



Published in final edited form as:

Nature. 2015 July 9; 523(7559): 212–216. doi:10.1038/nature14465.

## Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation

Matthew D. Schultz<sup>1,2,13,\*</sup>, Yupeng He<sup>1,2,\*</sup>, John W. Whitaker<sup>3,14</sup>, Manoj Hariharan<sup>2</sup>, Eran A. Mukamel<sup>4,5</sup>, Danny Leung<sup>6</sup>, Nisha Rajagopal<sup>6</sup>, Joseph R. Nery<sup>2</sup>, Mark A. Urich<sup>2</sup>, Huaming Chen<sup>2</sup>, Shin Lin<sup>7</sup>, Yiing Lin<sup>8</sup>, Inkyung Jung<sup>6</sup>, Anthony D. Schmitt<sup>6</sup>, Siddarth Selvaraj<sup>1</sup>, Bing Ren<sup>6,9</sup>, Terrence J. Sejnowski<sup>4,10,11</sup>, Wei Wang<sup>3,12</sup>, and Joseph R. Ecker<sup>2,11,#</sup>

<sup>1</sup>Bioinformatics Program, University of California, San Diego, La Jolla, California 92093, USA

<sup>2</sup>Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

<sup>3</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA

<sup>4</sup>Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

<sup>5</sup>Department of Cognitive Science, University of California, San Diego, La Jolla, California 92037, USA

<sup>6</sup>Ludwig Institute for Cancer Research, La Jolla, California 92093, USA

<sup>7</sup>Department of Genetics, Stanford University, 300 Pasteur Drive, M-344 Stanford, California 94305, USA

<sup>8</sup>Department of Surgery, Washington University School of Medicine, 660 S. Euclid Ave., Campus Box 8109, St. Louis, Missouri 63110, USA

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

#Correspondence and requests for materials should be addressed to J.R.E (ecker@salk.edu)..

\*These authors contributed equally to this work

### Author Contributions

B.R., T.J.S., W.W. and J.R.E designed and supervised research. S.L. and Y.L collected tissues. J.R.N. and M.A.U conducted MethylC-seq, RNA-seq, and genome sequencing experiments. D.L conducted ChIP-seq experiments. N.R performed ChIP-seq data analysis. M.D.S., Y.H., M.H and H.C performed sequencing data processing. J.W.W. performed motif prediction and mutation analysis. M.D.S. designed and implemented the methylation processing and analysis module. M.D.S, Y.H., J.W.W., M.H. and E.A.M performed statistical and bioinformatic analyses. M.D.S., Y.H., J.W.W. and J.R.E prepared the manuscript.

<sup>13</sup>Present Address: Human Longevity Inc., La Jolla, California 92121, USA

<sup>14</sup>Present Address: Discovery Science, Janssen Pharmaceutical of Johnson & Johnson, La Jolla, California, 92121, USA

The authors declare no competing financial interests.

### Author Information

The sequencing data sets generated for this study as well as those for the IMR90, H1, and H1 derived samples can be found at GEO under the accession number GSE16256. The sequencing data sets for the fetal tissues used in this study can be found at GEO under the accession number GSE18927. The sequencing data sets for the placental tissue used in this study can be found at GEO under the accession number GSE39777. The sequencing data sets for the neuronal and glial samples can be found at GEO under the accession number GSE47966 (NRN GSM1173776; GLA GSM1173777). The human tissue sequencing data generated for this study can be found at SRA under the project number SRP000941. Analyzed data sets can be obtained from [http://neomorph.salk.edu/human\\_tissue\\_methylomes.html](http://neomorph.salk.edu/human_tissue_methylomes.html).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature)

<sup>9</sup>University of California, San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, La Jolla, California 92093

<sup>10</sup>Division of Biological Sciences, University of California at San Diego, La Jolla, California 92037, USA.

<sup>11</sup>Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

<sup>12</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA

## Summary

Understanding the diversity of human tissues is fundamental to disease and requires linking genetic information, which is identical in most of an individual's cells, with epigenetic mechanisms that could play tissue-specific roles. Surveys of DNA methylation in human tissues have established a complex landscape including both tissue-specific and invariant methylation patterns<sup>1,2</sup>. Here we report high coverage methylomes that catalogue cytosine methylation in all contexts for the major human organ systems, integrated with matched transcriptomes and genomic sequence. By combining these diverse data types with each individual's phased genome<sup>3</sup>, we identified widespread tissue-specific differential CG methylation (mCG), partially methylated domains, allele-specific methylation and transcription, and the unexpected presence of non-CG methylation (mCH) in almost all human tissues. mCH correlated with tissue-specific functions, and using this mark, we made novel predictions of genes that escape X-chromosome inactivation in specific tissues. Overall, DNA methylation in multiple genomic contexts varies substantially among human tissues.

---

To better understand the variability of DNA methylation across human tissues, we obtained post-mortem samples of 18 tissue types from 4 individuals (5 singletons, 8 duplicates, and 5 triplicates; Fig. 1a; Methods; Supplementary Table 1) and performed deep transcriptome (36 mRNA-seq samples; 120-475 million reads per sample), base-resolution methylome (36 MethylC-seq<sup>4</sup> samples; 30x-80x genome coverage per sample), and genome sequencing (4 whole genome sequences; 20x-45x genome coverage per sample). We focused our initial analysis on cytosines in the CG context and used a previously published method<sup>2</sup> to identify differential methylation (Methods). We found that 15.4% (4,073,896 of 26,474,560 sites tested) of CG sites in these experiments are strongly differentially methylated (DMS; minimum methylation difference = 0.3; Extended Data Fig. 1a), which is similar to a previous study<sup>2</sup>. To identify differentially methylated regions (DMRs), we combined sites within 500bp of one another and found 1,198,132 DMRs. Even with these stringent criteria, 719,837 (60.1%) of the DMRs we identified were novel<sup>2,5</sup>.

As expected, hypomethylation at DMRs correlated with tissue-specific functions<sup>2,6</sup>. For example, strongly hypomethylated DMRs in aorta overlap with aorta-specific super enhancers<sup>7</sup> around MYH10, a gene involved in blood vessel function<sup>8</sup> (Fig. 1b). To further validate our DMRs, we performed hierarchical clustering on their weighted methylation levels<sup>9</sup> (Methods; Fig. 1c; Extended Data Fig. 1b, c). Tissues that were part of the same

organ system clustered together (e.g., heart and muscle tissues). We compared these results to a clustering of differentially expressed genes identified in the transcriptomes and found a similar separation of organ systems (Methods; Fig. 1d; Extended Data Fig. 1d). Furthermore, GREAT<sup>10</sup> analysis on the most hypomethylated tissue-specific DMRs revealed many tissue-specific functions (Extended Data Fig. 1e, f; Methods; Supplementary Information; Supplementary Table 2-3).

