DNA methylation in mice is influenced by genetics as well as sex and life experience

Sara A. Grimm, Takashi Shimbo, Motoki Takaku, James W. Thomas, Scott Auerbach, Brian D. Bennett, John R. Bucher, Adam B. Burkholder, Frank Day, Ying Du, Christopher G. Duncan, John E. French, Julie F. Foley, Jianying Li, B. Alex Merrick, Raymond R. Tice, Tianyuan Wang, Xiaojiang Xu, NISC Comparative Sequencing Program, Pierre R. Bushel, David C. Fargo, James C. Mullikin & Paul A. Wade

DNA methylation is an essential epigenetic process in mammals, intimately involved in gene regulation. Here we address the extent to which genetics, sex, and pregnancy influence genomic DNA methylation by intercrossing 2 inbred mouse strains, C57BL/6N and C3H/HeN, and analyzing DNA methylation in parents and offspring using whole-genome bisulfite sequencing. Differential methylation across genotype is detected at thousands of loci and is preserved on parental alleles in offspring. In comparison of autosomal DNA methylation patterns across sex, hundreds of differentially methylated regions are detected. Comparison of animals with different histories of pregnancy within our study reveals a CpG methylation pattern that is restricted to female animals that had borne offspring. Collectively, our results demonstrate the stability of CpG methylation across generations, clarify the interplay of epigenetics with genetics and sex, and suggest that CpG methylation may serve as an epigenetic record of life events in somatic tissues at loci whose expression is linked to the relevant biology.
Methylation of cytosine in the context of the simple palindromic dinucleotide 5′ CG 3′ represents the most common form of DNA modification in mammals. Maintenance of DNA methylation states following DNA replication constitutes an essential mechanism wherein daughter cells inherit cell-type specific epigenetic programs. The global pattern of DNA methylation is reprogrammed during genesis of germ cells and also during very early embryogenesis, establishing a common epigenetic slate for development and differentiation, raising questions regarding the extent to which DNA methylation patterns in offspring resemble those in parents. Nonetheless, evidence exists that DNA methylation patterns may be, to some extent, under genetic control, suggesting a mechanistic basis for similarity between parents and offspring.

The relationship between local DNA methylation and transcription factor–DNA interactions appears to be complex. Biochemical and genomic analyses have defined multiple transcription factors whose productive interaction with local DNA sequence is blocked by cytosine methylation that occurs within cognate recognition sequences and other transcription factors whose binding is facilitated by DNA methylation. So-called pioneer transcription factors are widely believed to have the inherent capacity to penetrate local chromatin-based barriers to binding, giving them the capacity to direct alterations in cell identity. Further, transcription factor binding has been posited as a mechanism wherein local CpG dinucleotides are protected from action of DNA methyltransferases, leading to local hypomethylation. These observations suggest that different transcription factors may influence, or be influenced by, local DNA methylation patterns in different ways. The downstream output of gene transcription is also likely to be influenced in a complex manner dependent on rate-limiting transcription factors.

Here, we address the relationship of DNA methylation patterns in somatic tissue across generation using inbred mouse strains in a genetic model system. Our findings demonstrate thousands of local sites where different strains of inbred mice, grown in identical conditions, differ in DNA methylation pattern. These genotype-dependent differences in local DNA methylation are preserved on parental alleles in hybrid F1 progeny, suggesting linkage to DNA sequence. We suggest that the linkage of DNA methylation state to DNA sequence results, in part, from its relationship to transcription factor biology. In some cases, genetic control of transcription factor binding correlates with differential methylation in our genetic system, as observed for other epigenetic marks and as has been reported for DNA methylation.

In other cases, it seems likely that local DNA methylation influences the quality of transcription factor interaction with local DNA sequences, either in a positive or negative manner. Furthermore, in comparison of animals of different sex and life history, we find that major life events such as pregnancy may leave a DNA methylation signature in nonreproductive somatic tissues at loci whose expression is linked to the relevant biology.

Results

A genetic system for study of DNA methylation. To address the degree of similarity of DNA methylation patterns in a somatic tissue when comparing parents to offspring, we crossed C57BL/6N and C3H/HeN mice (subsequently referred to as B6 and C3) in both directions to derive offspring (both male and female F1s) from a total of six crosses (three of each type). Animals were reared in a controlled environment and were sacrificed at identical ages to minimize confounding variables. We chose livers as a somatic tissue of interest and prepared genomic DNA for further analysis. We performed whole-genome sequencing (30–35× genome coverage, Supplementary Data 1) and constructed local genome assemblies for B6 and C3, from which we identified approximately 2.8 million autosomal single-nucleotide variations (SNVs) from the reference (mm9) genome.

