Exploring how much complex trait variation is captured by DNA methylation in epigenome-wide association studies

Thomas Battram1 Tom R. Gaunt1, Doug Speed2, Nicholas J. Timpson1, Gibran Hemani1

1MRC Integrative Epidemiology Unit, University of Bristol

2Aarhus Institute of Advanced Studies, Aarhus University, Denmark

Word count: 4761

Submitting to: Nucleic Acids Research

## Abstract

Following several years of epigenome-wide association studies (EWAS), for a majority of traits the number of sites with strong evidence for association with DNA methylation variation is small. Is this due to a lack of relevance of the assayed sites, or a need for larger sample sizes to detect many small effects? Here we performed EWAS and variance component analysis on 418 traits using blood samples from 940 individuals of the ALSPAC study. At least one EWAS hit was found in 17 traits. We then estimated the proportion of phenotypic variation captured by all 421,693 DNA methylation markers (h2EWAS) and found there was little evidence any trait had an h2EWAS > 0. The trait with the greatest evidence for h2EWASestimates being above zero was having smoked cigarettes regularly (h2EWAS = 0.4, FDR-corrected p-value = 0.1). The h2EWAS value was predictive of the number of EWAS hits (AUC = 0.7). Modelling the contributions of the methylome on a per-site versus a per-region basis gave varied h2EWAS estimates (r=0.48) but neither approach obtained substantially higher model fits across all traits. Our analysis indicates that most complex traits do not heavily associate with the markers commonly measured in EWAS within blood, and that using h2EWAS is a reasonable way to prioritise traits that are likely to yield associations.

Following several years of epigenome-wide association studies (EWAS), traits analysed to date tend to show limited evidence of association signal given the power of studies undertaken, Given the properties of methylation data and the prior hypotheses of this form of regulation as a mediating factor for biological relationships, this is unexpected. Here, we have formally estimated the proportion of phenotypic variation captured by 421,693 blood derived DNA methylation markers (h2EWAS) across 17 traits selected on the basis of \*\*. We found there was little evidence any trait had an h2EWAS > 0. Despite this, there were singular examples of stronger association. The trait with the greatest evidence for h2EWASestimates being above zero was having smoked cigarettes regularly (h2EWAS = 0.4, FDR-corrected p-value = 0.1). Furthermore, h2EWAS was predictive of the number of EWAS hits (AUC = 0.7). Modelling the contributions of the methylome on a per-site versus a per-region basis gave varied h2EWAS estimates (r=0.48), but neither approach obtained substantially higher model fits across all traits. Based on traits examined here, our analyses suggest that complex traits have limited global evidence for association with DNA methylation markers commonly measured in EWAS. Looking at specific associations and despite there being less association than expected, h2EWAS appears to be a reasonable way to prioritise traits that are likely to have stronger association with blood derived DNA methylation status.

\*\* Tom – I have tried to reword here… it is not so easy as you have a bipolar abstract containing both “there is nothing” and then “let’s rank by this to find”… The existing abstract does need work though

## Introduction

Epigenome-wide association studies (EWAS) aim to assess the association between phenotypes of interest and DNA methylation across hundreds of thousands of CpG sites throughout the genome [1], [2]. Many of the recent EWAS resulted in few sites across the genome with strong evidence for association and the proportion of total trait variance associated with these sites is small [1]. There is a need to have a global view of the contribution of DNA methylation to complex traits in order to interpret these results.

There are multiple possible reasons for there being few EWAS signals. Firstly, DNA methylation varies between cells and tissues, thus any changes related to a trait may occur in any number of tissues. Currently, because of ease of access and cost, the most common tissue used for EWAS is blood, which may not capture changes in DNA methylation related to the trait of interest [1], [2]. Secondly, the commonly used technologies probe a small percentage of the total number of potentially methylated sites. In the absence of full knowledge of the correlation structure across methylation site variation, it is therefore difficult to fully understand coverage in current measures. Two more possibilities are that DNA methylation variation is actually not associated with the traits studied or that the associations are many and but individually too small to detect with current sample sizes.

The total contribution of genetic variation to complex trait variance has long been estimable through heritability studies [3]. Arguably, the finding that most traits had a heritable component provided an *a priori* justification for the pursuit of gene mapping endeavours that eventually gave rise to GWAS. When GWAS was performed in small sample sizes and returned few hits, a simple interpretation was that larger sample sizes were needed to detect many small effects. Recent practices of using hundreds of thousands of samples which yield large numbers of GWAS hits have given credibility to this model (Box 1).

