

DNA methylation sites in early adulthood characterised by pubertal timing and development: A twin study

Authors: Emir Sehovic^{1,2}, Stephanie M. Zellers³, Markus K. Youssef⁴, Aino Heikkinen³, Jaakko Kaprio³, Miina Ollikainen^{*3,5}

Affiliations:

1. Department of Life Sciences and Systems Biology, University of Turin, 10100 Turin, Italy.
2. Cancer Genomics Lab, Fondazione Edo ed Elvo Tempia, 13900 Biella, Italy.
3. Institute for Molecular Medicine Finland, University of Helsinki, 00290 Helsinki, Finland.
4. Department of Immunobiology, University of Lausanne, CH-1066 Epalinges, Switzerland
5. Minerva Foundation Institute for Medical Research, 00290 Helsinki, Finland

*Corresponding author – contact: miina.ollikainen@helsinki.fi

ABSTRACT

Background

Puberty is a highly heritable and variable trait, with environmental factors having a role in its eventual timing and development. Early and late pubertal onset are both associated with various diseases developing later in life, and epigenetic characterisation of pubertal timing and development could lead to important insights. Blood DNA methylation, reacting to both genotype and environment, has been associated with puberty; however, such studies are relatively scarce. We investigated peripheral blood DNA methylation profiles (using Illumina 450K and EPIC platforms) of 1539 young adult Finnish twins associated with pubertal development scale (PDS) and pubertal age (PA).

Results

Fixed effect meta-analysis of the two platforms on 347521 CpGs in common identified 58 CpG sites associated ($p < 1 \times 10^{-5}$) with either PDS or PA. All four CpGs associated with PA and 45 CpGs associated with PDS were sex specific. Thirteen CpGs had a high heritability (h^2 : 0.51-0.98), while one CpG site (mapped to *GET4*) had a high shared environmental component accounting for 68% of the overall variance in methylation at the site. Utilising twin discordance analysis, we found 6 CpG sites (5 associated with PDS and 1 with PA) that had an environmentally driven association with puberty. Furthermore, genes with PDS- or PA-associated CpGs were linked to various developmental processes and diseases, such as breast, prostate and ovarian cancer, while methylation quantitative trait loci (meQTLs) of associated CpG sites were enriched in immune pathways developing during puberty.

Conclusions

By identifying puberty-associated DNA methylation sites and examining the effects of sex, environment and genetics, we shed light on the intricate interplay between environment and genetics in the context of puberty. Through our comprehensive analysis, we not only deepen the understanding of the significance of both genetic and environmental factors in the complex processes of puberty and its timing but also gain insights into potential links with disease risks.

Keywords: Puberty, Twins, DNA methylation, meQTL, Heritability, Young adults

BACKGROUND

Puberty is a necessary developmental phase between childhood and adulthood, enabling sexual reproduction. In addition to pubertal timing and pubertal development differing between the sexes, they are highly variable between individuals of the same sex. Early pubertal development is associated with an increased risk of many diseases in later life, such as breast, endometrial and prostate cancer (1–3). On the other hand, late pubertal development is associated with asthma in

both sexes and cervical cancer in females (3). Nevertheless, the mechanisms behind these links remain elusive. Thus, understanding puberty from a genetic and epigenetic perspective may provide further insight into this important period of life and its links with disease risk in later life.

Pubertal timing and development have high heritability, ranging from 37% to 91% (4–7). Studies have identified numerous genetic loci associated with pubertal timing (1,8,9), with a large proportion of the implicated genes involved in neural processes. These identified genetic signals partially explained some diseases, such as breast and prostate cancers, linked to the timing of puberty (1,8,9). Additionally, recent evidence indicates a shift towards an earlier onset of puberty in girls, suggesting that key nongenetic factors play a role in pubertal timing (10,11). Indeed, prenatal growth, diet, maternal smoking, psychological distress in childhood, and exposure to endocrine-disrupting chemicals such as phytoestrogen have been associated with early onset of puberty (12). In addition, although mainly explained through genetic overlap with pubertal timing, high BMI has been associated with early puberty (13,14). Epigenetic marks react to the environment and may provide a mechanism for the environment to act on genome function and ultimately lead to phenotypic consequences. Indeed, DNA methylation, the most studied epigenetic mark, has been associated with pubertal onset (15–17). As the focus of these studies was mainly the change in methylation before and after pubertal timing, there is still a need to identify consistent CpG sites associated with pubertal timing or development in (early) adulthood, where the methylation profile relevant to puberty is more stable than that within early adolescence and could therefore be more reliably linked to disease risk in adulthood. Considering the high heritability of puberty and the wide heritability range of DNA methylation at each CpG site (18), it is necessary to investigate the variance components of DNA methylation sites associated with puberty and to analyse them in the context of methylation quantitative trait loci (meQTLs) (19).

In this epigenome-wide association study (EWAS), we profiled DNA methylation in the blood of young adults to further understand pubertal timing, pubertal development and potential diseases they associate with. We investigated genome-wide DNA methylation and its association with pubertal timing and pubertal development in both sexes. As puberty differs between males and females, we additionally stratified the analyses by sex. We also employed twin study designs to evaluate heritability and within-pair differences and to explore qualitative and quantitative sex differences in DNA methylation at the associated CpG sites. The twin analyses coupled with meQTLs were utilised to gain insight into the genetic and environmental factors affecting methylation values of CpG sites associated with puberty.

MATERIALS AND METHODS

Cohorts

This study was based on the participants of two longitudinal and comprehensive birth cohorts of Finnish twins, namely, the FinnTwin12 (twins born in 1983-1987) and FinnTwin16 study (twins born in 1974-1979). Both cohorts have been described in detail elsewhere (20,21). In brief, FinnTwin12 was initiated when the twins were 11-12 years of age, and the follow-up questionnaires were sent at ages 14, 17, and 24, while the FinnTwin16 baseline questionnaire

was at the age of 16, with follow-up at the ages of 17, 18, 25, and 35. A subsample of these twins participated in on-site visits, where biological samples were collected, and anthropometric measures were taken (N = 1539, 54.6% female). The current study included both monozygotic (MZ) and dizygotic (DZ) twin pairs who had completed the questionnaires regarding puberty (see below) and had blood DNA methylation data (Illumina 450K or EPIC) in early adulthood available.

Data on Puberty

The pubertal development scale (PDS) was measured twice in the FinnTwin12 cohort, at ages 12 and 14 years, based on the discrete self-reported markers of puberty: growth spurt, body hair, skin changes, voice change (boys)/breast change (girls) and facial hair (boys)/menarche (girls). PDS was calculated by summing the markers of puberty to obtain a total score, and then the total score was divided by five (22). All individuals with any missing value on PDS variables were excluded from the analysis. Self-reported pubertal age (PA), defined as the age of menarche (girls) or voice break (boys), was obtained from FinnTwin12 participants at the age of 17 and FinnTwin16 participants at the age of 16. All individuals with missing or inconsistent information on PA were excluded. Moreover, individuals who reported that they did not have menarche/voice break yet were coded as the event happening at the age of 17 (13).

Other Phenotypic Data

For each twin included in the analysis, we also extracted the following items: sex, zygosity, age at questionnaire wave as well as age, smoking status and alcohol consumption at the time of blood sampling. Smoking status was categorised into current, former (abstinent for at least six months) and never smokers, while alcohol consumption was calculated in units of ethanol g/day during the past week, as described previously (23).

DNA Methylation Data

Blood samples for DNA methylation analyses were collected during in-person visits in adulthood (age range: 21-33.7 and 23.3-42.7 for FinnTwin12 and FinnTwin16, respectively). High molecular weight DNA extracted from whole blood by standard protocols was bisulfite converted with the EZ-96 DNA Methylation-Gold Kit (Zymo) according to the manufacturer's instructions. Methylation was quantified using Illumina Infinium HumanMethylation450 and EPIC BeadChip platforms that cover more than 450,000 and 850,000 CpG sites, respectively.

The DNA methylation data were preprocessed in the R package *meffil* (24). Bad quality samples were excluded based on the following criteria: i) sex mismatch, ii) median methylation vs. unmethylated signal > 3 standard deviations (SD), iii) failed control probe metrics and if >20% of probes per sample had iv) detection p value > 0.05 and v) bead number < 3. To remove technical variation between the samples, functional normalisation including the control probe principal components was performed, followed by bad quality probe removal: i) detection p value > 0.05, ii) bead number < 3, iii) sex chromosome probes and iv) ambiguously mapped probes (25,26). BMIQ normalisation implemented in the R package *wateRmelon* (27) was then performed to adjust for type 2 probe bias. The same pipeline was used for 450K and EPIC data. The data preprocessing resulted in 390304 and 765385 probes in 450K and EPIC data,

respectively. Thirty-nine and 85 samples were excluded after preprocessing in 450K and EPIC data, respectively. Beta values representing DNA methylation were used in all analyses.

Statistical Analysis

All analyses were performed in R software (version 4.1.1) (28).

