Epigenome-wide analyses of well-being through direct epigenetic measurement and Mendelian Randomization
Abstract

We performed an epigenome wide association studies (EWAS) meta-analysis of well-being, where individual differences in CpG site methylation in whole blood are associated with individual differences in well-being. We control for two well-known confounders of epigenetic associations, smoking and BMI. However, we are aware of the effect that potential unmeasured confounders could have on our results as well as uncertain of the direction of causation of the association between well-being and CpG methylation. To guard against unmeasured confounding and to infer a direction of effect we perform Mendelian Randomization, specifically we perform summary-based Mendelian Randomization (SMR). We perform SMR in which the (cis) QTL effect of SNPs on methylation (cis-mQTL), and a large GWAS of wellbeing are combined to infer the (causal) effect of CpG methylation on well-being. We perform SMR leveraging cis-mQTLs discovered in both blood and brain tissues and compared results between tissues, and between SMR and EWAS. We found a high consistency of direction of effect ($r > .9$) between SMR results where the QTL is discovered in different whole blood datasets as well as high consistency between whole blood and fetal brain datasets ($r = .72$). However, when comparing the direction of effect between our EWAS and SMR results, no notable correlations were observed. Our results indicate that if the aim is to increase our understanding of the functional consequences of epigenetic changes on wellbeing, SMR may be preferred over EWAS in whole blood. If, however, the aim is to identify ways in which well-being Is itself a driver of environmental influences on differences in DNA methylation, possibly effecting gene-expression, a sufficient powered EWAS study will provide valuable information. The concurrent use of Mendelian Randomization and epigenome-wide association analysis proved to be a potent combination to further increase our understanding of the relation between well-being and CpG methylation.
Introduction

Well-being is linked to numerous determinants and behaviors across the life course, such as income and employment, health, neighborhood environment (e.g. green space), air pollution, smoking, stress, alcohol use and several social factors such as friendship patterns (see review Diener and colleagues\textsuperscript{22}). These exposures and behavioral characteristics are potential candidate drivers of differential epigenetic patterns between individuals having higher or lower levels of well-being. To our knowledge, our previous work is the only study that investigated the association between methylation differences and phenotypic variation in well-being ($N = 2,519$), reporting two CpG sites (cg10845147, $p = 1.51 \times 10^{-8}$ and cg01940273, $p = 2.34 \times 10^{-8}$) reached genome-wide significance after Bonferroni correction\textsuperscript{23}. Gene ontology (GO) analysis highlighted enrichment of several central nervous system categories among higher-ranking methylation sites. However, replication of these results is warranted in larger samples.

Despite these positive outcomes from the epigenome-wide association studies (EWAS), there are interpretational problems which may complicate distilling etiology and biology from epigenetic studies\textsuperscript{24–29}. The foremost interpretational difficulty is the uncertainty about cause and effect, e.g. does methylation causally influences complex trait outcomes, is the causal effect reverse, or does a third trait influences both methylation levels and traits? For instance, a recent study found using a stepwise Mendelian Randomization analysis, that differential methylation is the consequence of inter-individual variation in blood lipid levels and not vice versa\textsuperscript{30}. Considering the tissue-specific nature of epigenetic processes, a second important consideration for EWAS is the assessment of methylation of trait relevant tissue. Empirical results suggest that easily accessible tissues, such as whole blood, cannot be used to address questions about inter-individual epigenomic variation in inaccessible tissues, such as the brain. Hannon et al. explored covariation between tissues and found that, for the majority of the genome, a blood-based EWAS for traits where brain is presumed to be the primary tissue of interest will provide limited information relating to underlying processes\textsuperscript{31}. This finding is enforced by another study, which found that only 7.9% of CpG probes, obtained in a sample of epilepsy patients, showed a substantial and statistically significant correlation between blood and brain tissue\textsuperscript{32}.