As heritability estimates arise from the relationship between phenotypic variability and genetic relatedness (known in families and across matrices of more distantly related individuals), there is no simple corollary to similarly evaluate the total contribution of methylome variation to complex traits. Hence, Interpretation of the paucity of EWAS hits is difficult because there is no knowledge of the total contribution of methylation variation to the trait. However, analogous to the calculation of genetic heritability estimates which have been expanded estimates across non-familial population-level data (SNP heritability), the total contribution of methylation markers to complex traits could be estimated in a similar way and potentially beneficial in interpreting EWAS results in a more comprehensive manner (See Box 2 for a simple explanation of SNP heritability and its application to DNA methylation).

SNP heritability estimates are sensitive to assumptions of the underlying genetic architecture and there are different ways in which to model the contribution of each SNP to the overall genetic component. The original model of calculating h2SNP introduced by Yang et al. assumes that each variant has an effect that is independent of the regional linkage disequilibrium (LD) structure as each variant is unweighted (the blanket model), and this effectively assumes regions of high LD contribute more to phenotypic variance [4]. Speed et al. proposed a new model, which considered the LD between SNPs so that each region of high LD can effectively be counted as a singular effect (the grouping model)[5]. Finding which models fit the data better helps ensure a more accurate estimation of the proportion of DNA methylation association with a trait, but they could also be biologically informative.

Gene regions are methylated in a coordinated fashion, which is associated with changes in gene expression [6], [7], with a tendency for promotor regions to be unmethylated and gene body regions to be methylated when gene expression is activated [7]. This, amongst other complex patterns of gene regulation, induces a correlation structure within EWAS data, and it is not clear whether a single site is driving an association and neighbouring sites are consequentially correlated, or if the cumulative contributions of all neighbouring sites associate with the regulatory process. In EWAS, it is commonplace to collapse DNA methylation sites into groups based on proximity and if they share the same direction of association and potentially magnitude of association, this is often called differentially methylated region (DMR) analysis [8]. This, however, does not explain whether the sites within groups are acting independently and cumulatively or as a single entity. A scenario could arise such that sites are identified in EWAS within a short distance of each other and near to a gene linked to the trait of interest. Changes in DNA methylation at these sites are associated with changes in gene expression of the nearby gene. In this scenario, you might expect the changes in DNA methylation to act in tandem and thus the grouping model would fit this data better. However, we cannot be certain that the sites in this scenario are indeed acting completely (or at all) as one, and so the blanket model may actually be better suited. **Figure 1** shows a representation of how the differences in models apply to DNA methylation data at a single small region using the scenario just described as an example. Of course, there are far more scenarios possible and furthermore, the models aren’t restricted to a single small region in the genome. They apply to all sites, as do the DMR methods used in EWAS. Thus, by applying both methods to DNA methylation data across multiple phenotypes and comparing their utility we can gain insight into how DNA methylation operates across gene regions. Furthermore, it is important to find the model that best fits the data to help prevent biased estimates.

This study aims to estimate h2EWAS values across a plethora of traits and test their ability to predict successful identification of sites in an EWAS at a given p-value threshold, which would suggest these values can be used to assess future trajectories. To do this we perform hundreds of EWAS studies and evaluate if h2EWAS estimates are predictive of the number of sites identified by the EWAS at various P value thresholds. We also compare the performance of different models underlying h2EWAS estimates to infer likely methylation architecture of complex traits.

Box1: The argument for increasing sample size for EWAS

The need for larger sample sizes in GWAS has been empirically demonstrated across a broad range of traits. For height and body mass index (BMI), the number of associations dramatically increased from 12 to 3290 and from one to 941, respectively after increasing sample sizes by ~670,000 [27]–[29]. This trend can be seen for many traits. Similar to early GWAS, many EWAS are discovering few sites strongly associated with complex traits. However, an example that suggests promise for increasing sample sizes for EWAS is seen with BMI, where an EWAS of 459 individuals identified just five sites, but increasing the sample size to over 5,000 led to identification of 278 sites [30], [31]. While we can continue to improve sample sizes in EWAS, there is a need to determine the upper limit of the information we can obtain from EWAS of complex traits like BMI. Furthermore, the BMI EWAS example may be unrepresentative of other traits, so having a corollary test for estimating h2SNP for DNA methylation would help us understand if we’re capturing relevant information from the current arrays we are using in EWAS. Such information could inform future study designs in terms of growing sample sizes with the current assays available versus designing new assays.

Box2: Applying SNP heritability estimator methods to DNA methylation

Methods used to estimate h2SNP use restricted maximum likelihood (REML) tests to estimate the proportion of variance attributable to these genetic variants. Essentially this assesses whether individuals that are genetically similar are more likely to be phenotypically similar. If those individuals that have a high genetic overlap tend to correlate strongly phenotypically compared to those that don’t have high genetic overlap, then the phenotype of interest will have a high h2SNP. Unlike genetic variants, DNA methylation is responsive to the environment [1] and determining causal directionality between DNA methylation markers associated with traits is not trivial [32]–[34]. Therefore, estimating the proportion of trait variation captured by DNA methylation variation (which will henceforth be denoted as h2EWAS) using the same techniques will ascertain effects going in both directions as well as associations due to confounding. The combination of these mechanisms may increase power to detect trait-DNA methylation association, and could be the reason that so many DNA methylation markers are found in small EWAS compared to similarly sized GWAS [31].



