nature genetics

DNA methylation profiles in monozygotic and dizygotic twins

Zachary A Kaminsky[1,2], Thomas Tang[1], Sun-Chong Wang[1,3], Carolyn Ptak[1,2], Gabriel H T Oh[1,2], Albert H C Wong[1,2], Laura A Feldcamp[1,2], Carl Virtanen[4], Jonas Halfvarson[5,6], Curt Tysk[5,6], Allan F McRae[7], Peter M Visscher[7], Grant W Montgomery[7], Irving I Gottesman[8], Nicholas G Martin[7] & Art Petronis[1,2]

Twin studies have provided the basis for genetic and epidemiological studies in human complex traits[1,2]. As epigenetic factors can contribute to phenotypic outcomes, we conducted a DNA methylation analysis in white blood cells (WBC), buccal epithelial cells and gut biopsies of 114 monozygotic (MZ) twins as well as WBC and buccal epithelial cells of 80 dizygotic (DZ) twins using 12K CpG island microarrays[3,4]. Here we provide the first annotation of epigenetic metastability of ~6,000 unique genomic regions in MZ twins. An intraclass correlation (ICC)-based comparison of matched MZ and DZ twins showed significantly higher epigenetic difference in buccal cells of DZ co-twins ($P = 1.2 \times 10^{-294}$). Although such higher epigenetic discordance in DZ twins can result from DNA sequence differences, our *in silico* SNP analyses and animal studies favor the hypothesis that it is due to epigenomic differences in the zygotes, suggesting that molecular mechanisms of heritability may not be limited to DNA sequence differences.

Twin research has been of fundamental importance in human studies for two main reasons. First, comparison of phenotypic concordance rates in MZ twins versus DZ twins is a powerful strategy to estimate heritability. Second, phenotypic discordance in MZ co-twins has traditionally indicated roles of environmental factors. Countless twin studies have been done over the last century on almost every trait imaginable but primarily on human disease[2]. Nearly universally, MZ twins show various degrees of discordance, generally lower in comparison to discordance in DZ twins. These observations have provided the basis for the current paradigm of human normal and morbid biology, which focuses on DNA sequence variation and environmental differences. In the last decade, however, evidence has been accumulating that epigenetic modifications of DNA and histones can have a primary role in phenotypic outcomes, including human disease[5]. DNA methylation shows only partial stability, which could be caused by a wide variety of factors, including developmental programs, environment, hormones and stochastic events[6–9]. Such epigenetic metastability may result in substantial epigenetic differences across genetically identical organisms[10]. Several studies have identified epigenetic differences, either at selected genes of MZ twins[11–14] or in the overall epigenome[15]. Despite this promising start, no locus-specific epigenome-wide studies have yet been conducted to catalog the extent of this phenomenon, and few have been done in tissues other than peripheral blood cells.

In this study, we mapped MZ twin DNA methylation differences in white blood cells (WBC) ($N = 19$ pairs), buccal epithelial cells ($N = 20$ pairs) and gut (rectum) biopsies ($N = 18$ pairs) by interrogation of the unmethylated genome on the 12K CpG island microarray[3]. We first ensured that the microarray technology identifies actual DNA methylation differences between MZ co-twins rather than artifactual differences due to technical variation. For this, four parallel enrichments of the unmethylated fraction of genomic WBC DNA were done from the DNA stock of the same individual. DNA samples from eight MZ twins (four pairs) were compared against themselves (to measure technical variation) or the respective co-twin (to measure biological variation). The biological variation significantly exceeded the technical variation in all cases ($P = 1.4 \times 10^{-238}$, $P = 1.1 \times 10^{-202}$, $P = 2.1 \times 10^{-7}$, $P = 2.6 \times 10^{-39}$), indicating that the detected MZ co-twin differences are genuine (**Fig. 1**). The technical variation (σ^2) was consistent between all self–self hybridizations, and the degree of biological variation varied significantly between twin pairs (**Fig. 1**). Biological variation was detectably higher than technical variation in all tissues (**Supplementary Note** online). Furthermore, microarray validation done by sodium bisulfite sequencing and pyrosequencing (**Supplementary Fig. 1** online) indicated that the microarray signals detected reflect the actual DNA methylation status in the tested samples. For WBC-based analyses, we also conducted a spot-wise correlation between cell subfraction counts and confirmed that the

[1]Centre for Addiction and Mental Health, Toronto, Ontario M5T 1R8, Canada. [2]University of Toronto, Toronto, Ontario M5S 1A1, Canada. [3]Institute of Systems Biology and Bioinformatics, National Central University, Chungli 320, Taiwan. [4]University Health Network Microarray Centre, Toronto, Ontario M5G 1L7, Canada. [5]Division of Gastroenterology, Department of Medicine, 701 85 Örebro University Hospital, Örebro, Sweden. [6]School of Health and Medical Sciences, Örebro University, 701 85 Örebro, Sweden. [7]Queensland Institute of Medical Research, Brisbane 4029, Queensland, Australia. [8]University of Minnesota, Minneapolis, Minnesota 55454, USA. Correspondence should be addressed to A.P. (arturas_petronis@camh.net).