However, there is reason for optimism, as recent studies utilizing GTEx data showed that the genetic correlation of gene expression between tissues in local regions (i.e. $< 1$MB of the
transcription start site) is much higher than in distal regions\textsuperscript{33,34}. This optimism is further supported by a recent study that found there is no evidence for tissue relevant eQTLs enrichment for associations with complex traits\textsuperscript{35}. In this context, the question arises whether this holds for methylation QTLs (mQTLs) and to what extent the cis-genetic effects on DNA methylation in blood differ from those in brain. Capitalizing on this strong, cross tissue local genetics effects on methylation levels, Zhu et al. developed summary-based Mendelian Randomization (SMR), to infer the effect of eQTLs and mQTLs on complex traits\textsuperscript{36}. 

Mendelian Randomization relies on the presence of genetic variants which confer a risk for an exposure of interest (CpG methylation in this case), as people do not self-select into a particular genotype group at birth, the genotype which indexes variation in CpG methylation can be considered random with respect to the outcome (well-being). Thus Mendelian Randomization offers a pseudo controlled experiment of the effect of variation in CpG methylation on well-being. SMR integrates summary-level data from GWAS together with data from eQTL or mQTL studies to identify genes whose expression levels, or CpG sites who’s methylation level, are associated with a complex trait due to pleiotropy. Pleiotropy in this case refers to a single causal variant underlying differences in gene expression/DNA methylation and phenotypic variation, which is of more biological interest than linkage, where, in the case of two distinct causal variants, one affects gene-expression or CpG methylation and the other trait variation. The observed Pleiotropy between a CpG site and a trait will likely be caused by an effect of CpG methylation on the outcome. A previous study\textsuperscript{37} leveraged SMR to infer the relationship between CpG methylation (either in blood or brain) in over 40 different complex traits. Their results were highly consistent between both tissues, specifically, because local cis genetic regulation of methylation does not differ strongly across blood and brain tissues. The cross tissue stability in cis-regulation was supported by a recent study\textsuperscript{38} that reported a high cis-genetic correlation ($r = .78$) between CpG methylation in brain and blood samples. Thus, several empirical results seem to support that while the relation between complex trait and methylation is tissue dependent, and individual differences in methylation (directly measured) do not correlate strongly between blood and brain tissues, local genetic regulation of methylation level is correlated across tissues.

The present study assessed methylation differences associated with differences in well-being using two study-designs.

1) \textit{Epigenome-wide Association meta-Analysis}
We performed large association meta-analyses of well-being and genome-wide DNA methylation in whole blood (Illumina 450K array) samples of adult participants from twelve population-based cohorts (Supplementary Table 1-3). We performed two EWAS meta-analyses; (1) A basic model not corrected for smoking behavior and body mass index (BMI; \( N = 9,496 \)) and (2) An adjusted model corrected for smoking behavior and BMI (\( N = 8,463 \)).

2) Summary-based Mendelian Randomization (SMR) with genome-wide mQTL data

We performed a genome-wide meta-analysis (GWAMA) of well-being (Supplementary Table 4) and integrated the results with four publically available mQTL datasets (three whole-blood and one fetal brain dataset) using SMR\(^{39,40,41}\). CpG sites identified using this approach might provide important leads to design further functional studies to understand the mechanisms by which DNA variation leads to complex trait variation. Besides identifying CpG probes associated with well-being, an important aim of the current study was to assess the concordance between an EWAS, where a direct association between well-being and CpG methylation is tested, and SMR where the local genetic effects on methylation are used to infer which CpG sites effect well-being.
Results

Epigenome-Wide Meta-Analyses.

Genome-wide DNA methylation analyses were performed for our basic model ($N = 9,496$) and adjusted model ($N = 8,463$). Cohort specific EWAS summary statistics were combined in a fixed effects meta-analysis adjusted for test-statistic bias and inflation$^{42}$ (Bayesian estimates of bias and inflation from all analyses are provided in Supplementary Table 3). Our basic model (not adjusted for smoking and BMI) identified two probes (cg19275632; $P < 9.84 \times 10^{-9}$ and cg14535274; $P < 8.73 \times 10^{-8}$) significantly ($P < 1.38 \times 10^{-7}$) associated with well-being (Figure 1A). However, when adjusting for smoking and BMI, no genome-wide significant SNPs were observed (Figure 1B). As expected, a significant correlation between the $Z$-statistics of the basic –and adjusted model was observed ($r = .98, P < 2.2 \times 10^{-16}$; Supplementary Fig 1).

