Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press



Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver

Alayne L. Brunner, David S. Johnson, Si Wan Kim, et al.

Genome Res. published online March 9, 2009
Access the most recent version at doi:10.1101/gr.088773.108

P<P Published online March 9, 2009 in advance of the print journal.

Open Access Freely available online through the *Genome Research* Open Access option.

License Freely available online through the Genome Research Open Access option.

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



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

Copyright © 2009 by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver

Alayne L. Brunner,[1,5] David S. Johnson,[1,5,6] Si Wan Kim,[1,5] Anton Valouev,[2] Timothy E. Reddy,[3] Norma F. Neff,[1] Elizabeth Anton,[1] Catherine Medina,[1] Loan Nguyen,[1] Eric Chiao,[1] Chuba B. Oyolu,[1] Gary P. Schroth,[4] Devin M. Absher,[3] Julie C. Baker,[1,8] and Richard M. Myers[1,7,8]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; [2]Department of Pathology, Stanford, California 94305, USA; [3]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA; [4]Illumina, Inc., San Diego, California 92121, USA

To investigate the role of DNA methylation during human development, we developed Methyl-seq, a method that assays DNA methylation at more than 90,000 regions throughout the genome. Performing Methyl-seq on human embryonic stem cells (hESCs), their derivatives, and human tissues allowed us to identify several trends during hESC and in vivo liver differentiation. First, differentiation results in DNA methylation changes at a minimal number of assayed regions, both in vitro and in vivo (2%–11%). Second, in vitro hESC differentiation is characterized by both de novo methylation and demethylation, whereas in vivo fetal liver development is characterized predominantly by demethylation. Third, hESC differentiation is uniquely characterized by methylation changes specifically at H3K27me3-occupied regions, bivalent domains, and low density CpG promoters (LCPs), suggesting that these regions are more likely to be involved in transcriptional regulation during hESC differentiation. Although both H3K27me3-occupied domains and LCPs are also regions of high variability in DNA methylation state during human liver development, these regions become highly unmethylated, which is a distinct trend from that observed in hESCs. Taken together, our results indicate that hESC differentiation has a unique DNA methylation signature that may not be indicative of in vivo differentiation.

[Supplemental material is available online at www.genome.org. Raw Methyl-seq sequence data are available from NCBI's Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) (accession no. SRA008154). Processed Methyl-seq data are available for download and visualization at http://genome-test.cse.ucsc.edu/cgi-bin/hgTrackUi?db=hg18&g=wgEncodeHudsonalphaMethylSeq. Gene expression data can be downloaded from NCBI's GEO (http://www.ncbi.nlm.nih.gov/geo/) (accession no. GSE14966). Additional information and files can be found at http://myers.hudsonalpha.org.]

During embryonic development, somatic cells are specified through gene expression programs to give rise to all the tissues in the body. Human embryonic stem cells (hESCs) have become an important tool for studying development because they can be differentiated in vitro toward many specific cellular types, which require a precise program of regulation of gene expression (Fritsch and Singer 2008). DNA methylation at CpG dinucleotides has long been considered a key mechanism of transcriptional regulation (Holliday and Pugh 1975; Razin and Cedar 1991; Reik et al. 2001; Bird 2002; Jaenisch and Bird 2003; Morgan et al. 2005; Nafee et al. 2008; for review see Lister and Ecker 2009). DNA methylation is known to target the inactive X chromosome, transposons, imprinted loci, promoters, and coding regions (Goll and Bestor 2005; Chang et al. 2006; Rollins et al. 2006; Mandrioli 2007; Weber et al. 2007; Sha 2008; Suzuki and Bird 2008). Because of numerous studies showing it associated with heterochromatin, DNA methylation has gained a reputation as a permanent, silencing mark (Jaenisch 1997; Walsh et al. 1998; Bird and Wolffe 1999; Siegfried et al. 1999; Fuks et al. 2003; Hashimshony et al. 2003; Zhang et al. 2005; Reik 2007; Esteller 2008; Tiwari et al. 2008). However, recently, regions of methylated DNA have been correlated with the tissue-specific expression of several genes and with active coding regions across the genome (Kumar and Biswas 1988; Ngo et al. 1996; Grunau et al. 2000; Imamura et al. 2001; Kroft et al. 2001; Kusui et al. 2001; Futscher et al. 2002; Song et al. 2005; Butta et al. 2006; Da et al. 2006; Fujii et al. 2006; Lavelle et al. 2006; Douet et al. 2007; Hellman and Chess 2007; Kitamura et al. 2007; Shen et al. 2007; Zilberman et al. 2007; Yagi et al. 2008). Additionally, rather than a permanent mark, methylation has been shown to be dynamic, capable of temporally changing at gene promoters (Weiss and Cedar 1997; Kangaspeska et al. 2008; Metivier et al. 2008). This complex range of DNA methylation

[5]These authors contributed equally to this work.
Present addresses: [6]Gene Security Network, Inc., Redwood City, CA 94063, USA; [7]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA.
[8]Corresponding authors.
E-mail rmyers@hudsonalpha.org; fax (256) 327-0978.
E-mail jbaker@stanford.edu: fax (650) 725-1534.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.088773.108. Freely available online through the *Genome Research* Open Access option.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

targets has raised new questions about its role in early development, particularly in regulating gene expression during tissue differentiation (Suzuki and Bird 2008).

The global distribution of DNA methylation patterns in human tissues is only beginning to be uncovered, and its role during tissue differentiation in early human development remains uncharacterized (Rakyan et al. 2004; Bibikova et al. 2006; Eckhardt et al. 2006; Keshet et al. 2006; Rollins et al. 2006; Schumacher et al. 2006; Hellman and Chess 2007; Ladd-Acosta et al. 2007; Shen et al. 2007; Weber et al. 2007; Bollati et al. 2008; Illingworth et al. 2008; Rauch et al. 2008; Kaminsky et al. 2009). Deconstructing the functional role for DNA methylation in development has remained a challenge because of the large number of potential CpG targets in the genome (Suzuki and Bird 2008). Here, we present a new technique, Methyl-seq, that assays more than 250,000 methyl-sensitive restriction enzyme cleavage sites, which, when grouped, represent more than 90,000 regions in the nonrepetitive human genome; 35,528 of these regions are within annotated CpG islands, while the remaining 55,084 non-CpG island regions are distributed across the genome in promoters, genes, and intergenic regions. We used Methyl-seq to assay the DNA methylation status of these regions in human ES cells, hESC-derived cells, and fetal and adult liver.

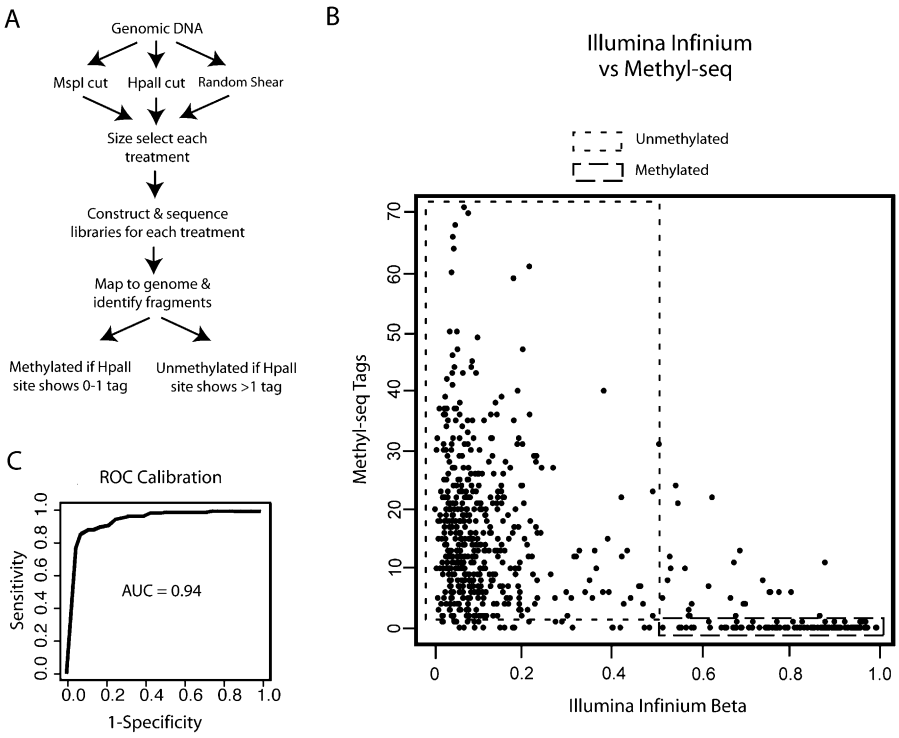


Figure 1. The Methyl-seq method and validation. (*A*) Workflow for the Methyl-seq assay. (*B*) Comparison of the number of Methyl-seq tags and Illumina beta values at 160 CpGs for four samples: HCT116, H9 hESC, H9 endoderm, and adult liver. HCT116 was used in Methyl-seq development and validation. There is good correlation between methylated calls (zero or one Methyl-seq tags and high beta values) and, likewise, unmethylated calls (low beta values and high numbers of sequencing tags) (for binary comparison, see Supplemental Fig. 7; *P*-values $<10^{-8}$). (*C*) Receiver operator characteristic (ROC) curve generated from Methyl-seq and Illumina Infinium data. The area *under* the curve is 0.944.