EWAS Designs

The EWAS analyses were performed using the *limma* package (29). The EWAS on PDS was performed separately on individuals who reported the PDS items at ages 12 and 14. Both sex-stratified and combined models for PDS were performed. The EWAS on PA was performed on the merged cohorts of individuals who reported their PA at age 17 in FinnTwin12 and 16 in FinnTwin16. The EWAS on PA was performed separately in males and females. All EWAS designs are reported in the Supplementary Methods (see Additional file 1). For both 450K and EPIC data, the cell type proportions were calculated using *Flow-Sorted.Blood*. *EPIC* (30) R package, which is based on a modified version of the Houseman algorithm (31).

We evaluated various combinations of covariates by utilising the AIC to determine the best model fit for the data. This was done using the “selectModel” function from the *limma* package. For each EWAS run, we evaluated the test statistics inflation and bias using the *BACON* package (32). Additionally, for visual investigation of inflation and bias, QQ plots and test-statistic histograms were used, respectively. The QQ and Manhattan plots were created using the *qqman* package (33).

All EWAS models were corrected for smoking, alcohol consumption, age (at the time of blood sampling), cell type proportions, date of BeadChip run and row of the array slide. Additionally, the family ID was included as a random effect to account for the relatedness of twins within a pair. The models that included both sexes were corrected for sex, and the models on PA were corrected for the cohort.

EWAS Meta-Analysis

The EWAS results from the 450K and EPIC platforms, based on the same design, were merged using a meta-analysis. A fixed effect meta-analysis with inverse-variance weights was performed using the “rma” function from the *metafor* package (34). The standardised effect sizes and sampling variances were calculated using the “escalc” function. The computed metric was the partial correlation coefficient (35).

The input parameters were the t-statistic, sample size and number of covariates in the regression model. The output parameters were the standardised meta-analysed effect size, p value, Q heterogeneity estimate and its corresponding p value. We corrected the heterogeneity p values using the Benjamini-Hochberg (BH) method. All CpG sites with a corrected Q p value < 0.01 were considered heterogeneous, and a random-effect meta-analysis was performed on such CpG sites.

For the discovery of CpG sites, we used two significance cut-offs on the p values obtained from the meta-analysis: the cut-off recommended for the 450K platform 2.4×10^{-7} (36) and a

suggestive p value cut-off at 1×10^{-5} (37). We annotated all CpG sites with a standardised effect size larger than the absolute value of 0.13 using the *IlluminaHumanMethylation450kanno.ilmn12.hg19* package (38).

Analyses of Genetic and Environmental Influences on CpG Sites

To follow up on results from the EWAS, we ran four sets of analyses to further understand the genetic and environmental influences on CpG sites significantly associated with puberty. These analyses, described in detail below, include univariate twin modelling, discordant twin pair analyses, meQTL analysis and analysis of sex differences in CpG sites. Broadly speaking, each analysis provides information on the sources of variation underlying methylation at each CpG site.

Univariate Twin Modelling

CpG sites associated with PDS or PA with a p value lower than the suggestive cut-off ($p < 1 \times 10^{-5}$) that met the model assumptions and with methylation value SD > 0.05 were assessed by twin modelling (see details in Supplementary Methods - Additional file 1). Univariate twin modelling was performed by the *openMx* R package (39) with age at blood sampling and sex included as covariates to estimate variance components for additive genetic effects (A), common environmental effects (shared exposures and experiences within each twin pair - C), genetic dominance effects (D), and unique environmental effects (nonshared exposures and experiences - E). All potential models (i.e., ACE, ADE, AE, CE, DE, E) were compared to determine the best-fitting set of variance components based on the log-likelihood test.

Discordant Twin Pair Analyses

CpG sites associated with PDS or PA with a p value lower than the suggestive cut-off ($p < 1 \times 10^{-5}$) were assessed by discordant twin pair analysis. The discordant twin pair design tests whether cotwins who differ in pubertal age or pubertal development also differ in methylation value at each CpG site while naturally controlling for all genetic and environmental confounders shared by the twins within a pair. Discordant twin pair analyses offer the most powerful statistical approach when conducted within MZ twin pairs since the twins in a pair share 100% of their genomes. Here, we additionally pooled MZ and DZ twin pairs together to increase power due to increased sample size compared to the within-pair analyses among MZ pairs only.

We performed a mixed effects model that decomposes the individual-level effect identified in the EWAS into between-pair and within-pair effects. We included a random effect of family and fixed effects of age at blood sampling, sex, and platform as covariates. The between-pair predictor was defined as the average of the exposure within a twin pair. The within-pair predictor (i.e., discordance) was defined as each twin's value on the exposure subtracted from their cotwin's value. All pairs were included in the analysis without setting any threshold for within-pair discordance, as they still contribute information to the between-pair effect. Discordant twin pair analyses were conducted in the same subsample for which significant EWAS results were identified.

meQTL Analysis

In the context of this study, on the one hand, meQTL SNPs could represent a genetic mechanism that explains the high heritability of a CpG site and, on the other hand, could additionally provide functional information on the SNP-trait association. We used the Genetics of DNA Methylation Consortium (GoDMC) meQTL database (40,41) to explore whether the CpG sites associated with PDS or PA are established meQTL. For each CpG site, we extracted all genome-wide significant ($p < 5 \times 10^{-8}$) meQTLs.

Sex Differences

Sex differences at the PDS- or PA-associated CpG sites were tested in two ways: differences in methylation values and differences in twin model parameters between the sexes. To test for the differences in DNA methylation value medians between males and females within CpG sites associated with PDS or PA, we performed a Wilcoxon two-sample test using the “wilcox.test” function and corrected the p values using the Bonferroni method. CpG sites with a p value < 0.01 were considered differentially methylated between the sexes. Potential sex differences were also evaluated for all CpGs included in the univariate twin modelling analyses. We explored quantitative and qualitative sex differences, where quantitative sex differences refer to cases where the same genes influence a trait (here, methylation at a particular CpG site) in both sexes but in different magnitudes, and qualitative sex differences refer to cases where different sets of genes influence a trait for each sex (i.e., genes specific to one sex). Importantly, this analysis cannot identify which specific genes influence the trait in each sex, only whether the set of genes as a whole is the same or different between sexes.

An omnibus test was performed to explore whether there were any sex differences in any parameter in the model. The omnibus test compares a model in which every parameter is freely estimated for males and females to a nested model in which all parameter estimates are simultaneously fixed to be equal for both sexes. For all CpG sites with a significant omnibus test, we evaluated each parameter individually to determine which differed between the sexes. The specific sex differences were evaluated by comparing the model in which one parameter was constrained to equality between sexes to a model in which the parameter was freely estimated. P values were generated from likelihood ratio tests and were corrected using the Bonferroni method (function “p.adjust” in R). The significance cut-off for these adjusted p values was < 0.01 for all tests.

Pathway Analysis

Ingenuity pathway analysis (IPA) software from QIAGEN Inc. (42) was used to assess whether the PDS- and PA-associated CpG sites and meQTLs were enriched in specific gene ontologies or pathways. The IPA was performed on gene annotated CpG sites with a standardised effect size $> |0.13|$, using the “Human Methylation 450 v1-2” and the “Ingenuity Knowledge Base” as the reference gene set for CpGs and meQTL SNPs, respectively. The analyses were based only on *Homo sapiens*, and we interpreted the results based on the p value of the associations (significance cut-off was 0.05) with a focus on the canonical pathways and linked diseases or biological functions (“Diseases or function annotation”).

RESULTS

To identify blood DNA methylation profiles in early adulthood associated with PDS or PA, we performed sex-stratified meta-EWAS on Finnish twins with sample sizes stratified by the array platforms shown in Supplementary Table S1 (see Additional file 3). Furthermore, we also performed EWAS models on PDS using the full sample to determine sex-independent CpG sites. The workflow of the current study is presented in Figure 1. The average age at which the methylation data were obtained was 22.7 years (SD: 1.7) for FinnTwin12 and 28.1 years (SD: 4.3) for the FinnTwin16 cohort. The PDS was obtained from the FinnTwin12 cohort at mean ages of 11.4 (SD: 0.3) and 14.0 (SD: 0.1), while the PA was obtained at mean ages of 16.2 years (SD: 0.1) and 17.6 years (SD: 0.2) from the FinnTwin16 and FinnTwin12 cohorts, respectively. The PDS and PA of the selected individuals (individuals with methylation data included in the study) did not significantly differ from the complete FinnTwin12 and FinnTwin16 cohorts (43). Detailed descriptive statistics on PDS, PA and other relevant variables can be seen in Table 1. A total of 347521 CpG sites were included in the EWAS meta-analyses, with no evidence of inflation or bias in any of the EWAS models (across both platforms, the bias test statistic ranged from -0.213 to 0.142, while the inflation test statistic ranged from 0.9 to 0.995). Overall low heterogeneity was observed between the two platforms (< 0.07% heterogeneous CpGs across all models).

