Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm

Bo Zhang,1,16 Yan Zhou,2,3,16 Nan Lin,4,16 Rebecca F. Lowdon,1,16 Chibo Hong,5 Raman P. Nagarajan,5 Jeffrey B. Cheng,6 Daofeng Li,1 Michael Stevens,1 Hyung Joo Lee,1 Xiaoyun Xing,1 Jia Zhou,1 Vasavi Sundaram,1 GiNell Elliott,1 Junchen Gu,1 Taoping Shi,1,17 Philippe Gascard,7 Mahvash Sagaroudinia,7 Thea D. Tlsty,7 Theresa Kadlecak,8 Arthur Weiss,8 Henriette O’Geen,9 Peggy J. Farnham,10 Cécile L. Maire,11 Keith L. Ligon,11,12 Pamela A.F. Madden,13 Angela Tam,14 Richard Moore,14 Martin Hirst,14,15 Marco A. Marra,14 Baoxue Zhang,2,18 Joseph F. Costello,5,18 and Ting Wang1,18

1–15[Author affiliations appear at the end of the paper.]

DNA methylation plays key roles in diverse biological processes such as X chromosome inactivation, transposable element repression, genomic imprinting, and tissue-specific gene expression. Sequencing-based DNA methylation profiling provides an unprecedented opportunity to map and compare complete DNA methylomes. This includes one of the most widely applied technologies for measuring DNA methylation: methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq), coupled with a complementary method, methylation-sensitive restriction enzyme sequencing (MRE-seq). A computational approach that integrates data from these two different but complementary assays and predicts methylation differences between samples has been unavailable. Here, we present a novel integrative statistical framework M&M (for integration of MeDIP-seq and MRE-seq) that dynamically scales, normalizes, and combines MeDIP-seq and MRE-seq data to detect differentially methylated regions. Using sample-matched whole-genome bisulfite sequencing (WGBS) as a gold standard, we demonstrate superior accuracy and reproducibility of M&M compared to existing analytical methods for MeDIP-seq data alone. M&M leverages the complementary nature of MeDIP-seq and MRE-seq data to allow rapid comparative analysis between whole methylomes at a fraction of the cost of WGBS. Comprehensive analysis of nineteen human DNA methylomes with M&M reveals distinct DNA methylation patterns among different tissue types, cell types, and individuals, potentially underscoring divergent epigenetic regulation at different scales of phenotypic diversity. We find that differential DNA methylation at enhancer elements, with concurrent changes in histone modifications and transcription factor binding, is common at the cell, tissue, and individual levels, whereas promoter methylation is more prominent in reinforcing fundamental tissue identities.

[Supplemental material is available for this article.]
Integrating MeDIP-seq and MRE-seq is expected to improve our ability to
accurately measure DNA methylation differences. This is corroborated by an increase of MRE-seq signal, suggesting that the inference is reflecting a biological event (we infer that this region is differentially methylated). For example, a decrease of MeDIP-seq signal could indicate that the combination of these two methods showed promise in identifying differentially methylated regions (DMRs) that are enriched in CpG islands and are associated with gene expression. This is consistent with previous findings that MeDIP-seq and MRE-seq are the result of CpG methylation. We further assumed that, within the same region of interest, MeDIP-seq signal is proportional to the number of methylated CpGs in any given region. This assumption was made by previously published tools (Pelizzola et al. 2008; Chavez et al. 2010; Maunakea et al. 2010), which produces similar results as Batman but with higher computational efficiency.

MRE-seq is a complementary approach to MeDIP-seq that identifies unmethylated CpG sites in the restriction sites for multiple methylation-sensitive restriction enzymes (Harris et al. 2010; Maunakea et al. 2010). By using simple heuristics, we demonstrated that the combination of these two methods showed promise in identifying differentially methylated regions (DMRs) as well as intermediate or monoallelic methylation (Harris et al. 2010). Here, we further explore and leverage the complementary nature of MeDIP-seq and MRE-seq by integrating them in a statistical framework. Our approach is based on the principle that all observed genome-wide measurements (MeDIP-seq, MRE-seq, WGBS, etc.) are derived from methylation states of the sample. We infer methylation states from the observed data, which are sequencing reads aligned to the reference genome. However, all current approaches to assessing DNA methylation have their own inherent errors and biases. Because MeDIP-seq and MRE-seq are independent, complementary measurements of the same methylation states, our confidence in inferring methylation states should increase when results from these two methods are integrated (Stevens et al. 2013). For example, a decrease of MeDIP-seq signal could reflect a biological event (we infer that this region is demethylated) or could be a methodological artifact; but if it is corroborated by an increase of MRE-seq signal, then the inference of demethylation is much more likely to be accurate. Thus, integrating MeDIP-seq and MRE-seq is expected to improve our ability to detect DMRs accurately.

Here, we describe a novel statistical framework which we call "M&M" (for integration of MeDIP-seq and MRE-seq) that detects DMRs. M&M explicitly models the relationship between DNA methylation level, CpG content, and expected MeDIP and MRE reads in any given genomic context. By analyzing WGBS, MeDIP-seq, and MRE-seq data for the same DNA samples, we show that M&M outperforms MEDIPS in detecting DMRs. We applied M&M to 19 human samples representing nine cell types from four tissues (embryonic stem cells, breast, blood, and brain) which we assayed with MeDIP-seq and MRE-seq. Our results revealed a large, definitive panel of known and mostly novel tissue type-, cell type-, and individual-specific DNA methylation differences. Consistent with expectations, we identified enrichment of DMRs in promoter regions of genes with tissue-specific functions. Importantly, we identified a large number of DMRs that are not methylated in tissues where the same local region also harbored enhancer chromatin signatures. These enhancer-marked DMRs comprised 30% of the tissue-specific DMRs, >70% of the cell type-specific DMRs, and >40% of the individual-specific DMR landscape.

**Results**

**Summary of the M&M algorithm**

Differentially methylated regions (DMRs) are defined as any genomic region where the overall CpG methylation levels are statistically significantly different between cell populations of two samples being compared. The M&M algorithm identifies DMRs by computing a probability score for the difference in DNA methylation for any given genomic region based on observed MeDIP-seq and MRE-seq measurements. We made several simple assumptions and definitions. First, we only considered CpG methylation and made the reasonable assumption that all signals obtained from MeDIP-seq and MRE-seq are the result of CpG methylation. We note that methylation of cytosines in the non-CpG context (i.e., CHG and CHH) is rare in somatic cells but is more common in embryonic stem cells, albeit at low levels at any given site, and is associated with highly methylated CpGs (Lister et al. 2009). The biological significance of CHG and CHH methylation in mammalian cells is yet to be determined. Our statistical model is general enough to incorporate non-CpG cytosine methylation, but to facilitate comparisons with existing tools, we only considered CpG methylation in this study. Second, we assumed that MeDIP-seq signal is proportional to the number of methylated CpGs in any given region. This assumption was made by previously published tools (Pelizzola et al. 2008; Chavez et al. 2010; Maunakea et al. 2010), and we confirmed that the rule, in general, holds (Supplementary Fig. S1A). Third, we assumed that MRE-seq signal is proportional to the number of unmethylated CpGs at the enzyme recognition sites (defined as MRE sites) (Supplementary Fig. S1B). We further assumed that, within the same region of interest, methylation levels of CpGs in MRE sites reflect levels of nearby CpGs that are not within the MRE sites. Finally, we defined methylation level (µ) as the proportion of methylated CpGs versus total CpGs in a given region. Thus, observed MeDIP-seq and MRE-seq data become a function of methylation level, CpG content, and MRE site content of a given genomic region. MeDIP-seq signal and MRE-seq signal are related by the methylation level, µ, of the region, with their expectations proportional to µ and (1 − µ), respectively. When comparing two samples in the same genomic region, we are testing the null hypothesis that methylation levels are the same between the samples. This hypothesis is conditioned on the observed MeDIP-seq and MRE-seq data, given the CpG content and MRE-site content; CpG and MRE-site content are fixed for any
specific genomic region when SNP, mutation, or copy number differences are not considered. When genetic variation data is available for the sample, corrections can be made to reflect known variation. To better formulate the problem, we illustrate the algorithm by taking a window-based approach and partitioning the reference genome into \( B \) equally spaced, nonoverlapping windows (typically 500 bp in size). We only considered windows that contain CpG sites. For the \( i \)-th bin, \( i = 1, \ldots, B \) window, let \( m_i \) denote the number of CpGs and \( k_i \) denote the number of CpGs in MRE sites. Let \( X_{i1} \) and \( X_{i2} \) denote the MeDIP-seq read counts of the two samples being compared. Since many CpGs are not in MRE sites, MeDIP-seq read counts can be estimated, but their ratio can. We then used a conditional test for the sample, corrections can be made to reflect known variation. When genetic variation data is available, let \( T \) be a Poisson random variables with expected values \( E \left( X_{ji} \right) = \lambda_{ji} \) and \( E \left( Y_{ji} \right) = \gamma_{ji} \), where \( j = 1, 2 \) refers to the two samples being compared. Let \( l_1 = \sum_{i=1}^{B} \lambda_{ji} \) and \( l_2 = \sum_{i=1}^{B} \gamma_{ji} \). We then modeled the expected values of \( X_{ji} \) and \( Y_{ji} \) as

\[
E (X_{ji}) = \lambda_{ji} = \frac{\mu_j m_i}{S_1} L_{i1} \quad \text{and} \quad E (Y_{ji}) = \gamma_{ji} = \frac{1 - \mu_j m_i}{S_2} L_{i2},
\]

where \( S_1 = \sum_{i=1}^{B} \mu_j m_i, S_2 = \sum_{i=1}^{B} (1 - \mu_j) m_i, \) and \( \mu_1 \) and \( \mu_2 \) are the unknown methylation levels of the two samples in the \( i \)-th window.