Results

Methyl-seq as a technique for identifying methylation patterns in the genome

Methyl-seq combines DNA digestion by a methyl-sensitive enzyme with next-generation (next-gen) DNA sequencing technology. It is based upon the historically most robust assay for DNA methylation—methyl-sensitive restriction enzymes (Cedar et al. 1979). Here, we used the isoschizomers HpaII and MspI, which recognize the same cleavage site (5′-CCGG-3′); MspI digests at the site regardless of methylation status, while HpaII digests only unmethylated sites. In Methyl-seq, genomic DNA is digested with HpaII or MspI, ligated to next-gen sequencing adapters, and subjected to size selection to isolate small gel fragments (~100–350 bp, including the adapters, corresponding to genomic fragments ~30–260 bp in length), to generate a fragment library for next-gen sequencing (Fig. 1A; Supplemental Fig. 2; Supplemental Tables 1, 2; Methods). The fragment library is sequenced to obtain millions of short sequence tags that are 25–35 bp in length ($\sim 3 \times 10^6$ tags per HpaII library and 1×10^7 tags for the MspI library). Because MspI digests at both methylated and unmethylated sites where HpaII is blocked from digesting methylated sites, sequence tags present in MspI libraries but not in HpaII libraries are derived from methylated regions. Conversely, sequence tags that occur in HpaII libraries come from at least partially unmethylated regions. Analysis of these sequence tags is described in the Supplemental Material, but briefly, aligned sequence reads are matched to in silico-predicted digestion sites in the genome. Only those digestion sites that have four or more sequenced MspI reads are considered "assayable" and used in the analysis (Supplemental Methods; Supplemental Fig. 5). Additionally, a size bias of the fragments clearly exists within the Illumina sequencing (formerly known as Solexa sequencing) platform, so that reads smaller than 100 bp are preferentially sequenced (see size distribution in Supplemental Fig. 4 and detailed discussion in Supplemental Methods). Instead of sequencing deeper to increase coverage, we chose to focus only on the digestion sites from 35- to 75-bp fragments, whose sequences were enriched in our sequencing reads, to minimize the effects of this bias. By using this approach, we identified 255,266 assayable MspI digestion sites, which we grouped into 90,612 regions based on digestion site proximity of 35–75 bp. These regions are spread throughout the genome and cover all chromosomes (see Table 1, Supplemental Fig. 6). Most of these regions are composed of two digestion sites (69,778 of the 90,612 regions), while other regions contain as many as 18 digestion sites over a distance of up to 700 bp (for number of digestion sites per region, see Supplemental Table 3). These regions can be grouped into the mutually exclusive categories of annotated promoters (15%), untranslated regions (UTRs; 7%), exons (13%), introns (31%), and intergenic segments (34%) (Table 2). Although Methyl-seq requires HpaII sites, and therefore CpGs, in close proximity, which should bias our method toward CpG islands, more than half of the regions identified by the method are outside annotated

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 1. Distribution of the 90,612 Methyl-seq regions by chromosome (hg18)

chr1	7946
chr2	5923
chr3	3665
chr4	3212
chr5	3715
chr6	3765
chr7	4914
chr8	3794
chr9	4331
chr10	4315
chr11	5060
chr12	3847
chr13	1911
chr14	2813
chr15	2536
chr16	5167
chr17	5778
chr18	1814
chr19	6674
chr20	3129
chr21	1505
chr22	3055
chrX	1717
chrY	26

islands (55,084 of 90,612). Nevertheless, Methyl-seq is still able to effectively assay 65% of all annotated CpG islands in the human genome (17,945 of 27,639) (Bird 1987; Gardiner-Garden and Frommer 1987). In addition, the assayable regions cover a range of CpG densities, as illustrated with the overlap of Methyl-seq regions with promoters previously defined as high-CpG promoters (HCPs), intermediate-CpG promoters (ICPs), and low-CpG promoters (LCPs) (Table 2; Weber et al. 2007). This suggests that Methyl-seq can assay a range of genomic elements, allowing a broader survey of regions than classic methylation studies limited to CpG islands and promoters.

In each Methyl-seq experiment, we infer the methylation status from the presence or absence of HpaII tags at each digestion site. We called regions with an average tag count of greater than one tag per digestion site as unmethylated and called the remaining regions as methylated. We validated these methylation calls by three separate techniques: Illumina Infinium HumanMethylation27 genotyping, qPCR, and MeDIP (Weber et al. 2005; Gebhard et al. 2006; Zhang et al. 2006). For the sake of brevity, we describe here only the results obtained by the bisulfite genotyping, as such methods are often considered a gold standard for which to judge DNA methylation (Paz et al. 2003; Clark et al. 2006). The validation results obtained from qPCR and MeDIP are comprehensively detailed in the Supplemental Text and Supplemental Figures 8 and 9. All three methods strongly validate the Methyl-seq results and together provide high confidence for the technique. To validate Methyl-seq with regions detected by bisulfite conversion, we used the Illumina Infinium HumanMethylation27 platform, in which the methylation status of more than 27,000 CpG residues can be detected by bisulfite treatment followed by robust, sequencing-based genotyping. We compared the 160 CpG sites in the genome that are present on the Illumina HumanMethylation27 platform and are also assayable by Methyl-seq, and the results from the two methods were highly concordant. The presence of Methyl-seq tags is correlated with low beta scores in the Illumina HumanMethylation27 assay (unmethylated), and the absence of Methyl-seq tags is correlated with high HumanMethylation27 beta scores (methylated) (Fig. 1B). When bisulfite methylation calls were binarized so that HumanMethylation27 beta scores of greater than 0.5 were considered methylated and those less than 0.5 were considered unmethylated, we found 92.5% agreement across seven samples from four different cell types (Supplemental Fig. 7). From this analysis, we created a receiver operating characteristic (ROC) curve, where the area under the curve is 0.944 (Fig. 1C), indicating very low false positive and false negative rates in the Methyl-seq assay.

DNA methylation changes minimally during in vitro differentiation

The nature and extent of epigenetic regulation of lineage differentiation during human development is mostly unknown. However, mouse studies suggest a critical role for DNA methylation during development (Okano et al. 1999). To address how DNA methylation contributes to human somatic differentiation on a genome-wide level, we applied Methyl-seq to developing human tissues and cells (Fig. 2A; Supplemental Table 1). In our initial experiments, we asked how DNA methylation patterns change during the differentiation of hESCs into definitive endoderm and other derivatives. To this end, we generated a homogeneous (98% pure) population of definitive endoderm from H9 hESCs according to previously reported protocols, which include FACs sorting with CXCR4 after a 5-d culture regime in the presence of activin (see Methods) (D'Amour et al. 2005). These sorted cells express high levels of endoderm markers, including *SOX17*, *FOXA2*, *GATA4*, *GATA6*, and *CXCR4* (Fig. 2B). Applying Methyl-seq, we

Table 2. The breakdown of the 90,612 Methyl-seq regions into specific annotated genomic elements and binding locations

CpG island	35,528 (39%)
Non-CpG island	55,084 (61%)
Promoter (−1000 bp to TSS)	13,575 (15%)
5′ UTR	4366 (5%)
Coding exon	11,986 (13%)
Intron	28,035 (31%)
3′ UTR	1653 (2%)
Intergenic region	30,997 (34%)
Bivalent	5200 (6%)
H3K4me3	27,537 (30%)
H3K27me3	2366 (3%)
Neither	55,509 (61%)
HCP	14,330 (16%)
ICP	1254 (1.2%)
LCP	385 (0.4%)
7× Regulatory potential	75,297 (83%)

CpG island coordinates were obtained from the UCSC genome browser where they were calculated for hg18 by using the definition from Gardiner-Garden and Frommer (1987), where islands have a GC content of 50% or greater, a length greater than 200 bp, and a ratio greater than 0.6 of observed number of CG dinucleotides to the expected number on the basis of the number of Gs and Cs in the segment (1987). Coordinates for promoters and other genic elements were calculated using the transcription start sites and gene annotations from the Known Genes file, compiled by the UCSC Genome Browser. Coordinates for strong CpG islands or high-CpG promoters (HCPs), weak CpG islands or intermediate-CpG promoters (ICPs), and sequences with no local enrichment of CpGs or low-CpG promoters (LCPs) were from Weber et al. (2007). Coordinates for regions of 7× regulatory potential, or regions (average score >0.5) conserved in human, chimpanzee, macaque, mouse, rat, dog, and cow, were obtained from King et al. (2005). Histone ChIP-chip binding data was from Pan et al. (2007), where "bivalent" regions were regions that overlapped with both H3K4me3 and H3K27me3 peaks. Methyl-seq regions overlapping with H3K4me3 peaks only were called H3K4me3; regions overlapping with H3K27me3 peaks only were called H3K27me3; and all unassigned regions were defined as "neither" (Pan et al. 2007).

