



Extensive variation and low heritability of DNA methylation identified in a twin study

Kristina Gervin, Martin Hammerø, Hanne E. Akselsen, et al.

Genome Res. 2011 21: 1813-1821 originally published online September 26, 2011

Access the most recent version at doi:[10.1101/gr.119685.110](https://doi.org/10.1101/gr.119685.110)

References This article cites 27 articles, 2 of which can be accessed free at:
<http://genome.cshlp.org/content/21/11/1813.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2011 by Cold Spring Harbor Laboratory Press

Research

Extensive variation and low heritability of DNA methylation identified in a twin study

Kristina Gervin,¹ Martin Hammerø,¹ Hanne E. Akselsen,¹ Rune Moe,¹ Heidi Nygård,¹ Ingunn Brandt,² Håkon K. Gjessing,^{2,3} Jennifer R. Harris,² Dag E. Undlien,¹ and Robert Lyle^{1,4}

¹Department of Medical Genetics, Oslo University Hospital and University of Oslo, 0407 Oslo, Norway; ²Division of Epidemiology, Norwegian Institute of Public Health, 0403 Oslo, Norway; ³Department of Public Health and Primary Health Care, University of Bergen, 5020 Bergen, Norway

Disturbance of DNA methylation leading to aberrant gene expression has been implicated in the etiology of many diseases. Whereas variation at the genetic level has been studied extensively, less is known about the extent and function of epigenetic variation. To explore variation and heritability of DNA methylation, we performed bisulfite sequencing of 1760 CpG sites in 186 regions in the human major histocompatibility complex (MHC) in CD4⁺ lymphocytes from 49 monozygotic (MZ) and 40 dizygotic (DZ) twin pairs. Individuals show extensive variation in DNA methylation both between and within regions. In addition, many regions also have a complex pattern of variation. Globally, there appears to be a bimodal distribution of DNA methylation in the regions, but a significant fraction of the CpG sites are also heterogeneously methylated. Classification of regions into CpG islands (intragenic and intergenic), 5' end of genes not associated with a defined CpG island, conserved noncoding regions, and random CpG sites shows region-type differences in variation and heritability. Analyses revealed slightly lower intra-pair differences among MZ than among DZ pairs, suggesting some genetic influences on DNA methylation variation, with most of the variance attributed to nongenetic factors. Overall, heritability estimates of DNA methylation were low. Our heritability estimates are, however, somewhat deflated due to the presence of batch effects that artificially inflate the estimates of shared environment.

[Supplemental material is available for this article.]

The human epigenome is the state of all epigenetic modifications in a cell. It is under constant influence of the environment, is highly dynamic, and differs between cell types. DNA methylation of cytosines in a CpG-context is an epigenetic modification essential for mammalian development. It has been implicated in diverse processes such as genomic imprinting, X-chromosome inactivation, cell differentiation, chromosome structure, and the regulation of gene expression. The majority of CpGs in the human genome are clustered into CpG islands (CGIs), which are often colocalized with gene promoters. CGIs are generally unmethylated to allow transcription of associated genes, whereas the majority of CpGs in intergenic and coding regions are methylated. Regional differences in DNA methylation are necessary for regulation of gene expression and to maintain genome integrity (Bird 2002; Weber and Schubeler 2007).

Phenotypic differences between individuals result from genetic, epigenetic, and environmental influences. Recently we have gained considerable knowledge about genetic variation in the form of single-nucleotide polymorphisms (SNPs) and copy number variants (CNVs) and how these can affect disease susceptibility (The Wellcome Trust Case Control Consortium 2007, 2010). However, little is known about the normal range of epigenetic variation and the extent to which it contributes to phenotypic differences and disease susceptibility, although some studies have revealed inter-individual epigenetic variation (Flanagan et al. 2006; Bock

et al. 2008; Kaminsky et al. 2009; Schneider et al. 2010). Research is needed to gain a better understanding of normal DNA methylation variation and to understand the degree to which this variation is explained by genetic (heritable) influences. Although MZ twins are genetically identical, they do not show complete concordance for complex traits (Boomsma et al. 2002). This, in conjunction with reports of epigenetic differences in twins, highlights the importance of determining the degree to which MZ discordance can be explained by epigenetic variation (Fraga et al. 2005; Kaminsky et al. 2009).

In this study, we explore DNA methylation in a large set of healthy MZ and DZ twins to investigate inter-individual variation, patterns of DNA methylation, and the heritable basis of epigenetic differences in the major histocompatibility complex (MHC). The MHC is located on chromosome 6p21.3 and is a particularly polymorphic and gene-dense region, containing genes involved in several aspects of the immune response (Traherne 2008). We focus on the MHC because it is important in the development of autoimmune diseases that show heritable effects but that also show a large degree of discordance rates in MZ twin pairs (Hewagama and Richardson 2009). Furthermore, the MHC was also part of a pilot for the human epigenome project (HEP) which helps to complement our search for functional effects of methylation in this candidate region (Rakyan et al. 2004).

We isolated lymphocyte subpopulations and studied single-cell types to overcome the problem of epigenetic heterogeneity in whole blood. We quantified DNA methylation at 1760 CpG sites in 186 regions of the MHC in isolated CD4⁺ cells from 178 individuals (49 MZ and 40 DZ twin pairs). To our knowledge, this is the largest twin study to survey DNA methylation at the sequence level.

⁴Corresponding author.

E-mail Robert.Lyle@medisin.uio.no.

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.119685.110>.