Shotgun whole-genome bisulfite sequencing was performed (see Methods) on all 24 animals, collecting approximately 1.8 × 10^12 nucleotides after filtering, mapping and deduplication (Supplementary Data 1). This data set represents total read depth of approximately 150× genome coverage each for B6, C3, and the hybrid F1 progeny, B6C3F1 and C3B6F1 (Supplementary Data 1; Supplementary Figure 1). Global methylation levels were similar to previously reported values, (Supplementary Figure 1, Supplementary Table 1, Supplementary Table 2). In the strains utilized for this study, we identified 18.8 million autosomal CpG dinucleotides that were conserved in both our local genomes as well as the reference genome; we limited all further analysis to these 18.8 million CpGs.

Methylation variability at the individual CpG level. As one goal of our study was to compare CpG methylation patterns across groups of animals, we began our analysis by assessing the extent of gross methylation variability across the animals in our study. We calculated the mean and standard deviation of the percent of methylated alleles within biological replicate groups (e.g., B6 animals) for 20,000 randomly selected CpGs with different coverage levels (Fig. 1a). CpG sites with less than 10× coverage had a mean standard deviation of slightly greater than 10%. This value declined with increasing read depth, approaching a limit of approximately 5% (Fig. 1a). Coverage depth beyond 11–20× coverage provided minimal improvement in standard deviation, suggesting that this level of variability is an inherent property of the biological system or is technical in nature.

Even accounting for CpG sites lacking sufficient sequencing depth, there exist outliers with high standard deviation, suggesting a subset of CpGs may be inherently variable in methylation level across the population of alleles sampled. We sought to define the number of such outliers by plotting standard deviation (SD) across all animals (Fig. 1b). While approximately 95% of all CpG sites had SD less than 10%, a subset of CpGs were more highly variable (Fig. 1b). These CpGs are located in regions with higher CpG density than randomly selected controls, have intermediate levels of CpG methylation, and tend to be excluded from CpG islands (Fig. 1c), similar to previous reports. While this set of CpGs may have some unique biology, we could not identify a logical rationale for their exclusion from subsequent analyses.

Differentially methylated regions by genotype. To assess the impact of genetics on DNA methylation patterning, we sought to identify regions where methylation differed between parents, to ask whether the pattern in F1 offspring resembled one parent or differed from both. Currently, there is not a consensus statistical method for identification of differentially methylated genomic regions in whole-genome bisulfite sequencing data. Therefore, we employed two different approaches which use fundamentally different strategies. DSS uses a beta-binomial hierarchical model with a parametric statistical test. Metilene employs a scoring method to identify maximal between-group methylation differences in genomic regions of minimum length in combination with a nonparametric test. After employment of each tool, we refined the results with the following filters: (1) minimum methylation difference > 20%, (2) minimum of five CpG sites per DMR, (3) disregard any DMRs where one group has read depth > 500×, and (4) disregard any DMRs where the ratio of average
The DMRs identified each computational method (Fig. 2a; Supplementary Figure 3A). With most DMRs called by Metilene also called by DSS (Fig. 2b; Supplementary Figure 3B). When we assessed the genomic intervals; the vast majority are less than 1 kb in size (Table 1). Both computational methods called DMRs of each polarity (B6 promoters; roughly 70% of all DMRs are more than 10 kb away from the nearest RefSeq transcription start site (Fig. 2c; Supplementary Figure 3C). Consistent with the minimum methylation difference specified as a parameter in DMR calling, most DMRs have a substantial difference in methylation between the two strains (Fig. 2d; Supplementary Figure 3D). CpG content in DMRs was consistently higher than in random genomic intervals (Fig. 2e; Supplementary Figure 3E) although it did not reach the CpG density characteristic of CpG islands. We compared the DMRs with existing data on histone modifications and nuclease accessibility, finding that about 20% of DMRs overlap with DNase hypersensitive sites as measured in C57BL/6J liver (Fig. 2f; Supplementary Figure 3F), suggesting overlap with regulatory DNA. Further, nearly one-half of all DMRs also overlap with regulatory DNA, most likely at enhancers, a pattern observed in other types of comparisons.