Received 18 April 2008; accepted 10 October 2008; published online 18 January 2009; doi:10.1038/ng.286

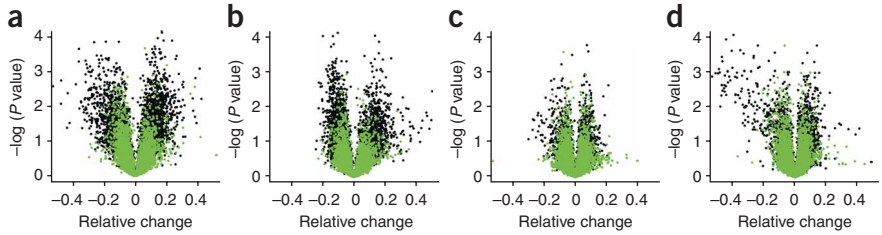

Figure 1 Volcano plots of four MZ twin versus co-twin WBC DNA methylation profile comparisons (black), with overlay of four matched twin DNA versus self comparisons (green) for each set of MZ twins. The *x* axis represents the mean relative change across the four replicas. The *y* axis represents the $-\log_{10}$ of the *P* value from a paired *t*-test. Higher significance denotes a higher consistency between replicates. (**a**–**d**) Significant variation in the spread of detected biological difference exists between twin pairs (Kruskal-Wallis $\chi^2 = 16.3$, d.f. = 3, $P = 0.001$) with a symmetrical large (**a** and **b**), symmetrical small (**c**) and asymmetrical (**d**) variation of the DNA methylation profile between co-twins. For each twin pair, a nonparametric Ansari-Bradley test demonstrated that levels of variance (σ^2) in the MZ twin–co-twin comparison were significantly larger than σ^2 in the self–self comparisons (twin set A: variance ratio = 2.91, $P = 1.4 \times 10^{-238}$; set B: 2.14, $P = 1.1 \times 10^{-202}$; set C: 1.12, $P = 2.1 \times 10^{-7}$; set D: 2.63, $P = 2.6 \times 10^{-39}$). Degrees of technical variation were not significantly different between groups (Kruskal-Wallis $\chi^2 = 1.81$, d.f. = 3, $P = 0.62$).

differences observed in WBC samples were not due to cell subfraction differences (**Supplementary Note**).

In the microarray-based studies, we detected a large degree of MZ co-twin DNA methylation variation in all tissues investigated. We used an intraclass correlation coefficient (ICC) to measure MZ co-twin variation for each unique genomic region, where an ICC range from +1 to –1 denotes high to low epigenetic similarity between co-twins relative to the variation between unrelated pairs. ICC distributions for each MZ twin group are depicted in **Supplementary Figure 2** online. We estimate that we have interrogated ~10,000–22,000 (~0.45–1%) of the 2.2 million *Hpa*II sites in the genome[3]. For each tissue, we generated an ICC-based annotation of MZ co-twin DNA methylation variation across ~6,000 unique DNA loci (**Fig. 2** (WBC) and **Supplementary Fig. 3** online (Buccal and Gut); all annotations are located online; see URLs section in Methods). We observed that ICC profiles were more similar across tissues of the same individual as compared to that of unrelated individuals (**Supplementary Note**). Notably, DNA methylation profiles in the buccal epithelial cells from monochorionic MZ twins were significantly more variable within pairs than those from dichorionic MZ twins (mean difference = 0.37 ± 0.0057, $P < 9.9 \times 10^{-324}$), which cannot be explained by technical differences between the hybridization batches of each group (**Supplementary Fig. 4** online). Chorionicity information was only available for the buccal and WBC samples; all WBC of MZ twins were dichorionic to avoid *in utero* twin blood transfusion effects. Dichorionic MZ twins are believed to result from a splitting of the blastomere