Results

Methylation in the MHC

Globally, DNA methylation for the selected regions in the MHC is characterized by a bimodal distribution, tending to show either low or high methylation (Fig. 1). This is consistent with previous studies (Eckhardt et al. 2006; Zhang et al. 2009). Comparisons of DNA methylation levels across the MHC in 89 unrelated individuals (one randomly selected twin from each twin pair) reveal considerable variation at the level of individual CpGs (Fig. 1A). We divided the amplicons into four different region types based on their sequence or biological characteristics: CpG islands (CGI), 5' ends of genes (SPRIM), conserved noncoding regions (CNC), and randomly selected regions (RAND) (Supplemental Tables S1, S2). CGIs, SPRIMs, and CNCs are regulatory regions, and DNA methylation at these locations is thought to be a key epigenetic mechanism involved in the transcriptional control of associated genes. CGIs are typically located in, or close to, transcription start sites (TSSs) of genes and are generally unmethylated to allow transcription. Wherever possible, one functional amplicon was designed to cover as many CpGs as possible in each of these four types of regions. RAND regions were chosen to provide an overview of the level of DNA methylation in typically CpG-poor parts of the MHC. We observe that DNA methylation within the CGIs, SPRIMs, and CNCs shows a strong bimodal distribution, with a larger portion of amplicons in CNCs being more methylated. In contrast, RAND amplicons are highly methylated (Fig. 1A,B). These results are consistent with known DNA methylation profiles in respectively CpG-rich and CpG-poor regions in the genome (Weber et al. 2007; Bock et al. 2008; Irizarry et al. 2009).

The results presented here are based on bisulfite sequencing of DNA isolated from cultured cells. Although culturing can have a potential effect on DNA methylation, it has not been extensively studied. To address this issue and to assess the level of technical noise as a potential contributor to the observed variation, we conducted two sets of control experiments. In the first set of control experiments, we sequenced all amplicons in isolated CD4⁺ cells from one individual and compared the data from cultured and uncultured cells (correlation = 0.97, Spearman). Further, to measure the reproducibility of the culturing, we selected one twin, made three cultures of CD4⁺ cells, and compared the data from all amplicons from the three cultures to one another. The replicate-to-replicate variance is generally low (Supplemental Fig. S1). In order to confirm the robustness of our study, we ran a second set of control experiments by replicating eight samples (4 MZ twin pair) to assess the level of technical and biological variation. Data from the 4 MZ pairs were compared against themselves or against the respective co-twin (Kaminsky et al. 2009). A one-sided Ansari-Bradley test showed that variance in self-self comparison is significantly smaller than the twin-co-twin comparison (*P*-values: 5.5×10^{-12} , $<2.2 \times 10^{-16}$, $<2.2 \times 10^{-16}$, and $<2.2 \times 10^{-16}$ for four twin pairs).

Region-specific variation in DNA methylation

We detect a considerable amount of variation in DNA methylation among individuals in the 186 selected regions in the MHC (Fig. 1A). Interestingly, the overall distributions of DNA methylation values vary significantly by region types (Kolmogorov-Smirnoff test, *P* < 0.0001) (Fig. 1B; Supplemental Table S3; Supplemental Fig. S2). In contrast to the large differences in methylation values, the variance of methylation values for region types is more similar;

however, CGI and SPRIM remain significantly different from CNC and RAND (Kolmogorov-Smirnoff test, *P* < 0.02) (Fig. 1C; Supplemental Table S3; Supplemental Fig. S2). These results suggest that differences in DNA methylation levels at the different region types are functional.

Methylation patterns in MZ and DZ twin pairs

Next we looked at within-pair methylation differences for MZ pairs (*n* = 49), DZ pairs (*n* = 40), and randomly selected pairs of individuals, matched for age and sex (RZ) (*n* = 40). Within-pair differences tend to be small but are largest among RZ pairs and larger for DZ than MZ pairs (Fig. 2). Consistent with promoters being generally unmethylated, there is very little variation in methylation at CGI and SPRIM regions between MZ and DZ pairs. In contrast, CNC and RAND show greater zygosity effects with larger intra-pair differences in DZ compared to MZ pairs (Fig. 2). Intra-class correlations (ICCs) were computed to test for differential co-twin resemblance in DNA methylation among MZ, DZ, and RZ pairs. These correlations, graphed as box plots in Figure 3, show the degree of within-pair resemblance for methylation of individual CpG sites across all regions (Fig. 3A) and stratification by region type (Fig. 3B). Overall, the ICCs are quite high, and the pattern of ICC (MZ>DZ>RZ) (Fig. 3A) is consistent with the hypothesis that genetic effects contribute to DNA methylation variation. However, the importance of these heritable effects varies by type of region as reflected by differences in the magnitude of the between-pair differences across the regions (Fig. 3B). Interestingly, SPRIM and CNC have higher correlations than CGI, implying a functional constraint on methylation variation at these regions.

Heritability of DNA methylation

Estimates of heritability were derived on the basis of the within-position correlations of methylation in the MZ versus DZ twins. Figure 4 presents the distributions of the MZ-DZ differences in methylation correlation values by all positions together (Fig. 4A) and by the four region types separately (Fig. 4B). The superimposed line shows the expected density curve under the model of no genetic influence on methylation; shifts of the MZ-DZ distributions to the right relative to this curve suggest genetic effects, with greater shifts reflecting greater values of heritability. Overall, our findings reveal low heritability, ranging from 2% to 16% across the four types of regions (Fig. 4C).