Under this model, we detected DMRs by testing for all \( i = 1, \ldots, B, \)

\[
H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2,
\]

which is equivalent to testing

\[
H_0 : \mu_1 (1 - \mu_2) = \mu_2 (1 - \mu_1) \quad \text{versus} \quad H_1 : \mu_1 (1 - \mu_2) \neq \mu_2 (1 - \mu_1).
\]

From Equation 1, we can rewrite Equation 3 as

\[
H_0 : c_1 \lambda_{i1} \gamma_{i2} = c_2 \lambda_{i2} \gamma_{i1} \quad \text{versus} \quad H_1 : c_1 \lambda_{i1} \gamma_{i2} \neq c_2 \lambda_{i2} \gamma_{i1},
\]

where \( c_1 = (S_1 L_{i2})/(S_1 L_{i1}) \) and \( c_2 = (S_2 L_{i2})/(S_2 L_{i1}) \) can be estimated from the data. Note that \( l_1 \) and \( l_2 \) can be estimated from the observed read counts, whereas \( S_1 \) and \( S_2 \) cannot be directly estimated, but their ratio can. We then used a conditional test based on the test statistic:

\[
T_i = c_1 X_{i1} Y_{i2} - c_2 X_{i2} Y_{i1}.
\]

Let \( n_i \) be the sum of the observed MeDIP-seq and MRE-seq read counts in the \( i \)-th bin. Based on Agresti (2007), given \( n_i \) the joint distribution of \( X_{i1}, X_{i2}, Y_{i1}, \) and \( Y_{i2} \) is a multinomial distribution (Supplemental Notes), which allows deriving the \( P \)-value defined as

\[
p_i = P(|T_i| > |t_i| |X_{i1} + X_{i2} + Y_{i1} + Y_{i2} = n_i),
\]

where \( t_i \) is the observed value of \( T_i \). For windows in which only MeDIP-seq data are available, let \( T_i = c_1 X_{i1} - X_{i2} \). Then, the \( P \)-value is given by \( p_i = P(|T_i| > |t_i| |X_{i1} + X_{i2} = n_i) \) with \( t_i \) the being the observed value of \( T_i \). In this case, our method reduces to the SAGE test (Robinson and Oshlack 2010). When the total read count \( n_i \) is large, we can achieve accurate analytical approximation to the discrete \( P \)-value in Equation 5 by normal approximation to estimate \( P \)-values analytically (Lehmann and Romano 2005) (Supplemental Notes).

Finally, since we applied M&Ms to genomic windows genome-wide, the genome-wide false discovery rate (FDR) was controlled using the group Benjamini-Hochberg method previously described in Hu et al. (2010).

The overall flow of the M&M algorithm is illustrated in Supplemental Figure S2 to facilitate understanding. Additional details of the M&M algorithm are described in Supplemental Notes.

**Benchmarking M&M’s performance**

Because M&M implements a novel test statistic, we evaluated its sensitivity, specificity, and reproducibility on multiple DNA methylomes from human tissues and populations of cells strongly enriched for individual cell types. We tested the performance of M&M against MEDIPS.

We generated complete DNA methylation data for 19 human samples (Supplemental Table 1) as a part of the NIH Roadmap Epigenomics project (Bernstein et al. 2010). Tissue and primary cell types included embryonic stem cells (H1 ESCs), fetal brain tissue, neural stem cells (neurosphere cultured cells, ganglionic eminence derived), adult breast epithelial cells (luminal epithelial cells, myoepithelial cells, and a stem cell-enriched population), unfractioned peripheral blood mononuclear cells (PBMCs), and adult immune cells (CD4+ naive and memory and CD8+ naive cells). All samples were assayed by both MeDIP-seq and MRE-seq. For H1 ESC, two biological replicates were obtained. In addition, we obtained WGBS data for H1 ESCs (Lister et al. 2009). We also generated a second WGBS data set for short-term cultured human fetal neural stem cells (HuFNSC02, neurosphere cultured cells [NSCs], ganglionic eminence derived, fetal age of 21 wk) for which we also generated MeDIP-seq and MRE-seq data. We compared M&M’s performance against that of MEDIPS by applying M&M and MEDIPS (which uses MeDIP-seq data only) for pairwise comparisons between the two H1 ESC replicates and between H1 ESCs and fetal NSCs. All tests were performed on 500-bp-sized, nonoverlapping windows genome-wide (a total of 5,313,352 windows; windows without CpGs in the hg19 build of the human genome were not considered). In each pairwise comparison, M&M and MEDIPS generated a \( P \)-value for each window, which was used to determine if the region within the window exhibited differential methylation between the two samples. In addition, for each DMR, the relative methylation status for the two samples was also determined, i.e., which sample was relatively hypermethylated and which sample was relatively hypomethylated.

We then examined the distribution of \( P \)-values across the different comparisons. In Figure 1A, we plotted histograms of all \( P \)-values generated by M&M when comparing the two H1 ESC biological replicates and when comparing H1 ESCs and fetal NSCs. The \( x \)-axis denotes negative log\(_{10} \) transformed \( P \)-values, and the \( y \)-axis denotes the log\(_{10} \) transformed number of DMRs at each \( P \)-value cutoff. Similarly, in Figure 1B, we plotted \( P \)-values from the same comparisons made by MEDIPS. At any reasonable cutoff, M&M and MEDIPS both predicted more DMRs between H1 ESCs and fetal NSCs than between the two H1 ESC replicates, consistent with our expectations. Because the H1 ESC samples are biological replicates, this comparison can be used to estimate the number of false positives at any \( P \)-value cutoff. At a \( P \)-value less than 1 \( \times 10^{-10} \), M&M reported 70 DMRs, while MEDIPS reported 2066 DMRs. Thus, the false positives rate was 0.43% for M&M and 18.51% for MEDIPS. Using the same \( P \)-value cutoff for the comparison between two different cell types, i.e., H1 ESCs and fetal NSCs, M&M reported 16,398 DMRs, while MEDIPS reported...
Figure 1. Benchmarking the performance of M&M. (A) The distribution of P-values generated by M&M when comparing two H1 ESC biological replicates (blue area) and when comparing H1 ESC and fetal NSC (red area). At a P-value cutoff of less than $1 \times 10^{-10}$ (green line), M&M predicted 70 DMRs between the two H1 samples, and 16,398 DMRs between H1 ESC and fetal NSC. (B) The distribution of P-values generated by MEDIPS for the same comparisons as in A. At a P-value cutoff of less than $1 \times 10^{-10}$ (green line), MEDIPS predicted 2066 DMRs between the two H1 ESC replicates, and 11,162 DMRs between H1 ESC and fetal NSC. (C) Whole-genome bisulfite sequencing (WGBS) data were used to validate DMRs predicted by M&M between H1 ESC and fetal NSC. DMRs predicted by M&M were ranked according to their P-values, then average DNA methylation levels for each of the top 1000 significantly hypermethylated DMRs (red) and the top 1000 significantly hypomethylated DMRs (blue) in fetal NSC were computed using WGBS data from the same two samples (H1 ESC and fetal NSC). Distribution of the DNA methylation level differences was plotted for hypermethylated DMRs and hypomethylated DMRs separately. The gray area represents the distribution of DNA methylation differences in the whole-genome background, calculated at 500-bp-window resolution using the same WGBS data sets. (D) Same as C, except that DMRs were predicted by MEDIPS. (E) DNA methylation differences between H1 ESC and fetal NSC were calculated using WGBS data for individual CpGs within the top 500, 1000, 2000, 5000, and 10,000 hypermethylated and hypomethylated DMRs (predicted by M&M, at varying cutoffs). These values were plotted as a boxplot. (F) Same as E, except that DMRs were predicted by MEDIPS. (G) Concordance between M&M (red) or MEDIPS (blue) predicted DMRs and differential methylation for these regions calculated from WGBS data. DMRs predicted by M&M and MEDIPS were ranked based on their P-values. At different cutoffs, DMRs were determined to be concordant with WGBS data (if differences in WGBS data were greater than 0.1 and were in the correct direction). (H) Reproducibility of DMR predictions in M&M (red) and MEDIPS (blue). DMR discovery was performed between two cell types from the same individual and repeated in a second individual. DMRs identified in each individual were ranked according to their P-values and intersected between the two individuals. The percentages of overlapping DMRs at different cutoffs were plotted.
11,162; only about 70 DMRs called by M&K from the H1 vs. fetal NSC comparison were expected to be false positives, while about 2066 of the DMRs called by MEDIPS could be false positives. These numbers suggest that M&K has higher specificity compared to MEDIPS.