**Parental DNA methylation is recapitulated in progeny.** DNA methylation scores at the DMRs identified by comparison of B6 vs. C3 parental animals were calculated and visualized by hierarchical clustering of all study animals including progeny (Fig. 3a; Supplementary Figure 4A). The study animals cluster by genotype, with hybrid F1 animals occupying a position intermediate between the parental extremities. DNA methylation scores at the DMRs identified by comparison of B6 vs. C3 parental animals were calculated and visualized by hierarchical clustering of all study animals including progeny (Fig. 3a; Supplementary Figure 4A). The study animals cluster by genotype, with hybrid F1 animals occupying a position intermediate between the parental extremities.
between the parental strains. The DNA methylation levels were visualized across all animals by calculating the average weighted methylation score for each DMR, binning the data into quartiles, and visualizing as box and whisker plots (Fig. 3b; Supplementary Figure 4B). In all cases, DNA methylation level in the F1 progeny, regardless of the direction of the genetic cross, is intermediate between the parental levels. The simplest explanation for these results is that F1 progeny recapitulate DNA methylation patterns in cis on alleles inherited from their respective parents. To directly visualize allele specificity of DNA methylation in F1 progeny at select DMRs, we examined individual bisulfite reads where a DMR CpG and a diagnostic sequence difference between strains fell on the same sequence read (or read pair). We found that the DNA methylation level of parental alleles in F1 animals appears nearly identical to that found in the parents in the overwhelming majority of cases; exemplar DMRs are depicted (Fig. 3c; Supplementary Figure 4C). These data demonstrate that DNA methylation patterns are recreated on respective parental alleles in somatic tissue in offspring regardless of which germ line they pass through.

Recapitulation of methylation in an allele-specific manner implies a linkage of methylation pattern to genetic information. Accordingly, we assessed the relationship of DMRs with strain-specific differences in DNA sequence. DMRs, when compared to random genomic intervals of the same size, have a higher frequency of local SNVs (Fig. 3d; Supplementary Figure 4D). Further DMRs that do not contain a SNV are, in general, closer to the nearest SNV than comparably sized random genomic intervals. For the DMRs identified by DSS, approximately 2/3 of all DMRs contain a SNV and approximately 90% fall within 1 kb of a SNV (Fig. 3e); for DMRs identified by Metilene, more than half contain a SNV and more than 80% are within 1 kb of a genetic difference between strain (Supplementary Figure 4E). These findings illustrate the relationship between local differences in DNA sequence and allele-specific methylation in offspring, as well as support a determining role for local DNA sequence in the establishment or maintenance of local epigenetic information.

### Table 1 Characteristics of differential methylation

| Comparison                      | Tool   | DMCs | DMRs | DMRs by type | DMCs in DMRs | % DMCs in DMRs | % DMR CpGs that are DMCs | DMR % of autosomes |
|---------------------------------|--------|------|------|--------------|--------------|----------------|--------------------------|------------------|
| B6 vs. C3                       | DSS    | 72,275 | 6380 | B6 > C3 3166 | 47,029       | 65.1           | 65.5                     | 0.143            |
|                                 |        |       |      | C3 > B6 3214 |              |                |                          |                  |
|                                 | Metilene | 185,194 | 2569 | M > F 534    | 23,839       | 12.9           | 78.8                     | 0.042            |
|                                 |        |       |      | F > M 1471   |              |                |                          |                  |
| Male vs. female                 | DSS    | 12,224 | 1575 | M > F 245    | 8532         | 69.8           | 58.4                     | 0.036            |
| (B6 + C3 animals)               |        |       |      | F > M 172    |              |                |                          |                  |
|                                 | Metilene | 51,675 | 439  | M > F 415    | 3426         | 6.6            | 81.1                     | 0.008            |
|                                 |        |       |      | F > M 668    |              |                |                          |                  |
| Male vs. female                 | DSS    | 7400   | 1083 | M > F 25     | 5108         | 69.0           | 51.8                     | 0.020            |
| (B6 animals)                    |        |       |      | F > M 45     |              |                |                          |                  |
|                                 | Metilene | 37,327 | 70   | M > F 25     | 653          | 1.8            | 54.6                     | 0.002            |
|                                 |        |       |      | F > M 45     |              |                |                          |                  |
| Male vs. female                 | DSS    | 11,142 | 1539 | M > F 548    | 7495         | 67.3           | 53.1                     | 0.029            |
| (C3 animals)                    |        |       |      | F > M 991    |              |                |                          |                  |
|                                 | Metilene | 62,785 | 93   | M > F 29     | 1018         | 1.6            | 64.4                     | 0.003            |
|                                 |        |       |      | F > M 64     |              |                |                          |                  |
| Male vs. female                 | DSS    | 6007   | 720  | M > F 65     | 4053         | 67.5           | 59.1                     | 0.017            |
| (F1 animals)                    |        |       |      | F > M 65     |              |                |                          |                  |
|                                 | Metilene | 32,574 | 207  | M > F 26     | 1655         | 5.1            | 83.2                     | 0.004            |
|                                 |        |       |      | F > M 181    |              |                |                          |                  |
| Parental female                 | DSS    | 2257   | 305  | p > v 10     | 1467         | 65.0           | 51.7                     | 0.006            |
| vs. virgin female               |        |       |      | v > p 10     |              |                |                          |                  |
|                                 | Metilene | 17,798 | 68   | p > v 2      | 437          | 2.5            | 61.8                     | 0.001            |