In this study, we chose to process samples from each twin pair together throughout the experiments and analytical phases in order to avoid potential influences of batch effects on within-pair differences in methylation. However, if batch effects are present, then this can artificially inflate the MZ and the DZ correlations and thereby lead to an overestimation of shared environment (*c*²). Since we are not able to disentangle the variance component in methylation explained by common environment from that due to these potential batch effects, we elect to present only heritability estimates.

Discussion

This study surveyed DNA methylation in the MHC in a single cell-type (CD4⁺ lymphocytes) from a sample of healthy MZ and DZ twins. Direct bisulfite sequencing reveals substantial inter-individual variation in DNA methylation. Classification of the 186 regions into CpG islands, 5' regions, conserved noncoding, and

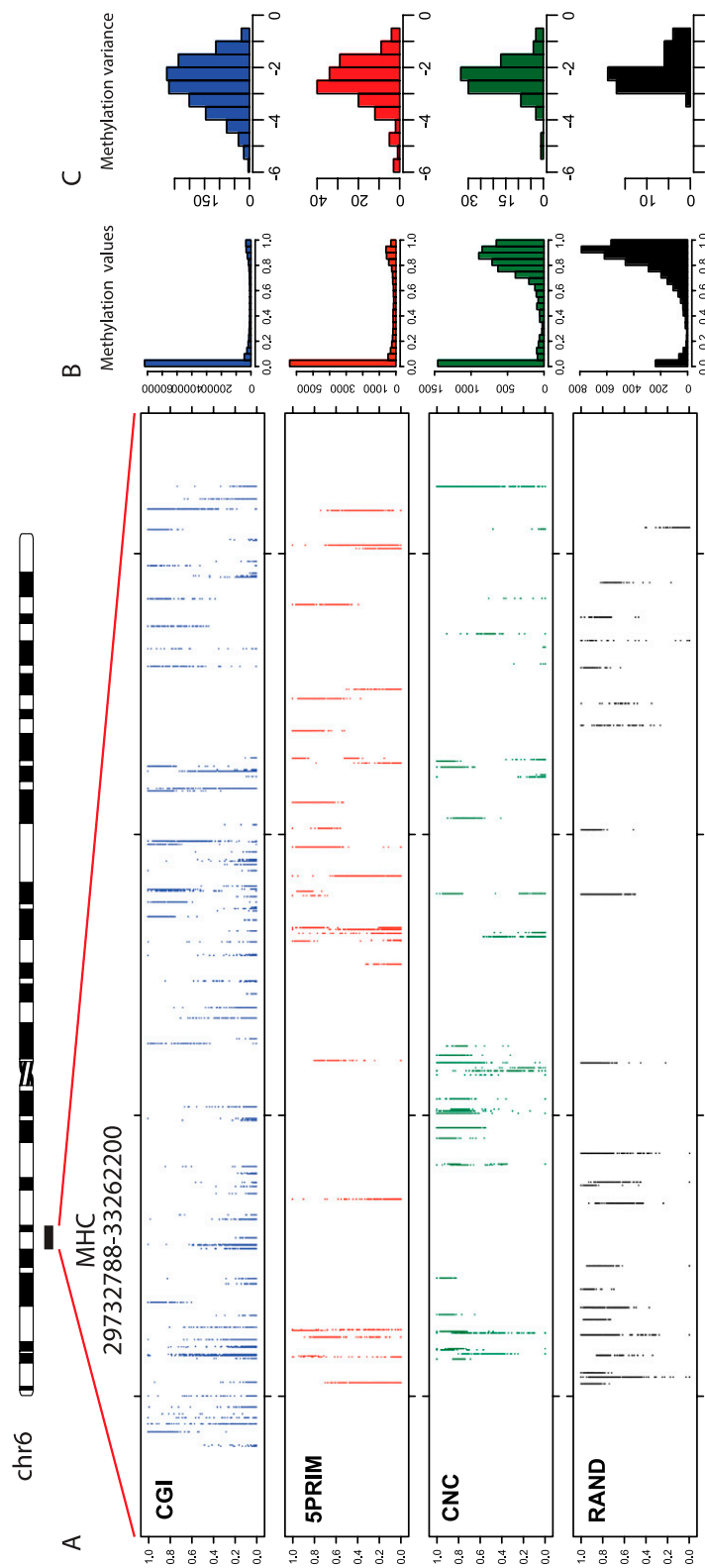
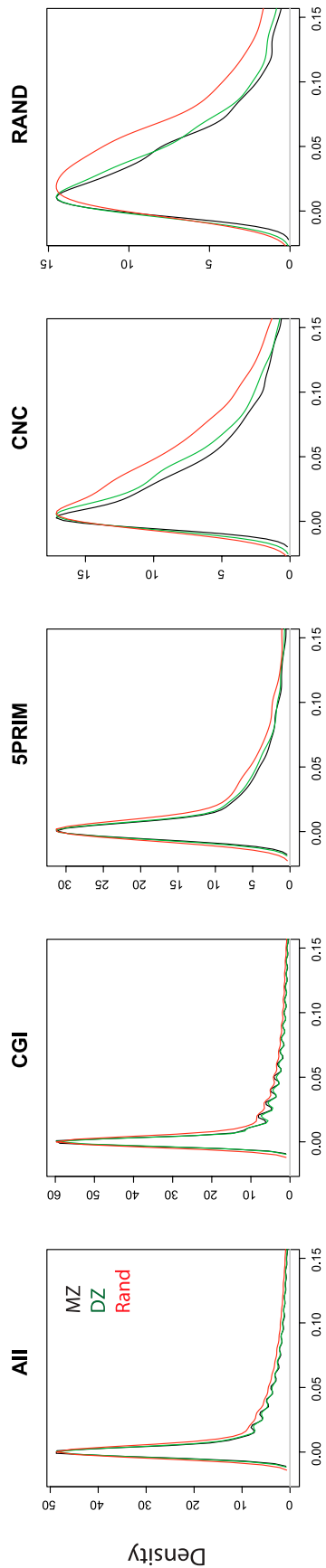