To compare the sensitivities of the methods, we examined the enrichment of individual CpGs with significantly different methylation levels within the predicted DMRs. We focused again on the comparison between H1 ESC and fetal NSC samples because WGBS was available for both samples from which we could derive methylation levels at single CpG resolution. In this pairwise comparison, we used M&K or MEDIPS to define any DMR in which fetal NSCs had a higher methylation level than H1 ESC as a hypermethylated DMR, and any DMR in which fetal NSCs had a lower methylation level than H1 ESC as a hypomethylated DMR. Based on ranked P-values, we used the top 1000 predicted hypermethylated DMRs and top 1000 hypomethylated DMRs for this comparison. Using the WGBS data, we derived methylation levels for individual CpGs located within the predicted DMRs. We then calculated methylation level differences by subtracting the individual CpG methylation values in H1 ESCs from their values in fetal NSCs. The histograms of individual CpG methylation level differences were plotted for both hypermethylated DMRs and hypomethylated DMRs, as shown in Figure 1, C and D for M&K and MEDIPS, respectively. Compared to the background methylation level differences between the two cell types, the top 2000 DMRs predicted by M&K were enriched for differentially methylated CpGs. While MEDIPS also enriched for differentially methylated CpGs, it did so to a much lesser degree than M&K. The trend remained the same when we compared differing numbers of top predicted DMRs (Fig. 1E,F).

We then analyzed the concordance between these DMR predictions with the WGBS data. For any predicted DMR, we defined it as concordant if it was predicted as a hypermethylated (or hypomethylated) DMR by M&K or MEDIPS and the averaged differences of WGBS methylation values across all CpGs in the DMR were greater than 0.1 (or less than 0.1; fetal NSC WGBS values minus H1 ESC WGBS values). Otherwise, the predicted DMR was called a discordant prediction. The rates of concordance for both M&K and MEDIPS were plotted for the top DMRs generated at increasingly relaxed statistical cutoffs (Fig. 1G). The high concordance between M&K’s prediction and actual CpG methylation differences inferred from WGBS data was robust regardless of the P-value used. Furthermore, M&K’s concordance rate was higher than that of MEDIPS.

We also examined the reproducibility of DMR predictions between biological replicates. We performed comparisons using the same two cell types isolated from two different individuals—breast luminal epithelial cell samples (RM066BreLum and RM070BreLum) and breast myoepithelial cell samples (RM066BreMyo and RM070BreMyo). The comparison between two cell types from one individual should enrich for DMRs underlying cell type specificity, and these DMRs should be identified again in the comparison between the same two cell types of another individual. We examined reproducibility by intersecting the DMRs from both individuals at multiple P-value cutoffs. M&K had three- to fourfold higher reproducibility than MEDIPS in this analysis (Fig. 1H).

In addition to these evaluations, we also examined the agreement among DMRs detected by M&K, MEDIPS, and by WGBS, between H1 ESC and fetal NSC (Methods). Of the top 10,000 DMRs predicted by each method, M&K and WGBS overlapped by 4224, while MEDIPS and WGBS overlapped by 2979 (Supplemental Fig. S3A). As expected, the average DNA methylation difference (calculated by using WGBS data) was the greatest for the DMRs predicted by two methods; interestingly, DMRs predicted by M&K only and DMRs predicted by WGBS only had almost identical average DNA methylation differences, while those predicted by MEDIPS only had smaller DNA methylation differences (Supplemental Fig. S3B).

Overall, we conclude that M&K has high specificity, sensitivity, and reproducibility, and exhibits superior performance in terms of these metrics when compared to a recently published MeDIP-seq analysis method, MEDIPS. We hypothesize that the improved prediction of DMRs when using the M&K algorithm likely results from the integration of complementary measurements of the underlying methylation state. We note that the comparison between M&K and MEDIPS was on different grounds. Adding MRE-seq data to MEDIPS did not further improve MEDIPS’ performance (Supplemental Notes; Supplemental Fig. S4); however, MEDIPS was not designed to work on MRE-seq data. M&K’s superior performance is likely due to both having complementary data sets and a new statistical model designed specifically for this scenario.

Detecting tissue-specific DMRs across four tissue types

We applied M&K to understand how DNA methylation underlies identity at three levels: tissue types, different cell types within tissues, and matched cell types from different individuals. We generated 19 methylinomes from embryonic stem cells, adult blood cells, adult breast cells, and fetal brain cells, representing four tissue types. We plotted the P-value distributions generated by each pairwise, genome-wide M&K comparison on 500-bp-sized windows (Fig. 2A; Supplemental Fig. S5). These distributions suggested that methylation differences between tissues outnumber differences between cell types of the same tissue or between the same cell types from two individuals, at least in the context of the current study.

We used a subset of the above pairwise comparisons to define known and novel tissue-specific DMRs. We identified genomic windows in which DNA methylation levels were similar between cell types from the same tissue but different from all cell types from the three other tissues. We required a window to have a Q-value of less than 1 × 10−5 in all comparisons between any cell type of one tissue to all cell types in the other three tissues but to have a Q-value of greater than 1 × 10−5 in all intra-tissue cell-type comparisons. Based on these criteria, a total of 2775 DMRs were defined as tissue-specific DMRs (Table 1; supporting website http://epigenome.wustl.edu/MnM/). Methylation levels of these DMRs clearly delineated the tissue types, as illustrated by biclustering analysis of MeDIP-seq and MRE-seq in these DMRs (Fig. 2B).

We hypothesized that these tissue-specific DMRs underlie important tissue-specific functions. Therefore, we examined their genomic distribution, chromatin patterns, and the functional enrichment of genes near or containing these DMRs. Of the 721 H1 ESC-specific DMRs, >80% were hypermethylated (Fig. 3A). Forty-eight percent of these overlapped CpG islands, and 23% overlapped gene promoters (Fig. 3A). By our definition, H1 ESC hypermethylated DMRs were hypomethylated in blood, breast, and fetal brain samples. Intriguingly, when we examined the histone modification profiles at H1 ESC-hypermethylated DMRs in blood, breast, and brain samples, we found that >50% were enriched for H3K4me3 signal (a promoter-associated histone modification), while only a small fraction (5%) was enriched for H3K4me1 signal (an enhancer-associated histone modification) (Fig. 4A; Table 1). This suggested that many H1 ESC hypermethylated DMRs were
associated with genes that were expressed in differentiated cells but repressed in H1 ESC cells. The apparent gain of H3K4me3, in the absence of gain of H3K4me1, in differentiated cells suggested that the up-regulation of expression of these genes relies on a key mechanism that is promoter-, rather than enhancer-, dependent. These DMRs represent a class of DNA-methylation-silenced promoters that are not marked by bivalent domains in H1 ESC. Genes associated with H1 ESC-specific hypermethylated DMRs enriched for zinc finger DNA binding proteins based on GREAT analysis (Fig. 3B; McLean et al. 2010), while H1 ESC-specific hypomethylated DMRs enriched for “target of Nanog” (Fig. 3C) and enriched for H3K4me3 in ESC (Fig. 4A). Some of the H1 ESC-specific hypermethylated genes may encode general differentiation factors (Supplemental Fig. S6). Interestingly, H1 ESC-specific hypermethylated DMRs displayed a moderate level of H3K4me1 enrichment, which may correlate with a transcriptionally poised state (Fig. 4A).

In contrast to ESCs, the majority of tissue-specific DMRs identified in blood, breast, and fetal NSC samples were hypomethylated (Table 1; Fig. 3A). Analysis of histone modification profiles for these regions revealed enrichment of H3K4me3 and H3K4me1 in the corresponding samples. GREAT analysis revealed that genes associated with tissue-specific DMRs strongly enriched for functions relevant to each tissue type (Fig. 3C). For example, fetal brain hypomethylated DMRs enriched for “neural tube patterning” ($P < 1 \times 10^{-15}$) and “spinal cord development” ($P < 1 \times 10^{-15}$), whereas breast hypomethylated DMRs enriched for “mammary gland epithelium development” ($P < 1 \times 10^{-15}$) and blood hypomethylated DMRs enriched for “immune response” ($P < 1 \times 10^{-15}$) (Fig. 3C). Interestingly, blood hypomethylated DMRs displayed enrichment of H3K4me3 and H3K4me1 signals in ESC, breast, and fetal brain samples, suggesting that these DMRs were regulatory regions that were specifically turned off in blood cells but were active or permissive for activity in other cell types (Fig. 4A). Representative genes included HOXAS and ISYNA1 (Supplemental Fig. S7).

These data suggested a strong connection between tissue-specific DNA methylation and tissue-specific gene activity. When hypomethylated, DMRs were almost always associated with tissue-specific gene regulatory elements. As expected, many DMRs in tissue-specific genes occurred at promoters, while others appeared to be associated with enhancers. The majority of tissue-specific DMRs were hypermethylated in embryonic stem cells. They became unmethylated in differentiated cell types, and 41% acquired a promoter-associated histone mark (H3K4me3), while 30% acquired an enhancer-associated histone mark (H3K4me1) (Fig. 4A). These epigenetic changes underscored the importance of DNA methylation in tissue differentiation.