Fig. 1. Manhattan plots of the EWAS analyses. (a) EWAS from the basic model, and (b) EWAS from the adjusted model (corrected for smoking and BMI). The $x$-axis represents the chromosomal position of the CPG sites, and the $y$-axis represents the significance on a $-\log_{10}$ scale. Each approximately independent genome-wide significant association (“lead CPG site”) is marked by Δ.
SMR analyses using methylation QTLs

We applied the SMR approach to test the association between DNA methylation probes and well-being, using mQTLs identified in a dataset of methylomic variation in whole blood and imputed SNP genotypes from the Lothian Birth Cohort (N = 1366) in conjunction with a multivariate GWAMA of well-being (N_{obs} = 491,455; Supplementary Fig 2 and Supplementary Table 5). The first stage of the SMR analysis identifies the most significantly associated SNP for a DNA methylation site (that is also present in the GWAMA dataset) as an instrumental variable for testing for association with well-being. This approach yielded 3 significant associations (P < 5.65 X 10^{-07} corrected for 88,531 tests; Supplementary Table 6 and Fig 2A) between well-being and DNA methylation probes.

Because the associations can be driven by highly correlated yet different causal variants for well-being and DNA methylation, also known as linkage, the second stage in SMR tests for heterogeneity in the association analysis by performing a heterogeneity in dependent instruments (HEIDI) test. The 3 significant associations survived the HEIDI test (P > 0.05) and can be described as pleiotropic.

Replication in two independent whole-blood mQTL datasets

We were able to test for replication of the SMR results with mQTLs generated from two independent datasets (Aberdeen N = 639 and University College London N = 665). Using the same strategy as above, we identified five associations between well-being and DNA-methylation (Fig 2B and Supplementary Table 6) in the Aberdeen dataset (P_{bonf} < 1.2 X 10^{-06} corrected for 41,803 tests). Moreover, one out of three significant associations from the discovery dataset were genome-wide significant in the Aberdeen dataset cg07879825), whereas one association at chromosome 3 lies within 10 kb (cg11645453) form the discovery dataset. A significant correlation of the Z-statistics (r = 0.93; P < 2.2 X 10^{-16}) indicates a large agreement in the direction of effect between both datasets (Supplementary Fig 3A).

Using UCL as the second replication dataset, we identified two associations between well-being and DNA-methylation (P_{bonf} < 1.66 X 10^{-06} corrected for 30159 tests; Figure 2C and Supplementary Table 6). None of the significant associations from the discovery datasets were significant in the UCL dataset, although the association at chromosome 3 lies in close
proximity to the association found in the discovery dataset (< 10 kb). However, the high significant correlation \(r = 0.91, P < 2.2 \times 10^{-16}\) between the Z-statistics is indicative of a large concordance between the LBC and UCL summary statistics (Supplementary Fig 3B).

Fig. 2. Manhattan plots of the SMR results. (a) LBC (whole blood) (b) Abderdeen (whole-blood), (c) University College Londen (whole blood), and (d) Human Developing Brain Resource. All plots in all panels are based on the same set of SNPs. The x-axis represents the chromosomal position, and the y-axis represents the significance on a \(-\log_{10}\) scale. Each approximately independent genome-wide significant association is marked by Δ.

**mQTL in whole blood versus brain**

Given the tissue-specific and developmentally dynamic nature of gene regulation, we were interested in examining the consistency of the SMR findings in a different tissue. To this end, we repeated the SMR analysis on mQTLs in a dataset of human fetal brain derived from the Human Developing Brain Resource (HDBR; \(N = 166\))\(^{31}\). We identified one association at chromosome 3 between well-being and DNA-methylation lying within 10 kb from the association present in our discovery dataset \((P_{\text{bonf}} < 6.58 \times 10^{-6}; \text{Supplementary Table 6})\) and a large correlation between Z-statistics of HDBR and LBC \((r = 0.72, P < 2.2 \times 10^{-16}\), indicating consistency between both summary statistics (Supplementary Fig 3C).
Direct epigenetic measurement versus Mendelian Randomization

