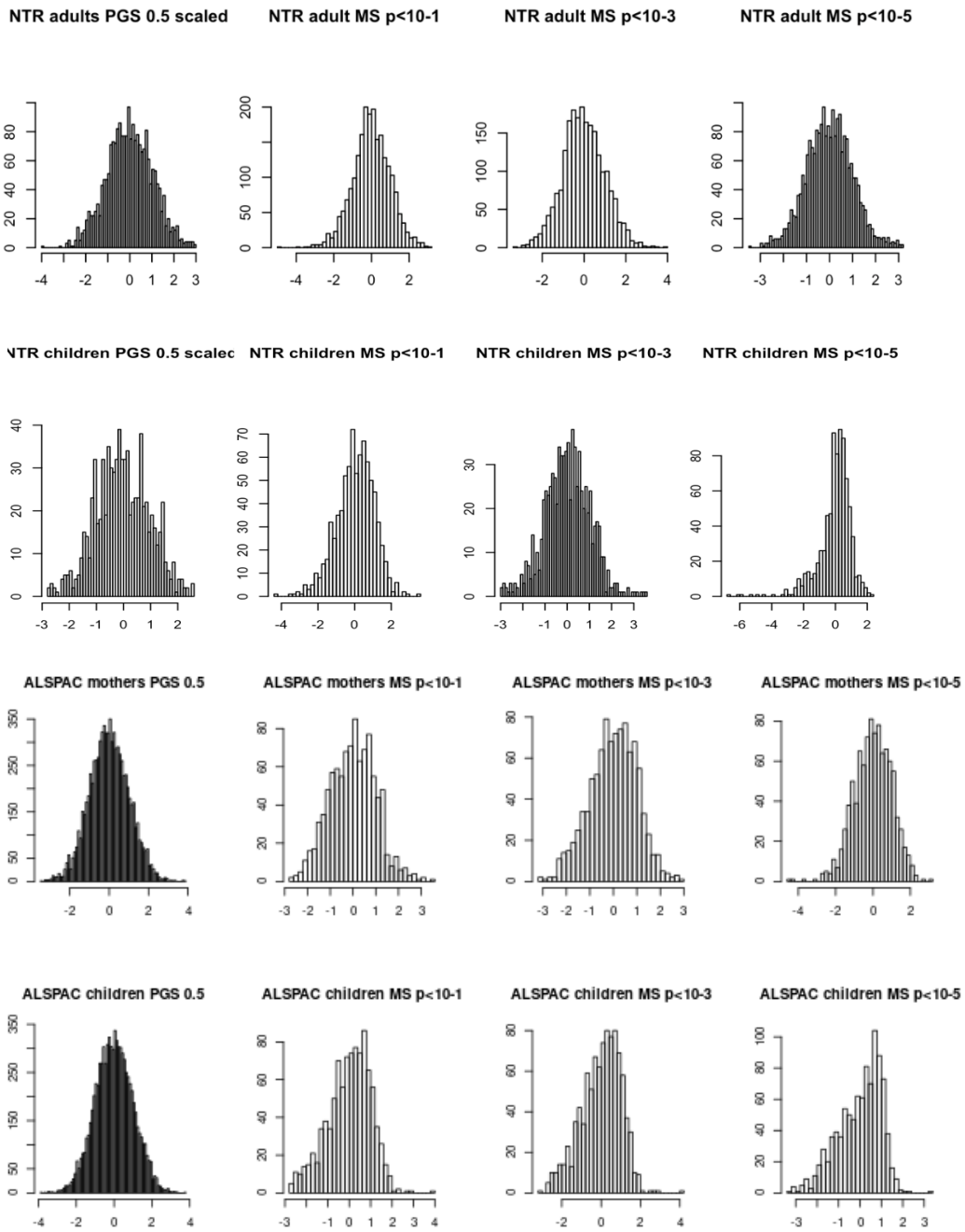


Table S29. Performance of polygenic and methylation scores of left-handedness

Cohort	N	Combined model					
		PGS + MS $p<1x10^{-1}$		PGS + MS $p<1x10^{-3}$		PGS + MS $p<1x10^{-5}$	
		PGS R^2% (p)	MS R^2% (p)	PGS R^2% (p)	MS R^2% (p)	PGS R^2% (p)	MS R^2% (p)
		Same-age same tissue (whole blood)					
NTR Adults	2198	0.06 (0.471)	0.00 (0.851)	0.06 (0.482)	0.06 (0.506)	0.16 (0.482)	0.06 (0.271)
ALSPAC Adults	574	0.37 (0.187)	-0.001 (0.998)	0.4 (0.177)	0.38 (0.493)	0.12 (0.216)	-0.17 (0.423)
		Same-age different tissues (buccal cells and whole blood)					
NTR Children	799	0.02 (0.748)	0.01 (0.921)	0.48 (0.251)	0.02 (0.870)	0.27 (0.261)	0.01 (0.928)
ALSPAC Children	630	0.35 (0.460)	0.0004 (0.978)	0.36 (0.455)	0.02 (0.873)	0.41 (0.438)	1.28 (0.103)

PGS, Polygenic Scores; MS, Methylation Scores; R^2%, variance explained by score in percentages; p, p-value; α=0.05/4=0.0125
The numbers of SNPs for PGSs and CpGs for MSs at 3 thresholds ($p<1x10^{-1}$, $p<1x10^{-3}$, $p<1x10^{-5}$) are indicated in Supplemental table S27.
See calculation of explained variance in Appendix 2.