This result was further supported by chromatin state annotation of genomic sequences predicted to be tissue-specific DMRs (Fig. 4B). We obtained chromatin state transition maps generated by chromHMM using nine cell lines (Ernst et al. 2011; Ernst and Kellis 2012; Methods). Almost all tissue-specific DMRs across ESC, fetal brain, breast, and blood were annotated as regulatory elements, including promoters, enhancers, and insulators. The only exception was fetal brain-specific hypomethylated DMRs—while most of these were marked by H3K4me1 in fetal brain samples, 60% did not have chromHMM annotation. This may be explained

| Table 1. Tissue-specific DMRs |
|-----------------------------|
|                            | ESC | Adult blood | Adult breast | Fetal brain | Union |
| Total DMRs                  | 721 | 725         | 578          | 751         | 2775  |
| Hypomethylated DMRs         | 593 | 229         | 115          | 0           | 937   |
| Hypomethylated DMRs peak    | 128 | 496         | 463          | 751         | 1838  |
| DMRs with H3K4me3           | 55% | 54%         | 33%          | 20%         | 41%   |
| DMRs with H3K4me1 peak      | 5%  | 36%         | 51%          | 30%         | 30%   |

Figure 2. M&M analyses of DNA methylation differences across multiple tissue types, cell types, and individuals. (A) P-value distributions of M&M predictions between tissue types (green lines), cell types (blue lines), and individuals (red lines). (B) Biclustering analysis of tissue-specific DMRs. (Left panel) Based on RPKM values of MeDIP-seq; (right panel) based on RPKM values of MRE-seq.
by the lack of a neural cell type among the nine cell lines used to produce the chromatin state map (Methods). Interestingly, promoters were more enriched in hypermethylated DMRs, while epigenetically defined enhancers dominated the hypomethylated DMR list (Fig. 4B).

Finally, gene expression data also supports a strong association between tissue-specific DNA methylation and tissue-specific gene activity. By using RNA-seq, we profiled transcriptomes of a subset of the samples. Expression levels of genes near tissue-specific DMRs were significantly higher in samples that were hypomethylated at these DMRs (Fig. 4C).

**Tissue-specific DMRs that span large chromosomal domains**

The majority of the tissue-specific DMRs we identified were relatively small in size, reflecting discrete regulatory elements such as enhancers. We also observed large domains of DNA methylation changes, some of which spanned over 75 kb in length. These distinct DMR patterns suggested that different underlying mechanisms could generate tissue-specific DMRs. We have summarized these large DMR domains in Supplemental Table 2. We describe two such examples below, with another four examples presented in Supplemental Figs. S8 and S9.

We discovered 18 breast-specific DMRs clustered in a 75-kb region on chromosome 22. This large region was hypomethylated in all breast-cell samples analyzed, as evidenced by decreased MeDIP-seq signal and increased MRE-seq signal (Fig. 5A; Supplemental Fig. S10A for bisulfite validation). This region spanned six CpG islands and five noncoding genes, including two long noncoding RNA genes, *LINC00899* and *LOC150381*, a putative coding gene *C22orf26*, and two isoforms of the tumor-suppressor miRNA *MIRLET7*, *MIRLET7A3*, and *MIRLET7B*. The *MIRLET7* family was discovered in *Caenorhabditis elegans* and is functionally conserved from worm to human. The human *MIRLET7* family includes 13 isoforms located on nine different chromosomes. Silencing of *MIRLET7* plays an important role in breast cancer progression (Yu et al. 2007), as reduced *MIRLET7* expression promotes cancer cell invasiveness and metastasis (Qian et al. 2011). We also examined the methylation state of this large region by WGBS from breast cancer cell line HCC1954 (Fig. 5A). Compared to normal human mammary epithelial cells (HMECs), this region was dramatically more methylated in the HCC1954 cancer cell line. This epigenetic event may reflect silencing of the pri-miRNA gene, *MIRLET7BHG*, that hosts the *MIRLET7* genes and potentially contribute to the invasiveness and increased proliferation previously reported in breast cancer cells.

A 740-kb region on chromosome 5 containing three Protocadherin (*PCDH*) gene families provided another interesting example where large domain changes consisted of many smaller, local changes (Fig. 5B; Supplemental Fig. S10B for bisulfite validation). Seventy-five of the 83 CpG islands in this region were specifically hypermethylated in H1 ESC. However, in differentiated tissues, these CpG islands gained a strong unmethylated signal, while maintaining a strong methylated signal (i.e., simultaneous high MeDIP-seq signal and high MRE-signal), indicating the CpG islands carry an intermediate methylation level. The *PCDH* gene...
family members belong to the cadherin superfamily and are present in all vertebrate genomes and highly conserved in mammals (Wu and Maniatis 1999). Most PCDH family members are clustered in three loci on chromosome 5, and share one highly conserved motif in their promoters (Wu et al. 2001). PCDH genes are known to play important roles in neuronal cell differentiation and brain development (Prasad et al. 2008; Garrett and Weiner 2009; Lin et al. 2010). Previous studies suggest that the expression of each PCDH member is monoallelic and regulated independently (Esumi et al. 2005; Kaneko et al. 2006), an observation that is consistent with our data, since an intermediate methylation level is a signature of monoallelic methylated sites. De novo methylation of the PCDH gene cluster is also associated with tumorigenesis (Novak et al. 2008; Dallosso et al. 2009), raising the possibility that establishing monoallelic methylation constitutes an important event in maintaining differentiated states. In contrast, promoters of the PCDH family are highly methylated in cells of all three germ layers differentiated from mouse ES cells but not in ES cells themselves (Singer 1988). Whether the acquisition by differentiated cells of intermediate DNA methylation patterning in this region is specific to humans and how this phenomenon evolved awaits further investigation.

Cell type-specific DMRs underlie enhancers associated with relevant pathways

Our data set includes three breast cell types (a breast stem cell-enriched population, luminal epithelial cells, and myoepithelial cells) and three blood T cell types (naive CD4+ T cells, memory CD4+ T cells, and naive CD8+ T cells). This presented a unique opportunity to discover cell type-specific DMRs and to compare their epigenomic signature to that of tissue-specific DMRs. To define such cell type-specific DMRs, we required a genomic window to have a Q-value of less than $1 \times 10^{-5}$ in all comparisons between two cell types of the same tissue in two independent biological replicates. This analysis revealed that the most striking feature of cell type-specific DMRs is the enrichment of an enhancer chromatin signature. We use the following example to illustrate this discovery.

We examined DNA methylation changes that were associated with maturation of naive CD4+ T cells into memory CD4+ T cells in the immune system. In addition to producing cytokines and chemokines, CD4+ T cells also act as mediators for other lymphocytes via cell-cell contact (Swain et al. 2006). During responses to antigens, most CD4+ T cells die within a few days after receiving...
Identification of tissue-specific DMRs spanning large chromosomal domains. (A) A breast-specific hypomethylated region containing multiple noncoding RNA genes. (Green box) 75-kb region hypomethylated in all breast cell types (luminal [Lum], myoepithelial [Myo], and stem cell-enriched [BSC]). (Red box) Hypermethylation events within the same region in the HCC1954 breast tumor cell line. (B) A large H1 ESC-specific hypermethylated chromosomal domain spanning the PCDHG gene cluster. (Orange box) H1 ESC-specific hypermethylated DMRs in the vicinity of the promoters of several PCDHG gene family members.
antigen stimulation, while only a small fraction survives. This small population corresponds to mature memory CD4+ T cells that contribute to later adaptive immune responses and reproduce rapidly upon restimulation of the same antigen. We compared the DNA methylomes of naive CD4+ T cells (CD4N) and memory CD4+ T cells (CD4M) from two individuals to identify cell type-specific DMRs (intra-CD4 DMRs). Compared to CD4N, CD4M cells showed hypomethylation in 349 genomic regions and hypermethylation in 287 regions (Fig. 6A). We detected enrichment of H3K4me1 signal in the majority of intra-CD4 hypomethylated DMRs in the samples where the DMRs are hypomethylated (62%), while a small fraction of DMRs displayed H3K4me3 signal (11%) (Fig. 6B; Table 2). The frequent overlap of intra-CD4 hypomethylated DMRs with enhancers was further supported by chromHMM annotation (Fig. 6D). Histone modification profiling supported that many of the intra-CD4 DMRs are regulatory sites. This is further supported by data from ENCODE, in that 88% of these DMRs directly overlapped DNase I hypersensitivity sites, and 17% directly overlapped EP300 binding sites in at least one of the cell lines assayed by ENCODE (Supplemental Fig. S11; The ENCODE Project Consortium et al. 2012; Thurman et al. 2012). Therefore, we reasoned that the intra-CD4 DMRs would harbor binding sites

**Figure 6.** Cell type-specific DMRs between CD4 naive cells and CD4 memory cells. (A) Genomic distribution of CD4 memory cell hypomethylated DMRs (green) and CD4 naive cell hypomethylated DMRs (red). (B) Histone modification profiles (H3K4me1 and H3K4me3) of DMRs between CD4 memory cells and CD4 naive cells. (C) Functional enrichment in CD4 memory cell (green) and CD4 naive cell hypomethylated DMRs (red). (D) ChromHMM regulatory function annotation of CD4 memory cell DMRs and CD4 naive cell DMRs. (E) TFBS enrichment of CD4 memory cell DMRs (green) and CD4 naive cell DMRs (red).
of relevant transcription factors. Indeed, by examining integrated ENCODE TFBS data, we found significant enrichment of many transcription factor-binding sites in these DMRs (Fig. 6E). Binding of transcription factors to genomic DNA motifs is associated with changes in the local epigenetic landscape (Asp et al. 2011; Stadler et al. 2011). Our data further support this type of association at DMRs that define different cell types within breast tissue.