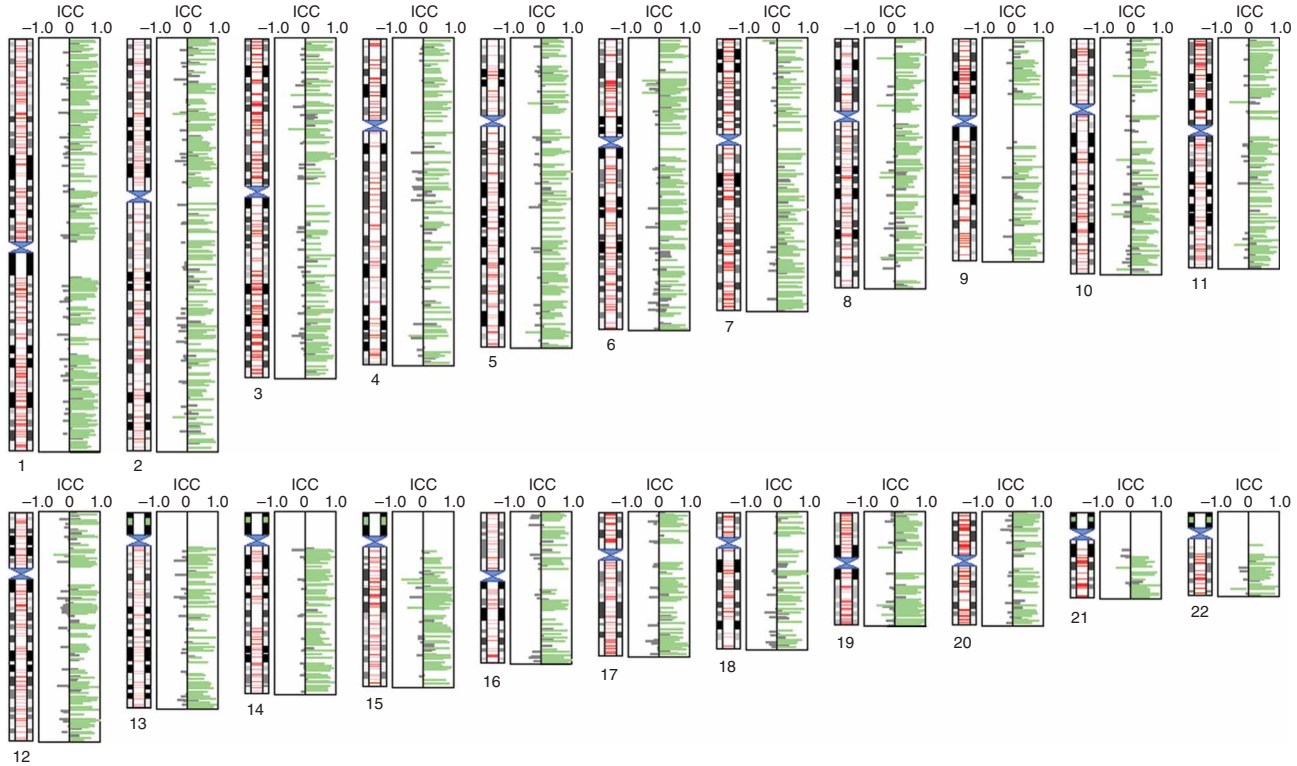

Figure 2 A chromosomal karyogram depicting degree of MZ co-twin similarity per interrogated locus in the WBC sample. Dark-to-light bars on the chromosomes represent chromosomal banding patterns as revealed by Giemsa staining, and red bars indicate regions of high microarray probe density. Bars to the right of each chromosome represent locus-specific ICCs depicting degrees of MZ co-twin epigenetic similarity. *P* values associated with the ICC statistic per locus were subjected to false discovery rate (FDR) correction for multiple testing. FDR-corrected *P* values below the level of $P < 0.05$ are depicted in green, and those with greater *P* values are shown in gray.

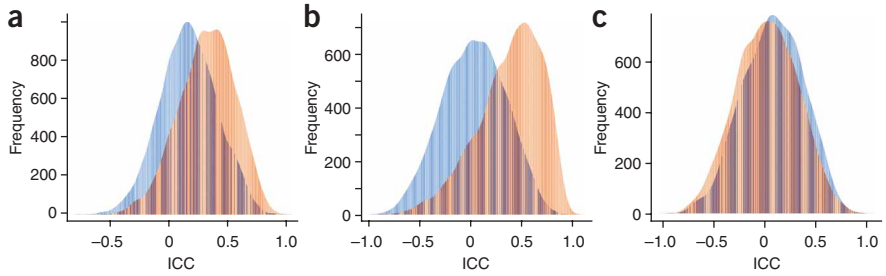

Figure 3 ICC distributions in buccal epithelial cells of MZ and DZ twins. (**a**) All MZ twins ($N = 20$ sets, red) and DZ twins ($N = 20$ sets, blue). (**b**) Dichorionic MZ twins ($N = 10$ sets, red) and matched DZ twins ($N = 10$ sets, blue). (**c**) Monochorionic MZ buccal samples ($N = 10$ sets, red) with matched DZ twins ($N = 10$ sets, blue).

within the first four days following fertilization, whereas monochorionic MZ twins arise after this point[16]. The varying degrees of epigenetic dissimilarity detected between these groups may reflect differences in epigenetic divergence among embryonic cells at the time the twin blastomeres separated.