To examine the relationship between methylation and transcription, we correlated the methylation levels of DMRs and the expression of the closest genes (Fig. 2a; Extended Data Fig. 2a, b; Methods). As expected, methylation in DMRs had a negative correlation with expression, and this correlation grew stronger closer to the transcription start site (TSS). The strongest negative correlation was not in gene promoters but downstream of the promoter up to 8kb away (intragenic vs. promoter median spearman correlation coefficient (SCC) difference -0.12; Mann-Whitney P-value 6.7e-17; Fig. 2a). This analysis shows that transcription is strongly associated with intragenic DMRs in the tissues we examined, extending similar observations in cancer methylomes<sup>11</sup>.

These intragenic methylation differences have previously been hypothesized to mark intragenic CG islands (CGIs) or CGI shores<sup>5,12-14</sup>. However, only a small fraction of intragenic DMRs fell in these features (19%; Extended Data Fig. 2c). In addition, predicted enhancers and putative promoters only accounted for 23% and 22% of intragenic DMRs, respectively, suggesting that the remaining DMRs, which we call undefined intragenic DMRs (uiDMRs), represent an unrecognized set of functional elements (35%; Extended Data Fig. 2c; Supplementary Information; Methods). The methylation level of these uiDMRs correlated strongly with the expression of the genes containing them. To examine their regulatory potential, we plotted their histone modification profiles (H3K4me1, H3K4me3, H3K27ac, H3K9me3, H3k27me3 and H3K36me3) derived from the same tissue samples<sup>15</sup> and found five classes: weak enhancer, promoter-proximal, transcribed, poised enhancer and unmarked. (Extended Data Fig. 2d-h, Extended Data Fig. 3a, b; Methods). Classes with strong, active histone modifications were moderately negatively correlated with expression (weak enhancer and proximal promoter uiDMRs; median SCC -0.31 and -0.16, respectively); whereas, uiDMRs with less active histone modifications exhibited a weak negative correlation (transcribed and poised enhancer uiDMRs). Notably, the correlation between expression and methylation at promoter-proximal uiDMRs was as strong as the correlation with intragenic DMRs that overlapped strong promoters (Extended Data Fig. 4; Methods), indicating that intragenic promoter and promoter-proximal sequences are more predictive of changes in methylation than those enriched for enhancer-like chromatin modifications.

In contrast, unmarked uiDMRs showed a weakly positive correlation with expression (Extended Data Fig. 4d). Interestingly, we found many of the motifs in tissue-specific uiDMRs were present in tissue-specific enhancers (e.g., HNF4a<sup>16</sup> in liver-specific uiDMRs), suggesting that these DMRs are tissue-specific regulatory elements (Methods; Supplementary Table 4-5). Recently, hypomethylated regions that appear inactive in adult tissues but active during fetal development were identified in mice<sup>6</sup>. We examined the DNase I hypersensitivity profiles of unmarked uiDMRs in matched fetal tissues<sup>17</sup> and found

an enrichment of hypersensitivity (Extended Data Fig. 5; Supplementary Table 6), suggesting that hypomethylation of inactive DMRs can be maintained at regions active earlier in development.

We next examined whether variation in methylation is associated with genetic variation across individuals, which has not been widely characterized in healthy primary tissues or using whole genome bisulfite sequencing<sup>18,19</sup>. To identify individual-specific DMRs, we used a method<sup>20</sup> that is sensitive to these differences unlike the methodology employed above (Methods). We first restricted our analysis to our triplicated samples and ranked DMRs by a tissue-specific methylation outlier score (MOS). We found a ~1.6-fold enrichment of SNPs associating with methylation changes in the top 2,500 MOS ranked DMRs in all tissues (Methods). We then used the Epigram pipeline<sup>21</sup> to predict tissue-specific methylation from DNA motifs in these DMRs and found them highly predictive (average area under the curve (AUC) 0.79; Methods). These full models used an average of 156 motifs; however, an average AUC of 0.74 was achieved using only 20 core TF motifs per tissue.

We then identified groups of corresponding motifs by clustering the sets of tissue-specific motifs (Methods). The motif groups were clustered by their tissue hypo- and hypermethylation specificities (Fig. 2b). 42 of 95 motifs only had hypomethylation specificity; for example, MEIS, which is involved in heart development<sup>22</sup>, is hypomethylated in left ventricle, right atrium and right ventricle. We also identified 34 motifs with tissue-dependent methylation specificity. Three of these motifs match TF families (FOX, HOX and GATA) and are most significantly enriched in hypomethylated regions, suggesting they are primarily involved in regulating hypomethylation.

Mammalian cells have high genome-wide levels of mCG, with the exception of a cultured human fetal fibroblast cell line (IMR90)<sup>4</sup>, cancer cells<sup>23,24</sup> and placenta (PLA)<sup>25</sup>. Surprisingly, large regions of the pancreatic methylomes (PA-2 and PA-3) were significantly hypomethylated (Extended Data Fig. 6a). We developed a method to identify PMDs genome-wide (Supplementary Tables 7-8; Methods) and found pancreatic PMDs were smaller than those in IMR90 and PLA (Extended Data Fig. 6b) and covered a smaller fraction of the genome (Fig. 2c). All pairs of PMDs overlapped significantly indicating that these regions are largely shared (>40% overlap; P-value < 0.001; Extended Data Fig. 6c).

Genes in samples with PMDs are transcriptionally repressed<sup>25,26</sup>, but these regions also show reduced expression in all of the tissues we surveyed whether or not a PMD is present (Fig. 2d). In both IMR90 and PA-2, these regions showed an enrichment in repressive modifications (H3K27me3 and H3K9me3; median difference 0.025 – 0.168 RPKM (reads per kilobase per million); Mann-Whitney P-value < 2.51e-161) and a depletion in active modifications (H3K4me1, H3K27ac, and H3K36me3; median difference 0.050 – 0.012 RPKM; Mann-Whitney P-value < 2.03e-53) compared to shuffled regions (Fig. 2e, f; Extended Data Fig. 6 d, e; Methods), which provides a potential mechanism for their repression. To try to account for this global hypomethylation, we plotted the expression levels of DNMT1, DNMT3A, DNMT3B and DNMT3L but found no systematic expression difference between samples with and without PMDs (Extended Data Fig. 7 a-d).

Previous studies have highlighted the existence of methylation outside of the CG context (mCH) in human embryonic stem cells<sup>4</sup>, brain<sup>1,20</sup> and at the promoter of the PGC-1 $\alpha$  gene in skeletal muscle<sup>27</sup>. We found evidence for appreciable amounts of mCH in many of these tissues (Fig. 3a; Extended Data Fig. 8a). A 5bp motif split the samples into two groups, one with mCH enriched in a TNCAC motif and another with mCH enriched in an NNCAN motif (where N is any base) (Methods). The TNCAC motif is highly similar to the one previously identified in purified glia (GLA) and neurons (NRN) (TACAC). These motifs are significantly different than the motif found in H1 embryonic stem cells (H1) and induced pluripotent stem cells (TACAG)<sup>4,26</sup> (Fig. 3b-d). We quantified the extent of mCH across these samples by plotting the distribution of methylation levels at mCH sites in the 25 samples with a TNCAC motif, which revealed a methylation level similar to that of GLA, NRN and H1 (Extended Data Fig. 8b)<sup>4,20</sup>. Most of the tissue types were consistently enriched for the TNCAC or NNCAN motif, but several (esophagus, lung, pancreas and spleen) had replicates which disagreed, suggesting that mCH is not homogeneously distributed across these tissues.