**Transcription factor–DNA interactions and DNA methylation.** As allele-specific histone modifications have been linked to genetic control of transcription factor–DNA interactions,20–24,30, we explored the possibility that a similar mechanism might correlate with the allele-specific DNA methylation observed here. We performed computational searches for known transcription factor binding sites at DMRs after extending each region of interest to a minimum size of 401 bp. A total of 20 of the approximately 140 motifs corresponding to transcription factors expressed in mouse liver were enriched over background in these genomic intervals (Fig. 4a, b; Supplementary Figure 5). The observed motifs were enriched in binding sites for transcription factors that function in liver development, hepatocyte differentiation, and liver homeostasis as opposed to transcription factors that function exclusively in other cell types. Enriched binding sites differ substantially in type of DNA-binding domain (Forkhead/winged helix; bZIP; homeobox; bHLH; Zinc finger).

We next considered what the relationship between transcription factor interaction and DNA methylation might involve. In principle, higher levels of DNA methylation at a DMR might negatively influence the capacity of a given transcription factor to bind its cognate recognition element7–11. Conversely, productive interaction of transcription factors with regulatory DNA could lead to local decreases in DNA methylation through physical hindrance of DNA methyltransferase15–19 or through chromatin-based mechanisms.20–24,34. Five enriched motifs (NRF1, Arnt:Ahr, CEBP, USF1, and USF2) have a CpG dinucleotide; binding of these transcription factors may be blocked by modification of the CpG within their consensus binding motif7–11. For the remaining motifs, the lack of CpG dinucleotides in the consensus binding element suggests that increased methylation within DMRs containing these motifs may result from loss of transcription factor binding.

Accordingly, we next asked whether SNVs within or adjacent to DMRs can interfere with DNA binding in a biochemical assay. Purified FoxA1 DNA-binding domain was assessed for the ability to bind to sites that fall within DMRs and have a variant sequence.
between the two strains within the binding site (Fig. 4c, e, f, Supplementary Figure 6). In most cases, recombinant FoxA1 DNA-binding domain exhibited a preference for the sequence found in the strain with lower methylation levels at the DMR tested. We asked whether these transcription factor binding results were reflected in vivo by performing ChIP-seq for FoxA1 in liver from B6 and C3 male animals. Analysis of binding at FoxA1 sites within or near a DMR that contain a SNV revealed that FoxA1 binding to these sites was impacted by the genetic differences between strain in the same manner as in the in vitro biochemical assays (Fig. 4d). Where the allele was more methylated in C3 animals, ChIP signal was higher in B6 animals and vice versa (Fig. 4e, f). At these loci, the ChIP-seq data are entirely consistent with genetic control of FoxA1 binding within/or near DMRs correlating inversely with local methylation status. These data suggest that local DNA sequence within a select group of genomic transcription factor binding sites has an influence on local DNA methylation. This finding is consistent with a multitude of literature reports linking local DNA methylation state to transcription factor binding.

Differentially methylated regions by sex and life experience. We capitalized on the design of our breeding experiment to ask whether autosomal DNA methylation patterns differ by sex. As pregnancy is known to impact liver biology in rodents, we first assessed whether the livers of dams studied here had normalized from pregnancy-associated hypertrophy by histologic examination of sections prepared from the right liver lobe of all study animals (see Methods). We observed no differences in histologic features in any group suggesting general features of cellular content are similar in all study animals. When comparing dams to virgin females or to males, we observed no difference in cell size, suggesting that hepatocytes in dams had returned to baseline at the cellular level at the time of sacrifice (Supplementary Figure 7). Accordingly, we utilized both DSS and Metilene to identify DMCs and differentially methylated regions (DMRs) upon comparison of the dams in our breeding experiment (both the C3 and B6 female animals) with the group of sires (both C3 and B6 male animals). DSS identified 12,224 DMCs based on sex which underlie 1575 DMRs; Metilene identified 51,675 DMCs and 439 DMRs (Table 1). The total number of DMRs, regardless of
analysis stream, were asymmetric in terms of polarity with female \( \gg \) male DMRs being found roughly twice as often as the converse (Supplementary Data 3). This dataset confirms that our animals have larger differences in autosomal DNA methylation pattern based on genotype than on sex in the tissue studied (i.e., 1575/439 DMRs based on sex vs. 6380/2569 DMRs based on genotype).