Figure 1. Overview of global DNA methylation across 186 amplicons in 89 unrelated individuals in the MHC. A high degree of variation in DNA methylation can be observed. (A) Extent of DNA methylation for different amplicon types in the MHC. (CGI) CpG islands, (5PRIM) 5' regions of gene (nonCG), (CNC) conserved noncoding sequence, (RAND) random. Y-axis: 0.0 unmethylated, 1.0 complete methylation; x-axis: position of CpG sites along the MHC. (B) Distributions of DNA methylation values for different region types showing that regions have different methylation patterns (all pairwise comparisons are significantly different). (C) Distribution of variance for methylation for region types. CGI and 5PRIM are significantly different from CNC and RAND (Kolmogorov-Smirnov test, $p < 0.02$) (Supplemental Table S3).



Intra-pair methylation difference

Figure 2. Absolute values of differences in DNA methylation between MZ and DZ pairs for the different classes of regions. Analysis of all regions together displays little difference between MZ and DZ pairs. Analysis of specific regions reveals no differences between MZ and DZ pairs for CGIs and 5PRIM, whereas CNCs and random sites (RAND) vary more, with DZ pairs showing larger differences.

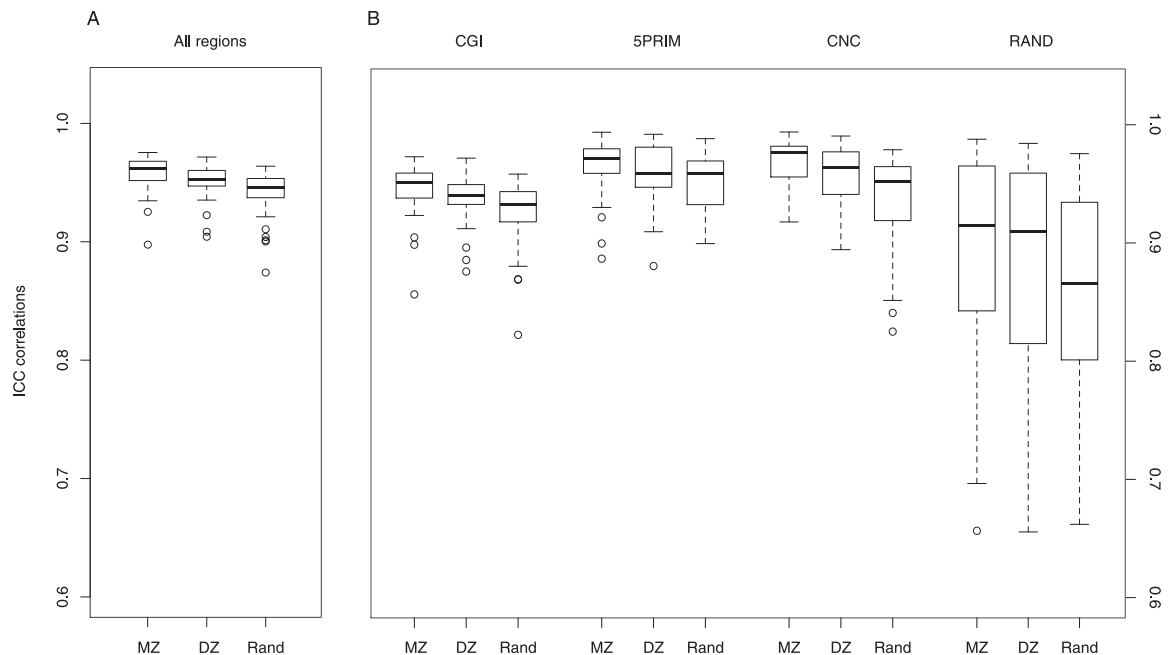


Figure 3. Box plots displaying intra-class correlations of DNA methylation overall (A) and in the different region types (B) calculated for MZ, DZ, and RZ. The correlations are computed for each twin pair separately, as correlations between all methylation values over all amplicons and CpGs for that twin pair.

random further reveals distinct patterns of DNA methylation, which may reflect the differential biological functions of these regions. CGI and 5PRIM are characterized by a bimodal distribution of DNA methylation, with a predominance of low methylation. This most likely reflects their localization in gene promoters and a role in regulation of gene expression, although there is no simple relationship between DNA methylation and gene expression. Many promoters are not associated with precisely defined CGIs but still have been shown to have methylation patterns which correlate with gene expression (Bernstein et al. 2007). CNCs also show a bimodal distribution of methylation but slightly more dispersed between high and low methylation. In contrast, the observed high methylation of amplicons in the RAND category reflects the global methylation status of noncoding DNA and is generally associated with repressive chromatin. The regions explored in this study may not be representative of the whole genome, but the study emphasizes the need for high resolution methods and large sample sets in order to explore epigenetic variation.

In contrast to other recent studies, we did not detect age-related changes in DNA methylation (Fraga et al. 2005; Boks et al. 2009; Schneider et al. 2010), probably because the age range in our twin sample (30–42 yr) was narrower.