To assess the concordance between an EWAS, where well-being and CpG methylation are directly correlated, and SMR where the local genetic effects on methylation are used to infer which CpG sites affect well-being, we correlated the Z-statistics that were present in both datasets. When including all corresponding CpG probes ($N = 70,564$), the Z-scores hardly correlated ($r = 0.02$, $P = 1.83 \times 10^{-10}$), which is indicative of little correspondence in direction of effects (Fig 3A). Next, we included only probes that survived the SMR HEIDI test ($P > 0.05$) and were present in both datasets ($N = 10,072$). The obtained correlation was still small but significant ($r = 0.02$, $P = 0.02$; Fig 3B). Finally, we tested the correlation between Z-statistics including probes with a SMR $P < 0.001$, $N = 57$) and found a small but non-significant correlation of $0.27$ ($P = .051$; Fig 3C).

Fig. 3. Correlation of the Z-statistics between LBC and EWAS. (a) correlation between Z-statistics LBC and EWAS (adjusted model for all corresponding CpG probed ($r = 0.02$, $P = 1.83 \times 10^{-10}$) (b) CpG probes surviving the HEIDI test ($r = 0.02$, $P = 0.02$), and (c) CpG probes surviving the Heidi test and $P_{SMR} < 0.001$ ($r = .27$, $P = .051$), In all plots, LBC is plotted at the x-axis.
Discussion

This study is one of the first large-scale investigations into epigenome-wide analyses of well-being through direct epigenetic measurement (EWAS) and summary based Mendelian Randomization (SMR). In the EWAS meta-analysis \( (N = 8,463) \), no genome-wide significant methylation hits were identified after correcting for multiple testing, smoking, and BMI.

In the Summary-based Mendelian Randomization analyses, we identified three genome-wide significant associations by integrating summary data from a GWAMA of well-being and the publically available mQTL dataset including participants from LBC (whole-blood). We were able to replicate one out of three associations using Abderdeen as an independent dataset, while another association lies in close proximity to the discovery dataset. In the third dataset (UCL), none of the three associations replicated although one lies in close proximity to the discovery dataset. Nevertheless we found high correlations of Z-statistics between the analyses with different datasets suggesting a large concordance in direction of effect. This replication confirms that the instruments used to index CpG methylation were consistent across multiple datasets, though we were not able to replicate the effects on well-being as a second sufficient powered independent GWAS of well-being is not available.

Next, we were interested in examining the consistency of our SMR findings in different tissue. Therefore, we repeated the SMR analyses in a dataset of human fetal brain derived from the Human Developing Brain Resource. The majority of SNP-DNA methylation relationships identified for SMR analysis in whole blood using the LBC dataset are characterized by a consistent direction of effect when tested in fetal brain.

Finally, we assessed the concordance between the EWAS, where well-being and CpG methylation are directly measured and SMR analyses, where the local genetic effects on methylation are used to infer which CpG sites effect well-being. Interestingly, no notable correlations were observed, even when different \( P \)-value threshold were used.

Our SMR results are in concordance with Hannon et al.\(^ {37} \) who found evidence for pleiotropic effects between SNPs that cause variation in CpG methylation and trait-associated genetic variation in over 40 complex traits with robust GWAS data. Moreover, similar to what we report in the present study, they found that a significant proportion of the associations (99.2\%) with mQTLs using an independent whole-blood dataset had the same direction of association. The large consistency that we observe between associations with mQTL measured in whole-blood as well as in fetal brain (\( r = 0.72 \)) is in concordance with previous studies\(^ {34,37,38} \) and
suggests that the correlation between mQTLs in local regions (i.e. ~1MB of the transcription start site) is fairly high. In addition, the lack of correlation between direct epigenetic measurement and summary-based Mendelian Randomization is in agreement with two previous studies\textsuperscript{31,32} that found that most CpG probes are uncorrelated between whole-blood and brain tissues.