To examine the potential functional effect of mCH in adult tissues, we plotted the distribution of expression levels for various quantiles of gene body mCH as it was previously reported to be positively correlated with expression in H1<sup>4</sup> and negatively correlated with expression in neurons<sup>20</sup>. This analysis revealed a negative correlation between expression and mCH (Extended Data Fig. 8c; Methods). Next, we combined our replicates and clustered genes by the patterns of CAS methylation (where S is a G or C) in and around their gene body (Fig. 3e; Methods; Supplementary Information). To characterize the genes assigned to each cluster, we performed DAVID functional annotation clustering (Supplementary Table 9; Methods), which revealed several different classes. Clusters 1, 2, 11, 16 and 19 contained genes highly enriched for terms involved in basic cellular processes and had an active methylation state (i.e., hypermethylation in embryonic samples and hypomethylation in tissue and brain samples) across all samples. Clusters 5 and 6 were dominated by terms related to neuronal function and genes in this class were differentially methylated between neurons and glia and have inactive methylation states in other samples (i.e., hypomethylation in embryonic samples and hypermethylation in tissue and brain samples). Cluster 12 was enriched for heart and muscle related terms and its genes had an active methylation state in the three heart tissues as well as a weakly active methylation state in psoas but appeared inactive in other samples. Lastly, cluster 14 possessed an active methylation state in brain and tissue samples but were inactive in embryonic samples. Despite being inactive in the H1 samples, this class of genes was highly enriched for terms related to development.

To better define the transition of mCH motifs over development, we examined the ratio of the methylation level of CAC and CAG (mCAC and mCAG) sites in a variety of differentiated (tissues, NRN, and GLA), embryonic (H1), and embryonic derived cells (neural progenitor cells, NPC; mesendoderm MES; trophoblast-like TRO; mesenchymal stem cells, MSC)<sup>28</sup> samples (Fig. 3f). With the exception of brain cells, mCH levels drop during differentiation, and the mCAC/mCAG ratios revealed a shift in motif usage across developmental time (Fig. 3f); although, mCAC and mCAG within the same gene remain

tightly correlated in both early embryonic and differentiated tissues (Extended Data Fig. 8d, e).

Methylation has previously been shown to be predictive of genes escaping X chromosome inactivation (XI) in neurons<sup>20</sup>. We investigated this phenomenon in these samples by comparing the promoter mCG and gene body mCH of genes that had previously been identified to escape X chromosome inactivation<sup>29</sup> in 11 tissues with mCH (Fig. 4a). Female-specific promoter mCG hypomethylation and gene body mCH hypermethylation was present at escapee genes at a similar level as in neurons (Extended Data Fig. 9a)<sup>20</sup>. Utilizing these tissue methylomes, gene body mCH was appreciably predictive of biallelically expressed genes (AUC 0.89; Extended Data Fig. 9b; Methods). To a lesser extent, we observed female-specific promoter mCH and gene body mCG hypermethylation at escapee genes (Extended Data Fig. 9a, c, d). Although female-specific promoter mCG hypomethylation, promoter mCH hypermethylation and gene body mCG hypermethylation are somewhat predictive of XI escapees, female-specific gene body mCH hypermethylation is the most predictive feature of XI escapees (Extended Data Figure 9a, b-e). We detected significant female-specific mCH hypermethylation in 109 of 612 X-linked genes, including 9 genes hypermethylated in all 11 tissues and 72 genes that were significantly hypermethylated in only one tissue (Fig. 4b). Several genes such as *FUNDC1* showed female-specific hypermethylation in several tissues but not in neurons, suggesting a tissue-dependent regulation of the escape from X inactivation.

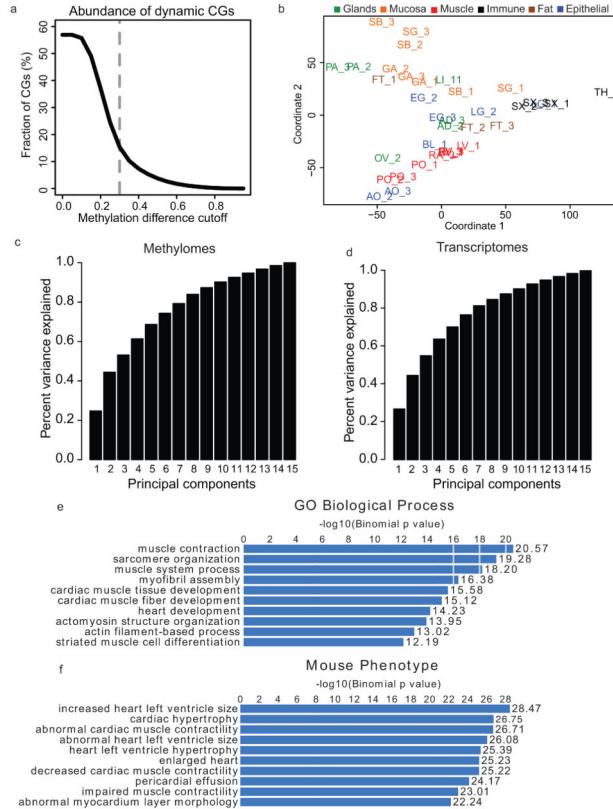
Allele-specific methylation (ASM) and expression (ASE) may also play a role in the regulation of autosomal genes. To examine these phenomena in human tissues, we combined the RNA-seq and MethylC-seq data sets with phased genotypes for each individual in this study<sup>3,15</sup> (Extended Data Fig. 10a; Methods). Using the triplicate tissue samples (FT, GA, PO, SB, and SX), we identified 8,464 - 48,560 ASM events in the CG context and 48 - 403 ASE genes across these tissues (Supplementary Table 10-11; Methods). We next looked for ASM events that varied across individuals within a tissue-type (tissue variable) and those that varied across a tissue-type within an individual (individual variable). Of the ASM events that varied, 4.1 - 7.5% and 54.5 - 70.0% were individual- and tissue-variable, respectively; whereas, of the ASE events that varied, 0.0 - 20.0% were individual-variable and 13.3 - 48.8% were tissue-variable (Fig. 4c; Methods). Of the ASE events, 38.4 - 87.4% had an ASM event within 100 kilobases, and of these sites, 76% had an ASM and ASE event that was matched (i.e., a DMR was hypomethylated on the same haplotype as the more highly expressed allele). Furthermore, we found that a larger fraction of ASE genes were observed near ASM events whether or not the events matched (Extended Data Fig. 10 b, c; Methods). These results demonstrate a link between allele specific methylation and expression in human tissues.