Functional enrichment analysis (Fig. 6C) of CD4 hypermethylated DMRs identified genes enriched for functions including lymphocyte differentiation and T cell receptor V(D)J recombination. These regions became methylated during the process of CD4+ T cell maturation. Increased CD4M DNA methylation was also observed in genes involved in apoptosis and cellular response to interleukin-4 (IL4). These DNA methylation events were consistent with transitions in cellular function during the maturation process. In contrast, DNA hypomethylation in CD4M was detected in the vicinity of genes involved in the immune system process and activation, and protein synthesis, including protein translation and elongation. Interestingly, some hypomethylated DMRs in CD4M were found to be close to genes that regulate the production of the Th2 cytokine interleukin-4, including the IL4 gene (Supplemental Fig. S12). As a key factor during CD4+ T cell maturation, IL4 induces long-term proliferation of neonatal T cells and stimulates production of other cytokines (Wu et al. 1994). Some memory CD4+ T cells produce IL4 and perform important immune regulatory functions (Cosmi et al. 2010; Xu et al. 2011).

Several genes important for CD4M function, including NDFIP1, EB13, SIVA1, and TNFRSF4 (Supplemental Fig. S12), displayed decreased DNA methylation at regions upstream of their respective promoter in CD4M, although the promoter itself was unmethylated in both CD4N and CD4M. These hypomethylated DMRs also gained H3K4me1 signal in CD4M (Supplemental Fig. S12). Therefore, epigenetic regulation of expression of these CD4+ T cell type-specific genes is likely to involve enhancers rather than promoters. Taken together, our data highlight that a majority of cell type-specific DMRs likely correspond to cell type-specific enhancer elements, while tissue-specific DMRs enrich primarily for gene promoters.

### Table 2. Cell type-specific DMRs

| Breast                                      | Blood                                      | Union |
|---------------------------------------------|--------------------------------------------|-------|
| DMR hypomethylated in luminal epithelial cells | DMR hypomethylated in CD4 naive cells     | NA    |
| DMR hypomethylated in myoepithelial cells   | DMR hypomethylated in CD4 memory cells     | NA    |
| DMRs with H3K4me3 peak                      | DMRs with H3K4me3 peak                     | NA    |
| DMRs with H3K4me1 peak                      | DMRs with H3K4me1 peak                     | NA    |

**DMRs between individuals overlap with gene regulatory elements**

Epigenetic polymorphisms, including DNA methylation differences between individuals, are increasingly associated with phenotypic diversity and disease susceptibility (Tost et al. 2006; Baranzini et al. 2010; Coolsen et al. 2011; Eichten et al. 2011; Gertz et al. 2011; Gervin et al. 2011). Unlike genetic polymorphisms such as SNPs and copy number variation, epigenetic polymorphisms can be influenced by both genetic and environmental determinants (Anway et al. 2005; Li et al. 2011; Crews et al. 2012; Skinner et al. 2012). In our study, we obtained biological replicates from two individuals for the breast, fetal brain, and blood data sets. We did not address the association between genotype and epigenotype in the current study. Rather, we sought to identify regions of the genome that are hotspots for individual-specific differential methylation by comparing DNA methylomes of the same cell types between different individuals (Table 3).

We identified 1032 DMRs between each pair of individuals (inter-individual DMRs) (Fig. 7A). We noticed that 389 of these DMRs overlapped with satellite DNA and microsatellite repeats. This class of DMRs could result from genetic polymorphism (i.e., copy number differences in satellite repeats) among the individuals and not epigenetic polymorphism (Haaf and Willard 1992) or could be known artifacts associated with mapping short reads to satellite repeats. Therefore, we excluded these regions from further analysis (Table 3).

The remaining 643 DMRs, when considered together, did not seem to associate with genes that enrich for any particular function. Nevertheless, more than half of these DMRs were annotated by chromHMM as regulatory elements (Fig. 7B). Interestingly, >40% of the inter-individual DMRs identified using the fetal brain samples (inter-brain DMRs) from two monozygotic twins displayed individual-specific H3K4me1 marks (Fig. 7C; Table 3). These inter-brain DMRs strongly enriched for association with genes in brain development (Fig. 7E) and also enriched for transcription factor binding sites (Fig. 7F). Taken together, we hypothesize that at least some of the inter-individual DMRs might influence gene expression in an individual-specific manner and therefore influence particular traits. For example, we found potentially regulatory DMRs in the introns of CYP2D6 and CYP2E1, both of which belong to the cytochrome P450 family and are implicated in metabolizing precarcinogens, drugs, and solvents to reactive metabolites (Agundez 2004; Bozina et al. 2009). Other examples included DMRs located near neuronal specific genes, e.g., FGFR3, a gene that plays an important role in neuronal development (Puligilla et al. 2007), NFIX, a gene that regulates expression of glial fibrillary acidic protein, GFAP (Singh et al. 2011), and NAV1, a member of the neuron navigator family (Supplemental Fig. S13; Maes et al. 2002).

### Table 3. Individual-specific DMRs

|                     | TC007 vs. TC009 (Blood) | RM066 vs. RM070 (Breast) | TwinA vs. TwinB (Fetal brain) |
|---------------------|-------------------------|--------------------------|-------------------------------|
| Total DMRs          |                         |                          |                               |
| DMRs with H3K4me3 peak | 310  (no data)         | 55  (no data)           | 307  (no data)              |
| DMRs with H3K4me1 peak | (no data)              | (no data)               | NA                            |

A complete list of tissue-specific, cell type-specific, and individual-specific DMRs is provided at the following supporting website: http://epigenome.wustl.edu/MnM/. H3K4me3 and H3K4me1 peaks were identified using MACS (Zhang et al. 2008). A DMR was defined to have histone peaks when at least 50% of the DMR overlapped with histone peaks.
DNA methylation plays important roles in cells, including the regulation of genes during development and disease (Robertson 2005; Lister et al. 2009; Deaton et al. 2011; Jones 2012). It has been increasingly associated with tissue-specific gene activity (Kitamura et al. 2007; Illingworth et al. 2008; Maunakea et al. 2010). The technology is now available for studying DNA methylation genome-wide, at high resolution and in a large number of samples, presenting an unprecedented opportunity to map DNA methylation differences with M&M.
differences between tissues and cells in healthy and diseased states and during development (Bock 2012).

Here, we introduce a novel statistical framework, “M&M,” to detect differentially methylated regions by integrating two next-generation sequencing-based technologies—MeDIP-seq and MRE-seq. MeDIP-seq, or methylated DNA immunoprecipitation followed by sequencing, uses an antibody that recognizes 5-methylcytosines in the genome. Because the MeDIP-seq protocol is straightforward and generates unbiased, cost-effective, and full-genome methylation levels, it is widely used for whole-genome DNA methylation analysis. Several computational tools, including Batman (Down et al. 2008), MEDIPS (Chavez et al. 2010), and MeQA (Huang et al. 2012), were developed to process and analyze MeDIP-seq data. MRE-seq was recently developed as a complementary technique to MeDIP-seq, such that it utilizes methylation-sensitive restriction enzymes to digest genomic DNA and generate DNA fragments where the free ends contain single unmethylated CpGs. Previously, we showed that by combining MeDIP-seq and MRE-seq we can generate genome-wide DNA methylation maps of very high coverage and can identify intermediate and/or monoallelic methylated regions (Harris et al. 2010). We now show that utilizing M&M to combine data from MeDIP-seq and MRE-seq in a statistically rigorous manner provides improved prediction of differentially methylated regions. Since these two methods assay for opposing methylation states, an increase in the signal when using one method is expected to be accompanied by a decrease in the signal when using the other method. The M&M algorithm models the expected MeDIP-seq and MRE-seq values in relation to the numbers of methylated and unmethylated CpGs in any given genomic interval, and unifies these two complementary data types by the relative methylation levels. M&M then detects DMRs by testing for the null hypothesis that the methylation levels of two samples are the same, given the observed MeDIP-seq and MRE-seq data. The complete M&M package is now part of Bioconductor under the name of “methylMnM.”

To benchmark the performance of M&M, we selected two samples (H1 ESC and fetal NSC) for which we had characterized genome-wide DNA methylation profiles using WGBS, MeDIP-seq, and MRE-seq. We applied M&M to discover DMRs between H1 ESC and fetal NSC and compared the results with those obtained by applying MEDIPS, one of the leading software tools currently used for analyzing MeDIP-seq data, on our MeDIP-seq data. M&M displayed superior sensitivity, specificity, and reproducibility as highlighted by a much higher concordance for methylation level differences when compared to those determined by WGBS data (Fig. 1C–E). The addition of MRE-seq data provided independent information about methylation state, which M&M takes advantage of by integrating MeDIP-seq and MRE-seq under the same statistical framework for more accurate DMR detection. It should be noted that the statistical framework can be generalized and can potentially integrate other types of DNA methylation data.

As part of the Roadmap Epigenomics Initiative, we produced a large collection of DNA methylomes using MeDIP-seq and MRE-seq, representing multiple tissue and cell types from multiple individuals. By applying M&M to selected data sets, we were able to create comprehensive lists of DMRs that are either tissue type-, cell type-, or individual-specific. This resource is publicly available through the Wash U Human Epigenome Browser (Zhou et al. 2011, 2013) and the UCSC Genome Browsers (http://VizHub.wustl.edu).