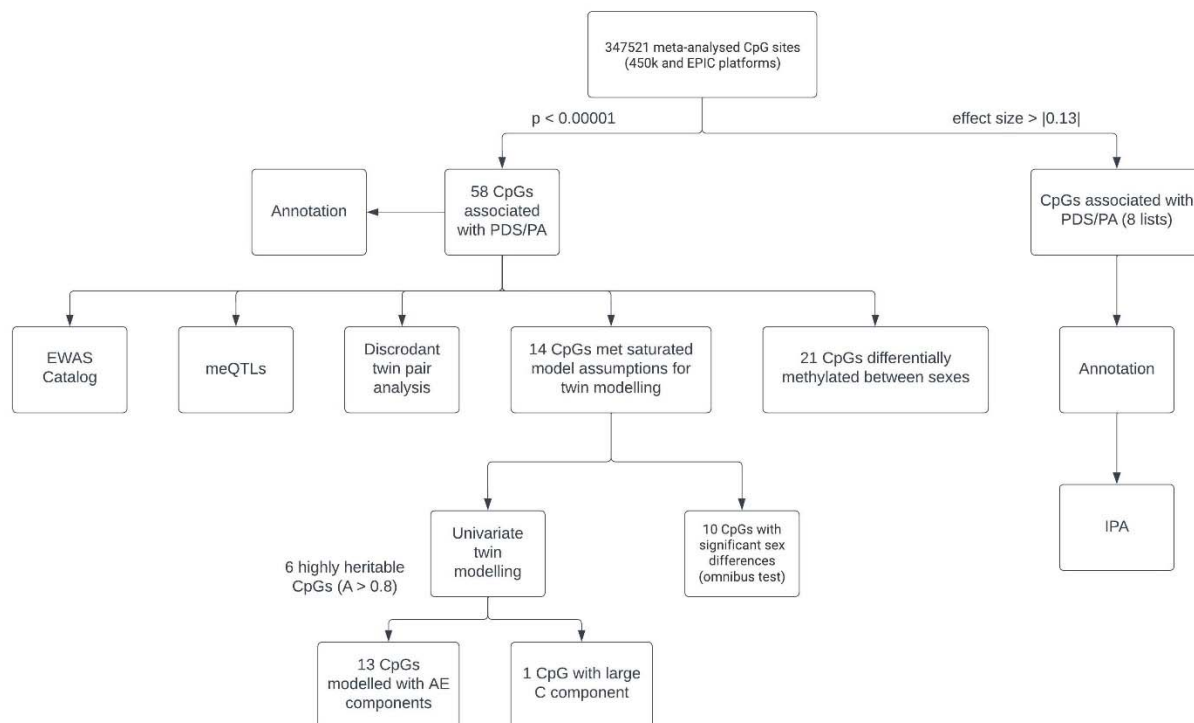


Figure 1. Workflow diagram of the study. Through a meta-analysis of 450K and EPIC platforms on 347521 CpGs, we identified CpG sites associated with PA or PDS at age 12 or 14 ($p < 0.00001$). We performed univariate twin modelling to assess heritability and to define the proportion of variance attributed to unique and shared environments and investigated potential sex differences in the significant CpG sites. We identified enriched pathways and linked diseases among CpG sites with an absolute effect size > 0.13 by IPA.

Puberty-associated CpG Sites

At the age of 12, PDS was associated with 9 CpG sites ($p < 1 \times 10^{-5}$) in both sexes combined, while 13 and 10 were exclusively associated with PDS in males and females, respectively (Table 2, Supplementary Figure S1 and S2 - Additional file 2). Two CpG sites in females (cg07581365, cg06988547) had a lower p value than the threshold of 2.47×10^{-7} . At the age of 14, no CpG sites were associated with PDS in the model including both sexes (Supplementary Figure S1 and S2 - Additional file 2). However, 12 and 10 CpG sites were associated with PDS in males and females, respectively (Table 2 and Figure 2). One CpG site in males (cg04239863) and one in females (cg20599748) were significantly associated with PDS ($p < 2.47 \times 10^{-7}$). A total of 12 of the 54 CpG sites associated with PDS showed significant heterogeneity between the 450K and EPIC platforms (Table 2), indicating lower reliability of the observed associations for those CpG sites. After performing a random effect meta-analysis for these CpG sites, none remained significantly associated with PDS.

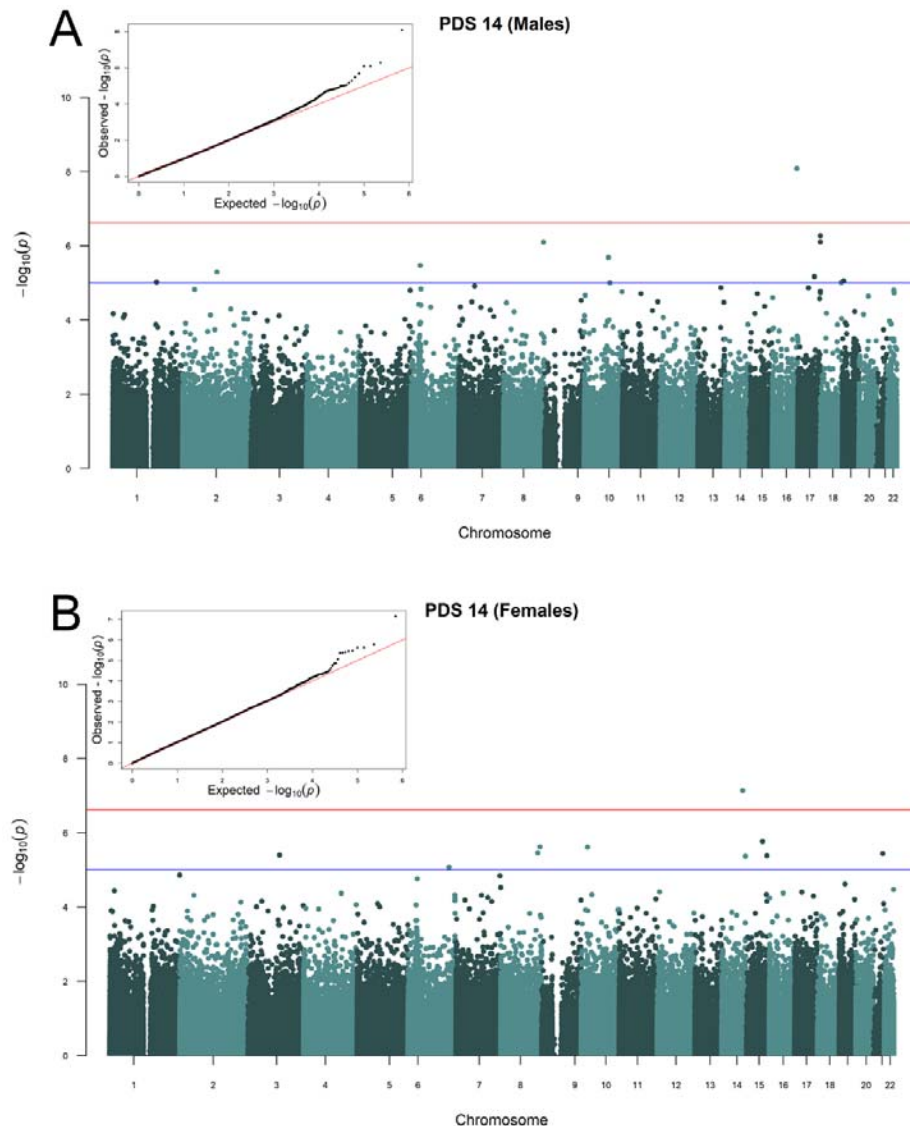


Figure 2. Manhattan plots of the meta-analyses. The Manhattan plots show significant associations between CpG methylation and PDS at age 14 in A) males and B) females. The red line indicates a significance cut-off of 2.4×10^{-7} as recommended for the 450K platform, while the blue line indicates a suggestive significance cut-off at 1.0×10^{-5} . Embedded (top left corner of Manhattan plots) are the QQ plots of meta-analysis p values.

Four CpG sites were associated with PA with $p < 1 \times 10^{-5}$: two unique CpGs in males and two in females (Table 2, Supplementary Figure S1 and S2 - Additional file 2). One of the CpG sites associated with PA in males (cg06096446) had a $p < 2.47 \times 10^{-7}$. None of the CpGs associated with PA were heterogeneous between the two platforms.

Genetic and environmental effects underlying the associations

First, as puberty itself as well as DNA methylation are highly heritable, we wanted to investigate the heritability and shared/nonshared environmental components explaining the variance of the 58 CpG sites associated with PDS or PA in early adulthood. We identified 14 CpGs for which both assumptions of the twin models were met, and the CpG had methylation SD > 0.05 (Supplementary Table S2 - Additional file 3). With respect to the univariate models, 13 out of 14 CpG sites were best modelled with only additive genetic (A) and unique environmental components (E) (Table 3). Heritabilities for the AE CpG methylation values ranged from 0.51-0.98, indicating strong additive genetic influences on these CpGs (average heritability was 0.79). Only one of the 14 CpGs (cg15030014, associated with PDS at age 12 in both sexes) also required the inclusion of a shared environmental component (C), indicating strong familial resemblance on this CpG due to nongenetic sources in addition to additive genetic influences (A=0.24, C=0.68).