Exemplar DMRs are illustrated in browser format for each analysis stream (Fig. 5a, Supplementary Figure 8A). Like regions that differ in comparison across genotype, sex-dependent DMRs are relatively small, the overwhelming majority under 1 kb in size (Fig. 5b, Supplementary Figure 8B). Sex-dependent DMRs, like genotype-dependent DMRs, fall distant from mapped transcription start sites with greater than 60% falling more than 10 kb from the nearest TSS (Fig. 5c; Supplementary Figure 8C). Differences in methylation level between the two groups compared is similar in sex \( \gg \) DMRs and in genotype \( \gg \) DMRs (Fig. 5d; Supplementary Figure 8D). Like their genotype-driven counterparts, sex \( \gg \) DMRs have higher CpG density than random controls but lack the

---

**Fig. 3** DNA methylation patterns at DSS genotype-DMRs are recapitulated on parental alleles in F1 progeny. **a** Heatmap view of weighted methylation scores per B6-vs.-C3 DSS DMR (\( N = 6380 \)) per animal. Hierarchical clustering of animals performed by R package "amap" (cluster with method = euclidean and link = average). DMRs were split by direction (B6 > C3: red bar, or C3 > B6: green bar) then sorted by average methylation score over all 24 animals. **b** Distribution of average weighted methylation scores, with DMRs (\( N = 6380 \)) split into quartiles according to decreasing methylation score of the hypermethylated parental genome. The comparison between B6 and F1 or between C3 and F1 are significant at \( p < 1 \times 10^{-80} \) (Mann-Whitney) in all quartiles. **c** Methylation in F1 progeny at the read level for an exemplar DSS DMR (B6 \( \gg \) C3). Each box represents the collection of read fragments (for a given F1 genotype) that could be unambiguously assigned as originating from either the B6 parent (left side) or the C3 parent (right side) based on the presence of a diagnostic SNV. Each row within a box and column represents one sequenced fragment, with each CpG site indicated by a circle colored according to the following scheme: yellow = unmethylated cytosine, blue = methylated cytosine, red = noncytosine, and empty/gray = no mapped base at given CpG site in given fragment. **d** Number of SNVs local to DMRs (\( N = 6380 \)), compared to size-matched random genomic regions (\( N = 6380 \)). \( p < 1 \times 10^{-300} \) (Mann-Whitney test). **e** Distribution of distance to the nearest SNV for DMRs (\( N = 6380 \)) or size-matched random genomic regions (\( N = 6380 \)). Distance = 0 indicates that at least one SNV is found within a region of a given type. In the box-and-whisker plots, the box depicts the 25th to 75th percentiles, the black dot is the median, the whiskers extend to data points up to 1.5*IQR beyond the box, and open gray circles are data points outside the whisker range.
Fig. 4 SNVs at DMRs affect DNA-binding activity of a tissue-specific transcription factor. 

** c Top 10 enriched TF motifs in B6 > C3 DSS DMRs (after expanding DMR size to a minimum of 401 bp) according to HOMER v4.9.1. b Top 10 enriched TF motifs in C3 > B6 DSS DMRs (after expanding DMR size to a minimum of 401 bp) according to HOMER v4.9.1. c Protein purification of FOXA1 DNA-binding domain (DBD). Purified FOXA1 DBD was analyzed by SDS-PAGE. The black arrow indicates FOXA1 DBD. d FOXA1 ChIP-seq signal at SNV-containing FOXA1 motifs. Occurrences of FOXA1 motifs (as defined by HOMER motifs FOXA1.LNCAP or FOXA1.MCF7) within 100 nt of a DMR were categorized by whether the B6 or C3 allele was preferred according to the HOMER position weight matrix (PWM). FOXA1 signal in the B6 and C3 animals is reported as input-subtracted ChIP-seq read count at 100 nt regions centered on the motif occurrences. The box depicts the 25th to 75th percentiles, the black dot is the median, the whiskers extend to data points up to 1.5*IQR beyond the box, and open gray circles are data points outside the whisker range. Reported p values are from Mann-Whitney test. e, f DNA-binding analyses of FOXA1 DBD with SNV-containing DNAs. The genomic positions and DNA substrates are indicated at the top of each panel with a representative FOXA1 motif. The position of the SNV is highlighted with light blue boxes. The band intensities were calculated using ImageJ software. Totally, 20 bp double-stranded DNAs were incubated with the FOXA1 DBD. The concentration of FOXA1 DBD was as follows: 0 μM, lanes 1, 6; 0.15 μM, lanes 2, 7; 0.3 μM, lanes 3, 8; 0.6 μM, lanes 4, 9; 1.2 μM, lanes 5, 10. The protein-DNA complex was separated by native-polyacrylamide gel electrophoresis. Browser tracks of ChIP-seq data for each locus is shown below the DNA-binding data.
Finally, like genotype–DMRs, sex–DMRs also colocalize to an appreciable extent with DNAse hypersensitivity and with the enhancer mark, H3K4me1 (Fig. 5f; Supplementary Figure 8F).