Using locus-specific DNA methylation information, we investigated whether the degree of co-twin epigenetic similarity is associated with functional genomic elements. In each tissue, we compared the distribution of ICCs of the CpG islands (CGIs) to that of all non-CGI loci. Promoters were investigated in an identical manner. We carried out six tests and corrected P values for multiple testing using the Bonferroni method. Both CGIs and promoters were less epigenetically variable in WBC-derived DNA (Wilcoxon rank-sum test, $\text{mean}_{\text{CGI}} = 0.43 \pm 0.0065$, $\text{mean}_{\text{Non-CGI}} = 0.39 \pm 0.0053$, $P = 1.5 \times 10^{-4}$ and $\text{mean}_{\text{Promoter}} = 0.43 \pm 0.0085$, $\text{mean}_{\text{Non-Promoter}} = 0.4 \pm 0.0048$, $P = 0.0077$; Bonferroni-corrected $P = 8.7 \times 10^{-4}$ and $P = 0.047$, respectively). Promoters also showed a trend toward being less epigenetically variable in gut tissue (Wilcoxon rank-sum test, $\text{mean}_{\text{Promoter}} = 0.11 \pm 0.0065$, $\text{mean}_{\text{Non-Promoter}} = 0.09 \pm 0.0037$, $P = 0.057$; Bonferroni-corrected $P = 0.34$). No statistically significant differences were detected in the buccal epithelial cells. The promoter and CGI probes were also subjected to the Gene Ontology (GO)-based analysis[17]. Most of the identified GO categories associated with epigenetically similar loci between co-twins (top fifth percentile of ICCs) had direct functional relevance to the tissue investigated (**Supplementary Table 1a** online). Our observations are consistent with an earlier study[7] where the fidelity of CpG methylation patterns was twice as high in promoter as opposed to nonpromoter regions. Taken together, greater epigenetic similarity between MZ co-twins at functionally important regions in comparison to the loci without clearly defined regulatory function suggests functional stratification of the epigenome. Epigenetically variable loci (bottom fifth percentile of ICCs) were associated with cell division processes (**Supplementary Table 1b**), which may reflect an early developmental epigenetic discordance as one of the hypothetical reasons of twin formation[16].

Cases of DNA sequence variation in MZ twins have been documented[18], but these are uncommon and unlikely to account for even a fraction of the MZ co-twin differences identified in our experiments. Further studies may include a more detailed annotation of epigenetic differences in MZ co-twins a search for disease-specific epigenetic changes in discordant MZ twins and a dissection of environment-induced versus stochastic epigenetic differences. As MZ twins reared apart are generally quite similar to MZ twins reared together according to an array of traits (electroencephalogram, IQ, personality, social attitudes)[19], we speculate that stochastic events in epigenetically determined phenotypic differences in MZ co-twins are much more important than environment.

The second part of our study focuses on comparisons of epigenetic similarities in MZ versus DZ twins, the same design that has been used in heritability studies. DNA methylation differences in buccal epithelial cells from 20 sets of MZ co-twins (described above) were significantly lower in comparison to 20 sets of DZ co-twins matched for age and sex (mean $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}} = 0.15 \pm 0.0039$, $P = 1.2 \times 10^{-294}$; **Fig. 3a**). All the effect observed was attributed to the ten sets of dichorionic MZ twins (mean $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}} = 0.35 \pm 0.0057$, $P < 9.9 \times 10^{-324}$; **Fig. 3b**), whereas the mean ICC of monochorionic MZ twins was close to 0 (**Fig. 3c**). In WBC from 19 sets of MZ twins (described above) and 20 sets of DZ twins matched for age, sex and blood cell count (total WBC count, neutrophil and lymphocyte fractions), MZ–DZ differences were much more subtle but still significant (mean $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}} = 0.0073 \pm 0.0034$, $P = 0.044$). The observed effect may have been diminished by our conservative efforts to bias against larger epigenetic MZ–DZ differences by selecting matched DZ twins with smaller co-twin cell subfraction differences as compared to the MZ twins. For buccal tissue, a locus-specific annotation of $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}}$ values representing dichorionic MZ co-twin similarity relative to DZ co-twin similarity is provided (**Fig. 4**; see **Supplementary Fig. 5** online for WBC and monochorionic buccal samples).

All techniques for enrichment of differentially methylated DNA sequences for microarray-based DNA methylation profiling can potentially be confounded by DNA sequence variation. In our experiments, SNPs within *Hpa*II restriction sites may have caused enrichment differences, which would then result in larger variation in DZ twins. In addition, DNA sequence variants may influence the epigenetic status, as in the literature, there are several examples of DNA allele or haplotype association with specific epigenetic profiles[13,20,21]. Alternatively, DZ twins may show more epigenetic differences than MZ twins because the former originate from different zygotes carrying two different epigenetic profiles, whereas the latter develop from the same zygote, and therefore should possess similar epigenomes at the time of blastocyst splitting. Although the experiments described below do not unequivocally prove this second hypothesis, we favor the idea of these zygotic epigenetic effects for reasons discussed below.

First, epigenetic profiles are not fully determined by DNA sequence; if that were the case, MZ twins would show no epigenetic differences. Therefore, the observed major, epigenome-wide differences in the buccal epithelial cells from MZ twins versus DZ twins are highly unlikely to be caused exclusively by DNA sequence differences in DZ twins. Furthermore, $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}}$ differences were tissue-specific, as the buccal epithelial cells from dichorionic MZ twins showed much larger MZ-DZ epigenetic differences in comparison to that of a subset of WBC obtained from the same individuals at the same time. As the DNA sequences should be identical (or nearly identical) between the tissues of the same organism, the tissue-specific $\text{ICC}_{\text{MZ}}\text{-ICC}_{\text{DZ}}$ differences argue against DNA sequence as a major controlling factor of epigenetic profiles.