Here we have presented the deepest set of base resolution maps of mCG and mCH to date along with chromatin modification states, haplotype-resolved genome sequences and transcriptional profiles for a large set of human tissues. These data sets allowed us to accurately identify *cis*-regulatory elements. Additionally, they revealed the existence of mCH genome-wide in a subpopulation of cells from differentiated human tissues, which appears to be repressive. Our analysis of genic mCH indicates that these genes are distinct



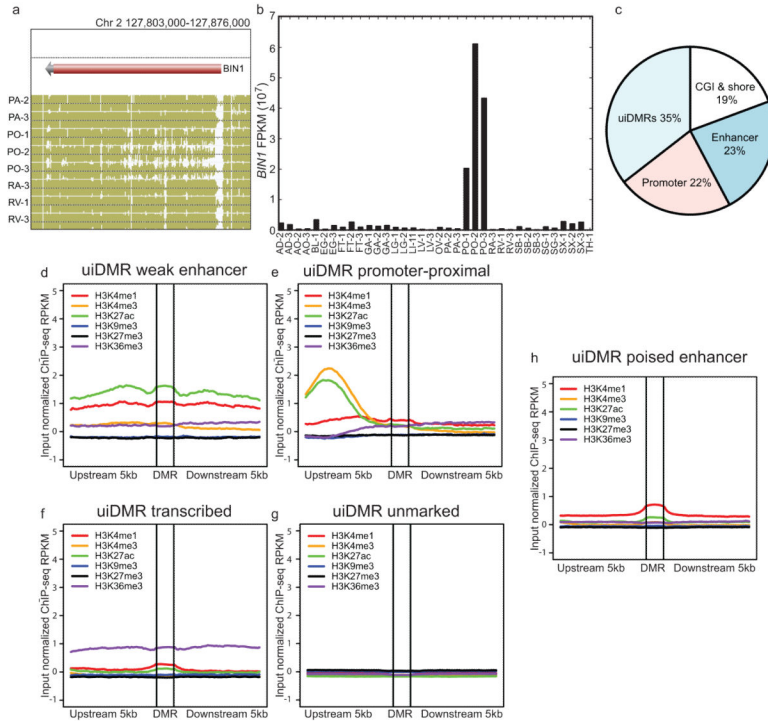
from those that were previously identified in embryonic stem cells and the brain and showed enrichment for a variety of functions, most surprisingly those involved in development. These analyses raise the intriguing possibility that mCH is utilized in adult stem cells<sup>30</sup> and could help to repress these genes as the cells transition into their differentiated role.

## Extended Data

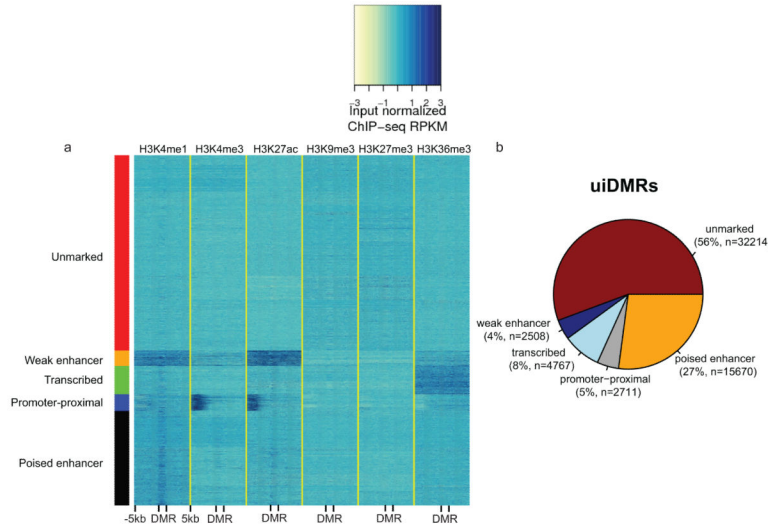


### Extended Data Figure 1. Identification of differentially methylated regions (DMRs) and Multidimensional Scaling Analysis

**a**, Line plot showing the fraction of differentially methylated CG sites (DMSs, dynamic CGs) out of all CG sites under various methylation difference cutoffs. The methylation difference of a CG site is defined in Ziller et al.<sup>2</sup> **b**, A plot of the first two principal components from the methylation level multi-dimensional scaling. Tissues are shaded by the organ group they belong to as in Figure 1c and 1d. **c-d**, Bar charts of the cumulative amount of variance explained by the first N principal components from the multi-dimensional scaling performed on the methylation levels of all DMRs (c) and the expression levels of all differentially expressed genes (d). **e**, A representative example of enriched GO biological process terms based on the most hypomethylated DMRs from LV-1. **f**, A representative example of enriched mouse phenotype terms based on the most hypomethylated DMRs from LV-1.



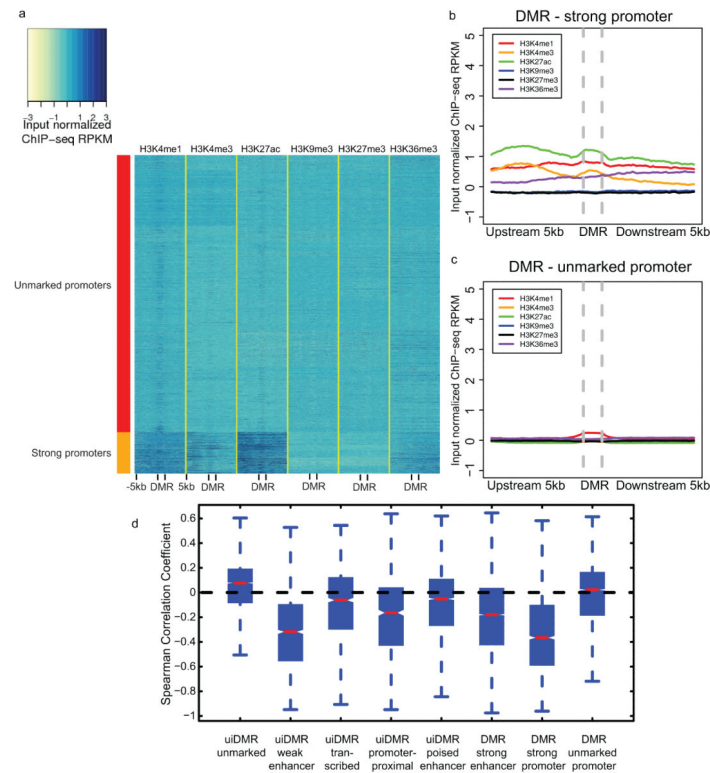
**Extended Data Figure 2. DMRs and their correlation with transcription**  
**a**, A browser screenshot of an example DMR downstream of the TSS. **b**, Expression level of the *BIN1* gene which contains the DMR in (a). **c**, The percentages of hypomethylated intragenic DMRs in each class of genomic features. **c-h**, Histone modification profiles of five categories of uiDMRs.