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

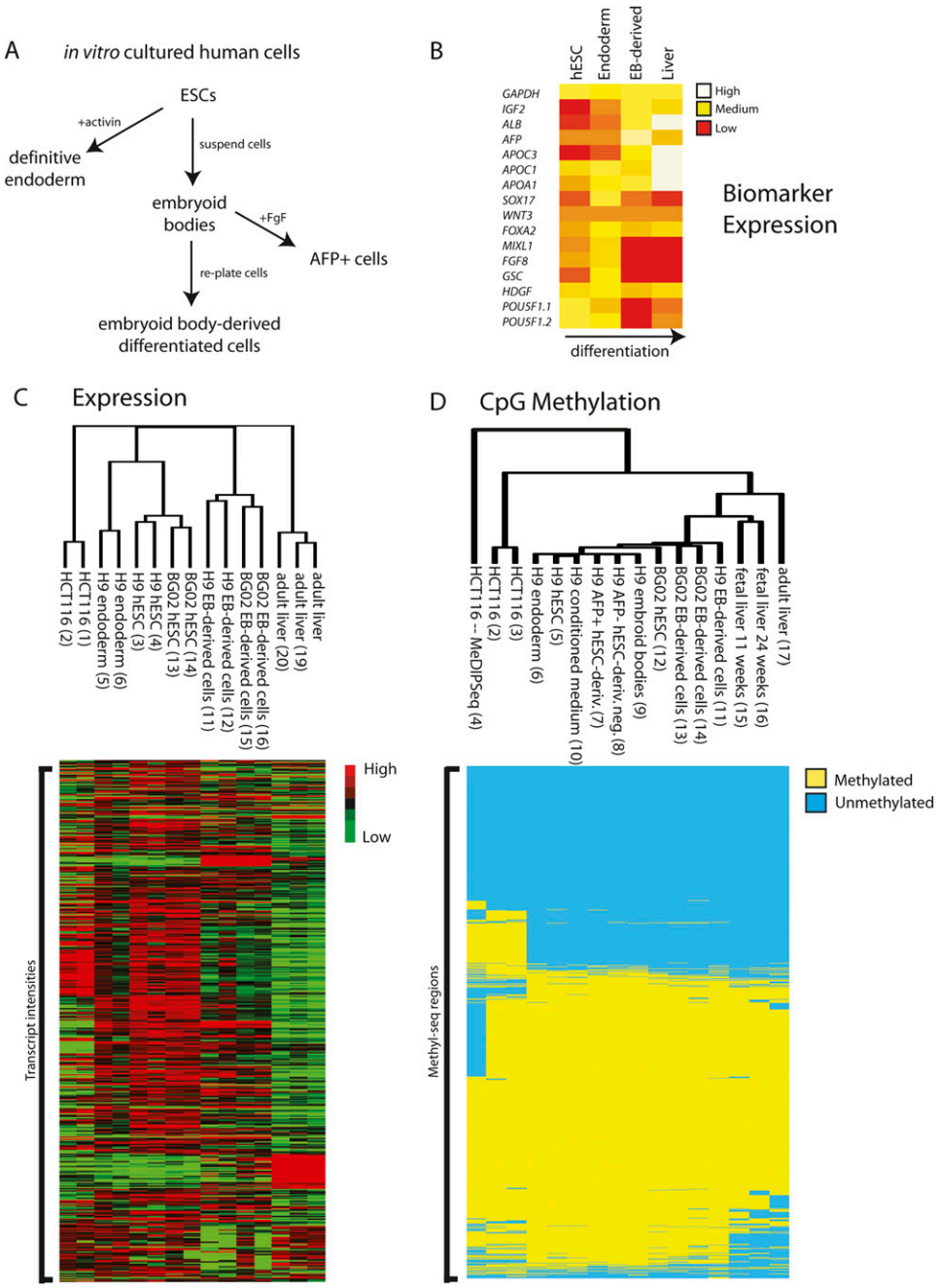


Figure 2. Expression and methylation changes in cells derived from hESCs and human tissues. (*A*) hESC in vitro differentiation scheme. (*B*) Expression of key genes in the liver differentiation pathway measured with expression arrays. Expression values are the average of intensities across all biological replicates and both parental hESC lines for each particular tissue. (*C*) Hierarchal clustering by genes and by samples of log-transformed, normalized expression data for key tissues in the endoderm-lineage model. Shown are a set of 2700 genes that clustered well. Sample numbers are shown in parentheses (see Supplemental Table 1). (*D*) Hierarchal clustering of the 90,612 Methyl-seq regions for each of the tissues assayed. HCT116 was used in Methyl-seq development and validation and shows reproducibility between biological replicates and concordance with MeDIPSeq data (for more information, see Supplemental Text; Supplemental Fig. 9). Library numbers are shown in parentheses (see Supplemental Table 2).

compared the DNA methylation profiles of the undifferentiated H9 hESC line and CXCR4-sorted definitive endoderm cells (Supplemental Table 2; Libraries 5, 6, respectively). Although these hESCs and definitive endoderm populations have markedly different genome-wide transcript profiles (Fig. 2C), we found that, upon surveying a total of 90,612 regions, only a small fraction of the regions change in their methylation state following differentiation of H9 hESC to definitive endoderm: 1017 (1.1%) become methylated and 682 (0.8%) become demethylated (see Supplemental Table 5C). These results suggest that significant methylation changes are not needed during endoderm formation to control gene expression, which is consistent with previous evidence showing that embryos lacking de novo methyltransferases are viable during gastrulation (Okano et al. 1999).

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

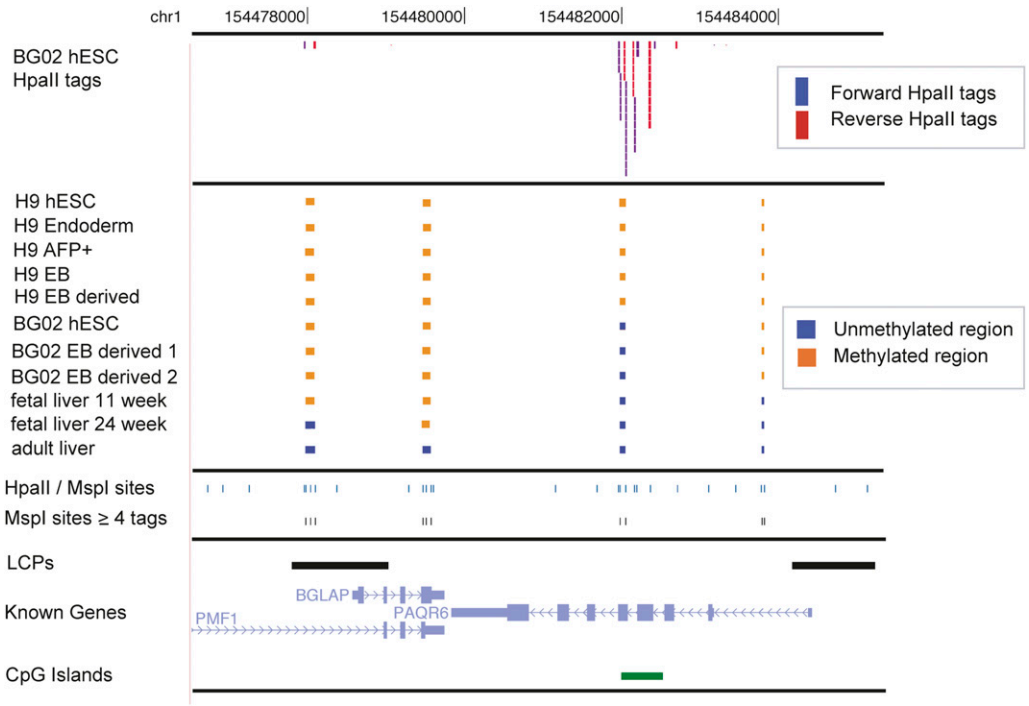


Figure 3. Browser shot of ~9 kb of human chromosome 1 (hg18). *Top* data track shows sequence tags for the BG02 hESC HpaII library. Tags oriented in the forward direction are in blue; reverse direction, in red. The second grouping of data tracks illustrate the methylation calls for the Methyl-seq regions. Orange boxes represent methylated regions, and blue boxes are unmethylated regions. The third group of data tracks shows all of the predicted HpaII/MspI digestion sites (CCGG). The track labeled "MspI sites ≥4 tags" shows the Methyl-seq assayable sites, above the threshold of four. The last group of data tracks shows the regional annotations. LCPs are the low density CpG promoters, as predicted by Weber et al. (2007). The known genes and CpG islands were obtained from the UCSC Genome Browser. This window was chosen as an example of unmethylated and methylated regions, and shows regions with in vivo methylation changes: fetal liver 11 wk methylated vs. fetal liver 24 wk unmethylated (region 1); fetal liver methylated vs. adult liver unmethylated (region 2); a CPG island showing methylation differences between H9 samples and BG02 samples (region 3); and a region with differences between in vitro (methylated) and in vivo (unmethylated) samples (region 4).

Based on these results, we hypothesized that a cell type representing a later stage of development might exhibit more differences in DNA methylation than we observed between endoderm and hESCs. To test this notion, we derived several different cell types from hESCs, including alpha-fetoprotein (AFP)-positive-sorted cells with hepatocyte-like characteristics (Chiao et al. 2008), AFP-negative-sorted cells with highly heterogeneous mixture of cell types, embryoid bodies (EBs), and heterogeneous populations of cells derived from EBs after 20 d of plating (EB-derived) (Fig. 2A). By using Methyl-seq, we found that DNA methylation changed at only a small percentage (2%–5%) of regions during hESC differentiation, regardless of heterogeneity (see Fig. 3 for example of regional methylation data). In the most homogeneous population (the AFP positive), we found 2628 differences (2.9%) compared with the naïve hESC state (Supplemental Table 5E). Similar numbers were observed in EBs, EB-derived, as well as in the AFP-negative populations. More than half of the changes (1732) are shared between the derived cells, even though they are comprised of unique cell types (Fig. 4B; Supplemental Table 5F,H). These same changes occur in endoderm cells and EBs (see Fig. 4A,C). Furthermore, of the 1011 changes common to both AFP differentiation and endoderm differentiation, only 182 are absent in the differentiation toward AFP-negative cells, indicating that a similar group of changes is occurring in hESC differentiation regardless of the derived cell type (Supplemental Table 5D,J). To test whether these changes are specific to the particular hESC line used, H9, or represent a general hESC trend, we also tested several of these differentiation regimes on another hESC line, BG02. We found that many of the changes occurring upon differentiation are common between samples derived from either hESC line (Supplemental Table 5C–W). The high frequency of overlapping methylation changes in the derived heterogeneous populations, which are frequently composed of beating cardiomyocytes and other tissue types, strongly suggests that similar changes in methylation occur during hESC differentiation regardless of induced cell type.