**Figure 1** Comparison of the grouping and blanket models in the context of the relationship between DNA methylation and gene expression. Both regions are exactly the same, the only difference is how each model assumes the methylation sites should be treated. The grouping model down-weights the contribution of correlated CpGs, effectively grouping them, and the blanket model assumes each CpG independently associates with a trait. As seen here, the grouping of correlated CpG sites may not be the correct thing to do as some of the sites may be acting independently of their correlated partners.

## Methods

### Study sample(s)

#### Avon Longitudinal Study of Parents and Children (ALSPAC)

All data for the study came from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. ALSPAC recruited pregnant women in the Bristol and Avon area, United Kingdom, with an expected delivery date between April 1991 and December 1992 (http://www.bris.ac.uk/alspac/). Over 14,000 pregnancies have been followed up (both children and parents) throughout the life-course. Full details of the cohort has been published previously [9], [10]. This study uses phenotypic and DNA methylation data from the mothers (N = 940).

Continuous and binary phenotypes measured in mothers were extracted from the cohort. A summary of the phenotypes is present in the **Supplementary Material** and full details of all the data is available through a fully searchable data dictionary: http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/

Phenotype data were extracted using the ALSPAC R package (github.com/explodecomputer/alspac) and went through various quality control steps, which are detailed in the **Supplementary Material** and summarized in **Supplementary Figure 1**.

All continuous traits were rank-normalised for further analyses. A Shapiro-Wilk test of normality was performed on these rank-normalised traits and for those with some evidence of non-normality (P < 0.05), we re-examined the distribution of those traits by eye to ensure it was approximately normal. It was found that any non-normality of phenotype distributions corresponded to an inflation of zero values. These traits were removed and overall there were 2423 traits left for analyses. These traits do not necessarily represent independent phenotypes and as such we wanted to prevent correlated traits skewing results. The absolute Pearson’s correlation coefficient between each trait was subtracted from one (1 –[r]). Then traits were greedily selected where 1 –[r] < 0.4 with any other trait. This left 418 traits, which consisted of ~30% clinically measured variables (including roughly 50 metabolites and some anthropometric traits), ~25% health related questions (for example “have you ever had asthma?”), ~40% behavioural and social traits (for example educational attainment variables, use of pesticide, and having pets), and ~5% of traits were related to the partner or child of the participant (for example the employment status of the partner). Plots showing the correlation between all the phenotypes as well as with just the “independent” traits can be seen in **Supplementary figure 2-3.**

Ethical approval for ALSPAC was obtained from the ALSPAC Ethics and Law Committee and from the UK National Health Service Local Research Ethics Committees. Written informed consent was obtained from both the parent/guardian and, after the age of 16, children provided written assent.

### DNA methylation data

DNA methylation was measured using the Illumina Infinium HumanMethylation450 (HM450) BeadChip. Before use, the data went through quality control and were normalised separately to the phenotype data. Full details can be found in the **Supplementary Material**.

DNA methylation data generated from blood collected at a single clinic visit was used for each of the participants.

Probes were excluded if they were present on either of the sex chromosomes, a SNP/control probe, had a detection p value < 0.05 across over 10% of samples or were identified as problematic by Zhou et al. [11]. This left 421,693 CpG sites for analyses.

Before analysis a linear regression model was fitted with beta values for methylation (which ranges from 0 (no cytosines methylated) to 1 (all cytosines methylated)) as the outcome against batch variables (plate ID in ALSPAC) modelled as a random effect to help remove the effects of batch in the subsequent analyses.

Cell proportions (CD8+ and CD4+ T cells, B cells, monocytes, natural killer cells, and granulocytes) were estimated using an algorithm proposed by Houseman et al. [12].

### REML analysis

Using LDAK [13] the relationship between the methylomes (as measured by the HM450 BeadChip) of 940 individuals was estimated by producing a DNA methylation relationship matrix (MRM). This matrix was used as input for the REML analysis to estimate the proportion of a trait’s variation that was explained by DNA methylation (h2EWAS). Age, the top 10 ancestry principal components, and derived cell proportions were added as covariates to the model.

When producing the MRM, probes were scaled by their observed variance and the weighting of each probe was based on the variance of DNA methylation at that site using the formula below:

where is the variance of methylation at CpG .