**Extended Data Figure 3. Classification of uiDMR histone profiles and uiDMR properties**  
**a**, heatmap of the histone modification profiles for the five types of uiDMRs. The profiles were plotted for each mark across the DMR and the 5kb upstream and downstream and the colors of each cell indicate the input normalized ChIP-seq RPKM. The colors on the left

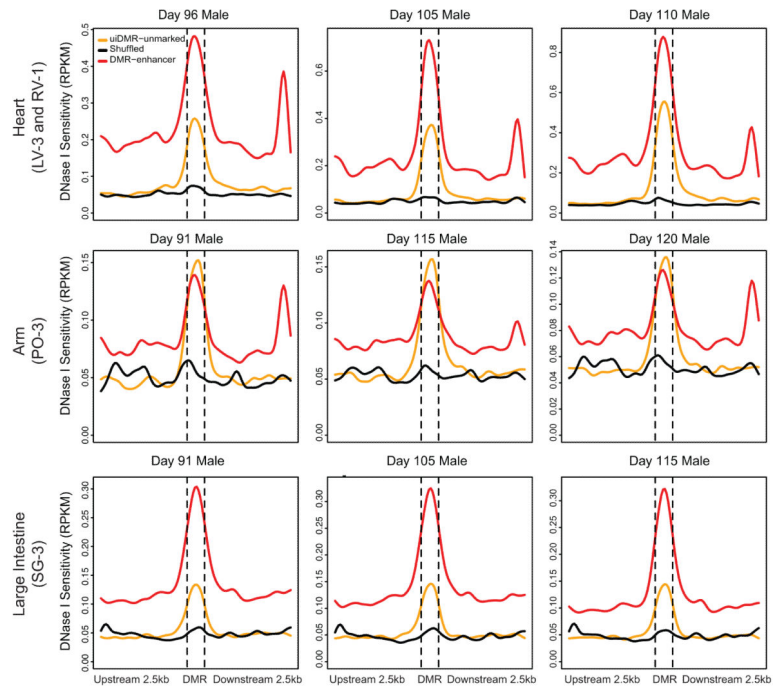


indicate the group of each profile assigned by k-means clustering (red, weak enhancer; orange, promoter-proximal; green, transcribed; blue, unmarked; black poised enhancer). **b**, A pie chart of the distribution of uiDMRs across the classes defined by k-means clustering.



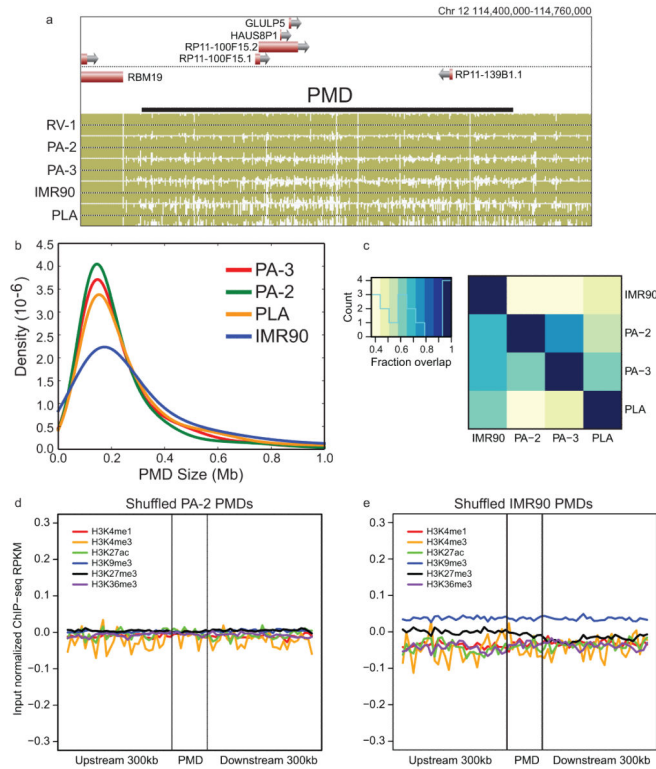
#### Extended Data Figure 4. Classification of promoter histone profiles

**a**, A heatmap of the histone modification profiles across strong (rows labeled with red) and unmarked (rows labeled with orange) promoters. The profiles were plotted for each mark across the promoter and the 5kb upstream and downstream and the colors of each cell indicate the input normalized ChIP-seq RPKM. **b-c**, The aggregate profiles for strong and unmarked promoters (b) and (c), respectively. **d**, The distribution of the Spearman correlation coefficients between the methylation level of different types of hypomethylated intragenic DMRs and the expression of the nearest gene. Notches indicate a confidence interval estimated from 1,000 bootstrap samples.



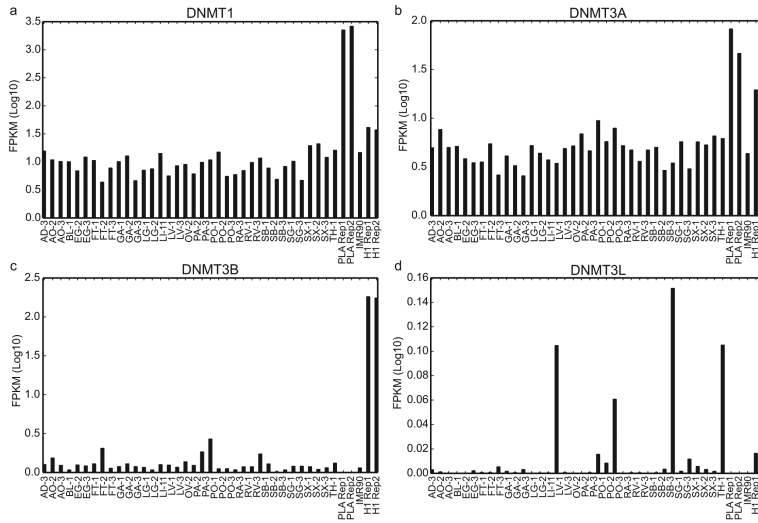
#### Extended Data Figure 5. uiDMR fetal DNase I profiles

DNase I profiles of various fetal tissues corresponding to the tissues presented in this study. The samples are arranged columnwise by age, and row-wise by fetal tissue. The uiDMR – unmarked line represents the DNase I profile of uiDMRs without histone modifications. The DMR – enhancer line represents the DNase I profile of DMRs that overlapped an enhancer in a matched tissue in this study (indicated in the row label in parentheses). The shuffled line represents the DNase I profile of uiDMRs randomly shuffled across the genome.