Only a small fraction (<5%) of the regions that we assayed by Methyl-seq showed changes in DNA methylation during in vitro differentiation. We asked whether this was unique to cells derived from hESCs or whether it was also true for in vivo human fetal differentiation. Thus, we measured human 11- and 24-wk fetal, as well as adult liver samples for their DNA methylation patterns by Methyl-seq (Supplemental Table 2; Libraries 15–17, respectively). Because the fetal liver also plays a role in hematopoiesis throughout gestation, it is a heterogeneous combination of erythrocytes and hepatocytes (MacSween et al. 2007). Conversely, the adult liver is composed of ~78% hepatocytes, making it one of the more homogeneous adult organ systems (Blouin et al. 1977; MacSween et al. 2007). Although we were concerned about the heterogeneity of the fetal liver, we were not able to separate these populations due to the limiting nature of the tissue (Dan et al. 2006; Tosh and Strain 2005).

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

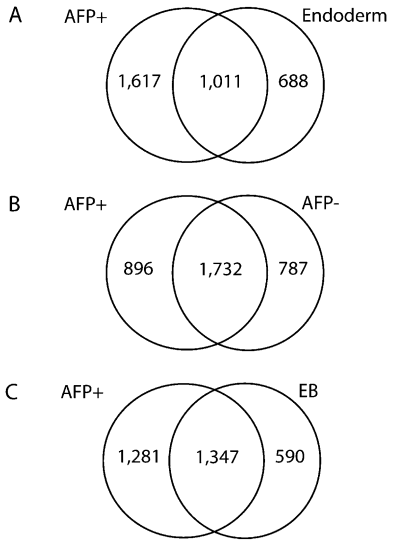


Figure 4. DNA methylation changes occurring between naïve H9 hESCs and various differentiated samples. (*A*) Venn diagram for methylation differences between H9 hESC (5) and H9 AFP+ hESC-derived cells (7) and differences between H9 hESC (5) and H9 Endoderm (6). (*B*) Venn diagram for methylation differences between H9 hESC (5) and H9 AFP+ hESC-derived cells (7) and differences between H9 hESC (5) and AFP- negative hESC-derived cells (8). (*C*) Venn diagram for methylation differences between H9 hESC (5) and H9 AFP+ hESC-derived cells (7) and differences between H9 hESC (5) and H9 embryoid bodies (EB) (9). Library numbers are in parentheses.

However, between 11 and 24 wk, the relative proportions of cell types remain similar, so we can still assess changes occurring between these two stages of liver development. When we applied Methyl-seq to these tissues, we found that 6034 regions changed their methylation status between 11- and 24-wk fetal liver, 10,237 regions changed between 11-wk fetal and adult liver, and 9359 changed between 24-wk fetal and adult liver (Fig. 3 Supplemental Table 5AG–AI). We also observed 12,934 regions with methylation differences between adult liver and hESCs, of which 5173 were common to fetal liver at 11 wk and 7215 were common to fetal liver at 24 wk (Supplemental Table 5Z,AB,AC). These results suggest that although these tissues are heterogeneous, the methylation patterns seen for all stages of liver have significant overlap and indicate progression of methylation changes over time in vivo.

Genomic context guides DNA methylation regardless of tissue type

We compared the DNA methylation status results of the 90,612 assayable regions in the distinct cells and tissues tested in our study. By comparing the relative percentage of methylation in these regions between samples, we found that hESCs and their derivatives have a larger fraction of methylated regions than do the in vivo fetal and adult tissues. The fraction of DNA regions methylated in hESCs and their derivatives (54%–59%) is significantly higher than in that of fetal liver, adult liver, and pancreas (40%–49%, $P = 5.5 \times 10^{-4}$, Mann-Whitney *U*-test; Supplemental Table 4A; data not shown). However, the relative percentage of methylation in the hESC derivatives does not appear to change substantially when compared with the initial hESC methylation levels, regardless of time in culture or heterogeneity of the population (see Supplemental Table 4A). Conversely, the global methylation percentage appears to decrease during liver development, with the 11-wk liver having the highest percentage of DNA methylation (49%) and the adult liver the lowest (43%), suggesting that in vivo liver development is biased toward demethylation events (Supplemental Table 4A).

We assigned the 90,612 assayable regions into distinct categories of genomic elements and then analyzed the Methyl-seq data in each sample according to each type of functional element. The annotated genomic elements are (1) predicted CpG islands (which overlap extensively with other classifications); (2) promoters and 5′-UTRs; (3) exons; (4) introns; (5) intergenic regions; (6) high density CpG promoters (HCPs); (7) intermediate density CpG promoters (ICPs); (8) low density CpG promoters (LCPs); (9) regions with predicted regulatory potential (7× RP); and (10) regions previously shown to be bound in hESCs by specific modified histone H3 proteins, including H3K4me3 (referred to as univalent regions), H3K27me3, and H3K4me3/H3K27me3 (referred to as bivalent regions) (Table 2; King et al. 2005; Bernstein et al. 2006; Pan et al. 2007; Weber et al. 2007). In Supplemental Table 4, B through N, we present the degree of DNA methylation for each genomic element in each sample. We found that, in every sample analyzed, regions that include CpG islands, promoters and both univalent and bivalent H3K4me3 domains are more highly unmethylated (less than 15% of regions methylated), whereas regions outside of CpG islands, including those occupied by H3K27me3, exons, introns, 3′ UTRs, and intergenic regions are more highly methylated (ranging from 39% to more than 78% of the sites are methylated; Fig. 5). Thus, different genomic functional elements vary in their degree of DNA methylation.

H3K4me3 and H3K4me3/H3K27me3 binding correlates highly with unmethylated promoters

Transcriptional promoters are critical components of gene regulation. To understand how DNA methylation is regulated at these elements and because H3K4me3 and H3K27me3 binding are correlated with distinct methylation status in hESCs (i.e., unmethylated and methylated, respectively), we focused on a comparison of DNA methylation and these binding events in promoter regions. Examining HCPs and LCPs with the occupancy of H3K4me3 and H3K27me3, we found that HCPs are highly enriched (82%) for H3K3me3 binding. Of these HCP univalent regions, 98% are unmethylated. When we examined the small fraction of these HCP regions that are methylated, we observed that most of them are located on the edge of unmethylated CpG islands. Therefore, most of these promoters show some unmethylated portion but are not fully unmethylated across their entirety. Although univalent domains are much less frequent in LCPs (13%), these few LCP univalent regions are also highly unmethylated (80%). This is in stark contrast to the LCP regions lacking both H3K4me3 and H3K27me3, where only 9% are unmethylated. Bivalent domains in both HCPs and LCPs show a similar trend; HCP bivalent regions are 95% unmethylated, and LCP bivalent regions are 71% unmethylated. Thus, univalent and bivalent domains are both strong indicators of regions of unmethylated DNA and are particularly informative at predicting unmethylated LCPs. This is not a hESC-specific phenomenon, as we observed the same trend in 11-wk fetal liver and in all hESC-derived cells, and a similar result was recently reported for mouse tissues (Meissner et al. 2008).

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

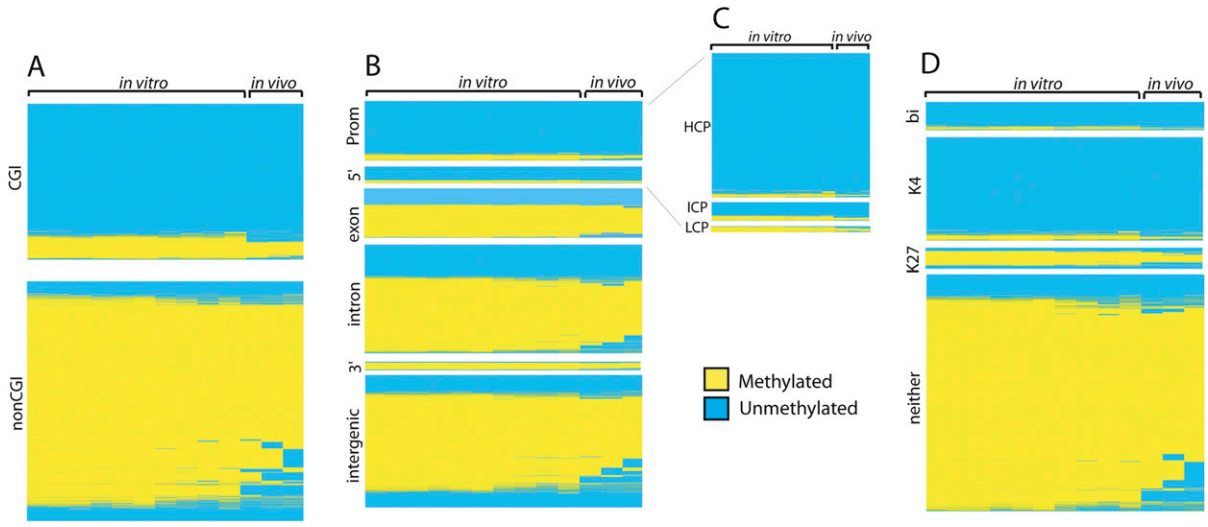


Figure 5. Hierarchal clustering of Methyl-seq regions for each of the tissues assayed. (*A*) Clustergram of Methyl-seq data reclustered by region for regions overlapping with annotated CpG islands (CGI) or regions outside of CpG islands (nonCGI). Samples are not reclustered from Figure 2D and are as follows. In vitro: H9 endoderm (6), H9 hESC (5), H9 conditioned medium (10), H9 AFP-positive hESC-derived cells (7), H9 AFP-negative hESC-derived cells (8), H9 embryoid bodies (9), BG02 hESC (12), BG02 EB-derived cells (13), BG02 EB-derived cells (14), H9 EB-derived cells (11); in vivo: fetal liver 11 wk (15), fetal liver 24 wk (16), adult liver (17). (*B*) Clustergram of Methyl-seq data reclustered by region overlap with gene structure annotations: promoters (Prom), 5′ UTRs (5′), coding exons (exon), introns (intron), 3′ UTRs (3′), and intergenic regions. Samples are again ordered as described in *A*. (*C*) Clustergram of Methyl-seq data reclustered by region overlap with promoter CpG density annotations: HCP, ICP, and LCP. Samples are ordered as described in *A*. (*D*) Clustergram of Methyl-seq data reclustered by region overlap with histone data: bivalent domains (bi), H3K4me3 (K4), H3K27me3 (K27), or regions bound by neither histone (neither). Samples are ordered as described in *A*.