DNA methylation has been investigated for its role in specifying tissue-specific gene expression (Song et al. 2005; Shen et al. 2007; Irizarry et al. 2009; Laurent et al. 2010; Maunakea et al. 2010; Liang et al. 2011; Andersen et al. 2012; Davies et al. 2012). It has long been appreciated that DNA methylation of non-CpG island gene promoters is a durable mechanism that results in repression of genes in specific tissues during development (Lamond and Stockdale 1989; Luo et al. 2000; Winders et al. 2004). Cancer cells also exploit a similar mechanism to turn off tumor suppressor genes (Mommparler and Bovenzi 2000; Robertson 2005; Daniel et al. 2011). Promoters and CpG islands have been, therefore, the primary focus of DNA methylation studies. Recently, cell type-specific “lowly methylated regions” (LMRs) distal to genes have been discovered in the mouse genome. These LMRs are created by binding of DNA binding factors outside of promoter regions (Stadler et al. 2011), highlighting the importance of comprehensive comparisons of DNA methylomes and DMR discovery beyond promoters and CpG islands.

Our analysis revealed distinct histone modification and gene regulatory signatures underlying tissue-, cell type-, and individual-specific DMRs. Consistent with our expectations, DNA methylation differences were strongly associated with gene regulatory elements. The majority of the DNA methylation differences were accompanied by differences in the active chromatin marks H3K4me3 and H3K4me1, suggesting that these DMRs overlap promoters and/or enhancers. Genes associated with DMRs strongly enriched for functions relevant to the tissue or cell types in any given comparison. Surprisingly, the proportions of promoter-like DMRs and enhancer-like DMRs were quite different between tissue-specific DMRs and cell type-specific DMRs—about 40% of tissue-specific DMRs were annotated as promoters, while >70% of cell type-specific DMRs were annotated as enhancers (Tables 1, 2). This signature suggests that different mechanisms may establish and maintain DNA methylation patterns during development and differentiation, resulting in different phenotypic consequences. DNA methylation of promoters likely has a strong impact on gene expression, equivalent to an on/off switch that keeps genes in an active state or shuts down gene expression completely. Thus, methylation of gene promoters can be a mechanism for major cell lineage determination. Our most significant finding was that many differentially methylated regions associate with tissue- and cell type-specific enhancers, as determined by histone modifications and chromHMM annotations (Tables 1, 2; Figs. 4, 6B,C, 7B,E). DNA methylation in enhancers may vary with binding of specific transcription factors (Stadler et al. 2011) and also influences transcription factor binding and modulates gene expression. Methylation of enhancers, thus, could provide a mechanism that allows cells to fine-tune gene activity and make a final lineage commitment or maintain a distinct cell fate.

We also identified 307 individual-specific DMRs when comparing the same tissue type (fetal brain tissue) between two monozygotic twins. Interestingly, these DMRs also enriched for strong, individual-specific H3K4me1 enhancer signal, highlighting the potential functional impact of epigenetic variation between genetically identical individuals. Epigenetic polymorphism has attracted enormous interest in recent years. Several recent studies investigating the connection between genetic variation and epigenetic variation revealed that epigenetic polymorphisms are likely the result of a complex interplay between genetic and environmental factors (Anway et al. 2005; Li et al. 2011; Crews et al. 2012; Skinner et al. 2012). Our finding provides additional support for the potential functional impact of epigenetic polymorphisms.

Leveraging new sequencing-based DNA methylome mapping technologies and a novel, integrative statistical framework, we mapped DNA methylation differences across multiple tissues and
cell types from multiple individuals at the whole-genome level. Differences in DNA methylation can be tightly associated with gene regulation, impacting regulatory elements such as promoters and enhancers, and defining the cellular epigenetic landscape. Interestingly, the DMRs identified in our study were enriched for evolutionarily conserved sequences (Fig. 8). Many of the functional elements in the human genome have been identified by comparative genomic approaches based on evolutionary principles (Lindblad-Toh et al. 2011). Sequence conservation correlates with conservation of some epigenetic marks, including DNA methylation, between species. Furthermore, conserved localization of epigenetic marks can be used to define regulatory DNA elements (Maunakea et al. 2010; Xiao et al. 2012). Enrichment of conserved sequences in DMRs suggests that these sequences may be under purifying selection and that their regulatory functions may be evolutionarily selected. It would be exciting to examine whether the same principle applies to tissue-, cell type-, and individual-specific DNA methylation for other species.

Methods

All data were obtained from the NIH Roadmap Epigenomics Mapping Centers’ repository for the human reference epigenome atlas (Bernstein et al. 2010). Experiments were performed under the guidelines of the Roadmap Epigenomics project (http://www.roadmapepigenomics.org/protocols). Specifically, MeDIP-seq and MRE-seq experiments were performed as described previously (Maunakea et al. 2010). ChIP-seq was performed as described in O’Geen et al. (2011). All data have been previously submitted to NCBI (Supplemental Table 1).

M&M statistical model

The complete M&M statistical model derivation, data processing and normalization, and false discovery rate control are described in Supplemental Notes.

Processing of MeDIP-seq and MRE-seq data

The reads were aligned with Bowtie (Langmead et al. 2009) to hg19. MRE reads were normalized to account for differences in enzyme efficiency. Scoring consisted of tabulating reads with CpGs at each fragment end (Maunakea et al. 2010).

Processing of Roadmap Epigenomics histone modification data

H3K4me3 and H3K4me1 ChIP-seq data for relevant cell types were produced as part of the Roadmap Epigenomics project and deposited to GEO (GSE16368). Mapped read density was generated from aligned sequencing reads using customized Perl scripts. Read density overlapping DMRs and their 5-kb upstream/downstream regions were extracted at 50-bp resolution as RPKM values, with histone input data subtracted.

ENCODE HMM chromatin state annotation

ChromHMM annotations for ENCODE cell lines (Ernst et al. 2011) were obtained from the UCSC Genome Browser (Rosenbloom et al. 2012). The nine cell lines are: H1 ESC, GM12878, K562, HepG2, HUVEC, HMEC, HSMM, NHEK, and NHLF. For each DMR, we examined overlapping annotation of “promoter,” “enhancer,” and “insulator” states in these chromHMM maps.

ENCODE DHS and EP300 binding sites data

ENCODE DNase I hypersensitive sites (DHS) and EP300 binding sites data were downloaded from the UCSC Genome Browser ENCODE data portal (Rosenbloom et al. 2012). DHS data contain 999,988 regions, and EP300 binding sites data contain 92,623 regions.

Whole-genome bisulfite sequencing data processing

The whole-genome bisulfite sequencing data for H1 ESC and human fetal brain NSC (sample HuFNSC02) were downloaded from GEO (GSE16368). For validation of methylation changes in each DMR between H1 ESC and HuFNSC02, the averaged methylation difference between all CpG sites in the DMR was calculated.

The whole-genome bisulfite sequencing data for normal breast cells (HMEC) and breast cancer cells (HCC1954) were downloaded from GEO (GSE29127). The raw data were processed and mapped to human genome hg19 assembly using Bismark (Krueger and Andrews 2011).

Figure 8. Sequence conservation of DMRs. Vertebrate phastCon scores were obtained at 100-bp resolution for each DMR and their respective upstream/downstream 5-kb regions. Averaged scores in each 100-bp window were plotted. (A) Conservation of tissue-specific DMRs. (B) Conservation of cell type-specific DMRs. (C) Conservation of individual-specific DMRs.
RNA-seq data processing
RNA-seq data of H1 ESC, fetal brain, myoepithelial cells, and PBMC were downloaded from GEO (GSE16368). Data were processed with Cufflinks (Trapnell et al. 2010) to obtain RPKM values for all RefSeq genes (annotation obtained from UCSC Genome Browser).

Detecting DMRs from WGBS data
WGBS-defined single CpG resolution DNA methylation levels of H1 ESC and fetal NSC were used as input. For each genomic interval, a Student’s t-test was performed on the two groups of methylation levels to obtain a P-value. DMRs were selected based on ranked P-values.

Applying MEDIPS
The MEDIPS package was downloaded from Bioconductor. The command lines utilized were as described in the MEDIPS manual (version 1.0.0) with genomic window size set to 500 bp. DMRs were selected by commands: MEDIPS.selectSignificants(), using the parameters suggested in the manual.

Applying MACS to histone data
The following parameters were used to apply MACS (Zhang et al. 2008) to histone data for the identification of peaks:

```
--petdist = 200; --bw = 300; --p1e-5; --slocal = 1000;
--llocal = 1000; --shiftsize = 100; --verbose = 2.
```

A DMR was defined as enriched for histone signal when at least 50% of the DMR overlapped with histone peaks.

ENCOD TFBS enrichment
ENCOD TFBS data were downloaded from the UCSC Genome Browser ENCOD data portal (Rosenbloom et al. 2012). The binding site enrichment score (ES) for each transcription factor with respect to DMRs was calculated as:

```
ES = ntfsbs/nDMR
```

where ntfsbs is the number of DMRs that contain experimentally annotated TFBS; nDMR is the total number of DMRs; ntfsbs is the number of genomic windows with annotated TFBS; Nall is the number of 500-bp windows in the human genome (hg19).