We observed a substantial range of variation in DNA methylation across individuals. In comparison with genetic variation, the potential functional outcome of epigenetic variation in shaping the phenotype has yet to be explored. In order to be able to interpret observed methylation patterns that could be associated with disease, it is first essential to reveal the normal range of variation. In this context, our study contributes to the understanding of how patterns of DNA methylation vary between individuals, how the distribution differs between distinct regions in the genome, and the influence of genetic and environmental factors on DNA methylation.

In general, epigenetic marks are considered heritable throughout mitotic divisions, whereas between generations these marks

are erased and re-established. Over the past few years, the environment has been shown to strongly influence the establishment and composition of the epigenome (Cropley et al. 2007; Majumdar et al. 2010; Fryer et al. 2011). The use of twins is an attractive model to investigate the heritability of DNA methylation by comparing the correlation of DNA methylation in MZ versus DZ twins. We observe larger within-pair differences for DZ than MZ twins in CNC and RAND regions compared to promoter regions (CGI and 5PRIM). This is in agreement with another study which also observed locus-specific differences in heritability (Kaminsky et al. 2009). The majority of variation in methylation in our study can be attributed to nongenetic factors, although the observed lower intra-pair differences among MZ than among DZ pairs suggest some genetic influence on DNA methylation variation. In addition, variance in DNA methylation can also be due to cumulative stochastic differences during epigenetic programming and replication throughout development. This variation would reduce the similarity between twins regardless of zygosity.

While the low replicate-to-replicate variance is reassuring, our analysis of the replicated four MZ twin pairs showed that technical variation is still a substantial part of the overall variation. Heritability estimates may thus be deflated since the denominator always contains biological variability plus technical noise, but comparing MZ with DZ twins is still valid since twin pair samples were always kept together during processing. Collectively, our findings indicate high reliability of our methylation results. We are unable to isolate variation in methylation that may be induced solely by culturing from other sources of technical noise. However, if culturing is the main source of random variation that may be affecting our methylation results, those culturing effects should be captured by our experiments. Finally, although not experimentally derived, we note that culture-induced variation would operate to make twins less similar, as we would expect these effects to occur randomly across both MZ and DZ pairs and thereby decrease the ICCs. In our models, this would be estimated as nonshared environmental