hESC differentiation is defined by significant changes in methylation status at LCPs and H3K27me3 and H3K4me3/H3K27me3-occupied regions

We next analyzed the DNA methylation changes that occur during differentiation in vitro and in vivo. We divided the 90,612 assayable Methyl-seq regions into those that are different and those that are the same between all samples. Seventy-five percent of the regions analyzed (68,116 of 90,612) showed no changes in DNA methylation in any of the cell types or tissues that we tested; 49% of these unchanged regions were always methylated, while 51% were always unmethylated, in all samples (Supplemental Table 5A). In agreement with the trends described above with other work, the regions that are always methylated in all samples tended to be located in non-CpG islands, exons, introns, intergenic regions, and regions not bound by either H3K4me3 or H3K27me3 (Fig. 5). In addition, the regions that are always unmethylated tended to be located in CpG islands, promoters (including HCPs and ICPs), and H3K4me3-bound regions. The remaining 25% (22,496) of the assayable regions show at least one difference in methylation status between all samples (Supplemental Table 5B). In Supplemental Table 5, C through AI, we provide details of the relative changes of these regions between hESCs, hESC-derived cells, and tissues. We found that, during hESC differentiation, only 120 regions change uniquely in all hESC-derived samples (Supplemental Table 5W). Seventy-seven (64%) of these regions become methylated, and 43 (36%) become unmethylated. Seven of these 120 regions are transcriptional promoters; interestingly, all of these promoters become methylated during hESC differentiation, in contrast to the nonmethylated state of most promoters in the genome (see Supplemental Table 5A).

We next identified regions in the genome that changed in DNA methylation state during differentiation. We observed an average of 1.3% of the regions undergoing de novo methylation and 1.6% undergoing demethylation (see Fig. 6A, total). Dividing these changes into different genomic contexts, we found that this trend is seen in all categories except for LCPs, H3K27me3, and bivalent domains. These elements are highly variable (Fig. 6A). H3K27me3-occupied regions and bivalent domains become substantially de novo methylated in the majority of differentiated profiles ($1.4 \times 10^{-27} < P < 1.2 \times 10^{-6}$ and $1.8 \times 10^{-23} < P < 6.5 \times 10^{-11}$, respectively; hypergeometric tests). LCPs are also extremely variable in DNA methylation state during hESC differentiation, but unlike bivalent and H3K27me3-occupied regions, the changes are significantly enriched in both directions ($1.5 \times 10^{-8} < P < 1.8 \times 10^{-3}$; Fig. 6A). Therefore, within the regions whose DNA methylation state changes during hESC differentiation, LCPs, bivalent domains, and H3K27me3 marks are significantly more represented than any other genomic region, indicating that these regions are "hotspots" for either de novo methylation or demethylation in hESC differentiation. Although it is unknown if these differentially methylated regions are functionally important, these regions may be prime targets for the regulation of unique cell fates.

Fetal liver differentiation characterized by demethylation

While hESC changes are defined by a combination of de novo methylation and demethylation enriched at particular genomic elements, fetal liver differentiation is characterized by substantial demethylation (Fig. 6B–D). For example, there are 6034 DNA methylation differences between 11- and 24-wk human fetal liver; 76% of these become demethylated by 24 wk (Fig. 6B; Supplemental Table 5AG). This finding is further supported by the analysis of regions that change uniquely within the developing and adult liver, but not within hESC or hESC derivatives (Fig. 6D; Supplemental Table 5AF). Of these 3050 genomic regions, almost all (99.7%) become demethylated. Similar to our findings with

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

hESC differentiation, significant methylation changes occur in LCPs, bivalent regions, and H3K27me3-bound regions. However, liver development consists of significantly fewer methylation changes in bivalent regions, suggesting these regions are far less variable during in vivo differentiation compared to in vitro differentiation ($1.5 \times 10^{-111} < P < 3.8 \times 10^{-66}$; hypergeometric test). Furthermore, whereas hESC differentiation showed significant numbers of de novo methylation events at H3K27me3-bound regions, the in vivo differentiation shows significant numbers of demethylation events occurring in these regions ($1.2 \times 10^{-11} < P < 8.1 \times 10^{-7}$, hypergeometric test). We obtained similar results for LCPs; these elements were significantly variable for methylation changes in both directions during hESC development. However, during fetal liver development, LCP regions are significantly enriched for demethylation ($P = 1.7 \times 10^{-13}$, hypergeometric test). Thus, while hESCs show variable methylation patterns (including both de novo methylation and demethylation) during differentiation, liver development is specifically characterized by demethylation during differentiation. However, the changes observed for both in vitro and in vivo differentiation occur in similar genomic contexts, including LCPs, bivalent regions, and H3K27me3-bound regions. This suggests that the DNA methylation changes during hESC differentiation are not an accurate reflection of the changes in DNA methylation in vivo, at least in liver development.

BG02 and H9 hESC lines share 49 specific DNA methylation regions

hESC lines exhibit different phenotypic responses to differentiation, and each line has a long history of being cultured under distinct conditions. In this study, we used two different lines, BG02 and H9, and the observations described above are consistent in both of these lines. Comparison of the assayable DNA methylation regions between BG02 and H9 revealed a larger number of differences between these two lines (2913) than occur between the naïve hESCs and their derivatives (see Supplemental Table 5T vs. Supplemental Table 5C–S,U). Most striking is that 78% of the differences between the two hESC lines occur where H9 is methylated and BG02 is not. These changes do not appear to influence gene expression, as the hESC expression profiles are highly similar. We suggest that these changes are the result of exposure to different culturing conditions and derivation strategies. Comparison of both BG02 with H9 to all other hESC-derived cells and in vivo liver samples revealed 49 hESC-specific regions that changed between all samples but were similar between BG02 and H9 (Supplemental Table 5AE). Twenty-three of these regions unique to BG02 and H9 surround genes, in coding regions, promoters, or introns; only one lies within a CpG island. The remaining 26 regions are in intergenic sequences and are mostly unmethylated (70%) in the hESCs. Several of the genes flanked by these regions are of notable developmental importance, including *NFATC2* and *IGF2*. Thus, although there are many differences in the DNA methylation signatures between BG02 and H9, there are only 49 regions showing shared methylation patterns between these cell lines. These regions may be useful biomarkers for the undifferentiated state.

Discussion

During mouse embryogenesis, a genome-wide loss of DNA methylation is observed in the preimplantation embryo (Monk et al. 1987). At the blastocyst stage, de novo methylation occurs within the inner cell mass, but not the trophectoderm (Santos et al. 2002). By implantation, DNA methylation in the genome is greatly elevated compared with preimplantation levels and is thought to be maintained throughout subsequent differentiation in somatic lineages (Jaenisch and Jahner 1984; Razin and Cedar 1991; Kafri et al. 1992; Jaenisch 1997). This logic has largely been extrapolated to include human development; although due to ethical restraints and tissue procurement difficulties, it remains unknown whether a similar pattern of DNA methylation occurs in the early human embryo. Furthermore, how DNA methylation changes during both human and mouse development in somatic lineages is not understood (Reik et al. 2001; Latham et al. 2008; Suzuki and Bird 2008). Several studies have shown that human and mouse somatic tissues have distinct DNA methylation patterns, which is consistent with DNA methylation guiding the repression or activation of tissue-specific loci. However, many of these studies measured only small numbers of loci and did not study the changes in the differentiating cells through developmental time (Yeivin and Razin 1993; Shiota et al. 2002; Kremenskoy et al. 2003; Shiota 2004). Recently, several groups have applied genomic techniques toward understanding DNA methylation in ES cells and differentiating tissues. Fouse et al. (2008) reported genome-wide mapping of DNA methylation patterns in proximal promoter regions in mouse embryonic stem cells (mESCs). They found that binding of POU5F1 (OCT4), NANOG, Polycomb group proteins (PcG), and H3K27 trimethylation is correlated with DNA methylation of promoters in mESC cells (Fouse et al. 2008). Furthermore, Meissner et al. (2008) recently developed a method based on bisulfite treatment and next-gen sequencing (RRBS) and applied it to measure DNA methylation in mESCs, neural progenitor cells derived from mESCs, and several mouse tissues. Their report provided insight into how DNA methylation becomes modified during the differentiation of mESCs into neural tissue, and demonstrated that CpGs are dynamic marks that undergo specific changes during differentiation. Furthermore, they showed that the DNA methylation status can be predicted based upon association of histone methylation patterns (Meissner et al. 2008). Based upon these studies and others, a clearer picture is emerging of how developmental processes are influenced by DNA methylation. However, there remain many questions that need to be addressed, particularly in human ESCs and within specific developmental lineages, which are likely to have unique DNA methylation signatures (Weiss and Cedar 1997; Shiota 2004; Suzuki and Bird 2008).