Second, to address the putative effects of differential digestion of polymorphic *Hpa*II restriction sites in DZ twins, we tried to perform a comparative analysis between *Hpa*II and its isoschisomer, *Msp*I, as has

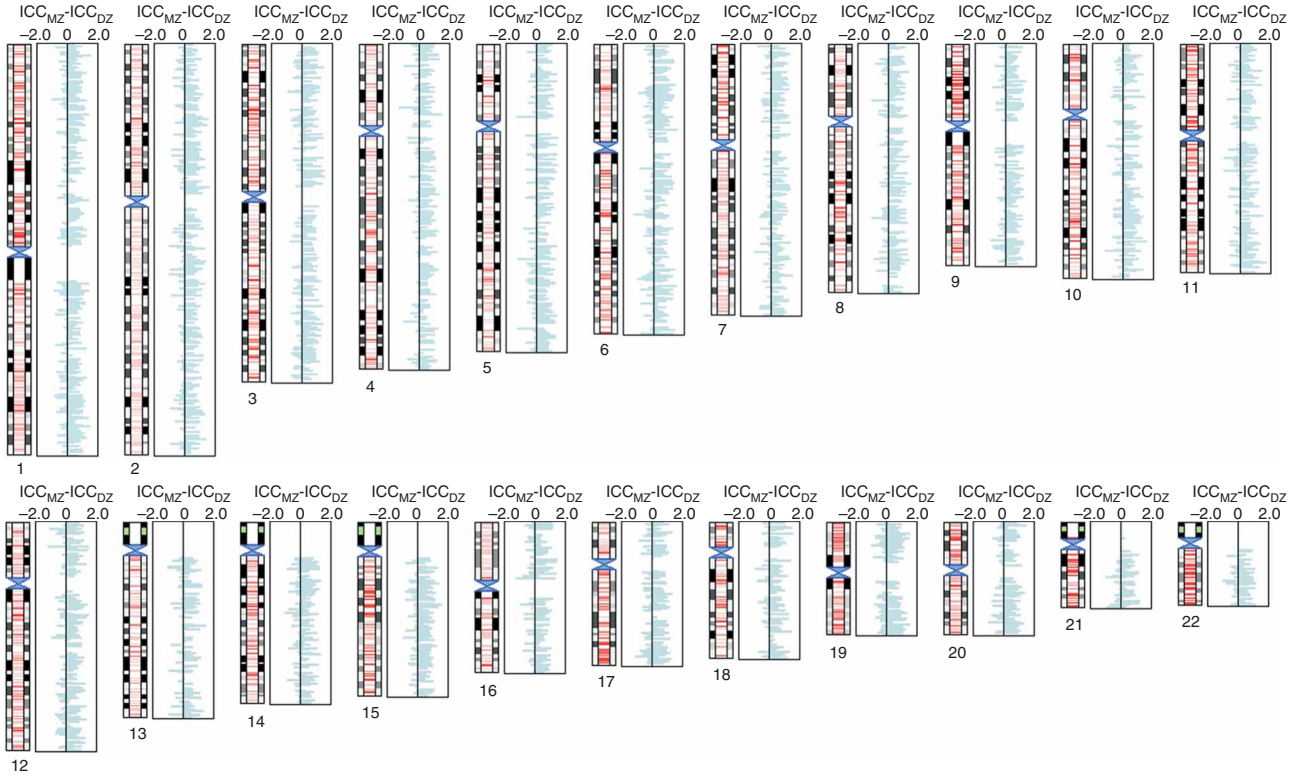

Figure 4 A chromosomal karyogram depicting degrees of dichorionic MZ co-twin similarity relative to DZ co-twin similarity per interrogated locus in the buccal sample. Blue bars to the right of each chromosome represent locus-specific ICC_{MZ}-ICC_{DZ} values.

been suggested in the HELP assay[22]; however, degrees of technical variation produced in *Msp*I-based experiments were markedly larger than those of *Hpa*II experiments (ratio of *Hpa*II/*Msp*I variance = 0.37, $P < 9.9 \times 10^{-324}$; **Supplementary Fig. 6** online). As a result, the two experiments were not directly comparable. Alternatively, we carried out an *in silico* analysis whereby the SNP and allele frequency information available in the dbSNP and HapMap databases were obtained to calculate heterozygosity quotients that represent the probability that a given probe would have a restriction site disrupted by a SNP. From the 6,405 and 5,917 unique sequences within the WBC and buccal datasets, 109 and 98 loci containing *Hpa*II SNPs were identified, respectively. For both datasets, there was no correlation of locus heterozygosity value with ICC_{MZ}-ICC_{DZ} value ($R = -0.0032$ and $P = 0.97$ for WBC; $R = 0.024$ and $P = 0.81$ for buccal cells). A similar analysis was done to address the epigenetic effects of SNPs in *cis* by extending the interrogated region to include all SNPs within 1 kb proximal to and including the probe sequence. Again, correlation analysis of heterozygosity values at 1,369 (WBC) and 1,284 (buccal) SNP containing loci showed no correlation with ICC_{MZ}-ICC_{DZ} value ($R = -0.019$, $P = 0.47$ (WBC), and $R = 0.033$, $P = 0.23$ (buccal cells)). These results are in agreement with a recent study that identified that only 0.16% of SNPs are associated with allele-specific DNA methylation changes[23].