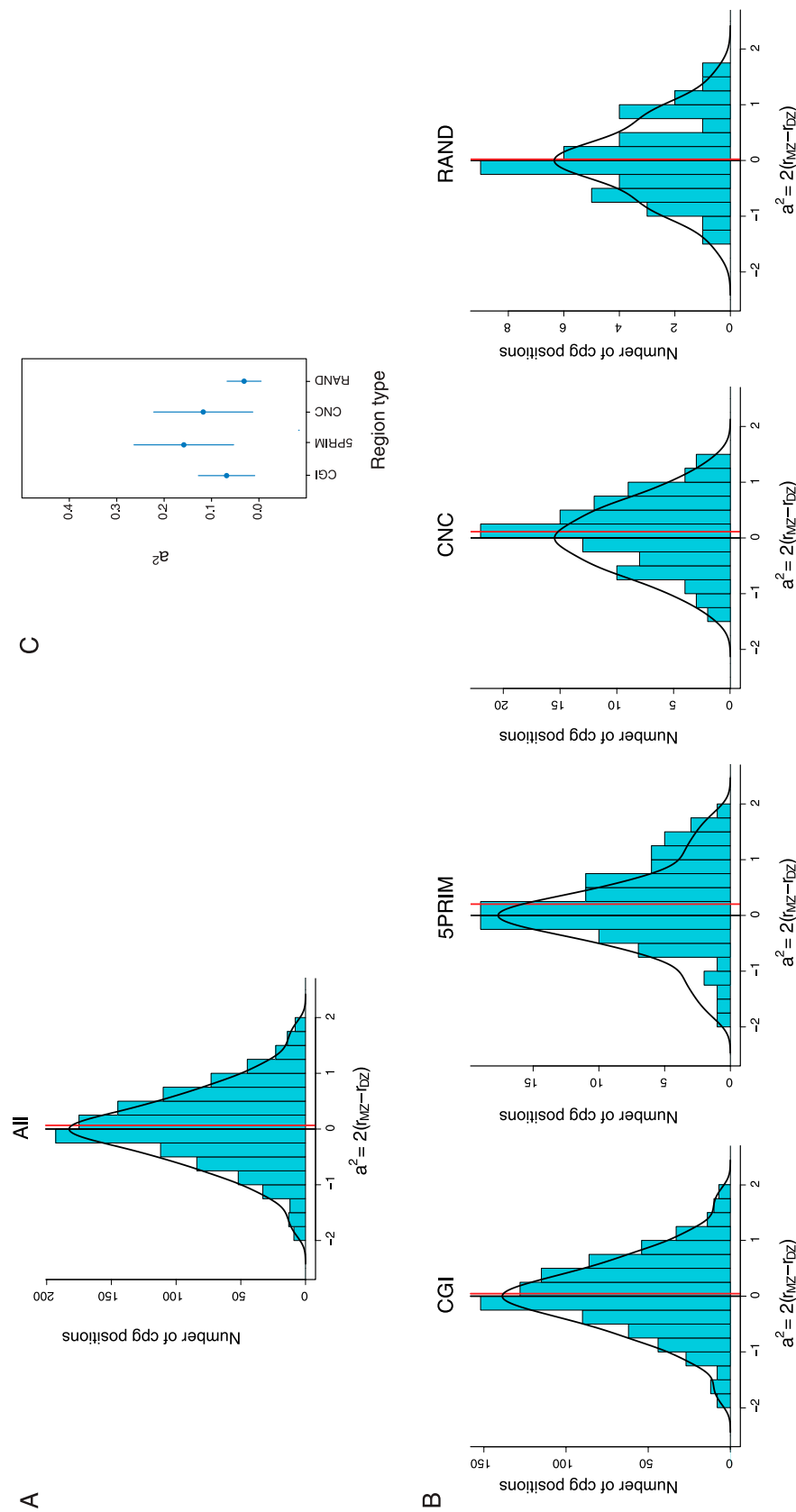


Figure 4. Histograms showing twin pair correlation of DNA methylation in MZ versus DZ twin pairs. MZ correlation minus DZ correlation is computed separately in each of 1156 individual CpGs in 175 amplicons (nonvariable positions are removed) and shown overall (A) and in different region types (B). The superimposed density curve (in black) shows what it would look like if there were no differences between MZ and DZ pairs, i.e., if genes had no influence, only a common environment. (C) Plots of confidence intervals from the estimated heritabilities for the various sequences.

variance, and these culturing effects would result in a downward bias of the heritability estimates.

In the original design, the samples from each twin pair were analyzed in the same experimental batches to ensure that within-pair differences are not confounded by batch effects. However, a drawback of this design is that the batch effects can artificially inflate twin resemblance leading to an overestimation of the effects of shared environment (c^2). This situation arises when members of a pair are perfectly correlated for a factor that is associated with the outcome under study. In the absence of any batch effects, our findings would reflect the upper limit of c^2 ; however our design does not disentangle variance in methylation explained by common environment from variance arising from potential batch effects. Furthermore, if batch effects are contributing to the overall variance, then this would deflate the a^2 estimates, and the results should be considered as lower limits. However, the batch effects seem to contribute far less to the overall methylation variability than do the amplicon-to-amplicon and CpG position-to-position.

Other studies examining heritability of DNA methylation reported greater heritability than we found (Heijmans et al. 2007; Boks et al. 2009). In addition, Kaminsky et al. (2009) indicated that the heritable basis of DNA methylation may not be due to genetic differences (Kaminsky et al. 2009). However, these studies used DNA extracted from heterogeneous cell populations (whole blood, buccal cells), suggesting cell-type differences in DNA methylation variation. Indeed, Kaminsky et al. (2009) noted differences between buccal and whole-blood cells. The present study explores a single cell-type isolated from whole blood, which consists of a complex composition of cells that are differentially methylated (Eckhardt et al. 2006). Thus, the possibility of masking genuine differences which can arise when averaging DNA methylation using pooled cell-types is diminished in our experiments.

One reason for our low estimates of heritability may be that the loci we studied are in the MHC, and in this region a particularly high level of the DNA methylation variation is due to environmental factors. Clearly, further work is needed on more cell types and loci to understand the genetic and environmental contribution to variation in DNA methylation.

Methods

Samples

The twins were recruited pairwise from the population-based study of Norwegian twins born 1967–1979 (Harris et al. 1995, 1997). Initial screening was conducted for the absence of health problems and diseases using self-reported health history data collected in 1998 (Harris et al. 2002). Consenting pairs were then invited to a clinical interview conducted by a nurse to determine if they remained illness-free. Additional information was also collected related to medication use, family health history, and lifestyle factors. Among the 354 pairs invited from the screening phase, 121 pairs consented, and 108 pairs participated. The data presented herein are based on 49 MZ and 40 DZ pairs. Sixty mL EDTA-blood was collected, either at Oslo University Hospital, Ullevål, or regional hospitals and transported by a field nurse or by post overnight at room temperature. The zygosity for all twin pairs was determined based on 13 microsatellites on chromosomes 13, 18, 21, X, and Y (Supplemental Table S4).

Cell separation

Peripheral blood mononuclear cells (PBMCs) were separated from whole blood using Lymphoprep (Axis-Shield). Different lympho-

cyte populations were then sequentially isolated in a semi-automated way using positive and negative isolation kits from Miltenyi on an autoMACS Pro separator. The PBMC fraction was divided into two aliquots: CD4⁺CD25⁺ cells were isolated from one fraction, and CD19⁺, CD8⁺, and CD4⁺ cells were sequentially isolated from the other fraction according to the manufacturer's instructions. Briefly, CD4⁺CD25⁺ cells were positively isolated from a depleted CD4⁺ fraction using the CD4⁺CD25⁺ regulatory T cell isolation kit and the Depl05 and Posseld2 programs on the autoMACS Pro separator. CD19⁺ cells were isolated by positive isolation using CD19⁺ Micro Beads and the Possel program. CD8⁺ cells were then positively isolated from the negative fraction using CD8⁺ Micro Beads and the Possel program. Finally, CD4⁺ cells were then separated from the negative fraction by negative isolation (i.e., by labeling all other cells but the CD4⁺ cells) using CD4⁺ T Cell Isolation kit II and the Deplete program. Only CD4⁺ cells were used in this study.

Cell culturing

CD4⁺ cells were cultured for 7 d with added exogenous rIL-2 (10 ng/ μ l) and Dynabeads CD3/CD28 T Cell Expander (Invitrogen) in an incubator using 1/2 bead per cell. Dynabeads were washed before use according to the manufacturer's instructions.

DNA isolation

DNA was isolated from cultured cells on a Gentra autopure LS (Qiagen) using the 2–5 $\times 10^7$ protocol. This resulted in high yields of pure DNA with an A_{260} – A_{280} between 1.7 and 1.9.