Although several reports have examined DNA methylation in individual human tissues by using different genomic technologies that assess promoters and CpG islands (Eckhardt et al. 2006; Ladd-Acosta et al. 2007; Shen et al. 2007; Weber et al. 2007; Illingworth et al. 2008; Rauch et al. 2008), our study provides the most comprehensive portrait to date of DNA methylation changes at a wide array of genomic elements during human cellular differentiation, including in hESCs and in vivo fetal tissues. It demonstrates that in vivo liver differentiation is characterized by demethylation through time; a theory that has been generally proposed for somatic differentiation but that has little supportive evidence (Weiss and Cedar 1997). Furthermore, these demethylation events are particularly enriched at H3K27me3-bound regions and LCPs, suggesting that these regions are key mediators in tissue-specific gene expression. However, it must also be noted that the methylation differences observed between the liver tissues of different developmental stages were from different

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

individuals, and it therefore remains a possibility that the differences observed were due to individual methylation variation and not a consistent developmental regime. More samples will need to be profiled in the future to address this issue. Our work also characterizes the methylation changes occurring in several hESC differentiation regimes, and using this extensive data set, we propose that the changes are composed of both de novo methylation and demethylation at particular genomic elements, including H3K27me3-bound regions and LCPs. We also found conclusively that hESCs are more methylated than other tissues, especially upon differentiation. This finding is consistent with the previous report that cultured cells show increased DNA methylation relative to primary tissues (Antequera et al. 1990), but contradicts reports that naïve hESCs and other mammalian cell lines capable of in vitro differentiation are hypomethylated (Razin et al. 1984, 1988; Young and Tilghman 1984; Jost and Jost 1994; Zvetkova et al. 2005; Bibikova et al. 2006).

The approach we describe here, Methyl-seq, offers a new tool for measuring the DNA methylation status of a very large number of regions of the human genome and any other vertebrate genome whose DNA sequence is known. Several other methods have been recently described, and all, including Methyl-seq, have strengths and weaknesses (Kremenskoy et al. 2003; Weber et al. 2005, 2007; Keshet et al. 2006; Zhang et al. 2006; Shen et al. 2007; Illingworth et al. 2008; Meissner et al. 2008). Methyl-seq allows measurement of the DNA methylation status of more than 90,000 regions, including a significant number that are not covered by any other single technique. The method is sensitive, highly specific with very low background, reproducible, and simple to execute. It is relatively inexpensive, requiring fewer reads on next-gen sequencers than RRBS (Meissner et al. 2008), and is likely to become less expensive as the sequencing technologies become more efficient and widespread. Despite these advantages, Methyl-seq has some key limitations that are important to keep in mind. First, Methyl-seq assays only the CpGs in a specific subset of HpaII restriction enzyme cleavage sites (those that appear within 35–75 bp of each other in the human genome). This significantly limits the number of CpGs assayed but nevertheless provides a snapshot of regional methylation patterns. Additionally, because Methyl-seq is sensitive in detecting small numbers of HpaII tags, this creates a problem measuring methylation quantitatively. While this may likely be improved in the future, we limited the analysis in this current study to binary calls of "methylated" and "unmethylated."

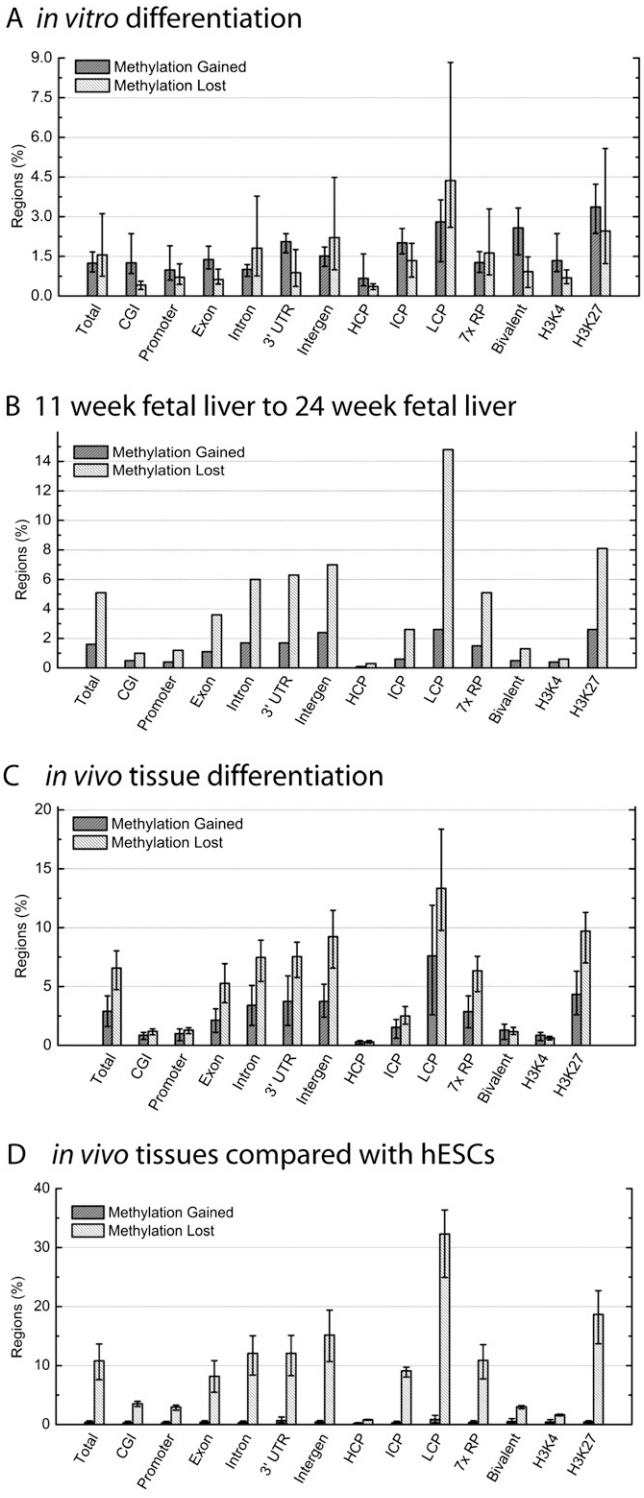


Figure 6. Percent of regions showing methylation changes by genomic feature. (*A*) Mean percentage of regions that gain or lose methylation in in vitro-differentiated cells compared with hESCs (*y*-axis), grouped by genomic features (*x*-axis). Categories are total changes, changes in CpG islands (CGI), promoters and 5′ UTRs (promoter), exons, introns, 3′ UTRs, intergenic regions (intergen), HCPs, ICPs, LCPs, 7× regulatory potential (7× RP), H3K4me3/H3K27me3-occupied regions (bivalent), H3K4me3-occupied regions (H3K4), and H3K27me3-occupied regions (H3K27). Error bars, maximum and minimum values across five differentiated sample comparisons: H9 hESC vs. H9 Endoderm (5→6), H9 hESC vs. H9 AFP+ hESC-derived cells (5→7), H9 hESC vs. H9 Embryoid bodies (5→9), H9 hESC vs. H9 EB-derived cells (5→11), BG02 hESC vs. BG02 EB-derived cells rep1 and rep2 (12→13,14). (*B*) Percentage of regions that gain or lose methylation in 11-wk fetal liver compared with 24-wk fetal liver (*y*-axis), grouped by genomic features (*x*-axis). Data are from the sample comparison between 11-wk fetal liver and 24-wk fetal liver (15→16). (*C*) Mean percentage of regions that gain or lose methylation during tissue differentiation (*y*-axis), grouped by genomic features (*x*-axis). Error bars, maximum and minimum values across three tissue sample comparisons: 11-wk fetal liver vs. 24-wk fetal liver (15→16), 11-wk fetal liver vs. adult liver (15→17), 24-wk fetal liver vs. adult liver (16→17). (*D*) Mean percentage of regions that gain or lose methylation in tissues compared with hESCs (*y*-axis), grouped by genomic features (*x*-axis). Error bars, maximum and minimum values across three tissue sample comparisons: H9 hESC and BG02 hESC vs. 11-wk fetal liver (5,12→15), H9 hESC and BG02 hESC vs. 24-wk fetal liver (5,12→16), H9 hESC and BG02 hESC vs. adult liver (5,12→17). Library numbers are in parentheses.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Because Methyl-seq currently uses "unmethylated" to describe any region showing HpaII fragments in the collection of next-gen sequencing reads, the quantitative methylation state of any individual DNA molecules are lost. This particularly limits the ability of Methyl-seq to detect partial methylation or imprinted loci. This is not likely to be a problem for most CpGs because numerous studies have shown that most CpGs are mostly methylated (>80%) or mostly unmethylated (<20%) (Meissner et al. 2008), and Methyl-seq is able to distinguish these quantitative differences. However, a more accurate interpretation of Methyl-seq regions showing HpaII tags may be "not fully methylated" as opposed to "unmethylated" to take into account that Methyl-seq calls are sensitive to any demethylation. Lastly, because the current Methyl-seq protocol relies on small HpaII fragments (grouped HpaII sites) as opposed to single cut sites, Methyl-seq does not assay most of the very low density CpG regions of the genome, including a number of low density CpG promoters. Because most of these promoters are assayed by Illumina's Infinium HumanMethylation27 platform, combining the methylation results from Methyl-seq and Illumina's Infinium HumanMethylation27 beadchips will provide a broad, comprehensive survey of methylation targets in the genome. Thus, when deep genomic bisulfite sequencing is not an option, Methyl-seq is a valuable addition to the mixture of emerging methods used to assay the methylation status of CpG dinucleotides in the human genome.