The higher the alpha value the more weight is given to probes with greater variance; an alpha value of -1 gives equal weight to probes with low and high variance. The alpha value of -0.25 was chosen because previous analysis by Speed et al. [13] suggested that this value was optimal for measuring h2SNP. Furthermore, it was hypothesised that probes with a greater variance would contribute more to trait variance. As the method was applied to DNA methylation data in this study, sensitivity analyses were conducted. MRMs were created specifying the alpha value at increasing increments of 0.25 from -2 to 0. The association between h2EWAS and number of EWAS hits as well as the difference in h2EWAS estimates for 10 randomly selected phenotypes was assessed for the varying alpha values.

The mean of the MRM diagonal should be 1 and the variance close to 0, as the diagonal values essentially represent the correlation between an individual’s methylome with itself. Although values are expected to vary slightly from 1. For the MRMs it was identified that some diagonal elements were very high (> 2), which caused the diagonal to have a high variance (0.13). To assess whether these values could skew results, we conducted sensitivity analyses removing individuals, with varying diagonal value cutoffs.

Like h2SNP estimates, h2EWAS estimates should range from zero to one. If a trait has a true h2EWAS value of zero, there is no association between the methylome and that trait, and if h2EWAS equals one then DNA methylation has the capacity to completely predict that trait. However, estimation of h2EWAS can be fairly imprecise and without constraining the software it’s possible to get estimates of h2EWAS that are outside 0-1 due to large standard errors. These point estimates have to be erroneous by definition.

Even though the grouping model effectively groups sites together, it is actually likely to increase the number of parameters because without the weightings imposed by this model, the blanket model essentially ignores sites that are not neighbouring others. Therefore, larger standard errors are expected with the grouping model. The grouping model applies a sliding window approach, with windows of 100kb, to capture the correlation between neighbouring sites and weight sites according to the correlation structure of the region. When applying the grouping model, the number of sites that were weighted were 45,863 (out of 421,693) and the number of sites neighbouring any single CpG site ranged from 29 to 28,217.

### Comparison of REML models

### Generating genetic principal components

Ancestry principal components were generated within ALSPAC mothers using PLINK (v1.9). Before analysis, genetic data went through quality control and were imputed, full details can be found in the **Supplementary Material**. After quality control and imputation, independent SNPs (r2 < 0.01) were used to calculate the top 10 ancestry principal components.

### Epigenome-wide association studies

EWAS were conducted for 418 non-independent traits within the ALSPAC cohort. For all traits, linear regression models were fitted with beta values of DNA methylation as the outcome and the phenotype as the exposure. Covariates included age, the top 10 ancestry principal components and cell proportions.

DMPs were extracted from the EWAS at P value thresholds ranging from 10-3 to 10-7. It was assessed whether h2EWAS could predict that the number of identified DMPs in an EWAS was greater than number of DMPs expected to be identified at a given P threshold defined as the number of sites tested multiplied by the threshold. The traits were also “pruned” in the same way as described above, to prevent including overly correlated traits and biasing results. The sensitivity and specificity of this prediction was calculated and a receiver operating characteristic (ROC) curve was plotted. At p-value thresholds of 10-6.5 and 10-7 there were too few EWAS hits, so these were removed from the analysis.

The association between the number of DMPs identified at P < 1x10-5 and h2EWAS values was assessed using a negative binomial hurdle model with the number of DMPs identified fitted as the outcome and h2EWAS as the exposure. The negative binomial hurdle Poisson regression model results are twofold. The first of which assesses whether there is an association between the binary trait of whether a DMP was identified by EWAS and h2EWAS. The second is a zero-truncated model, i.e. the zero values are removed from the model and the association between number of DMPs and h2EWAS is assessed.

All analyses were conducted in R (version 3.3.3) or using the command line software LDAK [13], GCTA [14], and PLINK [15]. For the EWAS analyses, the meffil R package was used [16]. A one-sided P value was used to assess if the h2EWAS for a trait was > 0, and two-sided P values were used for everything else.

Code used to perform analyses can be found here: github.com/thomasbattram/ereml

## Results

A flowchart showing our study design and giving a summary of the results is shown in **Figure 2**.

### 

**Figure 2** Study design with a summary of the results. ALSPAC = Avon Longitudinal Study of Parents and Children, QC = quality control, EWAS = epigenome-wide association study, MRM = methylation relationship matrix, AUC = area under curve.

### Estimating the proportion of phenotypic variance associated with DNA methylation

We used two models to estimate the total contribution of all DNA methylation sites to the variation (h2EWAS) for each of 418 traits within 940 individuals. The median for both models was zero with ranges of -0.4 to 0.4 and -0.5 to 0.4 for the blanket and grouping models respectively **Figure 3**. The trait with the greatest evidence for h2EWASestimates being above zero was having smoked cigarettes regularly (FDR-corrected P = 0.06 and 0.10 for the blanket and grouping models respectively). The correlation between the h2EWAS estimates of the two models was 0.48 and there was evidence that on average the estimates of the grouping model were higher (Paired t-test P = 3.5x10-5, **Figure 3**), but the mean difference between estimates was only 0.016.