Third, we investigated whether DNA variation may influence DNA methylation both in *cis* and in *trans* by methylation analysis of two strains of inbred (that is, nearly genetically identical) mice as compared to two strains of outbred (genetically nonidentical) mice. Mouse brains were subjected to 4.6K CpG island microarray-based DNA methylation profiling. First, we determined that the detected biological variation is significantly larger than technical variation in the mouse experiments ($P < 9.9 \times 10^{-324}$). We then compared the spot-wise distribution of within sibship DNA methylation variation (σ^2) between inbred and outbred mice at 2,176 unique genomic regions and did not detect any significant difference (mean difference = $2.1 \times 10^{-5} \pm 3 \times 10^{-4}$, $P = 0.68$) (**Fig. 5**). Although it is not completely clear to what extent mouse brain results can be extrapolated to human buccal cells despite their shared ectodermal origin, and although DNA variation in the outbred mice is less than that of unrelated humans (based on the Wellcome Trust study (see URLs section in Methods); our estimate is that in general, outbred mouse DNA heterozygosity is 2–4 times lower in comparison to unrelated humans), the impact of DNA polymorphisms on DNA methylation does not seem to be common.

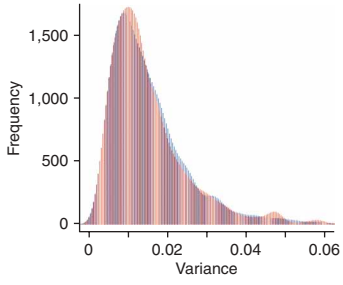

Figure 5 The spot-wise distributions of the within-sibship variance for both inbred (red) and outbred (blue) mice. A nonparametric comparison of the distributions with a paired Wilcoxon signed rank test did not identify any significant epigenetic difference between groups, despite the genetic variation within the outbred group (mean difference = $2.1 \times 10^{-5} \pm 3 \times 10^{-4}$, $P = 0.68$).

In the classical twin studies, greater phenotypic similarity among MZ twin pairs compared to DZ twins has been traditionally attributed to the degree of DNA sequence similarity. Our twin studies suggest that in addition to identical DNA, epigenetic similarity at the time of blastocyst splitting may also contribute to phenotypic similarities in MZ co-twins. By the same argument, DZ co-twins are more different from each other than MZ co-twins not only because they possess some DNA sequence differences (on average ~0.05%) but also because they originated from epigenomically different zygotes. In addition, epigenomic inheritance may explain the 'intangible variance', the concept that originated from the observation that regular (polyzygotic) inbred mice were much more different from each other than the MZ inbred mice of the same strain[24]. In conjunction with such findings, our data suggest that the phenotypic effects of the individual epigenomes of each zygote could be substantial.

METHODS

Twin sample. We investigated three cohorts of twins representing various tissues. WBC of 19 dichorionic MZ and 20 DZ twin pairs matched for age, sex and WBC count plus buccal epithelial cells from the 10 monochorionic MZ, 10 dichorionic MZ, and 20 DZ age- and sex-matched twin pairs were obtained from the Brisbane Adolescent Twin Study[25]. WBCs and buccal cells were obtained from the same individual for 10 dichorionic MZ and 10 DZ pairs. WBC samples were from twins 13.2 ± 1 y old (mean ± s.d.) and consisted of 20 females and 18 males. Monochorionic and dichorionic buccal epithelial cells both consisted of 10 males (aged 14 ± 0.77 y) and 10 females (all 14 y old); all were of European ancestry (mainly northern European ancestry). MZ and DZ twins in the WBC group were selected from several thousand sets of twins of the Australian Twin Registry using hematology report data. The percentage difference between cell subfraction counts for the whole WBC count, neutrophil and lymphocyte counts did not exceed 10%. The mean percentage difference in selected DZ twins was smaller than that of MZ twins to bias against the alternative hypothesis of more epigenetic variation in the DZ twin group. We determined zygosity by comparisons of nine microsatellite markers, which gave a probability of incorrect assignment of a DZ as an MZ of less than 0.0001. Gut biopsies from 18 pairs of MZ twins were obtained from a Swedish twin population with inflammatory bowel disease described previously[26]. Although all twin pairs had at least one twin affected with inflammatory bowel disease, we investigated biopsies from rectal mucosa, which were macroscopically not inflamed in any of the twins investigated. Written informed consent was obtained from all participants, and studies were approved by the local institutional review boards at participating institutions.