Genomic features
RepeatMasker annotations, CpG islands, genomic super duplications, 46-way phastCons, and refGene features (including 5’ UTR, exons, introns, and 3’ UTRs) were all downloaded from the UCSC Genome Browser (Kent et al. 2002; Meyer et al. 2012). Promoters were defined as 2.5 kb around the most 5’ transcription start site (2 kb upstream of and 0.5 kb downstream from TSS) of any refGene record. Intergenic regions were defined as regions between neighboring refGene loci.

Gene class enrichment analysis
Gene Ontology (GO) analyses for biological processes were performed using the GREAT package (McLean et al. 2010). Gene regulatory domains were defined by default as the regions spanning 5 kb upstream of and 1 kb downstream from the TSS (regardless of other nearby genes). Gene regulatory domains were extended in both directions to the nearest gene’s basal domain but no more than a maximum extension in one direction. Only categories that were below a false discovery rate of 0.05 were reported.

Bisulfite treatment and library construction for WGBS
One to 5 μg gDNA was sonicated to an approximate size range of 200–400 bp. Size selection was achieved by PAGE gel and yielded DNA fragments of 200–300 bp. DNA was quantified by fluorescent incorporation (Qubit, Invitrogen). The library preparation included end-repair and phosphorylation with NEBNextTM or Illumina Sample Prep Kit reagents and addition of an ‘A’ base to the 3’ end of the DNA fragments. Methylated adapters were ligated and size selection was performed to remove excess free adapters. The ligated DNA was quantified by Qubit, and ~100 ng DNA was used for bisulfite conversion. Methylated-adaptor ligated to unmethylated lambda-phae DNA (NEB) was used as an internal control for assessing the rate of bisulfite conversion. The ratio of target library to Lambda was 1600:1. Bisulfite conversion of the methylated adapter-ligated DNA fragments followed the FFPE Tissue Samples Protocol from Qijen’s Epitect Bisulfite Kit. Cleanup of the bisulfite-converted DNA was performed, and a second round of conversion was applied. Enrichment of adaptor-ligated DNA fragments was accomplished by dividing the template into five aliquots followed by eight cycles of PCR with adaptor primers. Post-PCR size-selection of the PCR products from the five reactions was achieved by PAGE gel. Following 100-hp paired-end sequencing on a HiSeq2000, sequence reads were aligned and processed through the Bismark pipeline.

Sample preparation
Blood
Buffy coats were obtained from the Stanford Blood Center (Palo Alto, CA). Blood was drawn and processed on the same day. Peripheral blood mononuclear cells (PBMC) were isolated by Histopaque 1077 (Sigma-Aldrich) density gradient centrifugation according to the manufacturer’s protocol. Further purification of CD4 memory, CD4 naïve, and CD8 naïve T lymphocytes was performed using a Robosep instrument and isolation kits for each subpopulation as listed below (STEMCELL Technologies). Total PBMC were karyotyped (Molecular Diagnostic Services, Inc) and analyzed for cell cycle. PBMC and T cell subpopulations were stained with antibodies and analyzed by FACS for purity. Cells were aliquoted for DNA and RNA samples and were washed in PBS. Cell pellets for RNA samples were resuspended in 1 mL TRIzol reagent (Invitrogen), and frozen at ~80°C. Cell pellets for DNA samples were flash-frozen in liquid nitrogen and stored at ~80°C. Reagents and antibodies:
- Anti-CD3 TRI-COLOR, Invitrogen
- Anti-CD4 PE, BD Biosciences
- Anti-CD8 FITC, BD Biosciences
- Anti-CD4 TRI-COLOR, Invitrogen
- Anti-CD45RO PE, Invitrogen
- Anti-CD45RA FITC, BD Biosciences
- Anti-CD8 TRI-COLOR, Invitrogen
- EasySep Human Memory CD4 T Cell Enrichment Kit,
- EasySep Human Naive CD4+ T Cell Enrichment Kit,
- Custom Human Naive CD8 T Cell Enrichment Kit,
- STEMCELL Technologies

Breast
Breast tissues were obtained from disease-free, premenopausal women undergoing reduction mammoplasty in accordance with
institutionally approved IRB protocol # 10-01563 (previously CHR # 8759-34462-01). All tissues were obtained as de-identified samples and linked only with a minimal data set (age, ethnicity, and in some cases, parity/gravidity). Tissue was dissociated mechanically and enzymatically, as previously described (Romanov et al. 2001). Briefly, tissue was minced and dissociated in RPMI 1640 with L-glutamine and 25 mM HEPES (Fisher, cat. # MT10041CV) supplemented with 10% fetal bovine serum (JR Scientific, Inc., cat. # J43603), 100 μM penicillin, 100 μg/mL streptomycin sulfate, 0.25 μg/mL fungizone, gentamycin (Lonza, cat. # CC4081G), 200 U/mL collagenase 2 (Worthington, cat. # CLS-2), and 100 μM hyaluronidase (Sigma-Aldrich, cat. # H3506-SG) at 37°C for 16 h. The cell suspension was centrifuged at 1400 rpm for 10 min, followed by a wash with RPMI 1640/10% FBS. Clusters enriched in epithelial cells (referred to as organoids) were recovered after serial filtration through a 150-μm nylon mesh (Fisher, cat. # NY9445658), and a 40-μm nylon mesh (Fisher, cat. # NY9660187). The final filtrate contained primarily mammary stromal cells (fibroblasts, immune cells, and endothelial cells) and some single epithelial cells. Following centrifugation at 1200 rpm for 5 min, the epithelial organoids and filtrate were frozen for long-term storage. The day of cell sorting, epithelial organoids were thawed out and further digested with 0.5 g/L 0.05% trypsin-EDTA and dispase-DNase I (STEMCELL Technologies, cat. # 7913 and # 7900, respectively). Generation of single-cell suspensions was monitored visually. Single-cell suspensions were filtered through a 40-μm cell strainer (Fisher, cat. # 087711), spun down, and allowed to “regenerate” in MEGM medium (Lonza) supplemented with 2% fetal calf serum for 60–90 min at 37°C. This “regeneration” step enables quenching of trypsin and re-expression of the cell surface markers prior to staining, as their extracellular domain had been cleaved by trypsin.

The single-cell suspension obtained as described above was stained for cell sorting with three human-specific primary antibodies, anti-CD10 labeled with PE-Cy7 (BD Biosciences, cat. # 541092) to isolate myoepithelial cells, anti-CD227/MUC1 labeled with FITC (BD Biosciences, cat. # 559774) to isolate luminal epithelial cells, or anti-CD73 labeled with PE (BD Biosciences, cat. # 550257) to isolate a stem cell-enriched cell population, and with biotinylated antibodies for lineage markers, anti-CD2, CD3, CD16, CD64 (BD Biosciences, cat. # 555325, 555338, 555405, and 555526), CD31 (Invitrogen, cat. # MHC3D115), CD45, CD140b (Biologend, cat. # 304003 and 323604) to specifically remove hematopoietic, endothelial, and leukocyte lineage cells, respectively, by negative selection. Sequential incubation with primary antibodies was performed for 20 min at room temperature in PBS with 1% bovine serum albumin (BSA), followed by washing in PBS with 1% BSA. Biotinylated primary antibodies were reveled with an anti-human secondary antibody labeled with streptavidin-Pacific Blue conjugate (Invitrogen, cat. # S11222). After incubation, cells were washed once in PBS with 1% BSA, and cell sorting was performed using a FACSAria II cell sorter (BD Biosciences).

Fetal brain

Post-mortem human fetal neural tissues were obtained from a case of twin nonsyndrome fetuses whose death was attributed to environmental/placental etiology. Tissues were obtained with appropriate patient consent according to Partner’s Healthcare/Brigham and Women’s Hospital IRB guidelines (Protocol #2010P001144). All samples and tissues were de-identified and linked only with a minimal data set (age, gender, brain location). Fetal brain tissue and fetal neural progenitor cells were derived from manually dissected regions of the brain (telencephalon), specifically the neocortex (pallium; GSM669614, GSM669615) and ganglionic eminences (subpallium; GSM669613). The tissues were minced and dissociated by combination of mechanical agitation (gentleMACS device) during enzymatic treatment with papain, according to the manufacturer’s protocol (Miltenyi Biotec, neural tissue dissociation kit #130-092-628). Cell suspensions were then washed twice in DMEM and plated at low density in human NeuroCult NS-A media (Stem Cell Technology, # 05751) supplemented with heparin, EGF (20 ng/mL) and FGF (10 ng/mL) in ultralow attachment cell culture flasks (Corning, #3814).

ESCO HI

Data were obtained from a previous publication (Harris et al. 2010).

Data access

M&M is an open source software. The source code, parameter sets, and instructions, as well as sample data sets, are available at http://epigenome.wustl.edu/MnM/. M&M has been released as an add-on package called “methylMnM” for R 2.12.1 and is freely available from Bioconductor (Gentleman et al. 2004; http://www.bioconductor.org/). A list of tissue-specific, cell-type specific, and individual specific DMRs can be found in the Supplemental Material and at http://epigenome.wustl.edu/MnM/.

