THLBB2021_PUB006: Epigenetic signature of early genome programming in identical twins

Cohort: Twin Study
Data set accession number in THL Biobank: THLBB2021_PUB006

Publication information

Identical twins carry a persistent epigenetic signature of early genome programming.
van Dongen J, Gordon SD, McRae AF, Odintsova VV, Mbarek h, Breeze CE, Sugden K, Lundgren S, Castillo-Fernandez JE, Hannon E, Moffitt TE, Hagenbeek FA, van Beijsterveldt CEM, Hottenga JJ, Tsai PC, BIOS Consortium*, Genetics of DNA Methylation Consortium*, Min JL, Hemani G, Ehli EA, Paul F, Stern CD, Heijmans BT Slagboom PE, Daxinger L, van der Maarel SM, de Geus EJC, Willemsen G, Montgomery GW, Reversade B, Ollikainen M, Kaprio J, Spector TD, Bell JT, Mill J, Caspi A, Martin NG & Boomsma DI.

Published article's journal and reference information

Nat Commun 12, 5618 (2021).
DOI: https://doi.org/10.1038/s41467-021-25583-7

Dataset availability information

The FTC DNA methylation data is deposited in THL Biobank Finland. For information on access and how to apply, see https://thl.fi/en/web/thl-biobank/for-researchers. The applicant for the data can reference the publication when asking for access. The webpage of the THL Biobank describes the exact procedure for accessing the data. The THL biobank grants access to qualified academic and commercial applicants with a scientifically justified study plan.

Participant and methylation quantification description

Twin participants
The Finnish Twin Cohort Study (FTC) is part of three longitudinal cohorts (Kaprio et al 2019; Kaprio, 2013), the Older Twin cohort56, FinnTwin16 (FT16) (Kaidesoja et al 2019), and FinnTwin12 (FT12) (Rose et al 2019). The Older Twin Cohort is comprised of 13,888 same-sex twin pairs born before 1958, while FT16 and FT12 are longitudinal studies of five consecutive birth cohorts (born in 1975–1979, n = 2800 pairs, and 1983–1987, n = 2700, respectively) of Finnish monozygotic and dizygotic twins who have completed surveys and interviews beginning in adolescence and into adulthood (FT16 at age 16, 17, 18, 24, 34; FT12 at age 11, 14, 17.5, 24). DNA extracted from whole blood of 2216 twin individuals were used in this study.

Participants were given information about the study procedures and design in their native language (Finnish or Swedish) and provided informed consent, following the principles of informed consent in the Declaration of Helsinki. All study procedures were approved by the ethics committees of Helsinki University Central Hospital (113/E3/2001, 249/E5/ 2001, 346/E0/05, 270/13/03/01/2008, and 154/13/03/00/2011).

Epigenome-wide association analysis
The EWAS included one randomly selected MZ twin from each twin pair, or the twin whose sample passed quality control in the case that one twin's sample was not of high quality, and all dizygotic twins (total n=1708).

White blood cell percentages were estimated using the updated IDOL libraries for 450k and EPIC data (Salas et al 2018).

EWAS analysis was performed with generalized estimation equation (GEE) models, which were fitted with the R package "gee" with DNA methylation β-value as outcome and the following predictors: zygosity, sex, age at DNA sampling, cellular composition (estimated proportions of CD8 T cells, CD4 T cells, natural killer cells, and neutrophils), smoking status, BMI, and array type. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the "exchangeable" option to account for the correlation structure within families. The following probes were removed: sex chromosomes, probes with a single nucleotide polymorphism (SNP) within the CpG site (at the C or G position) and ambiguous mapping probes reported by Chen et al. with an overlap of at least 47 bases per probe (Chen et al 2013).

Correlation analysis
Top CpG sites (n CpGs=226) were used in the correlation analyses that were performed for all available full twin pairs (n=1016 MZ, 752 SSDZ, 294 OSDZ). Methylation beta-value residuals after adjusting for covariates using *lm* function in R for each CpG were used to calculate Pearson correlations coefficients between the co-twins using *cor* R-function with use option set to *pairwise.complete.obs*.

References:
Kaprio, J. et al. The Older Finnish Twin Cohort - 45 years of follow-up. Twin Res. Hum. Genet. 22, 240–254 (2019).
Kaprio, J. The Finnish Twin Cohort Study: an update. Twin Res. Hum. Genet. 16, 157–162 (2013).
Kaidesoja, M. et al. FinnTwin16: A Longitudinal Study from Age 16 of a Population-Based Finnish Twin Cohort. Twin Res. Hum. Genet. 22, 530–539 (2019).
Rose, R. J. et al. FinnTwin12 Cohort: an updated review. Twin Res. Hum. Genet. 22, 302–311 (2019).
Salas, L. A. et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. 19, 1–14 (2018).
Chen, Y. et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8, 203–209 (2013).

Description of THLBB2021_PUB006 dataset

The following variables are included in this dataset:

Raw DNA methylation intensity data
Methylation beta-values
personID (recoded for each project)
familyID (recoded for each project)
zygosity (monozygotic = 1, same-sex dizygotic = 2, opposite-sex dizygotic = 3)
sex
age*
smoking (never=0, Current=1, former=2)*
BMI*
array type (450k/EPIC)
analysis sample was used for (EWAS/correlation/both)
Estimated proportions of CD8+ T-cells, CD4+ T-cells, Natural Killer cells and Neutrophils

* at the time of DNA sample collection for generation of methylation data