Bisulfite sequencing

Amplicons were designed using Methyl Primer Express v1.0 (ABI). Primers were selected to include at least three conversion positions to achieve specificity in the PCR and have no CpG sites to ensure amplification regardless of methylation status of the internal sequence (Supplemental Table S5). Wherever possible, amplicons containing >9 poly T repeats in the converted sequence were avoided due to problems associated with polymerase slippage and poor amplification. All primers were tested for specificity and the capability of producing sequences of high quality. For each sample, 3 μ g of genomic DNA was bisulfite-converted using MethylEasy Xceed (Human Genetic Signatures) according to the manufacturer's instructions. PCRs were performed in a final volume of 5 μ l in 384-well plates containing 1 \times Gold buffer (ABI), 4 mM dNTP mix (ABI), 1.5 mM MgCl₂, 0.5 U Gold polymerase (ABI), 5 pmol each of forward and reverse primer, and 10 ng of bisulfite-treated DNA. The cycling conditions were 95°C for 5 min, followed by five cycles of 95°C for 30 sec, 55°C for 2 min, and 72°C for 3 min, and then 30 cycles of 95°C for 30 sec, 65°C for 1 min, and 72°C for 3 min, and a final extension step for 85 min at 60°C. Unincorporated primers and dNTPs were enzymatically eliminated by adding 2 μ l ExoSAP-IT and incubating for 15 min at 37°C, followed by 15 min at 80°C. Amplicons were sequenced directly using a 1:32 dilution of BigDye Terminator v3.1 chemistry (ABI) in the forward direction using M13 primer, purified on Multiscreen₃₈₄-SEQ (Millipore), and separated on an ABI 3730 capillary sequencer. The level of methylation was quantified using Epigenetic Sequencing Methylation analysis software (ESME) (Lewin et al. 2004). We used a slightly modified version of ESME allowing analysis of amplicons with <3 CpG sites and the ABI KB basecaller.

Analysis of heritability

Selection of the specific model to analyze the sources of variation in twin data is guided by the pattern of ICCs across zygosity. MZ

twins are perfectly correlated for additive and nonadditive (dominance, epistasis) genetic effects, whereas DZ twins are correlated, on average, 0.5 for additive effects but significantly less for non-additive effects. Therefore, nonadditive genetic effects are typically modeled when the magnitude of the MZ correlations is greater than twice the DZ correlations. In contrast, DZ correlations greater than half the MZ correlations suggest the importance of additional sources of twin resemblance (beyond additive genetic effects), and shared environmental effects are then included in the model. The patterns of ICC in our data do not provide evidence of non-additive effects, and our analyses were, therefore, based on a so-called ACE model that includes additive effects of genes (A), effects of common environment (C), and effects of nonshared environment (E) (Neale and Cardon 1992). The model was estimated using equations that differentially parameterize the twin correlations for MZ and DZ pairs based on their differential degree of biological relatedness. Additive genetic influences on methylation are perfectly correlated between members of MZ pairs and are correlated 0.5 between members of DZ pairs. Shared environmental effects contribute to co-twin resemblance and are, by assumption, correlated 1.0 in both MZ and DZ pairs. Nonshared environmental influences contribute to intra-pair differences in both MZ and DZ pairs.

We use a standardized model, i.e., a model where the total methylation variance has been standardized to 1 within each CpG position. As a consequence, the variance of a methylation measurement can be decomposed as $1 = a^2 + c^2 + e^2$ for the ACE model, where a^2 refers to additive genetic variance, c^2 refers to variance due to shared environment, and e^2 refers to variance due to nonshared environment. The corresponding model expressions for twin pair correlations are $a^2 + c^2$ and $1/2 a^2 + c^2$ for MZ and DZ pairs, respectively. We estimated the MZ and DZ correlations using a standard estimator for the intraclass correlations r_{MZ} and r_{DZ} (Neale and Cardon 1992). The corresponding estimates for a^2 and c^2 were then obtained by solving the moment equations, giving $a^2 = 2(r_{MZ} - r_{DZ})$ and $c^2 = 2r_{DZ} - r_{MZ}$. One should note that the estimation is unrestricted in the sense that the models may well produce negative estimates for a^2 and c^2 . For instance, if there is no additive genetic component at a position, the estimates of a^2 in the ACE model will indeed often end up negative due to random variation in the samples. The model clearly predicts that $r_{MZ} \geq r_{DZ}$, and if $r_{MZ} \geq r_{DZ} \geq 1/2 r_{MZ}$, the ACE model will fit in the sense that both a^2 and c^2 will have nonnegative estimated values. The estimate of c^2 may be inflated in our model due to batching effects that arise when data from both twins in a twin pair were analyzed in the same batches. To separately assess batch effects, we resequenced eight randomly selected twin pairs over all amplicons as the first analyses. We then estimated the variance of the difference between first and second measurements on the same individuals. The variances were computed separately for each CpG position. This variance was compared to the variance of the intra-twin pair differences and the intra-pair differences in randomly paired individuals. The batch variance is almost as great as the variance in randomly paired individuals within a CpG position. While this is based only on eight twin pairs, it suggests that a batching effect is present and that the moderate common environment observed in the heritability analyses may be due to batching. We thus only present the a^2 estimates here and note that, without batching effects, the a^2 estimates might well be larger, since they measure the ratio of genetic variance to the total variance, including batch effects. A final caveat is that the models are usually applied to normally distributed outcomes, in which case the definitions of additive genetic effects are more clear-cut than in the case of a restricted outcome.