Second, as only 14/58 CpG sites met the twin modelling criteria, we applied a discordant twin pair study design to identify PDS- and PA-associated CpG sites with their methylation likely driven by nongenetic factors. Mixed effects models were run to decompose the effect identified in the EWAS into within-pair and between-pair associations. Significant within-pair associations are consistent with a nongenetic effect. Here, CpGs that were considered differentially methylated within the pairs had a significant within-pair difference in the pooled analysis (MZ and DZ) as well as a suggestive difference in the same direction ($p < 0.10$) in the MZ-only analyses. Full results are available in Supplementary Table S3 in Additional file 3. We identified six differentially methylated CpGs, five of which were associated with PDS-12 and one with PA. For three CpGs, the MZ twin with a higher PDS-12 score also had higher methylation than their cotwin (cg14609721, cg03468249, cg05797363). For two CpG sites, the MZ twin with a higher PDS-12 score had lower methylation than their cotwin (cg08091771, cg10669219). Finally, for one CpG, the MZ twin with older PA also had higher methylation than their cotwin (cg06096446). These effects suggest an environmentally driven relationship between early pubertal development and DNA methylation. This effect is consistent with a causal relationship but cannot determine the direction of effect (i.e., whether methylation causes earlier puberty or if early puberty causes methylation).

Third, to gain additional insight into the genetic basis of DNA methylation at the 58 CpG sites associated with PDS and PA, we looked for *cis* and *trans* meQTLs from the GoDMC database

affecting methylation at these sites. Thirty-nine of the 58 CpG sites had *cis*-meQTLs, and only one CpG had *trans*-meQTLs (Figure 3). The number of associated meQTLs ranged from 1 to 1676 at these 39 CpG sites. Additionally, we found 3 CpG pairs whose methylation was affected by a subset of common SNPs (Figure 3).

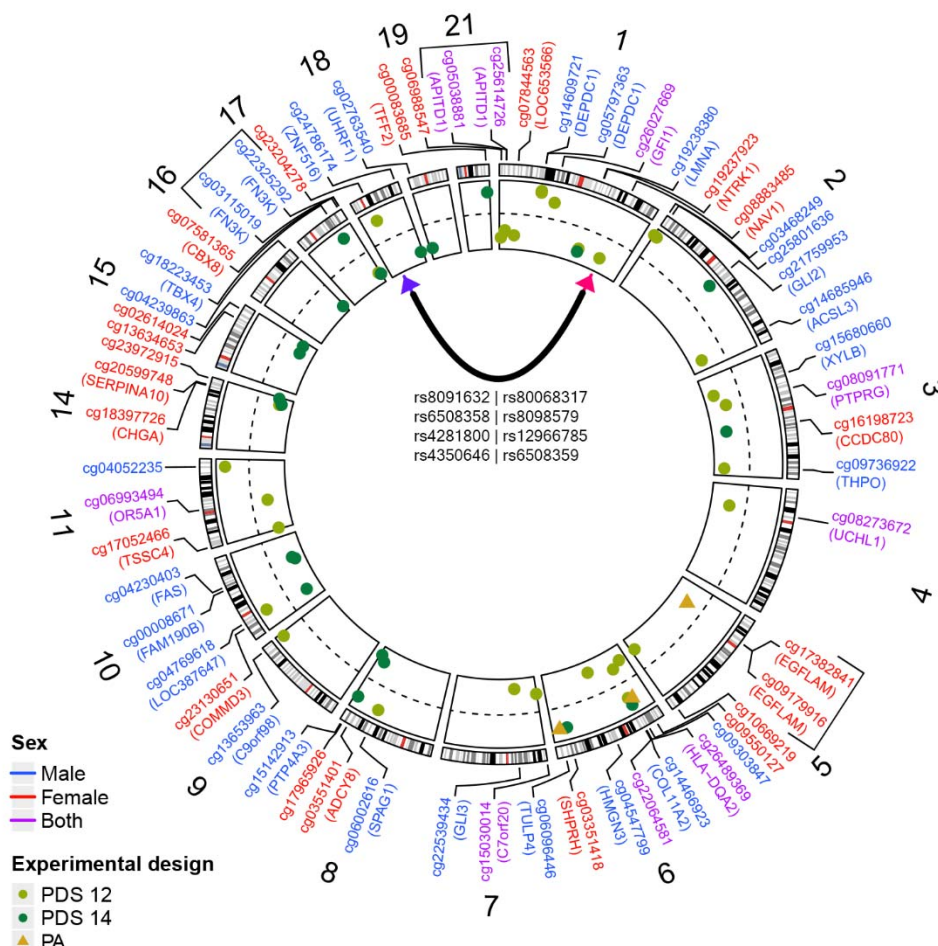


Figure 3. Circos plot showing the chromosomal locations of the 58 CpGs associated with PDS or PA. The inner circle shows the standardised effect sizes of the association (scale ranging from -0.3 to 0.3) with the dashed line representing 0. The shape of the points plotted on the effect size scale determines the variable analysed (PDS or PA) in the model in which the CpG was found, while the colour of the CpG site identifier shows whether the model was performed on males, females or both sexes. The 8 *trans*-meQTLs of CpG cg08883485 are depicted with the black line inside the circos plot (pink arrowhead pointing to the CpG site and purple arrowhead pointing to the genomic region of the 8 *trans*-meQTLs). The CpGs with common *cis*-meQTLs are shown with black square brackets.

Sex Differences in CpG sites

As we observed overall stronger associations in the sex-stratified models compared to models including both sexes and as there were no common significant CpG sites between the EWAS models in males vs females, we aimed to characterise potential sex differences and similarities in the identified CpG sites. Therefore, we performed a Wilcoxon two-sample test on CpG sites associated with PDS or PA and investigated sex differences by exploiting the properties of opposite-sex dizygotic twin pairs. Based on all included samples in this study, 21 CpG sites were

differentially methylated between males and females (Table 2). Three differentially methylated CpGs were obtained from models that included both sexes (associated with PDS at the age of 12).

To investigate sex differences utilising the twin modelling method, we first examined qualitative sex differences (i.e., whether different sets of genes impact CpG methylation in males and females). We also performed an omnibus model on all 14 CpGs included in the twin analyses to determine whether there are any quantitative sex differences (i.e., whether parameter estimates differ between males and females). Qualitative sex differences existed for none of the 14 CpGs, indicating that the same additive genetic influences impact methylation in both males and females. We obtained ten CpG sites (9 based on AE models and 1 ACE model) with quantitative sex differences (Table 3, Supplementary Table S4 - Additional file 3). Three CpGs exhibited significant differences in variance compositions. For both cg09179916 and cg02614024 (associated with PA in females and PDS at 14 in females, respectively), the additive genetic component was larger in females, and the nonshared environmental component was larger in males, but total variance did not differ between the sexes. For cg15030014 (associated with PDS at age 12 in both sexes), both the additive genetic component and the total phenotypic variance were larger in males. In general, we identified sex differences in both methylation and its underlying variation in CpGs associated with puberty in one or both sexes.

Enrichment Analysis

We obtained relevant canonical pathways and associated diseases/functions using IPA software on PDS- and PA-associated CpG sites with an effect size larger than the absolute value of 0.13 (see Supplementary Table S5 and S6 for the CpG sites - Additional file 3). CpGs associated with PDS at the age of 14 in males were enriched in sperm motility, axonal guidance signalling and synaptogenesis signalling canonical pathways, while in females, the pathway results were less conclusive. Full results of the canonical pathways from IPA for all models can be seen in Supplementary Table S7 (see Additional file 3).

CpGs associated with PA in males were enriched in adipogenesis, choline degradation and thyroid hormone biosynthesis canonical pathways, while in females, they were enriched in tyrosine degradation and choline biosynthesis. Across all sex-specific models, associated CpGs were linked to physiological system development, such as growth of muscle tissue or development of neural cells in PDS 14 models or connective tissue development and tissue morphology in PA (Supplementary Table S8 - Additional file 3). Breast, prostate and ovarian cancers were universally found to be linked to CpGs in the sex-stratified models on PDS both at ages 12 and 14. Intestinal and thyroid cancers were highly significantly linked diseases in CpGs associated with PDS at age 14 in males. In females, the highly significant diseases were cancer of the head (umbrella term) and gastric cancer (Supplementary Table S8 - Additional file 3).

To explore whether the SNPs underlying the methylation associations with puberty are enriched in any relevant processes and functions, we performed IPA on meQTLs of all 39 CpG sites. IPA was performed on 13787 unique meQTLs, and they were enriched in several immune pathways, such as antigen presentation and the T-helper 1 and T-helper 2 pathways, usually driven by genes linked to the human leukocyte antigen (HLA) complex (e.g., *HLA-DQA2* and *HLA-DPA1*). The multiple sclerosis signalling pathway was one of the highly enriched canonical pathways among

the meQTLs (Supplementary Table S9 - Additional file 3). Concerning the diseases and functions category of IPA, several diseases were highlighted, such as insulin-dependent diabetes mellitus, rheumatoid arthritis, and endocrine, thyroid and breast cancers.

DISCUSSION

In a sizable cohort of Finnish twins, we identified methylation values at 58 CpG sites to be positively or negatively associated with PDS or PA when assessed among individuals in early adulthood. We observed both sex-specific associations and associations common to both sexes, with overall stronger associations in sex-stratified models. On average, we identified stronger and more numerous associations with PDS than PA. Moreover, relatively few CpGs overlapped between any of the models.