There are several explanations to explain the lack of correlation between EWAS and SMR. First of all, epigenome-wide analysis of well-being through direct measurement provides the association of all epigenetic markers (genome-wide) with a trait of interest \textit{without} providing information about the possible mechanisms driving the association. There are five scenarios that theoretically can drive the association between DNA methylation and a trait; (1) There is an underlying single causal genetic variant that influence both DNA-methylation and trait, also known as pleiotropy, (2) There are 2 genetic variants (in high LD with each other) where on variant has an effect on DNA-methylation and one variant has an effect on trait variation, also known as linkage, (3) There is a causal relation where variation in CpG methylation will cause variation in well-being, inducing pleiotropy between the SNPs which influence CpG methylation, and well-being, (4) There were unmeasured confounding factors that have an effect on DNA methylation and trait variation, (5) There is reverse causation, where trait variation has an effect on differences in DNA-methylation instead of the other way around, and T (see Fig. 4 for graphical representation of the five scenarios). Summary-based MR in combination with the HEIDI test can distinguish between pleiotropic and linkage effects of trait associated genetic variation on DNA-methylation. Doing so, SMR provides useful information on the mechanism underlying the association between DNA methylation and trait variation (e.g. well-being). The absence of correlation between the EWAS and SMR results suggests that the associations we observed in our epigenome-wide analyses through direct epigenetic measurement are mainly driven by processes other than pleiotropy or a direct causal effect of CpG methylation on well-being. This finding is consistent with the presence of a direct relation between CpG methylation and well-being that varies over tissues (i.e is not present in white blood cells, but may still be present in brain cells).

The current EWAS of well-being, is one of the largest conducted today with a sample size of ~8600, but was not capable to identify probes significantly associated with well-being after adjusting for smoking and BMI. Future EWAS with larger samples that have sufficient statistical power should be able to identify CpG probes associated with well-being. However, if a tissue of convenience (i.e whole blood, buccal) is used, it should be expected that a
substantial portion of the findings may reflect confounding (other than those confounders corrected for in the model) or reverse causation where variation in well-being influences CpG methylation. Reverse causation in itself may be a very interesting mechanism which could hold clues to health outcomes which are a consequence of modified epigenetic states as a causal consequence of systematically lowered well-being. As biobanks increase their dense phenotyping, further EWAS studies may be able to interrogate multiple tissues in relation to well-being, and for individual loci identify the tissue of action. These studies should where possible leverage SMR which we have shown to be a valuable addition to EWAS in the context of psychological traits.

Conclusion:

We performed the largest EWAS of well-being to date and no genome-wide significant methylation hits were identified after correcting for multiple testing, smoking and BMI. Using summary Mendelian Randomization, we identified three associations (discovery dataset) where well-being and variable DNA methylation are pleiotropically associated with genetic-variation. Moreover, a high concordance in direction of effect was observed using three independent mQTL dataset measured in blood (2x) and brain (1x). Our results indicate that if the aim is to increase our understanding of the functional consequences of genetic risk variants for a complex trait and to facilitate the localization of specific genes within genomic regions identified by GWAS, SMR seems to be a promising way to go forward. If however, the aim is to identify environmental influences on the epigenome, a sufficient powered EWAS study might provide valuable information if brain tissue is not the only predominant tissue of interest. Combined use of the two designs may prove a potent cocktail able to identify correlation between CpG methylation and well-being while testing the exact nature of the observed correlation.
Fig. 4. Five scenarios that can drive the association between DNA methylation and a trait: (a) There is an underlying single causal genetic variant that influence both DNA-methylation and trait, also known as pleiotropy. (b) There are 2 genetic variants (in high LD with each other) where one variant has an effect on DNA-methylation and one variant has an effect on trait variation, also known as linkage; (c) There is a causal path from a genetic variant that influence DNA-methylation level an through DNA-methylation, trait variation will occur. (d) There were unmeasured confounding factors have an effect on DNA methylation and trait variation, and (e) there is reverse causation, where trait variation has an effect on differences in DNA-methylation instead of the other way around.
Methods

Epigenome-wide association study

Data on well-being, body mass index (BMI), smoking, white blood cell counts, and methylation level were available for 13 cohorts: ALPSAC ($N = 829$), QIMR ($N = 233$), DTR ($N = 1012$), FTC ($N = 593$), GENR ($N = 643$), KORAF4 ($N = 660$), LBC1921 ($N = 376$), LBC1936 ($N = 697$), LLD ($N = 730$), NFBC1966 ($N = 803$), NFBC1986 ($N = 593$), NAS ($N = 1195$), and NTR ($N = 2519$) (Supplementary Table 1). All participants provided written informed consent, and all contributing cohorts confirmed compliance with their local research ethics committees or institutional review boards.