The epigenetic state of undifferentiated stem cells has been extensively studied in recent years, but clearly the next frontier is to understand how this undifferentiated state is modified into specific tissue types (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006; Mikkelsen et al. 2007). Because DNA methylation is key to embryonic survival and differentiation, gaining information about how this process is modified in vivo and in vitro could be central to the future engineering of hESCs for therapeutic purposes (Keller 2005). Future studies involving hESCs should address the importance of the endogenous epigenetic state for tissue function. Further investigation will be required to determine whether DNA methylation changes are similar between somatic lineages throughout developmental time and to determine more concretely the epigenetic differences between in vitro and in vivo differentiation in distinct cell types.

Methods

Samples

Twenty samples (Supplemental Table 1) were used to obtain genomic DNA; we also purified RNA from 14 of these (see Supplemental Methods). Eight of the hESC in vitro samples were sorted by FACS prior to nucleic acid extraction.

In vitro differentiation

Definitive endoderm precursors were generated from hESCs by using activin A as previously described (D'Amour et al. 2005). EB-derived cells were obtained by plating clumps of undifferentiated hESCs in suspension to promote the formation of EBs. After 8 d, the EBs were plated onto tissue culture dishes coated with gelatin type A, and further differentiation was continued for 2 wk until a peak level of the hepatic protein AFP was detected. To generate AFP-positive fetal hepatocytes, we developed a hESC line expressing a reporter green fluorescent protein (GFP) from the AFP promoter. After 30 d of differentiation in the presence of FGF, AFP-positive cells were isolated by FACS (Chiao et al. 2008).

Methyl-seq and MeDIPSeq library construction

Methyl-seq and MeDIPSeq libraries were constructed by using the procedures outlined in Supplemental Protocols 1 and 2, respectively. Briefly, in Methyl-seq, 5 μg genomic DNA was digested with HpaII or MspI. Fragments were made into next-gen sequencing libraries by using adapters and reagents from Illumina's Library Construction Kit. After an initial round of PCR, DNA bands (fragments with adapters) of 100–350 bp were isolated by gel extraction (Supplemental Fig. 2). Size-selected libraries were then PCR-amplified prior to sequencing. For MeDIPSeq, 5 μg genomic DNA was digested with MspI and immunoprecipitated with anti-methylcytosine antibody (Calbiochem); fragments were made into Illumina libraries as described above.

Illumina sequencing and primary analysis

Libraries were sequenced on an Illumina 1G Genome Analyzer sequencing machine, and reads were aligned to the genome (UCSC hg18) by using the standard Illumina sequence analysis pipeline (25 nucleotides, two mismatches). We normalized all HpaII Methyl-seq libraries to contain between 3 million and 3.5 million usable reads (those that align to the genome and possess the CGG HpaII-cut signature on their 5′ ends; Supplemental Table 2; Supplemental Methods). The MspI library that we used contained 10 million usable reads.

Analysis for calling restriction enzyme digestion sites and region methylation status

Usable sequence reads were mapped to CCGG sites predicted in silico. Sites with four or more MspI tags occurring in either the forward or reverse direction were retained in the analysis. These "assayable" sites were then grouped with neighboring sites that were within 35–75 bp. Regions therefore represent any number of digestion sites (between two and 18) that have neighboring sites within 35–75 bp of each other. Methylation calls were made by using HpaII tag data from all assayable cut sites. The larger of either the forward read count or reverse read count for each site was averaged across each region. Regions that had an average of zero or one reads per digestion site were called methylated, and regions with more than one sequence read per site were considered unmethylated.

Illumina Infinium validation

We used Infinium HumanMethylation27 BeadChips (Illumina) to validate the methylation calls obtained from Methyl-seq. Genomic DNA from seven samples corresponding to four Methyl-seq libraries was treated with bisulfite and genotyped. Comparison of Infinium data and Methyl-seq tag counts was performed at 160 CpGs common to both methods. The ROC curve was generated with the ROCR package for R. Illumina's Infinium HumanMethylation27 data were used as the "truth" by dichotomizing the beta value with a threshold of 0.6 (i.e., >0.6 was methylated and <0.6 was unmethylated). Additionally, when the dichotomized Infinium beta values were compared with Methyl-seq data (zero to one tag methylated, >1 tags unmethylated), Fisher exact test on each sample produced *P*-values $<10^{-8}$.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Gene expression

We used HumanRef-8 v2 Expression BeadChips (Illumina) for mRNA expression analysis. Array intensities were rank-invariant normalized, log-transformed, clustered using Cluster 3.0, and visualized by using TreeView (Eisen et al. 1998; Saldanha 2004). Genes were filtered to exclude those found in <30% of the samples.

Acknowledgments

We thank Hunter Shain for assistance with the tissue culture work, Larisa Tsavaler at the Stanford Human Genome Center for performing the mRNA expression experiments, Anming Xiong and Jeffrey Glenn for providing the fetal liver samples, Arend Sidow for helping analyze the data, and Rami Rauch for helping organize and submit the genomic data. This work was supported by a grant from the California Institute of Regenerative Medicine to J.C.B., NIH grant 7U54HG004576 to R.M.M., and funds from the HudsonAlpha Institute for Biotechnology to R.M.M. A.L.B. was funded by a NIH grant to the Stanford Genetics and Developmental Biology Training Program.

References

Antequera, F., Boyes, J., and Bird, A. 1990. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* **62:** 503–514.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326.

Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E., et al. 2006. Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* **16:** 1075–1083.

Bird, A. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3:** 342–347.

Bird, A. 2002. DNA methylation patterns and epigenetic memory. *Genes & Dev.* **16:** 6–21.

Bird, A.P. and Wolffe, A.P. 1999. Methylation-induced repression: Belts, braces, and chromatin. *Cell* **99:** 451–454.

Blouin, A., Bolender, R.P., and Weibel, E.R. 1977. Distribution of organelles and membranes between hepatocytes and nonhepatocytes in the rat liver parenchyma. A stereological study. *J. Cell Biol.* **72:** 441–455.

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., and Baccarelli, A. 2008. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech. Ageing Dev.* **130:** 234–239.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349–353.

Butta, N., Larrucea, S., Alonso, S., Rodriguez, R.B., Arias-Salgado, E.G., Ayuso, M.S., Gonzalez-Manchon, C., and Parrilla, R. 2006. Role of transcription factor Sp1 and CpG methylation on the regulation of the human podocalyxin gene promoter. *BMC Mol. Biol.* **7:** 17. doi: 10.1186/1471-2199-7-17.

Cedar, H., Solage, A., Glaser, G., and Razin, A. 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* **6:** 2125–2132.

Chang, S.C., Tucker, T., Thorogood, N.P., and Brown, C.J. 2006. Mechanisms of X chromosome inactivation. *Front. Biosci.* **11:** 852–866.

Chiao, E., Elazar, M., Xing, Y., Xiong, A., Kmet, M., Millan, M.T., Glenn, J.S., Wong, W.H., and Baker, J. 2008. Isolation and transcriptional profiling of purified hepatic cells derived from human embryonic stem cells. *Stem Cells* **26:** 2032–2041.

Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L., and Frommer, M. 2006. DNA methylation: Bisulphite modification and analysis. *Nat. Protocols* **1:** 2353–2364.

Da, L., Li, D., Yokoyama, K.K., Li, T., and Zhao, M. 2006. Dual promoters control the cell-specific expression of the human cell death-inducing DFF45-like effector B gene. *Biochem. J.* **393:** 779–788.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23:** 1534–1541.

Dan, Y.Y., Riehle, K.J., Lazaro, C., Teoh, N., Haque, J., Campbell, J.S., and Fausto, N. 2006. Isolation of multipotent progenitor cells from human fetal liver capable of differentiating into liver and mesenchymal lineages. *Proc. Natl. Acad. Sci.* **103:** 9912–9917.

Douet, V., Heller, M.B., and Le Saux, O. 2007. DNA methylation and Sp1 binding determine the tissue-specific transcriptional activity of the mouse Abcc6 promoter. *Biochem. Biophys. Res. Commun.* **354:** 66–71.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., et al. 2006. DNA methylation profiling of human chromosomes 6, 20, and 22. *Nat. Genet.* **38:** 1378–1385.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95:** 14863–14868.

Esteller, M. 2008. Epigenetics in cancer. *N. Engl. J. Med.* **358:** 1148–1159.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2:** 160–169.

Fritsch, M.K. and Singer, D.B. 2008. Embryonic stem cell biology. *Adv. Pediatr.* **55:** 43–77.

Fujii, G., Nakamura, Y., Tsukamoto, D., Ito, M., Shiba, T., and Takamatsu, N. 2006. CpG methylation at the USF-binding site is important for the liver-specific transcription of the chipmunk *HP-27* gene. *Biochem. J.* **395:** 203–209.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. 2003. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* **278:** 4035–4040.

Futscher, B.W., Oshiro, M.M., Wozniak, R.J., Holtan, N., Hanigan, C.L., Duan, H., and Domann, F.E. 2002. Role for DNA methylation in the control of cell type specific maspin expression. *Nat. Genet.* **31:** 175–179.

Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. 2006. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* **66:** 6118–6128.

Goll, M.G. and Bestor, T.H. 2005. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74:** 481–514.

Grunau, C., Hindermann, W., and Rosenthal, A. 2000. Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. *Hum. Mol. Genet.* **9:** 2651–2663.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. 2003. The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.* **34:** 187–192.