We considered PDS at age 14 of particular importance; compared to the models at age 12, it is generally closer to pubertal development peak (44), while PA was reported retrospectively, making it more prone to reporting errors. Additionally, CpG sites with methylation associated with PDS at age 14 were frequently located in development-related genes. For instance, male PDS-related hypomethylated cg19238380 is located in the *LMNA* gene, which plays a critical role in the normal development of the peripheral nervous system and skeletal muscle (45,46). Additionally, the CpG site cg03551401, where methylation is negatively associated with PDS in females, is located on the *ADCY8* gene, which contributes to several brain functions, including learning, memory, and drug addiction (47).

Three CpGs, with an effect size larger than the absolute value of 0.13 but below the specified p value thresholds, were also found in two studies that investigated genome-wide DNA methylation profiles in peripheral blood before and after puberty (16,17). The three CpG sites are cg01794929, cg21361322 and cg15028548 (the former two mapped to the *SELM* gene and the latter to *ABI3BP*) and were associated in our study with PDS at 14 in females, PDS at 12 in females and PDS at 14 in males, respectively. Notably, the main difference between our study and the two mentioned studies is that we explored the methylation profiles of young adults explained by pubertal timing and development, while they investigated methylation profiles before and after puberty. This difference is reflected in the observed opposite directionality of effect sizes for the 3 CpG sites.

Overall, canonical pathway enrichment results when both sexes were considered together were less conclusive than sex-specific pathway analyses, in which multiple CpG sites were enriched in endocrine or neuronal development-related pathways. Additionally, an antagonistic enrichment related to choline and tyrosine metabolism was found between males and females among the CpG sites associated with PA. Furthermore, breast, ovarian and prostate cancers were linked to CpG sites across all models, which is consistent with studies showing associations between early puberty and these diseases (1,9). Previously reported diseases related to early puberty, such as endometrial cancer (1,9) or cardiovascular diseases (3), were also linked to CpG sites in some of the models in the current study, such as in the sex-stratified PDS at age 14.

In addition to characterising the genes relevant to the identified CpG sites, we also examined their underlying sources of variation via twin modelling, as both puberty (4–7) and DNA methylation are known to be heritable traits (18,19). The CpG sites reported in this study, which

were predominantly best modelled with additive genetic and unique environmental sources of variation, were mostly found to be highly heritable. Generally, methylation values of these CpGs were more heritable than pubertal development or PA themselves, with heritability ranging between 0.51-0.98 for AE CpGs, while for age at menarche or age at voice change, it ranges from 0.50-0.59 (4,5,48). Methylation at only one CpG site (cg15030014), located in the *GET4* gene, showed strong influence from the environment. Interestingly, *GET4* has an enhancer with a SNP (rs9690350) mapped to the gene region of *PDGFA*, which is associated with pubertal onset (49). The same CpG, as well as 47 other significant CpGs, are associated with age (50), which supports our findings on associations with pubertal timing. Additionally, they could be considered markers for epigenetic age acceleration in this critical period of life. It is unclear to what extent the high heritability of CpGs observed here may reflect genetic mechanisms that influence pubertal development, given that the blood sampling occurred after the completion of puberty.

To assess in more detail the potential genetic effects on puberty-associated CpG sites, we performed meQTL analyses. A total of 4507 unique SNPs associated with age of menarche in two large-scale genome-wide association studies (GWAS) (1,9) were identified as meQTLs in our study. They could provide a putative mechanistic explanation for PDS- or PA-associated SNPs identified in the GWA studies, as DNA methylation may impart a mechanistic link for these SNPs and explain their association with age at menarche. A nonzero proportion of SNPs reported by the two studies was found in all our designs, barring PDS at age 14 in males and females combined. The percentage of meQTLs associated with CpGs from the eight models, also found among the 4507 SNPs, ranged from 21.1% to 43.2% (26-1289 meQTLs). The meQTLs of highly heritable CpGs also had a substantial proportion of common meQTLs with GWAS on puberty, ranging from 1% to 36.3% (16-465 meQTLs). Hence, these meQTLs could, in part, reflect genetic mechanisms that influence pubertal development.

Pathway-level analyses on the meQTLs of CpG sites associated with puberty revealed an enrichment in HLA complex-related immune pathways such as antigen presentation and T-helper pathways. A positive association between HLA complex heterozygosity and later pubertal development was reported and was hypothesised to be an evolutionary trade-off between immunocompetence reflected in pathogen resistance and sexual maturation (51). Moreover, an increase in antigen presentation and T-helper responses has been reported as part of immune changes during pubertal development (52). Thus, pathway analyses on meQTLs underlying the methylation associations with puberty demonstrate their involvement in immune processes fully developed during puberty, which might be relevant in terms of susceptibility to various diseases later in life. Furthermore, the IPA of meQTLs revealed some of the same linked diseases as our IPA of differentially methylated CpG sites, such as endocrine, breast and thyroid cancers. Interestingly, the enriched immune pathways were also associated with the central nervous system and could be linked to susceptibility to multiple sclerosis (52), which was strongly enriched in the meQTL IPA analysis.

To rule out the effects of shared genotypes and environment on the observed associations, we performed within-pair analyses. We observed 6 CpG sites with their methylation differing within the pairs, indicating a causal relationship between PDS or PA and methylation at these CpG sites. Two of these CpGs were positively related to PDS, and three were negatively related to PDS, while only one CpG site (cg06096446) with high heritability, whose mapped gene

(*TULP4*) is associated with height (53), was positively related to PA. The same CpG site has also been associated with alcohol consumption and disorders in pregnancy (54,55). One of the negatively associated CpGs with a significant within-pair methylation difference (cg08091771) has been associated with maternal BMI and nitrogen dioxide exposure (56,57) and is mapped on *PTPRG*, which is associated with precocious puberty (58). These effects are consistent with a causal environmentally mediated relationship between earlier pubertal development and methylation, although our analyses cannot conclusively determine the direction of causality (i.e., whether methylation causes earlier puberty or whether earlier puberty causes methylation). The rest of the puberty-associated CpG sites did not differ within the pairs, suggesting that the relationship between PDS/PA and CpG methylation is due to genetic or environmental confounding.

Twin modelling results on quantitative sex differences and differential methylation results were concordant: all CpGs identified as having sex differences in the mean structure were also differentially methylated between the sexes. Furthermore, for CpGs with sex differences in variance, the sex with a larger variance from additive genetic sources was also the sex for which the EWAS association was identified. For example, cg09179916 was associated with PA in females, and females had a larger additive genetic component. This convergence of results could suggest a mechanistic role, but further analyses are necessary to determine if methylation at these CpG sites has a functional role related to puberty. In addition to the mechanistic role, the observed sex differences emphasise the importance of sex-stratified models for epigenetic studies of puberty.

We found an overall high heritability across the PDS- and PA-associated CpG sites, which for 6 CpGs was likely due to direct SNP effects observed as meQTLs. However, the methylation discordance within twin pairs revealed that, for some CpG sites, despite their high heritability, environmental factors play an essential role in shaping their methylation, possibly indicating the presence of gene-environment interactions. Finally, the observed sex differences indicate that the same genetic mechanisms, but with different magnitudes in males and females, affect the heritability of CpG sites associated with puberty, giving additional importance to investigating meQTLs in this context.

Our study has both strengths and limitations. The key strengths of our study are the relatively large sample size as well as the inclusion of twin pairs with longitudinal phenotype data on which we were able to separate genetic effects from environmental effects on methylation at the puberty-associated CpG sites. Importantly, previous studies reported no substantial difference in PA between twins and singletons (7,59). As the EWAS was performed on blood samples several years after puberty was completed, the main limitation of this study is that it is unclear whether the methylation profile associated with PDS or PA was formed before, during, or after puberty. Although not revealing clear causal characteristics, the methylation profile in the peripheral blood of young adults is associated with pubertal timing and development, making the associated genes, enriched pathways and linked diseases potential intriguing subjects for future investigation. Furthermore, incorrect reporting of pubertal age, especially in males, could be a source of noise and inconsistency within the analyses (60). In addition, while the self-reported PDS is not the gold standard for measuring pubertal development, it is considered reliable (61). Furthermore, clinical assessments, such as Tanner staging (62,63), are too costly for study designs and sample sizes such as ours. Another limitation was the use of two different

methylation profiling platforms, reflecting the technological development of array-based methylation analysis. This resulted in high heterogeneity of some of the CpG sites associated with PDS, denoting perhaps a less reliable association. Nevertheless, the overall heterogeneity between the 450K and EPIC platforms was very low.

CONCLUSIONS

By identifying CpG sites associated with PDS and PA, we found potential genes, pathways and diseases relevant to puberty in an epigenetic context and complemented the previously reported genetic characterisation of pubertal timing. Our comprehensive analyses on puberty-related CpGs, such as twin modelling, assessment of environmentally driven associations or genetics underlying the association, highlight the role of genetics and unique environment in pubertal timing and development. In addition, we provide evidence for DNA methylation being a putative mechanistic link between genotypes and puberty, as well as puberty-related diseases. Furthermore, our within-pair effects are consistent with a causal environmentally mediated relationship between pubertal development and methylation. To further advance our understanding, it will be essential to utilise longitudinal data from multiple cohorts, such as the Adolescent Brain Cognitive Development Study. By doing so, we could acquire additional insights into the chronological methylation patterns of CpGs associated with PDS and PA, enabling us to deepen our knowledge of the intricate mechanisms underlying puberty and its implications for health and disease.