Well-being (WB) Measurements

All questionnaires, except LifeLines Deep (LLD), were measures of happiness or satisfaction with life (Supplementary Table 2). LLD derived their questions from the positive-affect negative-affect (PANAS) questionnaire with questions focusing on ‘interested’, ‘enthusiastic’, ‘proud’ or ‘inspired’.

Participants inclusion criteria

We performed two EWAS meta-analyses. (1) The basic model without correcting for smoking and BMI ($N = 9,496$), and (2) the adjusted model corrected for smoking and BMI ($N = 8,463$). To be included in the two analyses, participants had to satisfy several criteria: (1) all relevant covariate data were available for each participant; (2) Participants passed the cohort-level methylation quality control and (3) Well-being was measured with either satisfaction with life measurements or happiness measurements. To be included in the adjusted model, covariate data on smoking and BMI should be present.

Epigenome-Wide Association study

To investigate associations between well-being and individual methylation markers the participating cohorts first performed cohort-level EWAS with a pre-specified analysis plan. As is standard, the EWAS was performed as a set of linear regressions in each cohort, one methylation marker at a time, with the methylation beta value ($0–1$) as the dependent variable. The key independent variable was WB. We estimated two regression models that differ in the set of covariates included. In the basic model, the covariates were age, sex,
imputed or measured white blood cell counts, technical covariates from the methylation array, and four genetic principal components to account for population stratification. In the adjusted model, we additionally controlled for BMI (kg/m$^2$), smoker status (three categories: current, previous or never smoker), and As BMI and smoking are correlated with WB$^{45,46}$ and known to be associated with methylation$^{8,47}$, the basic model may identify associations with WB that are actually due to BMI or smoking. Although the adjusted model reduces that risk, it may also reduce power to identify true associations with WB (by controlling for factors that are correlated with WB). We present the results for both models, but focus on the adjusted, more conservative, model.

**EWAS QC and meta-analysis**

Each participating cohort uploaded EWAS summary statistics to a central secure server for QC and meta-analysis. We removed probes with missing $P$-value, standard error, or coefficient estimate (Beta). Quantile-quantile (Q-Q) plots were made for the two models (basic versus adjusted). Additionally, we asked the participating cohorts to perform an EWAS on smoking and provide the corresponding $P$-values. From these $P$-values, also Q-Q plots were made for visual inspection. To account for test statistic bias and inflation we used the method described by Iterson et al.$^{42}$ as implemented in the R-package Bacon (Supplementary Table 3). We performed a sample-size-weighted meta-analysis of the cleaned results using METAL.$^{48}$ The two EWAS meta-analyses were performed on 395,764 methylation sites (CpG sites present in all datasets). For all samples included in the EWAS, white blood cell counts were measured with the standard white blood cell differential as part of the complete blood count (CBC). A CpG site was considered to be genome-wide significant at the stringent Bonferroni level ($\alpha = 0.05/395,764 = 1.26 \times 10^{-7}$).

**Summary-based Mendelian Randomization**

mQTL summary statistic were from four publically available data sets. Peripheral blood from McRae et al.$^{39}$ (Lothian Birth Cohort; $N = 1,366$), Hannon et al.$^{40}$ (Aberdeen; $N = 639$, University College London; $N = 665$). Fetal brain summary statics were available from Hannon et al.$^{31}$ (Human Developmental Biology Resource; $N = 166$).

We included in the SMR-analyses a multivariate GWAMA$^{43}$ of well-being measures. In the multivariate analyses, datasets from Okbay et al.$^{49}$(imputed to 1000G Phase 1 using the software tool DIST)$^{50}$ and UK Biobank (UK Biobank ID 20456 and 4526) were combined to
maximize power to identify genetic variants associated with well-being. Quality Control of UK Biobank data is described elsewhere. In total 491,455 observations and 7,123,275 SNPs (MAF > 0.01) were included in the analyses.