DNA methylation profiling. The unmethylated fraction of genomic DNA was enriched using the methylation-sensitive restriction enzyme *Hpa*II[4] and interrogated on Human 12K CpG island microarrays[3]. Enrichment of the unmethylated genome of MZ and DZ twin pairs and hybridization to the microarrays was carried out in a randomized fashion. We did two technical replicates for each enrichment and hybridization, after which we averaged the log ratios per each replicate to produce one value per individual per locus. All samples were hybridized against a common reference (reference 1) with the exception of 9 MZ and 10 DZ pairs in WBC, which were originally hybridized against a different common reference (reference 2) and later transformed to match reference pattern 1. Transformation was achieved by first obtaining a spot-wise log ratio of reference 2 relative to reference 1 through a comparison of two dye-swapped reference 1 versus reference 2 hybridizations. Log ratios from the 9 MZ and 10 DZ pairs originally hybridized with reference 2 were multiplied by the log ratio values of reference 1 versus reference 2 to obtain log ratio values relative to reference 2. This transformation was followed by between array normalization using the Limma package in Bioconductor. We created the reference pools by addition of equal quantities of the enriched unmethylated WBC DNA fraction from 10 MZ and 10 DZ pairs.

Animal studies. We extracted genomic DNA using standard phenol and chloroform methods from whole-brain tissue of four strains of mice: c57BL6 and FVB inbred strains and CF-1–1 and CD-1 outbred strains, all obtained from Charles River Laboratories International. Three litters consisting of three male mice per litter were kept in uniform environments and killed at postnatal day 43. We enriched the unmethylated fraction of genomic DNA and created the common reference pool in an identical manner to the human reference design studies. The microarrays used were mouse 4.6K CpG island microarrays, all produced during a single printing at the microarray facility of the University Health Network, Toronto. Hybridizations were carried out in batches of 18 microarrays consisting of one amplification set from one inbred and one outbred strain per day for a total of four hybridization days. We determined selection and order of hybridization at random through sorting on a random number generator.

Data analysis. All microarrays were scanned on the Axon 4000A scanner and cross-referenced to annotated GAL files using Genepix 6.0 software. Microarray GAL annotation was made available from the manufacturer and downloaded (see URLs section below). We carried out normalization procedures in Bioconductor using the Limma package. All arrays underwent log ratio–based normalization, background correction, print tip loess normalization and scale normalization between blocks. We removed low-quality flagged loci identified by Genepix. Microarray data were trimmed on the basis of the annotation information such that spot IDs containing mitochondrial DNA, translocation hot spots and repetitive elements, and those located on the X and Y chromosomes were removed. After trimming and removal of flagged loci, 6,405 (WBC), 5,918 (buccal cells) and 5,941 (gut biopsies) unique DNA sequences in humans and 2,176 DNA sequences in mice were used for subsequent statistical analyses.

All statistical tests were done in R. Using an Anderson-Darling test from the nortest package, we found that all distributions derived from microarray data rejected the null hypothesis of normality, and we subsequently evaluated them with nonparametric tests. All statistical tests done were two tailed and a $P < 0.05$ is considered significant. Unless otherwise specified, ± denotes the s.e.m. Data are located online; see URLs section below.

Spot-wise epigenetic variation. We calculated a spot-wise ICC according to the one-way consistency model using the irr package, designating co-twin pairs as a class. The ICC formula is $ICC = (MS_b - MS_w)/(MS_b + MS_w)$. Here MS_b stands for the between pair mean square and MS_w represents the within-pair mean square of the specified class. As the ICC approaches 1, the co-twins are more similar to each other than unrelated twin pairs are to each other, whereas as it approaches –1, the within–co-twin difference across the group is consistently larger in comparison to unrelated twin pairs. Each unique DNA region investigated by the microarray was treated as an independent measurement. To address the null hypothesis that there are no differences in the amount of DNA methylation variability between MZ and DZ twins, we evaluated the distributions of unique locus ICC between MZ and DZ twins in WBC cells with a paired Wilcoxon signed rank test. For buccal epithelial cells, the same hypothesis for monochorionic and dichorionic twins was evaluated in a similar manner.

For inbred and outbred mice, separately, a spot-wise distribution of within sibship epigenetic variation was created by taking the average of the variance produced by the three mice per sibship. To address the null hypothesis that there are no differences in the degrees of epigenetic variation between inbred and outbred mice, we compared these spot-wise distributions with a paired Wilcoxon signed rank test. Information on further analysis done on these datasets can be found in **Supplementary Methods** online.

Validation of the microarray findings. We validated the microarray findings using sodium bisulfite modification as done previously in our laboratory[27]. Sodium bisulfite modification was followed by interrogation of specific CpG sites by pyrosequencing[28] or direct cloning and sequencing. PCR amplicon, pyrosequencing and sequencing primers are provided in **Supplementary Table 2** online (further details are given in **Supplementary Methods**).