DECLARATIONS

Ethics approval and consent to participate

The Finnish Twin Cohorts' data collection and analysis were approved by the Ethics Committee of the Helsinki University Central Hospital (Dnro 249/E5/01, 270/13/03/01/2008, 154/13/03/00/2011). Written informed consent was provided by the participants prior to data collection.

Consent for publication

Not applicable

Availability of data and materials

The Finnish Twin Cohort datasets used in the current study are located in the Biobank of the Finnish Institute for Health and Welfare, Helsinki, Finland. All biobanked data are publicly available for use by qualified researchers following a standardised application procedure (<https://thl.fi/en/web/thl-biobank/for-researchers>).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme, Marie Skłodowska-Curie (grant number 859860). Data collection in the FinnTwin12 cohort has been supported by the Wellcome Trust Sanger Institute, the Broad Institute, ENGAGE – European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4-2007, grant agreement number 201413, National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to R J Rose; AA15416 and K02AA018755 to D M Dick; R01AA015416 to Jessica Salvatore), the Academy of Finland (grants 100499, 205585, 118555, 141054, 264146, 308248 to JK, and grants 328685, 307339, 297908 and 251316 to MO, and the Centre of Excellence in Complex Disease Genetics grants 312073, 336823, and 352792 to JK), and Sigrid Juselius Foundation (to MO).

Authors' contributions

MO, JK and ES conceptualised the study. ES performed the data analysis, was involved in the visualisation of results, and wrote the first draft of the manuscript. SMZ performed the twin analyses and participated in writing. MKY was involved in data analysis and visualisation. AH performed data preprocessing and normalisation and contributed to data analysis. MO and JK acquired the data, supervised the work and participated in writing. All authors critically revised the manuscript for important intellectual content and read and approved the final manuscript.

Acknowledgements

We express our gratitude to Teemu Palviainen for his assistance with data management and curation and to Mikaela Hukkanen for her advice and discussions on IPA. We also thank Dr. Giovanna Chiorino for her invaluable advice and insightful discussions. Finally, we thank the FIMM Technology Centre supported by HiLIFE and Biocenter Finland for DNA methylation data generation services, as well as all FinnTwin12 and FinnTwin16 participants and their families included in this study.

REFERENCES

1. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet.* 2017 Jun;49(6):834–41.
2. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet Oncology.* 2012 Nov;13(11):1141–51.
3. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep.* 2015 Sep;5(1):11208.
4. Anderson CA, Duffy DL, Martin NG, Visscher PM. Estimation of Variance Components for Age at Menarche in Twin Families. *Behav Genet.* 2007 Oct 3;37(5):668–77.
5. Hur YM. Common Genetic Influences on Age at Pubertal Voice Change and BMI in Male Twins. *Twin Res Hum Genet.* 2020 Aug;23(4):235–40.

6. Silventoinen K, Haukka J, Dunkel L, Tynelius P, Rasmussen F. Genetics of Pubertal Timing and Its Associations With Relative Weight in Childhood and Adult Height: The Swedish Young Male Twins Study. *Pediatrics*. 2008 Apr 1;121(4):e885–91.
7. Kaprio J, Rimpelä A, Winter T, Viken RJ, Rimpelä M, Rose RJ. Common genetic influences on BMI and age at menarche. *Hum Biol*. 1995 Oct;67(5):739–53.
8. Day FR, Bulik-Sullivan B, Hinds DA, Finucane HK, Murabito JM, Tung JY, et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun*. 2015 Dec;6(1):8842.
9. Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*. 2014 Oct;514(7520):92–7.
10. Papadimitriou A. The Evolution of the Age at Menarche from Prehistorical to Modern Times. *Journal of Pediatric and Adolescent Gynecology*. 2016 Dec;29(6):527–30.
11. Eckert-Lind C, Busch AS, Petersen JH, Biro FM, Butler G, Bräuner EV, et al. Worldwide Secular Trends in Age at Pubertal Onset Assessed by Breast Development Among Girls: A Systematic Review and Meta-analysis. *JAMA Pediatr*. 2020 Apr 6;174(4):e195881.
12. Fisher MM, Eugster EA. What is in our environment that effects puberty? *Reproductive Toxicology*. 2014 Apr;44:7–14.
13. Silventoinen K, Jelenkovic A, Palviainen T, Dunkel L, Kaprio J. The Association Between Puberty Timing and Body Mass Index in a Longitudinal Setting: The Contribution of Genetic Factors. *Behav Genet*. 2022 May;52(3):186–94.
14. Cousminer DL. Pubertal timing and body mass: Genes involved. *Current Opinion in Endocrine and Metabolic Research*. 2020 Oct;14:117–26.
15. Han L, Zhang H, Kaushal A, Rezwani FI, Kadalayil L, Karmaus W, et al. Changes in DNA methylation from pre- to post-adolescence are associated with pubertal exposures. *Clin Epigenet*. 2019 Dec;11(1):176.
16. Almstrup K, Lindhardt Johansen M, Busch AS, Hagen CP, Nielsen JE, Petersen JH, et al. Pubertal development in healthy children is mirrored by DNA methylation patterns in peripheral blood. *Sci Rep*. 2016 Jun 30;6(1):28657.
17. Thompson EE, Nicodemus-Johnson J, Kim KW, Gern JE, Jackson DJ, Lemanske RF, et al. Global DNA methylation changes spanning puberty are near predicted estrogen-responsive genes and enriched for genes involved in endocrine and immune processes. *Clin Epigenet*. 2018 Dec;10(1):62.
18. van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016 Apr 7;7(1):11115.
19. Villicaña S, Bell JT. Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol*. 2021 Apr 30;22(1):127.

20. Kaidesoja M, Aaltonen S, Bogl LH, Heikkilä K, Kaartinen S, Kujala UM, et al. FinnTwin16: A Longitudinal Study from Age 16 of a Population-Based Finnish Twin Cohort. *Twin Res Hum Genet.* 2019 Dec;22(6):530–9.
21. Rose RJ, Salvatore JE, Aaltonen S, Barr PB, Bogl LH, Byers HA, et al. FinnTwin12 Cohort: An Updated Review. *Twin Res Hum Genet.* 2019 Oct;22(5):302–11.
22. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolescence.* 1988 Apr;17(2):117–33.
23. Stephenson M, Bollepalli S, Cazaly E, Salvatore JE, Barr P, Rose RJ, et al. Associations of Alcohol Consumption With Epigenome-Wide DNA Methylation and Epigenetic Age Acceleration: Individual-Level and Co-twin Comparison Analyses. *Alcohol Clin Exp Res.* 2021 Feb;45(2):318–28.
24. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics.* 2018 Dec 1;34(23):3983–9.
25. Chen Yan, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013 Feb;8(2):203–9.
26. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 2017 Feb 28;45(4):e22.
27. Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013 Dec;14(1):293.
28. R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2022.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.* 2015 Apr 20;43(7):e47–e47.
30. Salas LA, Koestler DC. FlowSorted.Blood.EPIC [Internet]. Bioconductor; 2018 [cited 2022 Nov 16]. Available from: <https://bioconductor.org/packages/FlowSorted.Blood.EPIC>
31. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* 2018 Dec;19(1):64.
32. the BIOS Consortium, van Iterson M, van Zwet EW, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017 Dec;18(1):19.
33. D. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *JOSS.* 2018 May 19;3(25):731.
34. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software.* 2010 Aug 5;36:1–48.

35. Aloe AM, Thompson CG. The Synthesis of Partial Effect Sizes. *Journal of the Society for Social Work and Research*. 2013 Jan;4(4):390–405.
36. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. *Genet Epidemiol*. 2018 Feb;42(1):20–33.
37. Neumann A, Walton E, Alemany S, Cecil C, González JR, Jima DD, et al. Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. *Transl Psychiatry*. 2020 Dec;10(1):398.
38. Hansen KD. IlluminaHumanMethylation450kanno.ilmn12.hg19 [Internet]. Bioconductor; 2017 [cited 2022 Nov 23]. Available from: <https://bioconductor.org/packages/IlluminaHumanMethylation450kanno.ilmn12.hg19>
39. Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, et al. OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*. 2016 Jun;81(2):535–49.
40. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 2021 Sep;53(9):1311–21.
41. Genetics of DNA Methylation Consortium. Cis and trans meta-analysis results from genome-wide scans of 420,509 DNA methylation sites [Internet]. [cited 2023 Feb 26]. Available from: <http://mqtl.db.godmc.org.uk>
42. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014 Feb 15;30(4):523–30.
43. Dick DM, Rose RJ, Pulkkinen L, Kaprio J. Measuring Puberty and Understanding Its Impact: A Longitudinal Study of Adolescent Twins. *Journal of Youth and Adolescence*. 2001 Aug;30(4):385–99.
44. Wehkalampi K, Silventoinen K, Kaprio J, Dick DM, Rose RJ, Pulkkinen L, et al. Genetic and environmental influences on pubertal timing assessed by height growth. *Am J Hum Biol*. 2008 Jul;20(4):417–23.
45. Muchir A. Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). *Human Molecular Genetics*. 2000 May 22;9(9):1453–9.
46. De Sandre-Giovannoli A, Chaouch M, Kozlov S, Vallat JM, Tazir M, Kassouri N, et al. Homozygous Defects in LMNA, Encoding Lamin A/C Nuclear-Envelope Proteins, Cause Autosomal Recessive Axonal Neuropathy in Human (Charcot-Marie-Tooth Disorder Type 2) and Mouse. *The American Journal of Human Genetics*. 2002 Mar;70(3):726–36.
47. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D480–9.