There was little evidence that either of the models fit the data better (had higher likelihoods) across the 418 traits tested (difference in median likelihoods = 0.19, Wilcoxon’s paired ranked sum test P = 0.73). Further, the majority of h2EWAS estimate differences between the traits were small.



h2EWAS

h2EWAS

**Figure 3** A comparison of h2EWAS estimates given by applying REML using the blanket and grouping models across 418 traits. The blue dashed line is at x=y. Values with h2EWAS lower than 0 are due to imprecision in h2EWAS estimates as the true estimate cannot be negative. Smoked\_cigs\_reg = smoked cigarettes regularly. The h2EWAS of this phenotype has the greatest evidence for being above 0 for both the blanket and grouping model (P = 1.44x10-4 and P = 2.61x10-4, respectively).

### Sensitivity analyses when estimating the proportion of phenotypic variance associated with DNA methylation

After examination of the MRMs required to produce the h2EWAS estimates, we found that for both the blanket and grouping model we observed some unexpected values: 96 diagonal elements had values over 1.5 when using the blanket model, with the maximum value being 3.562. When assessing the impact of these potential outliers in the MRM to results we found that the median and range of h2EWAS estimates varied little (**Supplementary figure 4**). The likelihood of the models tended to be greater as more outliers were removed (lower threshold for classing a diagonal element as an outlier), but it still didn’t vary much (**Supplementary figure 5**).

The weight of predictors used to produce the MRMs was also examined. As more weight was given to sites where methylation variation was greater (increasing alpha value) the h2EWAS estimates were slightly higher (**Supplementary figure 6**). However, the likelihood tended to remain the same, the median likelihood had a range of 2 across the alpha values (**Supplementary figure 7**).

Results of sensitivity analyses are summarised in **Supplementary table 1 and 2**.

### EWAS analyses

In order to assess the association between h2EWAS and EWAS results, we performed EWAS of 418 traits. No associations were found at the strict P value cutoff of P < 2.4x10-10 (1x10-7 / 418). A total of 30 associations between traits and CpGs were identified at the conventional EWAS P value cutoff – P < 1x10-7. Of the traits tested, 17 had at least one EWAS hit, with the maximum number of CpGs associated with a trait being 13(smoked cigarettes regularly). As there were so few traits with any identified hits, we took forward results from the lenient P value threshold of P < 1x10-5, at which 356 traits had at least one EWAS hit.

As the distributions of hit count data was heavily right skewed with an inflation at 0 and 1 (**Supplementary figure 8**), to test the association between h2EWAS and number of DMPs we opted to test goodness of fit for variations of Poisson models. Of the 6 models tested, the negative binomial hurdle Poisson regression model fit the data best, full results can be found in **Supplementary table 3**. We found there was strong evidence for an association between number of EWAS hits and h2EWAS (**Figure 4**). There was some evidence of association between the presence of DMPs and h2EWAS (beta = 6.8, [95%CI 3.0, 10.5]) as well as some evidence of an association between number of DMPs (when the number is above 0) and h2EWAS (mean increase of 0.58, [95%CI 0.38, 0.79] DMPs when h2EWAS increases by 0.1).



h2EWAS

**Figure 4** Association between h2EWAS and number of DMPs identified in EWAS. The correlation between DNA methylation and the variance of traits (h2EWAS) was calculated using REML analysis using the blanket and grouping models. EWAS were conducted on all the same traits and the distribution of the number of DMPs identified at P < 1x10-5 and h2EWAS are plotted above. Any traits where the h2EWAS estimate is below 0 are coloured grey. The true h2EWAS value cannot be negative, but sample sizes in this analysis are small so the estimates are very imprecise.

The ability of h2EWAS estimated by both models to predict whether the number of DMPs identified was greater than expected was assessed at varying P value thresholds. ROC curves were produced and the area under the curve (AUC) ranged from 0.65 and 0.67 at P < 1e-6 to 0.79 and 0.71 at P < 1e-3 for the blanket and grouping models respectively and the predictive ability remained fairly stable as the threshold increased (**Figure 5**).



**Figure 5**. The ability of h2EWAS values to predict whether the number of differentially methylated positions identified in an EWAS is higher than expected by chance. ROC curves for h2EWAS values predicting number of DMPs at differing P value thresholds. AUC = area under the curve.

## Discussion

The genetic architecture of complex traits has been explored using various methods, these aim to inform researchers of how to design future studies that seek to discover new genetic variants related to their trait of interest. In this manuscript we apply one such method, estimating the predictive capacity of variants across a SNP-chip (h2SNP), to DNA methylation data measured in blood with the HM450 BeadChip across 418 independent traits.