Summary Based Mendelian Randomization Analyses

The method behind SMR is extensively described by Zhu et al. In short, the SMR test was developed to test the association between an exposure (e.g. DNA methylation) with an outcome (e.g. Well-being) using a genetic variant as the instrumental variable to remove non-genetic confounding. Let \( x \) be an exposure variable, \( y \) be an outcome variable, and \( z \) be an instrumental variable. The Mendelian Randomization (MR) estimate of the effect of exposure on outcome (\( \hat{b}_{xy} \)) is the ratio of the estimated effect of instrument on exposure (\( \hat{b}_{zx} \)) and that on outcome (\( \hat{b}_{zy} \)).

\[
\hat{b}_{xy} = \frac{\hat{b}_{zy}}{\hat{b}_{zx}}
\]

where \( \hat{b}_{zx} \) and \( \hat{b}_{zy} \) are available from mQTL and GWAMA summary data. One of the core assumptions for MR is that the instrument should be strongly associated with exposure. Therefore in the SMR analyses, only top mQTLs (at least \( P < 1 \times 10^{-5} \)) are included as instrument for an SMR analysis. A significant association detected by the SMR test above can result from either a pleiotropic model (i.e. the exposure and the outcome are associated by a single shared genetic variant) or a linkage model (two or more variants in LD affecting the exposure and outcome independently). To distinguish pleiotropy from linkage a heterogeneity in dependent instruments (HEIDI) test was developed to test against the null hypothesis that there is a single causal variant underlying the association (pleiotropy model). For the HEIDI test we used multiple SNPs (e.g. the top 20 associated mQTLSs after pruning SNPs for either too strong or too weak LD in a cis region) to detect whether the association patterns across the region are homogeneous or not (a homogeneous pattern indicates a single shared causal variant). Thus, we assess the difference between \( \hat{b}_{xy} \) estimated at the top significant associated instrument \( \hat{b}_{xy(o)} \) and \( \hat{b}_{xy} \) estimated at a less significant instrument \( \hat{b}_{xy(i)} \).
\[
\hat{d}_i = \frac{\hat{b}_{xy(i)}}{\hat{b}_{xy(o)}}
\]

Under the null hypothesis (pleiotropic model), \( \mathbf{d} = \mathbf{0} \). If \( \mathbf{d} \) significantly deviates from \( \mathbf{0} \), we reject the SMR association due to heterogeneity.

**SMR correlations between the different datasets**

As the number of methylation probes were different between the EWAS analyses and the SMR analyses, only corresponding probes were included for each analysis. SMR analyses output do not report \( Z \)-statistics. Therefore, we calculated for each SMR output:

\[
Z_{SMR} = \frac{\beta_{SMR}}{\sigma_{SMR}}
\]

The resulting \( Z_{SMR} \) test statistics were correlated (two sided) with each other as an indication for consistency of direction of effects between whole-blood mQTL dataset as well as consistency of effect between whole-blood and fetal brain mQTL datasets.

**Direct epigenetic measurement versus Mendelian Randomization**

To test the consistency of effect between the EWAS analyses and whole-blood mQTL SMR analyses we correlated the \( Z \) statistics of the LBC whole-blood mQTL dataset and \( Z \) statistics of the EWAS meta-analysis corrected for smoking and BMI. Correlation analyses were performed in R\(^{52} \).
References


Supplementary Information

Supplementary Fig. 1. Correlation of the Z-statistics between EWAS basic model and adjusted model. X-axis representing the CpG probes of the basic model (model 1) and the Y-axis representing the CpG probes of the adjusted model (corrected for smoking and BMI). Correlation is $r = 0.98$, $P < 2.2 \times 10^{-16}$. 
Supplementary Fig. 2. Manhattan plots of N-weighted GWAMA of well-being. The x-axis represents the chromosomal position, and the y-axis represents the significance on a $-\log_{10}$ scale. Each approximately independent genome-wide significant association (“lead SNP”) is marked by Δ.
Supplementary Fig. 3. Correlation of the Z-statistics between basic model and adjusted model. (a) correlation between Z-statistics LBC and Aberdeen ($r = 0.93; P < 2.2 \times 10^{-16}$), (b) correlation between LBC and UCL ($r = 0.91, P < 2.2 \times 10^{-16}$), and (c) correlation between LBC and fetal brain ($r = 0.72, P < 2.2 \times 10^{-16}$). In all plots, LBC is plotted at the x-axis.