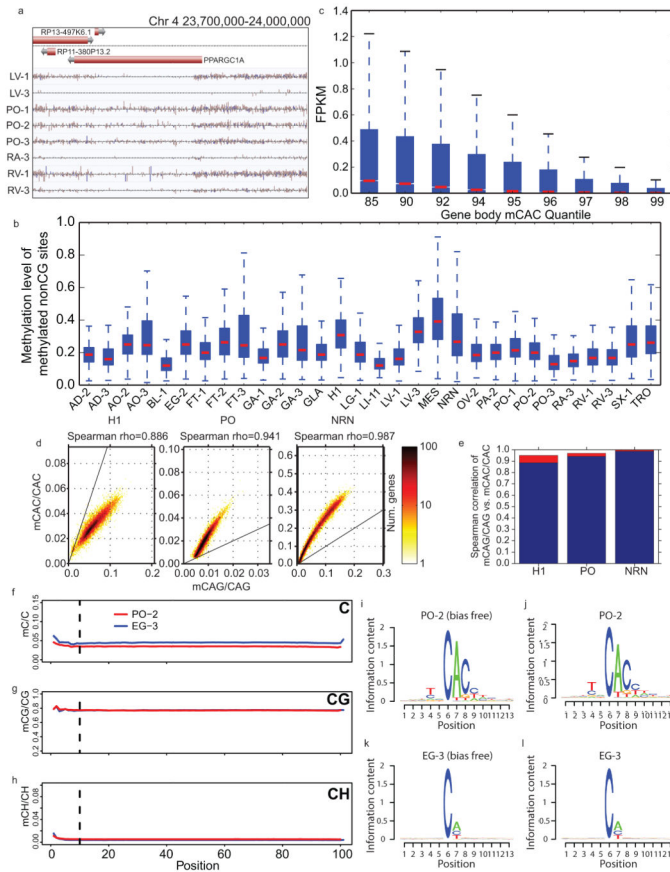
**Extended Data Figure 6. PMD Features**

**a**, A browser screenshot (see Figure 1 for description) of an example PMD found in IMR90, PLA, PA-2, and PA-3. RV-1 is included as a representative sample without PMDs. **b**, The distribution of sizes of PMDs in various samples. **c**, A heatmap representation of the overlap between various sets of PMDs. The denominator of the fraction of overlap is determined by the sample on the y-axis. **d-e**, ChIP-seq profiles of the PMD regions defined in PA-2 (**c**) and IMR90 (**d**) after shuffling.



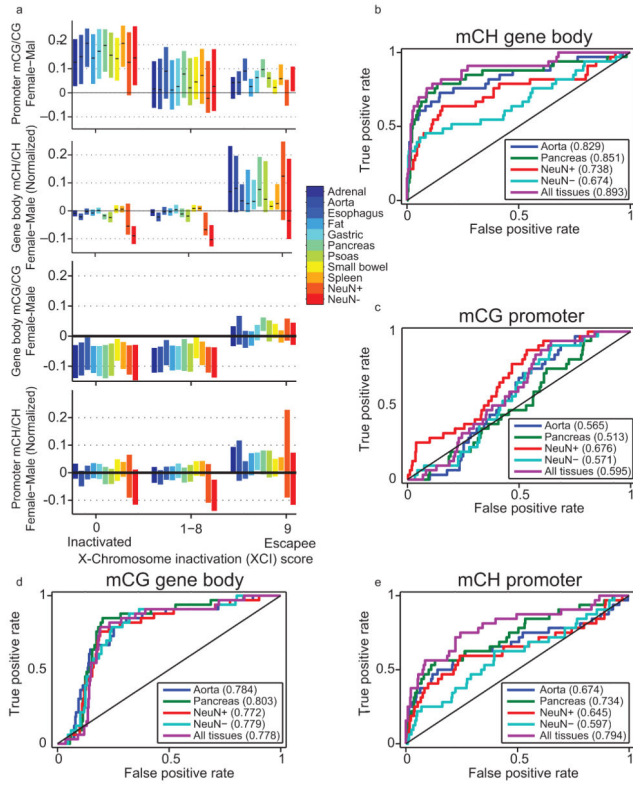
**Extended Data Figure 7. DNMT expression across tissues**

**a-d**, Bar plots of the expression (measured in log<sub>10</sub> FPKMs) of DNMT1 (a), DNMT3A (b), DNMT3B (c), and DNMT3L (d) across various samples.



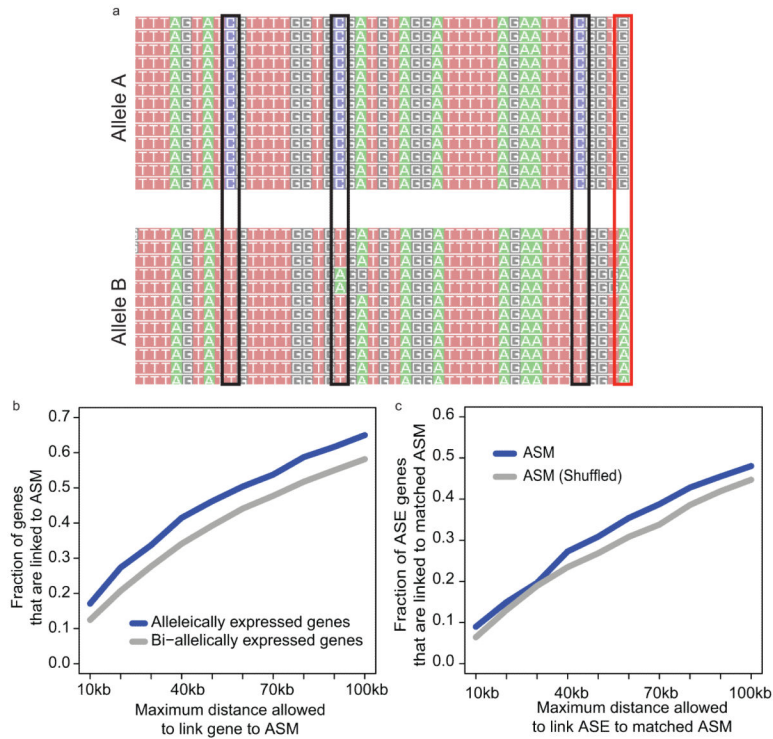
### Extended Data Figure 8. mCH distribution and correlation

**a**, A browser screenshot (see Figure 1 for description) of an example region with non-CG methylation (mCH). Purple and pink ticks are methylated CHG and CHH sites, respectively (H = A, C, or T). Ticks on the forward strand are projected upward from the dotted line and ticks on the reverse strand are projected downward. **b**, The distribution of methylation levels at mCH sites across all samples with a discernible TNCAC motif. Only mCH sites with at least 10 reads and a significant amount of methylation were considered. **c**, Boxplots of the expression values across different quantiles of CAC gene body methylation (Gene body mCAC). **d**, Scatterplot of mCAG vs. mCAC inside gene bodies. **e**, Bar plot of the correlation of mCAG and mCAC inside gene bodies (blue) and the theoretical maximal correlation (red) if mCAC and mCAG are perfectly correlated. **f-h**, The methylation levels of C (upper panel), CG (middle panel) and CH (lower panel) across the read positions for PO-2 (red line) and EG-3 (blue line). Vertical lines indicate the position (10<sup>th</sup> base from the beginning) where trimming was applied. **i**, mCH motif from PO-2 with the first 10 bases of each read trimmed. **j**, mCH motif from PO-2 without trimming. **k**, mCH motif from EG-3 with the first 10 bases of each read trimmed. **l**, mCH motif from EG-3 without trimming. The height of each letter represents its information content (i.e., prevalence).