Hellman, A. and Chess, A. 2007. Gene body-specific methylation on the active X chromosome. *Science* **315:** 1141–1143.

Holliday, R. and Pugh, J.E. 1975. DNA modification mechanisms and gene activity during development. *Science* **187:** 226–232.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6:** e22. doi: 10.1371/journal.pbio.0060022.

Imamura, T., Ohgane, J., Ito, S., Ogawa, T., Hattori, N., Tanaka, S., and Shiota, K. 2001. CpG island of rat sphingosine kinase-1 gene: Tissue-dependent DNA methylation status and multiple alternative first exons. *Genomics* **76:** 117–125.

Jaenisch, R. 1997. DNA methylation and imprinting: Why bother? *Trends Genet.* **13:** 323–329.

Jaenisch, R. and Bird, A. 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**(Suppl.)**:** 245–254.

Jaenisch, R. and Jahner, D. 1984. Methylation, expression and chromosomal position of genes in mammals. *Biochim. Biophys. Acta* **782:** 1–9.

Jost, J.P. and Jost, Y.C. 1994. Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair. *J. Biol. Chem.* **269:** 10040–10043.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. 1992. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Dev.* **6:** 705–714.

Kaminsky, Z.A., Tang, T., Wang, S.C., Ptak, C., Oh, G.H., Wong, A.H., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., et al. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet.* **41:** 240–245.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. 2008. Transient cyclical methylation of promoter DNA. *Nature* **452:** 112–115.

Keller, G. 2005. Embryonic stem cell differentiation: Emergence of a new era in biology and medicine. *Genes & Dev.* **19:** 1129–1155.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., et al. 2006. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38:** 149–153.

King, D.C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., and Hardison, R.C. 2005. Evaluation of regulatory potential and conservation scores for detecting *cis*-regulatory modules in aligned mammalian genome sequences. *Genome Res.* **15:** 1051–1060.

Kitamura, E., Igarashi, J., Morohashi, A., Hida, N., Oinuma, T., Nemoto, N., Song, F., Ghosh, S., Held, W.A., Yoshida-Noro, C., et al. 2007. Analysis of tissue-specific differentially methylated regions (TDMs) in humans. *Genomics* **89:** 326–337.

Kremenskoy, M., Kremenska, Y., Ohgane, J., Hattori, N., Tanaka, S., Hashizume, K., and Shiota, K. 2003. Genome-wide analysis of DNA methylation status of CpG islands in embryoid bodies, teratomas, and fetuses. *Biochem. Biophys. Res. Commun.* **311:** 884–890.

Kroft, T.L., Jethanandani, P., McLean, D.J., and Goldberg, E. 2001. Methylation of CpG dinucleotides alters binding and silences testis-specific transcription directed by the mouse lactate dehydrogenase C promoter. *Biol. Reprod.* **65:** 1522–1527.

Kumar, V. and Biswas, D.K. 1988. Dynamic state of site-specific DNA methylation concurrent to altered prolactin and growth hormone gene expression in the pituitary gland of pregnant and lactating rats. *J. Biol. Chem.* **263:** 12645–12652.

Kusui, C., Kimura, T., Ogita, K., Nakamura, H., Matsumura, Y., Koyama, M., Azuma, C., and Murata, Y. 2001. DNA methylation of the human oxytocin receptor gene promoter regulates tissue-specific gene suppression. *Biochem. Biophys. Res. Commun.* **289:** 681–686.

Ladd-Acosta, C., Pevsner, J., Sabunciyan, S., Yolken, R.H., Webster, M.J., Dinkins, T., Callinan, P.A., Fan, J.B., Potash, J.B., and Feinberg, A.P. 2007. DNA methylation signatures within the human brain. *Am. J. Hum. Genet.* **81:** 1304–1315.

Latham, T., Gilbert, N., and Ramsahoye, B. 2008. DNA methylation in mouse embryonic stem cells and development. *Cell Tissue Res.* **331:** 31–55.

Lavelle, D., Vaitkus, K., Hankewych, M., Singh, M., and DeSimone, J. 2006. Developmental changes in DNA methylation and covalent histone modifications of chromatin associated with the epsilon-, gamma-, and beta-globin gene promoters in *Papio anubis*. *Blood Cells Mol. Dis.* **36:** 269–278.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125:** 301–313.

Lister, R. and Ecker, J.R. 2009. Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Res.* (this issue). doi: 10.1101/gr.083451.108.

MacSween, R.N.M., Burt, A.D., Portmann, B., and Ferrell, L.D. 2007. *MacSween's pathology of the liver.* Elsevier Churchill Livingstone, London, UK.

Mandrioli, M. 2007. A new synthesis in epigenetics: Towards a unified function of DNA methylation from invertebrates to vertebrates. *Cell. Mol. Life Sci.* **64:** 2522–2524.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454:** 766–770.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452:** 45–50.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.

Monk, M., Boubelik, M., and Lehnert, S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99:** 371–382.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. 2005. Epigenetic reprogramming in mammals. *Hum Mol Genet* **14** (Spec. No. 1)**:** R47–R58.

Nafee, T.M., Farrell, W.E., Carroll, W.D., Fryer, A.A., and Ismail, K.M. 2008. Epigenetic control of fetal gene expression. *BJOG* **115:** 158–168.

Ngo, V., Gourdji, D., and Laverriere, J.N. 1996. Site-specific methylation of the rat prolactin and growth hormone promoters correlates with gene expression. *Mol. Cell. Biol.* **16:** 3245–3254.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99:** 247–257.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1:** 299–312.

Paz, M.F., Fraga, M.F., Avila, S., Guo, M., Pollan, M., Herman, J.G., and Esteller, M. 2003. A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res.* **63:** 1114–1121.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., et al. 2004. DNA methylation profiling of the human major histocompatibility complex: A pilot study for the human epigenome project. *PLoS Biol.* **2:** e405. doi: 10.1371/journal.pbio.0020405.

Rauch, T.A., Zhong, X., Wu, X., Wang, M., Kernstine, K.H., Wang, Z., Riggs, A.D., and Pfeifer, G.P. 2008. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci.* **105:** 252–257.

Razin, A. and Cedar, H. 1991. DNA methylation and gene expression. *Microbiol. Rev.* **55:** 451–458.

Razin, A., Webb, C., Szyf, M., Yisraeli, J., Rosenthal, A., Naveh-Many, T., Sciaky-Gallili, N., and Cedar, H. 1984. Variations in DNA methylation during mouse cell differentiation in vivo and in vitro. *Proc. Natl. Acad. Sci.* **81:** 2275–2279.

Razin, A., Levine, A., Kafri, T., Agostini, S., Gomi, T., and Cantoni, G.L. 1988. Relationship between transient DNA hypomethylation and erythroid differentiation of murine erythroleukemia cells. *Proc. Natl. Acad. Sci.* **85:** 9003–9006.

Reik, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447:** 425–432.

Reik, W., Dean, W., and Walter, J. 2001. Epigenetic reprogramming in mammalian development. *Science* **293:** 1089–1093.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. 2006. Large-scale structure of genomic methylation patterns. *Genome Res.* **16:** 157–163.

Saldanha, A.J. 2004. Java Treeview: Extensible visualization of microarray data. *Bioinformatics* **20:** 3246–3248.

Santos, F., Hendrich, B., Reik, W., and Dean, W. 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241:** 172–182.

Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T., et al. 2006. Microarray-based DNA methylation profiling: Technology and applications. *Nucleic Acids Res.* **34:** 528–542.

Sha, K. 2008. A mechanistic view of genomic imprinting. *Annu. Rev. Genomics Hum. Genet.* **9:** 197–216.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.P. 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* **3:** 2023–2036.

Shiota, K. 2004. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. *Cytogenet. Genome Res.* **105:** 325–334.

Shiota, K., Kogo, Y., Ohgane, J., Imamura, T., Urano, A., Nishino, K., Tanaka, S., and Hattori, N. 2002. Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells* **7:** 961–969.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., and Cedar, H. 1999. DNA methylation represses transcription in vivo. *Nat. Genet.* **22:** 203–206.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H., and Held, W.A. 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc. Natl. Acad. Sci.* **102:** 3336–3341.

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

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. 2008. PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* **6:** 2911–2927.

Tosh, D. and Strain, A. 2005. Liver stem cells: Prospects for clinical use. *J. Hepatol.* **42**(Suppl.)**:** S75–S84.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20:** 116–117.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and 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.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. 2007. Distribution, silencing potential and

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39:** 457–466.
Weiss, A. and Cedar, H. 1997. The role of DNA demethylation during development. *Genes Cells* **2:** 481–486.
Yagi, S., Hirabayashi, K., Sato, S., Li, W., Takahashi, Y., Hirakawa, T., Wu, G., Hattori, N., Hattori, N., Ohgane, J., et al. 2008. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res.* **18:** 1969–1978.
Yeivin, A. and Razin, A. 1993. Gene methylation patterns and expression. *EXS* **64:** 523–568.
Young, P.R. and Tilghman, S.M. 1984. Induction of alpha-fetoprotein synthesis in differentiating F9 teratocarcinoma cells is accompanied by a genome-wide loss of DNA methylation. *Mol. Cell. Biol.* **4:** 898–907.
Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126:** 1189–1201.
Zhang, Y., Fatima, N., and Dufau, M.L. 2005. Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. *Mol. Cell. Biol.* **25:** 7929–7939.
Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39:** 61–69.
Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J.E., Crompton, L.A., John, R., Feil, R., and Brockdorff, N. 2005. Global hypomethylation of the genome in XX embryonic stem cells. *Nat. Genet.* **37:** 1274–1279.

Received December 1, 2008; accepted in revised form March 6, 2009.