48. Kretsch N, Mendle J, Harden KP. A Twin Study of Objective and Subjective Pubertal Timing and Peer Influence on Risk-Taking. *J Res Adolesc.* 2016 Mar;26(1):45–59.
49. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *CP in Bioinformatics [Internet].* 2016 Jun [cited 2023 Mar 10];54(1). Available from: <https://onlinelibrary.wiley.com/doi/10.1002/cpbi.5>
50. Battram T, Yousefi P, Crawford G, Prince C, Sheikhal Babaei M, Sharp G, et al. The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res.* 2022 May 31;7:41.
51. Arnocky S, Hodges-Simeon C, Davis AC, Desmarais R, Greenshields A, Liwski R, et al. Heterozygosity of the major histocompatibility complex predicts later self-reported pubertal maturation in men. *Sci Rep.* 2021 Oct 6;11(1):19862.
52. Ucciferri CC, Dunn SE. Effect of puberty on the immune system: Relevance to multiple sclerosis. *Front Pediatr.* 2022 Dec 2;10:1059083.
53. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research.* 2023 Jan 6;51(D1):D977–85.
54. Kazmi N, Sharp GC, Reese SE, Vehmeijer FO, Lahti J, Page CM, et al. Hypertensive Disorders of Pregnancy and DNA Methylation in Newborns. *Hypertension.* 2019 Aug;74(2):375–83.
55. Dugué PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH, et al. Alcohol consumption is associated with widespread changes in blood DNA methylation: Analysis of cross-sectional and longitudinal data. *Addict Biol.* 2021 Jan;26(1):e12855.
56. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM, et al. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Hum Mol Genet.* 2017 Oct 15;26(20):4067–85.
57. de F C Lichtenfels AJ, van der Plaat DA, de Jong K, van Diemen CC, Postma DS, Nedeljkovic I, et al. Long-term Air Pollution Exposure, Genome-wide DNA Methylation and Lung Function in the LifeLines Cohort Study. *Environ Health Perspect.* 2018 Feb 6;126(2):027004.
58. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database.* 2016;2016:baw100.
59. Willemsen G, Odintsova V, de Geus E, Boomsma DI. Twin-Singleton Comparisons Across Multiple Domains of Life. In: Khalil A, Lewi L, Lopriore E, editors. *Twin and Higher-order Pregnancies [Internet].* Cham: Springer International Publishing; 2021 [cited 2023 Jan 12]. p. 51–71. Available from: https://link.springer.com/10.1007/978-3-030-47652-6_4
60. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, Hagen CP, Tinggaard J, Mouritsen A, et al. Validity of Self-Assessment of Pubertal Maturation. *Pediatrics.* 2015 Jan 1;135(1):86–93.
61. Koopman-Verhoeff ME, Gredvig-Ardito C, Barker DH, Saletin JM, Carskadon MA. Classifying Pubertal Development Using Child and Parent Report: Comparing the Pubertal Development Scales to Tanner Staging. *Journal of Adolescent Health.* 2020 May;66(5):597–602.

62. Tanner JM, Whitehouse RH. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood*. 1976 Mar 1;51(3):170–9.
63. van den Berg SM, Setiawan A, Bartels M, Polderman TJC, van der Vaart AW, Boomsma DI. Individual Differences in Puberty Onset in Girls: Bayesian Estimation of Heritabilities and Genetic Correlations. *Behav Genet*. 2006 Mar;36(2):261.

Additional Files

Additional File 1.docx - Supplementary methods on EWAS designs and twin modelling.

Additional File 2.pdf - Supplementary Figure S1. Manhattan plots on p values of meta-analysed EWAS models on 450K and EPIC platforms on PDS at age 12 A) in males and B) females, C) combined, on PDS at age 14 in D) males and females combined, and on PA in E) males and F) females.

Supplementary Figure S2. QQ plots on p values of meta-analysed EWAS models on 450K and EPIC platforms on PDS at age 12 A) in males and B) females, C) combined, on PDS at age 14 in D) males and females combined, and on PA in E) males and F) females.

Additional File 3.xlsx - Table S1. Sample sizes and final numbers of complete twin pairs included in 450K and EPIC platforms for EWAS designs on PDS and PA. **Table S2.** Details on univariate twin modelling results on the 14 CpG sites. The estimates for total phenotypic variance and mean, as well as the coefficients of the covariates (age and sex), are shown. **Table S3.** Individual associations and within-twin pair methylation differences of the 58 CpG sites identified in the EWAS. **Table S4.** Results of qualitative and quantitative sex differences for omnibus tests and specific parameters on CpG sites. **Table S5.** CpG sites associated with pubertal development scale (PDS) with a standardised effect size $> |0.13|$ stratified by the model. **Table S6.** CpG sites associated with pubertal age (PA) in males or females with a standardised effect size $> |0.13|$. **Table S7.** CpG sites associated with pubertal development scale (PDS) or pubertal age (PA) with a standardised effect size $> |0.13|$ enriched in canonical pathways from the Ingenuity Pathway Analysis. **Table S8.** CpG sites associated with pubertal development scale (PDS) or pubertal age (PA) with a standardised effect size $> |0.13|$ enriched in diseases and functions from the Ingenuity Pathway Analysis. **Table S9.** IPA results on meQTLs ($p < 5 \times 10^{-8}$) of CpG sites associated with pubertal development scale (PDS) or pubertal age (PA). Both the Ingenuity Canonical pathways and diseases/functions annotations are included.

Table 1. Characteristics of pubertal development scale (PDS) and pubertal age (PA), and covariates in the EWAS

A.	FinnTwin12 age 12			FinnTwin12 age 14					
	All	Males	Females	All	Males	Females			
PDS and its items									
Height growth spurt*	1.91; 0.80	1.79; 0.77	2.01; 0.81	2.43; 1.08	2.24; 0.80	2.57; 1.22			
Body hair growth*	1.56; 0.66	1.31; 0.51	1.74; 0.70	2.95; 0.72	2.60; 0.69	3.20; 0.63			
Skin changes*	1.49; 0.55	1.31; 0.48	1.63; 0.56	2.03; 0.53	1.94; 0.52	2.10; 0.53			
Breast development/ Voice change*	1.56; 0.67	1.05; 0.22	1.94; 0.64	2.56; 0.81	2.14; 0.95	2.86; 0.50			
Menarche/ Hair growth on face*	1.07; 0.43	1.02; 0.14	1.10; 0.55	2.50; 1.44	1.24; 0.46	3.42; 1.19			
PDS‡	1.52; 0.41	1.29; 0.26	1.68; 0.43	2.49; 0.65	2.03; 0.49	2.83; 0.54			
Age at questionnaire	11.41; 0.29	11.42; 0.28	11.41; 0.29	14.04; 0.08	14.04; 0.09	14.04; 0.07			
B.	Complete			FinnTwin12			FinnTwin16		
	All	Males	Females	All	Males	Females	All	Males	Females
PA and covariates in EWAS models									
Pubertal age	13.28; 1.30	13.93; 1.16	12.75; 1.16	13.27; 1.29	14.00; 1.11	12.75; 1.14	13.30; 1.31	13.85; 1.20	12.76; 1.19
Age at DNAm	25.17; 4.15	25.40; 4.06	24.96; 4.21	22.73; 1.71	22.77; 1.72	22.71; 1.71	28.07; 4.33	27.98; 4.04	28.16; 4.61
EtOH g/day	11.21; 15.33	15.48; 18.77	7.49; 10.14	11.20; 14.49	16.48; 18.39	7.28; 8.88	11.23; 16.29	14.47; 19.10	7.79; 11.74
Smoking (Never; Current;Former)	849; 572; 382	365; 297; 181	484; 275; 201	481; 352; 148	187; 172; 59	294; 180; 89	368; 220; 234	178; 125; 122	190; 95; 112

Values are given as the mean; SD, except for smoking.

*Self-reported evaluation of PDS items with the values ranging from 1 - 3 at age 12 and from 1 - 4 at age 14.