**Extended Data Figure 9. X chromosome inactivation**

**a**, Distributions of promoter CG methylation (mCG) levels (mCG/CG), gene body non-CG methylation (mCH) levels (mCH/CH), gene body mCG levels **and** promoter mCH levels in genes previously reported to express from only one allele (inactivated) or biallelically (escapee)<sup>63</sup>. Black ticks show median, and bars indicate 25-75<sup>th</sup> percentile range. Genes more prone to escaping inactivation have lower promoter mCG, higher gene body mCH, higher gene body mCG and higher promoter mCH in females. **b-e**, Discriminability analysis using **b**, gender-specific gene-body mCH, **c**, promoter mCG, **d**, promoter mCH and **e**, gene body mCG to predict the escapee status of X-linked gene, respectively. Among them, gene body mCH is the most predictive feature of chromosome X inactivation escapees.



### Extended Data Figure 10. Allele-specific Methylation and Expression

**a**, An example of allele-specific methylation (ASM). Reads that contain a heterozygous SNP (red box) are separated by allele. The number of methylated (reads containing Cs) and unmethylated (reads containing Ts) at adjacent CG sites (black boxes) and tested for differential methylation. **b**, Fraction of allele-specific expressed (ASE) genes (blue) and bi-allelically expressed genes (grey) that have at least one ASM event within a certain distance. Bi-allelically expressed genes were defined as genes that were covered by at least 10 reads and whose p-values given by binomial test for allelic expression were greater than 0.2 (i.e. no significance). **c**, Fraction of ASE genes that were linked to matched ASM event(s) (blue) and matched ASM events with their locations shuffled (grey). b-c are aggregated results using samples from triplicate tissues.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank R. J. Schmitz for critical reading of the manuscript. This work is supported by the NIH Epigenome Roadmap Project (U01 ES017166). E.A.M. was supported by National Institute of Neurological Diseases and Stroke grant (K99NS080911). J.R.E was supported by the Gordon and Betty Moore Foundation (GMBF3034). T.J.S and J.R.E are investigators of the Howard Hughes Medical Institute. S.L was supported by NIH fellowship grants F32HL110473 and K99HL119617. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper. URL: <http://www.tacc.utexas.edu>. The authors would also like to thank Mid-America Transplant Services, St. Louis, MO for their support of this research effort.



## References

1. Varley KE, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 2013; 23:555–567. [PubMed: 23325432]
2. Ziller MJ, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature.* 2013; 500:477–481. [PubMed: 23925113]
3. Selvaraj S, Dixon JR, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* 2013; 31:1111–1118. [PubMed: 24185094]
4. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009; 462:315–322. [PubMed: 19829295]
5. Irizarry RA, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet.* 2009; 41:178–186. [PubMed: 19151715]
6. Hon GC, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet.* 2013; 45:1198–1206. [PubMed: 23995138]
7. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–947. [PubMed: 24119843]
8. Yuen SL, Ogut O, Brozovich FV. Nonmuscle myosin is regulated during smooth muscle contraction. *Am J Physiol Heart Circ Physiol.* 2009; 297:H191–H199. [PubMed: 19429828]
9. Schultz MD, Schmitz RJ, Ecker JR. ‘Leveling’ the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* 2012; 28:583–585. [PubMed: 23131467]
10. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
11. Hovestadt V, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature.* 2014 doi:10.1038/nature13268.
12. Maunakea AK, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010; 466:253–257. [PubMed: 20613842]
13. Doi A, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet.* 2009; 41:1350–1353. [PubMed: 19881528]
14. Deaton AM, et al. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* 2011; 21:1074–1086. [PubMed: 21628449]
15. Leung D, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature.* 2015; 518:350–354. [PubMed: 25693566]
16. Parviz F, et al. Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat Genet.* 2003; 34:292–296. [PubMed: 12808453]
17. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–1195. [PubMed: 22955828]
18. Gutierrez-Arcelus M, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife.* 2013; 2:e00523. [PubMed: 23755361]
19. Liu Y, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol.* 2013; 31:142–147. [PubMed: 23334450]
20. Lister R, et al. Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science.* 2013
21. Whitaker JW, Chen Z, Wang W. Predicting the human epigenome from DNA motifs. *Nat. Methods.* 2014; 12:265–272. [PubMed: 25240437]
22. Stankunas K, et al. Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ Res.* 2008; 103:702–709. [PubMed: 18723445]
23. Hon GC, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 2012; 22:246–258. [PubMed: 22156296]
24. Berman BP, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet.* 2012; 44:40–46. [PubMed: 22120008]

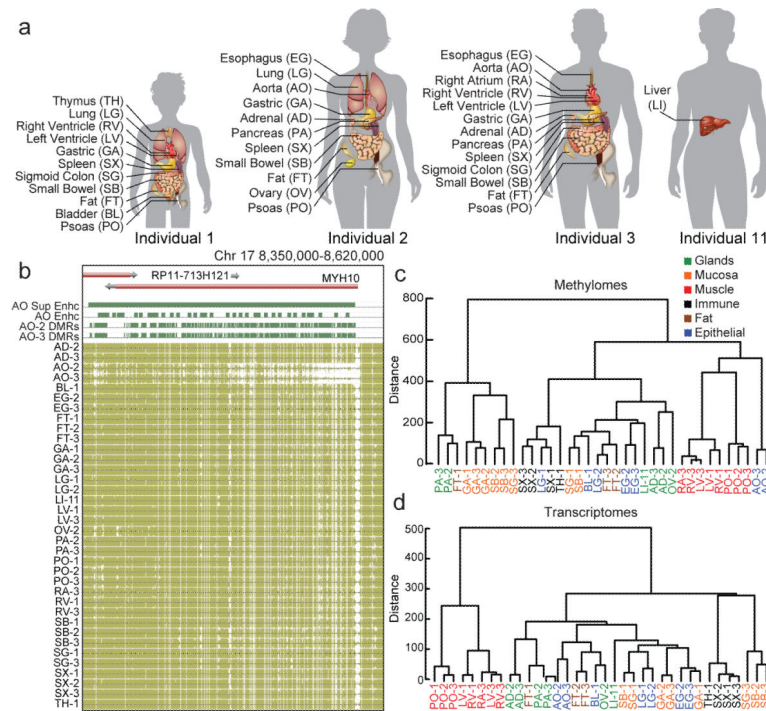
25. Schroeder DI, et al. The human placenta methylome. *Proc Natl Acad Sci U A*. 2013; 110:6037–6042.
26. Lister R, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*. 2011; 471:68–73. [PubMed: 21289626]
27. Barrès R, et al. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. *Cell Metab*. 2009; 10:189–198. [PubMed: 19723495]
28. Xie W, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013; 153:1134–1148. [PubMed: 23664764]
29. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005; 434:400–404. [PubMed: 15772666]
30. Wagers AJ, Weissman IL. Plasticity of adult stem cells. *Cell*. 2004; 116:639–648. [PubMed: 15006347]