We next performed clustering and visualization of DNA methylation levels of all 24 animals in the study at the sex–DMRs precisely as we did with genotype–DMRs (Fig. 6a; Supplementary Figure 9A). We observed that female > male DMRs behaved as expected; all female animals had high methylation levels while all male animals had lower levels (Fig. 6a, b; Supplementary Figure 9A, B). Surprisingly, the male > female DMRs had a different outcome. We observed, at this subset of loci, that female hybrid F1 animals had methylation levels that more closely resembled males than the parental female animals (Fig. 6a, b; Supplementary Figure 9A, B). Unlike the genotype–DMRs described above, there was no apparent allele specificity to DNA methylation at sex–DMRs (Fig. 6c; Supplementary Figure 9C), nor was there a proximity relationship to genetic differences between strain (Fig. 6d, e; Supplementary Figure 9D, E).

To investigate the unexpected behavior of sex–DMRs in female animals, we used the Genomic Regions Enrichment of Annotations Tool36 to predict functions of the male > female sex–DMRs. Genes predicted to be regulated by these loci were enriched in Gene Ontology terms related to pregnancy and lactation (Fig. 7a).

An exemplar gene from this category is the prolactin receptor gene located on chromosome 15 (Fig. 7b) which contains...
multiple male > female DMRs called by both analysis streams. Prolactin is known to be upregulated in liver in pregnant and lactating rodents. This finding is consistent with the design of our study wherein the female B6 and C3 animals analyzed were the dams of the female F1 animals (which were virgin females). These findings suggest that the class of male > female DMRs are dominated by genomic intervals in liver that lose DNA methylation downstream of pregnancy, lactation, or both.

We further explored the relationship of DNA methylation with pregnancy and lactation by defining DMRs in the comparison of dams (again including B6 and C3 animals) with daughters (F1 hybrid virgin females sacrificed at the same age). DSS identified 2257 DMCs and 305 DMRs in this comparison; Metilene identified 17,798 DMCs and 68 DMRs (Table 1). As anticipated, the polarity of DMRs was non-random, with 295 of 305 (and 66 of 68) DMRs having greater CpG methylation in virgin females than in dams (Supplementary Data 4). Of the 305 DMRs called by DSS in this comparison of dams versus F1 hybrid virgin females, 196 overlap M > F DMRs and 0 overlap F > M DMRs. This analysis suggests that DNA methylation patterns, when queried in an appropriate manner, can provide a record of life events such as pregnancy/lactation. This record appears to correlate with loci involved in the relevant biological response to pregnancy/lactation.

The sex–DMRs were also analyzed for enrichment of transcription factor motifs. The overwhelming majority of motifs...
identified in the male > female DMRs contain some version of the E-box sequence — CANNTG (Fig. 7c; Supplementary Figure 10A). Motifs in the female > male sex–DMRs are much more diverse, containing binding sites for homeobox factors, nuclear receptors, signal responsive transcription factors, bHLH, and bZIP factors (Fig. 7d; Supplementary Figure 10B). While many motifs enriched in this analysis do not contain CpG within their consensus motif, we note that Hox transcription factors can bind to methylated DNA, despite the lack of CG dinucleotides in their consensus recognition site11.
Discussion