Data cleaning

A large number of CpG positions exhibit little or no variation at all. At such positions, estimates of the intra-class correlation, and ultimately of the heritability, are unstable and thus likely to confuse the overall picture of heritability. To reduce the impact of such positions on our estimates, we studied the effect of removing positions with low information content. This was done using two different measures of methylation variability at a position—the entropy and the variance. High variance at a position typically means that the distribution has many zeros and ones. High entropy typically means that the distribution is “smeared out” and not focused on specific values. Low variance and/or entropy means that the distribution degenerates to a single value, usually zero or one. However, removing even a substantial amount of low variance or low entropy positions had little impact on the final heritability estimates as compared to the sampling uncertainty (results not shown). In the final heritability analysis, we removed the 580 CpG positions with lowest entropy (set at 20% of the maximal entropy observed at a position).

Data access

Sequence trace files are available from the authors on request.

Acknowledgments

This study was supported by the national Functional Genomics Programme (FUGE) of the Research Council of Norway. This research was also partly funded through the European Community’s Seventh Framework Programme (FP7/2007-2013), ENGAGE project (grant agreement HEALTH-F4-2007-201413), and Southern and Eastern Norway Regional Health Authority (grant agreement 51932).

Authors’ contributions: Study design: K.G., J.R.H., D.E.U., and R.L.; twin collection: H.N., I.B., and J.R.H.; cell separations: K.G., M.H., and R.M.; bisulfite sequencing: K.G., M.H., H.E.A., and R.M.; data analysis: K.G., H.K.G., and R.L.; and manuscript writing: K.G., M.H., H.K.G., J.R.H., D.E.U., and R.L.

References

- Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. *Cell* **128**: 669–681.
- Bird A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6–21.
- Bock C, Walter J, Paulsen M, Lengauer T. 2008. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. *Nucleic Acids Res* **36**: e55. doi: 10.1093/nar/gkn122.
- Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA. 2009. The relationship of DNA methylation with age, gender, and genotype in twins and healthy controls. *PLoS ONE* **4**: e6767. doi: 10.1371/journal.pone.0006767.
- Boomsma D, Busjahn A, Peltonen L. 2002. Classical twin studies and beyond. *Nat Rev Genet* **3**: 872–882.
- Cropley JE, Suter CM, Martin DI. 2007. Methyl donors change the germline epigenetic state of the A(vy) allele. *FASEB J* **21**: 3021–3022.
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al. 2006. DNA methylation profiling of human chromosomes 6, 20, and 22. *Nat Genet* **38**: 1378–1385.
- Flanagan JM, Popenkikyte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, Schumacher A, Zangeneh M, Lau L, Virtanen C, Wang SC, et al. 2006. Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* **79**: 67–84.
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci* **102**: 10604–10609.
- Fryer AA, Emes RD, Ismail KM, Haworth KE, Mein C, Carroll WD, Farrell WE. 2011. Quantitative, high-resolution epigenetic profiling of CpG loci

- identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics* **6**: 86–94.
- Harris JR, Tambs K, Magnus P. 1995. Sex-specific effects for body mass index in the new Norwegian twin panel. *Genet Epidemiol* **12**: 251–265.
- Harris JR, Magnus P, Samuelsen SO, Tambs K. 1997. No evidence for effects of family environment on asthma. A retrospective study of Norwegian twins. *Am J Respir Crit Care Med* **156**: 43–49.
- Harris JR, Magnus P, Tambs K. 2002. The Norwegian Institute of Public Health Twin Panel: A description of the sample and program of research. *Twin Res* **5**: 415–423.
- Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. 2007. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum Mol Genet* **16**: 547–554.
- Hewagama A, Richardson B. 2009. The genetics and epigenetics of autoimmune diseases. *J Autoimmun* **33**: 3–11.
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**: 178–186.
- Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GHT, Wong AHC, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, et al. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* **41**: 240–245.
- Lewin J, Schmitt AO, Adorjan P, Hildmann T, Piepenbrock C. 2004. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics* **20**: 3005–3012.
- Majumdar S, Chanda S, Ganguli B, Mazumder DNG, Lahiri S, Dasgupta UB. 2010. Arsenic exposure induces genomic hypermethylation. *Environ Toxicol* **25**: 315–318.
- Neale MC, Cardon LR. 1992. *Methodology for genetic studies of twins and families*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A, et al. 2004. DNA methylation profiling of the human major histocompatibility complex: A pilot study for the human epigenome project. *PLoS Biol* **2**: e405. doi: 10.1371/journal.pbio.0020405.
- Schneider E, Pliushch G, El HN, Galetzka D, Puhl A, Schorsch M, Frauenknecht K, Riepert T, Tresch A, Muller AM, et al. 2010. Spatial, temporal, and interindividual epigenetic variation of functionally important DNA methylation patterns. *Nucleic Acids Res* **38**: 3880–3890.
- Traherne JA. 2008. Human MHC architecture and evolution: Implications for disease association studies. *Int J Immunogenet* **35**: 179–192.
- Weber M, Schubeler D. 2007. Genomic patterns of DNA methylation: Targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**: 273–280.
- Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. 2007. Distribution, silencing potential, and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**: 457–466.
- The Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**: 661–678.
- The Wellcome Trust Case Control Consortium. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**: 713–720.
- Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, Ragozin S, Reinhardt R, Groth M, Walter J, et al. 2009. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet* **5**: e1000438. doi: 10.1371/journal.pgen.1000438.

Received January 6, 2011; accepted in revised form August 22, 2011.