URLs. The Krembil family epigenetics laboratory homepage, www.epigenomics.ca; Wellcome-CTC mouse strain SNP genotyping set, http://www.well.ox.ac.uk/mouse/INBREDS; microarray GAL annotation, www.microarrays.ca.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Professor V.M. Gindilis, an excellent scientist, dedicated teacher and creative twin researcher. We should like to thank S. Ziegler for technical assistance, A. Henders and M. Campbell for selection and preparation of twin DNA samples, and J. Chow for work generating the karyograms. This project was supported by the National Institute of Mental Health (R01 MH074127-01), the Canadian Institutes for Health and Research (CIHR) and the National Alliance for Research on Schizophrenia and Depression (NARSAD). A.P. is Senior Fellow of the Ontario Mental Health Foundation. Z.A.K. was supported by a CIHR graduate fellowship.

AUTHOR CONTRIBUTIONS

Study design: Z.A.K., S.-C.W., A.H.C.W., A.F.M., P.M.V., N.G.M. and A.P.; sample collection: G.W.M., N.G.M., J.H. and C.T.; animal preparation: L.A.F. and A.H.C.W.; sample preparation: Z.A.K. and C.P.; microarray enrichment and hybridization: Z.A.K. and C.P.; sodium bisulfite–based fine mapping: Z.A.K., C.P. and G.H.T.O.; statistical analysis: Z.A.K., T.T., S.-C.W., C.V., A.F.M. and P.M.V.; manuscript writing: Z.A.K., T.T., S.-C.W., C.P., G.H.T.O., A.H.C.W., L.A.F., C.V., J.H., C.T., A.F.M., P.M.V., G.W.M., I.I.G., N.G.M. and A.P.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Boomsma, D., Busjahn, A. & Peltonen, L. Classical twin studies and beyond. *Nat. Rev. Genet.* **3**, 872–882 (2002).
2. Martin, N., Boomsma, D. & Machin, G. A twin-pronged attack on complex traits. *Nat. Genet.* **17**, 387–392 (1997).
3. Heisler, L.E. *et al.* CpG island microarray probe sequences derived from a physical library are representative of CpG islands annotated on the human genome. *Nucleic Acids Res.* **33**, 2952–2961 (2005).
4. Schumacher, A. *et al.* Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**, 528–542 (2006).
5. Robertson, K.D. & Wolffe, A.P. DNA methylation in health and disease. *Nat. Rev. Genet.* **1**, 11–19 (2000).
6. Riggs, A.D., Xiong, Z., Wang, L. & LeBon, J.M. Methylation dynamics, epigenetic fidelity and X chromosome structure. *Novartis Found. Symp.* **214**, 214–225 (1998).
7. Ushijima, T. *et al.* Fidelity of the methylation pattern and its variation in the genome. *Genome Res.* **13**, 868–874 (2003).
8. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**(Suppl.), 245–254 (2003).
9. Jirtle, R.L. & Skinner, M.K. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **8**, 253–262 (2007).
10. Wong, A.H., Gottesman & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet.* **14**(Spec. No. 1), R11–R18 (2005).
11. Petronis, A. *et al.* Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr. Bull.* **29**, 169–178 (2003).
12. Kuratomi, G. *et al.* Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol. Psychiatry* **13**, 429–441 (2008).
13. Heijmans, B.T., Kremer, D., Tobi, E.W., Boomsma, D.I. & Slagboom, P.E. 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 (2007).
14. Oates, N.A. *et al.* Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am. J. Hum. Genet.* **79**, 155–162 (2006).
15. Fraga, M.F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA* **102**, 10604–10609 (2005).
16. Hall, J.G. Twinning. *Lancet* **362**, 735–743 (2003).
17. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
18. Bruder, C.E. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* **82**, 763–771 (2008).
19. Bouchard, T.J. Jr., Lykken, D.T., McGue, M., Segal, N.L. & Tellegen, A. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. *Science* **250**, 223–228 (1990).
20. Murrell, A. *et al.* An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum. Mol. Genet.* **13**, 247–255 (2004).
21. Flanagan, J.M. *et al.* Intra- and interindividual epigenetic variation in human germ cells. *Am. J. Hum. Genet.* **79**, 67–84 (2006).
22. Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* **16**, 1046–1055 (2006).
23. Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40**, 904–908 (2008).
24. Gartner, K. & Baunack, E. Is the similarity of monozygotic twins due to genetic factors alone? *Nature* **292**, 646–647 (1981).
25. Wright, M. & Martin, N. Brisbane Adolescent Twin Study: outline of study methods and research projects. *Aust. J. Psychol.* **56**, 65–78 (2004).
26. Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E. & Jarnerot, G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* **124**, 1767–1773 (2003).
27. Mill, J. *et al.* Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* **82**, 696–711 (2008).
28. Tost, J., El Abdalaoui, H. & Gut, I.G. Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* **40**, 721–722 724, 726 (2006).