List of affiliations

1Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri 63108, USA; 2Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin Province, 130024, China; 3School of Applied Sciences, Harbin University of Science and Technology, Harbin, 150080, China; 4Department of Mathematics and Division of Biostatistics, Washington University in Saint Louis, Saint Louis, Missouri 63130, USA; 5Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143, USA; 6Department of Dermatology, University of California, San Francisco, California 94143, USA; 7Department of Pathology, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143, USA; 8Howard Hughes Medical Institute, Division of Rheumatology, University of California, San Francisco, California 94143, USA; 9Genome Center, University of California Davis, Davis, California 95616, USA; 10Department of Biochemistry and Molecular Biology, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California 90089, USA; 11Department of Medical Oncology, Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; 12Departments of Pathology, Brigham and Women’s Hospital, Children’s Hospital Boston, and Harvard Medical School, Boston, Massachusetts 02115, USA; 13Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri 63110, USA; 14BC Cancer Agency, Canada’s Michael Smith Genome Sciences Centre, Vancouver, British Columbia, V5Z 4S6, Canada; 15Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada.

Acknowledgments

We thank collaborators in the Reference Epigenome Mapping Centers (REMC), Epigenome Data Analysis and Coordination Center, and NCBI who have generated and processed data used in this project. We acknowledge support from the NIH Roadmap Epigenomics Program, sponsored by the National Institute on
Drug Abuse (NIDA) and the National Institute of Environmental Health Sciences (NIEHS). J.F.C., M.H., and T.W. are supported by NIH grant SU01ES017154. T.W. is supported in part by Basil O’Connor Starter Scholar Research Award Grant no. S-FY10-491 from the March of Dimes Foundation, the Edward Jr. Mallinkrodt Foundation, P50CA134254, RO1HG007354, and a grant from the Foundation for Barnes-Jewish Hospital. J.B.C. is supported by a Career Development Award from the Dermatology Foundation.

B.Z. is supported by NIDA’s R25 program DA027995. N.L. is supported by NIDA's R25 program DA027995. K.L.L. and T.W. are supported by the National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium.

References

Agresti A. 2007. An introduction to categorical data analysis, 2nd ed. Wiley, Hoboken, NJ.

Andersen I.S, Reiner A.H, Aanes H, Alestrom P, Collas P. 2012. Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biol* 13: R65.

Anway MD, Cupp AS, Uzumcu M, Skinner MK. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308: 1466–1469.

Asp P, Blum R, Vethanath S, Paris F, Micsinai M, Cheng J, Bowman C, Kluger Y, Dynlacht BD. 2011. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc Natl Acad Sci USA* 108: E149–E158.

Baranzini SE, Mudge J, van Wulfsbergh JC, Khankharian P, Khebrtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, et al. 2010. Genomic, epigenetic and RNA sequences of monozygotic twin discordant for multiple sclerosis. *Nature* 464: 1351–1356.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28: 1045–1048.

Bock C. 2012. Analysing and interpreting DNA methylation data. *Nat Rev Genet* 13: 705–719.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Herrero J, Tomazou EM, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylation analysis. *Nat Biotechnol* 26: 779–785.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Herrero J, Tomazou EM, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylation analysis. *Nat Biotechnol* 26: 779–785.

Boyes J, Bird A. 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64: 1123–1134.

Botina N, Bradamante V, Lovric M. 2009. Genetic polymorphism of metabolic enzymes P450 (CYP) as a susceptibility factor for drug response, toxicity, and cancer risk. *Arх Хиг Рада Токсикол* 60: 217–242.

Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, Cardilicchia E, Frosali F, Qu W, Campbell MJ, Henders AK, Montgomery GW, Martin NG, Clark SJ. 2011. Impact of the genome on the epigenome is manifested in DNA methylation patterns of imprinted regions in monozygotic and dizygotic twins. *PLoS ONE* 6: e25590.

Cosmi L, Maggi L, Santafaz D, Capone M, Cardilicchia E, Frosali F, Querci V, Angeli R, Matucci A, Fambrini M et al. 2010. Identification of a novel subset of human circulating memory CD4+ T cells that produce both IL-17A and IL-4. *J Allergy Clin Immunol* 125: 222–230.

Crow D, Gillette R, Scarpino SV, Manikam M, Savenkov MI, Skinner MK. 2012. Epigenetic transgenerational inheritance of altered stress responses. *Proc Natl Acad Sci USA* 109: 9143–9148.
Discovering DNA methylation differences with M&M

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12: 996–1006.

Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, Figueroa ME, Glass JL, Chen Q, Montagner C, et al. Comparative isocitrate profiling of cytosine methylation: The HELP assay. Genome Res 16: 1046–1055.

Kitamura E, Igarashi J, Morohashi A, Hida N, Oinuma T, Nemoto N, Song F, Qin M, Mahmood S, Fujiwara K, Igarashi JR, Lee L, Ye Z, Ngo QM, et al. 2009. Human DNA methylomes at base resolution. Nature 458: 5868–5877.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. 2005. Reduced representation bisulfite sequencing for comparative high-throughput analysis. Genome Res 15: 2719–2728.

Molinaro AM. 2008. MEDME: An experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res 18: 1652–1659.

Prasad T, Wang X, Gray PA, Weiner JA. 2008. A differential developmental pattern of spinal interneuron apoptosis during synaptogenesis: Insights from genetic analyses of the protocadherin-y gene cluster. Development 135: 4153–4164.

Pulligilla C, Feng F, Ishikawa K, Bertuzzo S, Daboub A, Griffith AJ, Fritzsch B, Kelley MW. 2007. Disruption of fibroblast growth factor receptor 3 signaling results in defects in cellular differentiation, neuronal patterning, and hearing impairment. Dev Dyn 236: 1905–1917.

Qian P, Zuo Z, Wu Z, Meng X, Li G, Wu Z, Zhang T, Tan S, Pandey V, Yao Y, et al. 2011. Pivotal role of reduced let-7g expression in breast cancer progression and metastasis. Cancer Res 71: 1643–1647.

Robertson KD. 2005. DNA methylation and human disease. Nat Rev Genet 6: 597–610.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25.

Robinson MD, Statham AL, Speed TP, Clark SJ. 2010. Protocol matters: Which methylome are you actually studying? Epigenomics 2: 587–598.

Singer J. 1988. Cephalometric findings. Am J Orthod Dentofacial Orthop 94: 443–444.

Singh SK, Bhardwaj R, Wilczynska KM, Dumur CI, Kordula T. 2011. A complex of nuclear factor-I-X3 and STAT3 regulates astrocyte and glioma migration through the secreted glycoprotein YKL-40. J Biol Chem 286: 39993–39993.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: Provocative insights from epigenomics. Nat Rev Genet 9: 465–476.

Swain SL, Agrewala JN, Brown DM, Jelley-Gibbs DM, Golche S, Huston G, Jones SC, Kampscherer C, Lee WH, McKinnon KR, et al. 2006. CD4+ T-cell memory: Generation and multi-faceted roles for CD4+ T cells in protective immunity to influenza. Immuno Mol Med 21: 8–22.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Romanov SR, Kozakiewicz BK, Holst CR, Stampfer MR, Hauk K, Tlsty TD. 2011. Pivotal role of reduced let-7g expression in breast cancer progression and metastasis. Cancer Res 71: 1643–1647.

Wolfenden B, Granger-Guerrero C, Dabdoub A, Griffith AJ, Fritzsch B, Kelley MW. 2007. Disruption of fibroblast growth factor receptor 3 signaling results in defects in cellular differentiation, neuronal patterning, and hearing impairment. Dev Dyn 236: 1905–1917.

Wydas J, Melby PI. 1998. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 2: 1136–1143.

Wright OA, Davies JJ, Wittig D, Oakey EE, Haase J, Lam WL, Schubeler D. 2005. Chromosome-wide and promoter-specific analyses identify sites
of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37: 853–862.

Winders BR, Schwartz RH, Bruniquel D. 2004. A distinct region of the murine IFN-γ promoter is hypomethylated from early T cell development through mature naive and Th1 cell differentiation, but is hypermethylated in Th2 cells. *J Immunol* 173: 7377–7384.

Wu Q, Maniatis T. 1999. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97: 779–790.

Wu CY, Demeure CE, Gately M, Podlaski F, Yssel H, Kiniwa M, Delespesse G. 1994. In vitro maturation of human neonatal CD4 T lymphocytes. I. Induction of IL-4-producing cells after long-term culture in the presence of IL-2 or IL-12. *J Immunol* 152: 1141–1153.

Wu Q, Zhang T, Cheng JF, Kim Y, Grimmwood J, Schmutz J, Dickson M, Noonan JP, Zhang MQ, Myers RM, et al. 2001. Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res* 11: 389–404.

Xiao S, Xie D, Cao X, Yu P, Xie M, Chen CC, Musselman M, Xie M, West FD, Lewin HA, et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* 149: 1381–1392.

Xu X, Guo Z, Jiang X, Yao Y, Gao Q, Ding Y, Cao X. 2011. Regulatory dendritic cells program generation of interleukin-4-producing alternative memory CD4 T cells with suppressive activity. *Blood* 117: 1218–1227.

Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, et al. 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 131: 1109–1123.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149: 1368–1380.

Zhang Y, Liu T, Meyer CA, Eckehoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137.

Zhou X, Marticque B, Xie M, Li D, Sundaram V, Martin EA, Koebbe BC, Nielsen C, Hirst M, Farnham P, et al. 2011. The Human Epigenome Browser at Washington University. *Nat Methods* 8: 989–990.

Zhou X, Lowdon RF, Li D, Lawson H, Madden P, Costello JE, Wang T. 2013. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat Methods* 10: 375–376.

Received February 18, 2013; accepted in revised form June 13, 2013.