The results presented here demonstrate that epigenetic features such as DNA methylation are closely linked to DNA sequence with patterns in parents recapitulated on the relevant allele in offspring. Our finding that DMRs colocalize to a large extent with distal regulatory DNA and transcription factor binding motifs suggests a potentially causal relationship between DNA sequence-dependent alterations in the interactions of DNA-binding proteins with their respective target sequences and changes in local DNA methylation\(^\text{22,23,24,31}\). Further, the data suggest that differential CpG methylation, acting downstream of genetic changes in transcription factor–DNA interaction, demarcate enhancers that are active in only one strain, or on a single allele in F1 progeny. The differences in DNA methylation pattern between inbred strains, occurring largely at enhancer loci where DNA methylation pattern may be programmed by transcription factor action, resemble the differences between cell/tissue type within strain which also closely colocalize with enhancers\(^\text{32,32}\). Our data add to the growing body of evidence associating alterations in the quality of interaction of a transcript, factor with its DNA target, with changes in epigenetic features, including DNA methylation\(^\text{20-24,33}\). We observe evidence for differences in local DNA methylation correlating with binding quality for only a subset of transcription factors expressed in mouse liver. Thus, our findings refine available models, and suggest that not all transcription factors have the inherent ability to reprogram local epigenetic state.

The linkage of CpG methylation to local DNA sequence in somatic tissues hinges, in part, on the capacity of genetic information to influence DNA/protein interactions and recruitment of chromatin modifiers either in the tissue of interest or in a developmental precursor undergoing programmed refinement of the pattern of DNA methylation. This model has profound implications for interpretation of changing chromatin marks, including DNA methylation, in studies of animals derived from outbred populations, including humans. Our results further demonstrate that patterns of DNA methylation are remarkably similar across generation. We find that the overwhelming majority of CpG methylation is preserved in comparison of parent with offspring. However, we also find that appropriate comparisons can reveal DNA methylation changes that correlate with a major life event even weeks afterwards, in this case pregnancy and/or lactation. Of interest, the DNA methylation changes distinguishing dams from their virgin daughters were persistent when assessed 4 weeks after weaning and after cytologic alterations in hepatocytes characteristic of pregnancy had returned to baseline. While we have not identified, to date, the mechanistic basis behind focal loss of CpG methylation at a few hundred loci in female animals that have borne and nurtured offspring, it is tempting to speculate action of a transcription factor(s) acting to regulate expression level of genes integral to that biology.

Histology. Three representative sections were cut at 5 μm thickness from formalin-fixed, paraffin-embedded blocks. Section 1 was stained with hematoxylin and eosin (H&E). Female mice from three groups of animals: virgins (B6 and C3), breeders (B6 dams and C3 dams) and first generation (B6C3F1 and C3B6F1), were evaluated for hyperplasia and hypertrophy. For assessment of hyperplasia, liver sections were examined using a Leica-DMLB microscope with a 40x objective. Ten fields per slide were analyzed for hepatocyte mitotic figures. Hypertrophy assessment was based on hepatocyte density counts generated from the same, H&E- stained liver sections scanned on an Aperio ScanScope XT instrument (Vista, CA) using ImageScope software, (v11.2.0.780, Aperio). Hepatocyte nuclei from 10, nonoverlapping, 90 × 90 μm\(^2\) fields were analyzed per section to determine cell density.

Genomic DNA extraction. Frozen liver tissues were quickly pulverized with BioPulverizer (Bio Spec Products Inc.) on dry ice. DNA was extracted with Allprep DNA/RNA/Protein Mini Kit (Qiagen) according to the manufacturer’s protocol. DNA was further purified by phenol–chloroform extraction followed by ethanol precipitation.

Whole-genome bisulfite sequencing library preparation. DNA (1 μg) was spiked with 1 ng unmethylated lambda DNA (Promega), fragmented (average size; 300 bp), end-repaired, A-tailed, and adapter-ligated using TrueSeq DNA sample prep kit (Illumina) according to the manufacturer’s protocol. Adapter-ligated DNA was purified by QIAquick gel extraction kit (Qiagen). After clean up with AMPure XP beads, bisulfite conversion was performed using EpiTect Bisulfite kit (Qiagen) with the following thermal cycles, 95 °C 5 min, 60 °C 25 min, 95 °C 5 min, 60 °C 85 min, 95 °C 5 min, 60 °C 175 min, 95 °C 5 min, 60 °C 180 min. After clean up with AMPure XP beads, bisulfite converted DNA was amplified with PfuTurbo Cx Hotstart DNA Polymerase with following thermal cycles, 95 °C 5 min, 98 °C 30 s, 12 cycles of (98 °C 10 s, 65 °C 30 s, 72 °C 30 s), 72 °C 5 min. DNA was cleaned with AMPure XP beads, and stored at −30 °C until use. Sequencing was performed in Hiseq2000 using PE100 base format.