### Estimation of h2EWAS

The true h2EWAS of a trait gives the total predictive capacity of DNA methylation for that trait, which is equivalent to the proportion of that trait’s total variance that is associated with changes in DNA methylation. Knowing this information can help design future EWAS studies. A low value of h2EWAS doesn’t necessarily mean there is little correlation between DNA methylation and a trait, it could transpire that unmeasured sites contribute more to the association. It is important to remember that roughly 1.5% of CpG sites are targeted by the HM450 BeadChip and DNA can be methylated elsewhere (not at cytosine bases). Therefore, whole genome bisulphite sequencing, or a similar technique, may show that the variance of complex traits captured by DNA methylation is far higher. Furthermore, even if h2EWAS is low and the sites discovered already do not explain all of the h2EWAS estimate, there may still be value in increasing sample size to identify more DMPs as well as increase the precision of h2EWAS estimates. DMPs discovered may not be highly correlated with a trait, but this doesn’t mean the potential biological information gained isn’t valuable. For example, if a change in a the levels of protein X has a large effect on a trait and change in DNA methylation has a small effect on the levels of protein X, then the effect of that DNA methylation change on the trait will be small, but identifying that DMP could lead to discovering the importance of the protein. Another thing to consider is that DNA methylation is tissue and cell specific. This means, that h2EWAS may vary a lot depending on what tissue the methylation is measured in.

The true underlying genetic architecture of complex traits is still unknown, and therefore it is difficult to know the appropriate model to choose for estimating SNP-heritability and arguments for each model depending on this underlying genetic architecture are still being put forward [13], [17]–[19], thus the attempts made in this study to re-purpose genomic REML are likely to suffer the same flaws that are trying to be overcome in genetic data. With this in mind, in addition to the imprecise estimates of h2EWAS presented here (due to the small sample sizes of available data), we believe that individual trait h2EWAS values should be treated with caution. This doesn’t exclude the possibility that estimating h2EWAS may be useful and other methods are already being developed to measure the association between DNA methylation at all sites and complex traits [20].

### Future EWAS

It has become clear from the GWAS era of genetics, that for complex traits, such as coronary artery disease, many common genetic variants with small effects make up the genetic component of the trait [21], [22]. This suggests a large number of molecular pathways contribute to these traits. DNA methylation at CpGs is heritable [23], [24], thus it would be expected that the DNA methylation architecture of a trait will somewhat reflect the genetic architecture of the trait, although this has not been empirically tested.

Despite uncertainty of h2EWAS estimates for individual traits, we show h2EWAS has a modest ability to predict whether the number of EWAS hits will be greater than expected by chance at a given P value threshold. This predictive ability remained stable as the P value threshold for detection increased from P < 1x10-6 to P < 1x10-3. These results suggest that increasing sample sizes for traits which truly associate with DNA methylation should result in the discovery of more DMPs. Furthermore, these results support a model for which small changes in methylation at many CpGs across the genome are related to complex traits.

### Contributions of individual CpG sites

The original model for measuring h2SNP assumed all genetic variants contributed the same effect on a trait [4], Speed et al. offered an alternative model assuming a different underlying genetic architecture, whereby genetic variants in regions of high LD contributed less to the variance of a trait than more independent variants. Both groups have shown that the performance of the models depend on the alignment of the trait’s architecture with the models’ underlying assumptions. Previous literature has suggested that it is the methylation across groups of CpGs that may affect how other molecules interact with DNA and influence cellular functions such as gene expression [7]. Furthermore, CpGs are not randomly distributed throughout the genome – many exist in close proximity within “islands” or other regions, suggesting that grouping of the CpGs may have functionality. However, the most common method used in EWAS is to treat CpG sites as independent. Here, the models proposed by Speed et al. (the grouping model) and Yang et al. (the blanket model), when estimating h2EWASwere tested across 418 traits. The model fit the data better (had a higher likelihood) 216 times for the blanket model and 202 times for the grouping model. Thus for almost half the traits treating DNA methylation sites as independent seems to be preferable and even though there is correlation between CpG sites, which allows them to be grouped, it might be that in some groups of correlated sites, individual sites within the group contribute separately to trait variance.

It’s important to note that the grouping method takes into account correlation between CpGs within 100Kb of each other. Differential methylation at CpG sites may be correlated for a variety of biological reasons, for example, CpGs lying within a transcription factor binding site will be regulated together, but also, they will be correlated with CpGs that lie in other binding sites for that same transcription factor and these may be many megabases away. This is relevant to the relationship between DNA methylation and complex traits because transcription factor regulation might be the link between complex traits and DNA methylation. Even though grouping CpG sites might yet be the best way to model the relationship between DNA methylation and complex traits, the optimum way to group sites is unknown and will likely change depending on the trait of interest.

### Limitations

The main limitation of the study is the small sample size (N = 940) to estimate the h2EWAS. This meant the precision of our h2EWAS estimates were very low and so our power to assess their ability to predict number of DMPs and find individual trait h2EWAS values was low. To circumvent this problem, we assess trends across multiple traits and do not make strong conclusions for any one trait.