‡PDS was calculated as the sum of its 5 items divided by 5

Table 2. CpG sites significantly associated ($p < 1 \times 10^{-5}$) with pubertal development scale (PDS) and pubertal age

(PA)

Original model and CpG ID	Standardised Effect size*	P value	Q p value‡	Chr.	Gene	Mean (M)†	Mean (F)†	Wilcoxon two- sample test (p)§
PDS 12								
cg15030014	-0.175	2.85E-07	1.000	7	C7orf20	0.886	0.890	1.000
cg25614726	-0.169	7.61E-07	1.000	1	APITD1	0.378	0.399	2.31E-12
cg08273672	-0.161	2.59E-06	1.000	4	UCHL1	0.754	0.760	0.131
cg05038881	-0.161	2.65E-06	1.000	1	APITD1	0.430	0.451	3.32E-09
cg26027669	0.159	3.38E-06	1.000	1	GFI1	0.054	0.048	9.03E-14
cg22064581	0.156	5.30E-06	1.000	6		0.032	0.031	1.000
cg08091771	-0.156	5.57E-06	1.000	3	PTPRG	0.705	0.708	1.000
cg26489369	-0.155	6.19E-06	1.000	6	HLA-DQA2	0.950	0.952	0.228
cg06993494	-0.153	8.26E-06	1.000	11	OR5A1	0.966	0.967	0.132
PDS 12 - Males								
cg25801636	0.245	8.46E-07	0.007	2		0.949	0.949	1.000
cg13653963	0.242	9.10E-07	0.001	9	C9orf98	0.965	0.967	1.11E-04
cg14609721	0.249	1.04E-06	0.750	1	DEPDC1	0.019	0.018	0.098
cg03468249	0.250	1.16E-06	0.932	2		0.026	0.025	1.000
cg09736922	-0.245	1.99E-06	0.960	3	THPO	0.753	0.757	1.000
cg22539434	-0.246	2.03E-06	1.000	7	GLI3	0.963	0.962	0.937
cg15680660	-0.246	2.05E-06	1.000	3	XYLB	0.949	0.947	1.000
cg04547799	-0.234	5.80E-06	0.860	6	HMGN3	0.029	0.030	0.001
cg05797363	0.232	6.16E-06	0.700	1	DEPDC1	0.016	0.016	1.000
cg04052235	0.231	7.46E-06	0.722	11		0.422	0.433	1.49E-05
cg06002616	0.225	8.37E-06	0.039	8	SPAG1	0.074	0.071	1.000
cg04769618	0.230	8.70E-06	0.874	10	LOC387647	0.033	0.034	1.000
cg14685946	-0.222	8.73E-06	0.002	2	ACSL3	0.924	0.925	1.000
PDS 12 - Females								

cg07581365	-0.233	1.54E-07	1.000	17	CBX8	0.031	0.031	1.000
cg06988547	-0.231	1.87E-07	1.000	1		0.909	0.910	1.000
cg18397726	-0.217	1.38E-06	1.000	14	CHGA	0.949	0.949	0.016
cg07844563	-0.206	2.25E-06	0.001	1	LOC653566	0.959	0.956	0.038
cg08883485	-0.209	2.80E-06	0.660	1	NAV1	0.452	0.473	4.36E-05
cg23204278	0.211	2.87E-06	1.000	18		0.012	0.012	1.000
cg10669219	-0.207	4.34E-06	1.000	5		0.822	0.827	0.863
cg19237923	-0.204	5.39E-06	0.710	1	NTRK1	0.671	0.699	7.53E-39
cg09550127	-0.203	6.90E-06	1.000	6		0.929	0.936	1.24E-04
cg17052466	-0.201	8.25E-06	1.000	11	TSSC4	0.952	0.956	1.41E-06
PDS 14 - Males								
cg04239863	-0.273	8.15E-09	9.38E-09	16		0.831	0.841	0.004
cg03115019	-0.256	5.28E-07	0.040	17	FN3K	0.837	0.848	0.012
cg22325292	-0.252	7.82E-07	0.038	17	FN3K	0.719	0.729	1.000
cg15142913	-0.259	7.88E-07	0.952	8	PTP4A3	0.929	0.931	1.000
cg00008671	-0.247	2.06E-06	0.350	10	FAM190B	0.238	0.228	2.57E-04
cg14466923	0.232	3.40E-06	1.73E-04	6	COL11A2	0.943	0.943	1.000
cg21759953	0.242	5.09E-06	1.000	2	GLI2	0.917	0.918	1.000
cg18223453	0.227	6.73E-06	0.001	17	TBX4	0.682	0.735	3.59E-82
cg02763540	-0.235	8.71E-06	0.922	19	UHRF1	0.804	0.812	0.049
cg19238380	-0.235	9.49E-06	0.969	1	LMNA	0.473	0.487	1.63E-09
cg24786174	-0.236	9.83E-06	1.000	18	ZNF516	0.542	0.528	1.000
cg04230403	-0.224	9.98E-06	0.001	10	FAS	0.026	0.026	1.000
PDS 14 - Females								
cg20599748	-0.238	7.39E-08	0.390	14	SERPINA10	0.953	0.955	1.000
cg13634653	-0.212	1.70E-06	0.047	15		0.969	0.970	0.002
cg17965926	0.203	2.38E-06	5.18E-06	8		0.031	0.031	1.000
cg23130651	-0.203	2.40E-06	3.14E-06	10	COMMD3	0.883	0.871	9.54E-11
cg03551401	-0.208	3.46E-06	0.286	8	ADCY8	0.457	0.512	8.35E-53

cg00083685	0.206	3.62E-06	0.114	21	TFF2	0.924	0.927	0.497
cg16198723	-0.202	3.96E-06	0.001	3	CCDC80	0.600	0.598	1.000
cg02614024	-0.207	4.18E-06	0.733	15		0.194	0.200	1.000
cg23972915	-0.201	4.27E-06	4.14E-04	14		0.643	0.648	0.509
cg03351418	0.202	8.55E-06	1.000	6	SHPRH	0.841	0.842	1.000
PA - Males								
cg06096446	0.204	1.22E-07	1.000	6	TULP4	0.719	0.697	4.02E-07
cg09303847	0.173	9.85E-06	1.000	6		0.042	0.033	1.92E-20
PA - Females								
cg17382841	-0.166	2.74E-06	1.000	5	EGFLAM	0.807	0.792	0.001
cg09179916	-0.163	3.96E-06	1.000	5	EGFLAM	0.919	0.911	0.004

*Standardised effect sizes ranging from -1 to 1 obtained from the fixed-effect meta-analysis of 450K and EPIC datasets

‡Benjamini-Hochberg corrected p value of the heterogeneity test statistic (Q) on the meta-analysis between 450K and EPIC platforms

†Methylation Mean Beta Value

§The p value of the Wilcoxon two-sample test was corrected using the Bonferroni method

Table 3. Univariate and sex differences twin modelling of 14 CpG sites meeting the twin modelling assumptions.

CpG Name	MZ Twin Correlation	DZ Twin Correlation	Mean	SD	Best Model	Heritability [95% CI]	Shared Environment [95% CI]	Unique Environment [95% CI]	Sex Differences Identified
cg02614024	0.69	0.29	0.19	0.06	AE	0.67 [0.61, 0.72]	-	0.33 [0.28, 0.39]	Unstandardised genetic and environmental variance components
cg03115019	0.96	0.50	0.85	0.11	AE	0.96 [0.95, NA]	-	0.04 [NA, 0.05]	None
cg03551401	0.55	0.33	0.49	0.07	AE	0.51 [0.42, 0.58]	-	0.49 [0.42, 0.58]	Mean structure
cg05038881	0.83	0.44	0.44	0.06	AE	0.84 [NA, NA]	-	0.16 [0.13, NA]	Mean structure

cg06096446	0.89	0.48	0.70	0.09	AE	0.88 [0.86, 0.90]	-	0.12 [0.10, 0.14]	Mean structure
cg08883485	0.94	0.52	0.46	0.08	AE	0.94 [0.92, NA]	-	0.06 [NA, 0.08]	Mean structure
cg09179916	0.61	0.18	0.92	0.06	AE	0.60 [0.53, 0.67]	-	0.40 [0.33, 0.47]	Mean Structure, unstandardised genetic and environmental variance components
cg15030014	0.89	0.78	0.89	0.06	ACE	0.24 [0.17, 0.33]	0.68 [NA, 0.75]	0.08 [0.07, 0.10]	Unstandardised genetic variance component and total phenotypic variance
cg16198723	0.79	0.36	0.60	0.05	AE	0.78 [0.73, NA]	-	0.22 [NA, 0.27]	None
cg17382841	0.61	0.28	0.80	0.07	AE	0.64 [0.56, 0.70]	-	0.36 [0.30, 0.44]	Mean structure
cg18223453	0.81	0.48	0.71	0.05	AE	0.77 [0.72, 0.81]	-	0.23 [0.19, 0.28]	Mean structure
cg22325292	0.98	0.53	0.73	0.19	AE	0.98 [0.97, 0.99]	-	0.02 [NA, 0.03]	None
cg24786174	0.97	0.55	0.53	0.14	AE	0.97 [0.96, NA]	-	0.03 [NA, 0.04]	None
cg25614726	0.77	0.42	0.39	0.05	AE	0.78 [0.74, NA]	-	0.22 [0.19, 0.26]	Mean structure

Best twin model, standardised variance components and potential identified sex differences are shown