Author Manuscript

Author Manuscript

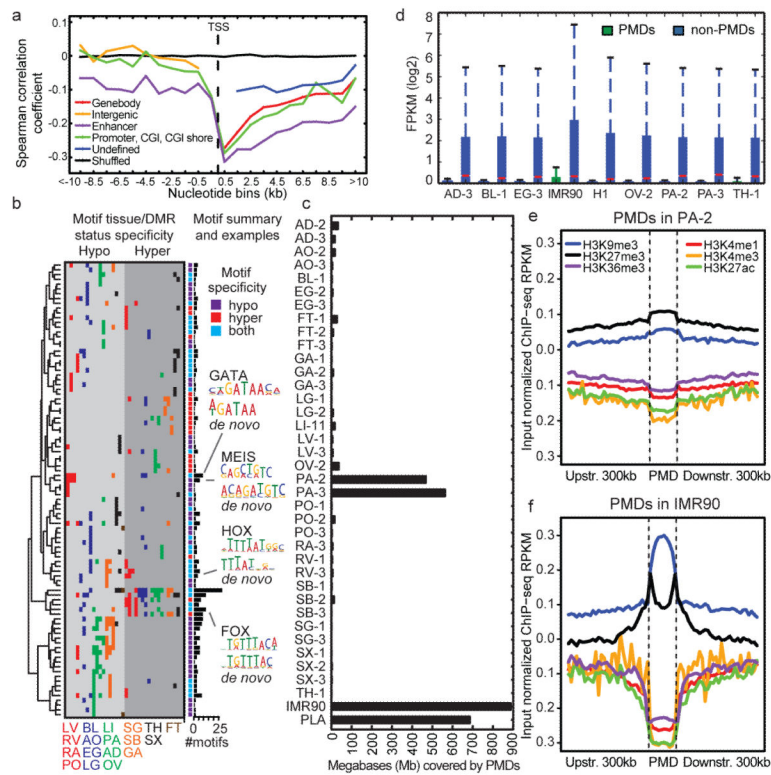
Author Manuscript

Author Manuscript



**Figure 1. The methylomes and transcriptomes of human tissues**

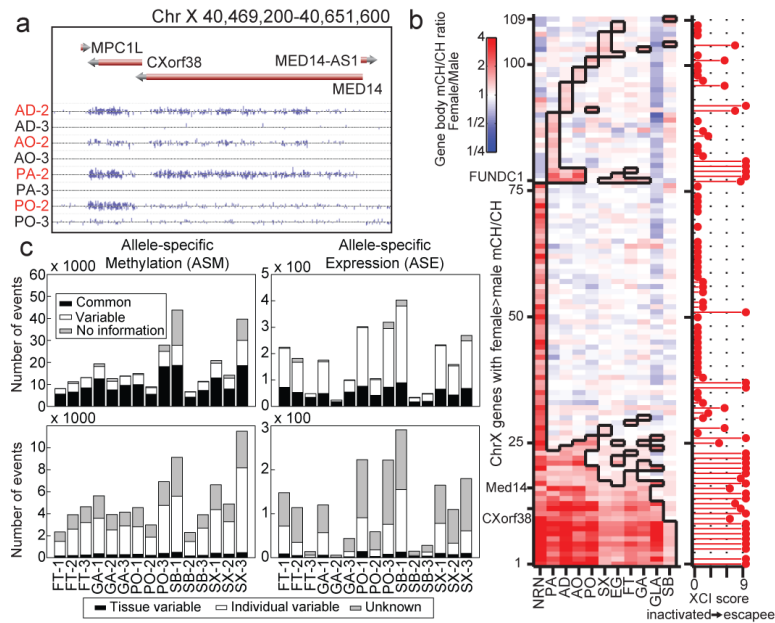
**a**, The tissues analyzed in this study. Samples are denoted by the two letter code in parentheses followed by an individual ID. **b**, Browser screenshot of an example DMR. The top track contains gene models. The following four tracks contain green blocks indicating the location of super enhancers, enhancers, and hypomethylated DMRs in aorta, respectively. The remaining tracks display methylation data from each sample. Gold ticks are CG sites with heights proportional to their methylation level. Ticks on the forward and reverse strand are projected upward and downward from the dotted line, respectively. **c-d**, Hierarchical clustering of DMR methylation levels (c) and expression levels of differentially expressed genes (d). Colors indicate organ systems each sample belongs to.



**Figure 2. DNA methylation and its relationship with gene expression**

**a**, The mean Spearman correlation coefficient at various distances between the methylation level of autosomal DMRs and the expression of the nearest gene. These correlations are shown for DMRs: overlapping genes (Genebody), overlapping enhancers (Enhancer), overlapping promoters or CpG islands (CGIs) or CGI shores (Promoter, CGI, CGI shore), not overlapping genes (Intergenic) and all remaining DMRs (Undefined). **b**, Heatmap showing each motif's tissue-specific methylation preference. The tissues are colored according to Fig 1c., and the ordering is listed at the bottom of the figure. The bar plot at the end of the panel shows the number of times the motif was present in the 20 motif models. **c**, The number of base pairs covered by PMDs in all samples. **d**, The distribution of expression inside and outside of PA-2 PMDs across various samples. Notches indicate a confidence interval estimated from 1,000 bootstrap samples. Each PMD boxplot consists of 3,627 genes and each non-PMD boxplot consists of 22,907 genes. **e-f**, Histone modification profiles in and around PMDs in PA-2 (e) and IMR90 (f).





**Figure 4. Allele-specific Methylation and Expression**

**a**, Browser screenshot of the increase in female mCH for a gene known to escape X chromosome inactivation (MED14). Sample names are colored by gender (male, black; female, red). **b**, Ratio of mCH level in female vs. male samples across genes with a significant difference in at least one sample. Cells boxed in black denote samples with a statistically significant difference between females and males. **c**, The number of ASM and ASE sites across the triplicated tissues. The top row depicts ASM events (left) and ASE events (right) which are allele-specific in all tissues (black), are variable across tissues (white), or do not possess enough data to tell (grey). The bottom row depicts the distribution of variable sites from the top row that vary by individual (white), tissue (black), or neither (grey).