As mentioned previously the HM450 BeadChip captures a small percentage of the total DNA methylome and h2EWAS estimates will likely vary upon assaying more DNA methylation sites. Furthermore, when measuring more sites, it might be that one of the models fits the data better. Nevertheless, the results of this study can still give evidence towards the hypothesis that differential methylation at many sites across the genome each contribute minimally to the overall association between the methylome and a complex trait.

Unlike germline genetic variants, there is intra-individual (between tissue) DNA methylation variation [1], [2]. Thus, it is to be expected that the variation of h2EWAS estimates across traits is partly a product of the tissue of choice (here blood). However, within the tissue biologically pertinent to the complex trait of interest, the number of pathways that associate with variation in that trait is likely to remain high, for example there are many processes affecting, or affected by, cancer development [26]. Thus, it would still be expected that differential methylation at many CpG sites each associate with a trait, but the effect sizes are small. DNA methylation also varies temporally, therefore a low h2EWAS at one timepoint may not apply to other timepoints.

The factors important for the correlation structure of DNA methylation data are less known than those for linkage disequilibrium structure of genetic variants. Therefore, when applying models, such as the grouping model here, that aim to account for correlation of neighbouring DNA methylation sites, we may be missing some of the important structure captured for example by trans-correlations (over 1Mb). A model that estimates h2EWAS by incorporating all of the underlying correlation of DNA methylation data may therefore outperform both models tested here.

## Conclusion

Overall, the number of traits with good evidence for h2EWAS > 0 was low (none met the threshold FDR < 0.05) and the values centred around 0, suggesting that for many traits DNA methylation variation as measured on the HM450 BeadChip in blood is of little relevance. However, these estimates varied greatly and therefore DNA methylation measured in this way will likely have relevance for some traits, for example smoking cigarettes regularly. Further, these estimates were correlated with the number of DMPs identified, suggesting that for traits whose variance associates with DNA methylation then increasing sample size will yield an increase in the number of CpGs identified in EWAS. We also provide evidence that there is value in assessing individual CpG-trait associations as opposed to groups of correlated CpG sites within 100Kb. However, this does not preclude the possibility that a more complex model of CpG site correlation may provide a better fit.

## Acknowledgements

We are extremely grateful to all the families who took part in the Avon Longitudinal Study of Parents and Children the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

## Funding

This work was partly supported by a Wellcome Trust PhD studentship to T.B. (203746). D.S. is funded by the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 754513, by Aarhus University Research Foundation (AUFF) and by the Independent Research Fund Denmark under Project no. 7025-00094B. This work was also supported by the UK Medical Research Council (MC\_UU\_00011/1, MC\_UU\_00011/4, MC\_UU\_12013/1, MC\_UU\_12013/2 and MC\_UU\_12013/4), which funds a Unit at the University of Bristol where TB, TRG, NJT and GH work. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. Methylation data in the ALSPAC cohort were generated as part of the UK BBSRC funded (BB/I025751/1 and BB/ I025263/1) Accessible Resource for Integrated Epigenomic Studies (ARIES) [http://www.ariesepigenomics.org.uk]. The phenotype collection was also in part funded by The British Heart Foundation (SP/07/008/24066), Roche Diagnostics and the National Institute for Health Research (NF-SI-0611-10196). A comprehensive list of grants funding is available on the ALSPAC website

(http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)

## References

[1] E. Birney, G. D. Smith, and J. M. Greally, “Epigenome-wide Association Studies and the Interpretation of Disease -Omics,” *PLOS Genet.*, vol. 12, no. 6, p. e1006105, Jun. 2016.

[2] V. K. Rakyan, T. A. Down, D. J. Balding, and S. Beck, “Epigenome-wide association studies for common human diseases,” *Nat Rev Genet*, vol. 12, no. 8, pp. 529–541, Aug. 2011.

[3] P. M. Visscher, W. G. Hill, and N. R. Wray, “Heritability in the genomics era - Concepts and misconceptions,” *Nature Reviews Genetics*. 2008.

[4] J. Yang *et al.*, “Common SNPs explain a large proportion of heritability for human height,” *Nat. Genet.*, 2010.

[5] D. Speed, G. Hemani, M. R. Johnson, and D. J. Balding, “Improved heritability estimation from genome-wide SNPs,” *Am. J. Hum. Genet.*, 2012.

[6] P. a Jones and G. Liang, “Rethinking how DNA methylation patterns are maintained.,” *Nat. Rev. Genet.*, vol. 10, no. 11, pp. 805–11, 2009.

[7] P. A. Jones, “Functions of DNA methylation: islands, start sites, gene bodies and beyond,” *Nat. Rev. Genet.*, vol. 13, no. 7, pp. 484–492, Jul. 2012.

[8] D. P. Chen, Y. C. Lin, and C. S. J. Fann, “Methods for identifying differentially methylated regions for sequence- and array-based data,” *Brief. Funct. Genomics*, vol. 15, no. 6, pp. 485–490, 2016.

[9] A. Fraser *et al.*, “Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort,” *Int. J. Epidemiol.*, 2013.

[10] A. Boyd *et al.*, “Cohort profile: The ’Children of the 90s’-The index offspring of the avon longitudinal study of parents and children,” *Int. J. Epidemiol.*, vol. 42, no. 1, pp. 111–127, 2013.

[11] W. Zhou, P. W. Laird, and H. Shen, “Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes,” *Nucleic Acids Res.*, vol. 45, no. 4, p. gkw967, Oct. 2016.

[12] E. Houseman *et al.*, “DNA methylation arrays as surrogate measures of cell mixture distribution,” *BMC Bioinformatics*, vol. 13, no. 1, p. 86, May 2012.

[13] D. Speed, N. Cai, M. R. Johnson, S. Nejentsev, D. J. Balding, and D. J. Balding, “Reevaluation of SNP heritability in complex human traits,” *Nat. Genet.*, vol. 49, no. 7, pp. 986–992, May 2017.

[14] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, “GCTA: A tool for genome-wide complex trait analysis,” *Am. J. Hum. Genet.*, vol. 88, no. 1, pp. 76–82, 2011.

[15] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, “Second-generation PLINK: rising to the challenge of larger and richer datasets,” *Gigascience*, vol. 4, no. 1, p. 7, 2015.

[16] J. L. Min, G. Hemani, G. Davey Smith, C. Relton, and M. Suderman, “Meffil: efficient normalization and analysis of very large DNA methylation datasets,” *Bioinformatics*, 2018.

[17] S. Gazal *et al.*, “Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection,” *Nat. Genet.*, vol. 49, no. 10, pp. 1421–1427, Sep. 2017.

[18] D. Speed and D. Balding, “Exposing flaws in S-LDSC; reply to Gazal et al.,” *bioRxiv*, p. 280784, Mar. 2018.

[19] D. Speed, J. Holmes, and D. J. Balding, “Evaluating and improving heritability models using summary statistics,” *bioRxiv*, p. 736496, Jan. 2019.

[20] D. T. Banos *et al.*, “Bayesian reassessment of the epigenetic architecture of complex traits,” *bioRxiv*, p. 450288, Jan. 2018.

[21] P. van der Harst and N. Verweij, “Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease.,” *Circ. Res.*, vol. 122, no. 3, pp. 433–443, Feb. 2018.

[22] P. M. Visscher *et al.*, “10 Years of GWAS Discovery: Biology, Function, and Translation,” *American Journal of Human Genetics*. 2017.

[23] J. van Dongen *et al.*, “Genetic and environmental influences interact with age and sex in shaping the human methylome,” *Nat. Commun.*, vol. 7, p. 11115, Apr. 2016.

[24] T. R. Gaunt *et al.*, “Systematic identification of genetic influences on methylation across the human life course,” *Genome Biol.*, vol. 17, no. 1, p. 61, 2016.

[25] C. Lövkvist, I. B. Dodd, K. Sneppen, and J. O. Haerter, “DNA methylation in human epigenomes depends on local topology of CpG sites,” *Nucleic Acids Res.*, 2016.

[26] D. Hanahan and R. A. Weinberg, “Hallmarks of cancer: The next generation,” *Cell*, vol. 144, no. 5. pp. 646–674, 2011.

[27] G. Lettre *et al.*, “Identification of ten loci associated with height highlights new biological pathways in human growth,” *Nat. Genet.*, 2008.

[28] T. M. Frayling *et al.*, “A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity,” *Science (80-. ).*, 2007.

[29] L. Yengo *et al.*, “Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry,” *Hum. Mol. Genet.*, 2018.

[30] K. J. Dick *et al.*, “DNA methylation and body-mass index: A genome-wide analysis,” *Lancet*, 2014.

[31] S. Wahl *et al.*, “Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.,” *Nature*, vol. 541, no. 7635, pp. 81–86, 2017.

[32] C. L. Relton and G. Davey Smith, “Epigenetic Epidemiology of Common Complex Disease: Prospects for Prediction, Prevention, and Treatment,” *PLoS Med.*, vol. 7, no. 10, p. e1000356, Oct. 2010.

[33] T. Tahara and T. Arisawa, “DNA methylation as a molecular biomarker in gastric cancer,” *Epigenomics*, vol. 7, no. 3, pp. 475–486, Jun. 2015.

[34] R. Kandimalla, A. A. van Tilborg, and E. C. Zwarthoff, “DNA methylation-based biomarkers in bladder cancer,” *Nat. Rev. Urol.*, vol. 10, no. 6, pp. 327–335, Jun. 2013.