Methods

Animal care. C57BL/6N and C3H/HeN mice were obtained from the NTP colony at Taconic Farms, Inc. (Germantown, NY). Animals were housed in an AAALAC accredited facility at Integrated Laboratory Systems (Durham, NC). Project-
SAM alignments for multiple libraries from the same animal were merged. Read pairs mapped to phage λ were used as a QC assessment to confirm that the observed bisulfite conversion was consistent. Read pairs mapped to the mm9 reference genome were used for downstream analysis.

Validation of genomic cytosine context. For each cytosine site in the mm9 reference genome, the fractions of mapped bases that are C at position N, G at position N + 1, and G at position N + 2 were calculated using WGBS data from all six B6 or from all six C3 parent animals. The cytosine context for a given genomic position was considered to be validated if the following criteria are met: (a) the expected cytosine context is consistent between the mm9 reference genome, the local C57Bl/6NJ assembly, and local CH/HeN assembly; (b) at least 75% of mapped bases at each of the N, N + 1, and N + 2 positions are consistent with the expected context using the B6 WGBS data; and (c) at least 75% of mapped bases at each of the N, N + 1, and N + 2 positions are consistent with the expected context using the C3 WGBS data. For CpG context, validation was required only on one of the two strands. For CHG and CHH context, each strand was evaluated independently. This process identified 18,887,127 validated autosomal CpG sites, 101,235,798 and 101,246,101 validated autosomal CHG sites (on the plus and minus strand, respectively), and 351,699,190 and 351,716,331 validated autosomal CHH sites (on the plus and minus strand, respectively). In this study, methylation analysis was limited to only autosomal cytosines with validated CpG context.

DMR detection. Using the DSS R package v2.15.0, DMCs were identified by DSS with the callDMR function (default parameters), and DMRs were identified with the callDMR function (pct.sig = 0.75, all other parameters default). DMC and DMR calls were also made via Metilene v.0.2–6 (-m 5 for DMR calls, all other the callDMR function (pct.sig = 0.75, all other parameters default), with a p-value threshold of 0.01 and mean methylation difference of 0.2 for DMCs and a q value of 0.05 for DMRs. All DMR calls from both tools were subject to additional filters, as described below. DMRs were required to contain at least five validated CpG sites and have a weighted methylation score

References
1. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492 (2012).
2. Schubeler, D. Function and information content of DNA methylation. Nature 517, 321–326 (2015).
3. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157, 95–109 (2014).
4. Bjornsson, H. T. et al. Intra-individual change over time in DNA methylation with familial clustering. J Am Med Assoc 299, 2877–2883 (2008).
5. Johnson, M. D. et al. Genetic analysis of the cardiac methylome at single nucleotide resolution in a model of human cardiovascular disease. PLoS Genet. 10, e1004813 (2014).
6. Orozco, L. D. et al. Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. Cell Metab. 21, 905–917 (2015).
7. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature 528, 575–579 (2015).
8. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482–485 (2000).
9. O’Malley, R. C. et al. Catrione and epicistrome features shape the regulatory DNA landscape. Cell 165, 1280–1292 (2016).
10. Hashimoto, H. et al. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. Mol. Cell 66, 711–720 (2017).
11. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356, 502 (2017).
12. Brandeis, M. et al. Methyl-dependent and spatial-specific DNA recognition by the orthologous transcription factors human AP-1 and Epstein–Barr virus Zta. Nucleic Acids Res. 45, 2503–2515 (2017).
13. Hashimoto, H. et al. Distinctive Klf4 mutants determine preference for DNA methylation status. Nucleic Acids Res. 44, 10177–10185 (2016).
14. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. Development 143, 1833–1837 (2016).
15. Bird, A. P. & Wolffe, A. P. Methylation-induced repression—belts, braces, and chromatin. Cell 99, 451–454 (1999).
16. Macleod, D., Charlton, J., Mullins, J. & Bird, A. P. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292 (1994).
17. Brandeis, M. et al. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438 (1994).
18. Mummmaneni, P., Yates, P., Simpson, J., Rose, J. & Turker, M. S. The primary function of a redundant Sp1 binding site in the mouse aprt gene promoter is to block epigenetic gene inactivation. Nucleic Acids Res. 26, 5163–5169 (1998).
19. Zhang, Y. et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional programming. Nat. Genet. 48, 1003–1013 (2016).
20. Kasowski, M. et al. Extensive variation in chromatin states across humans. Science 342, 750–752 (2013).
21. Kilpinen, H. et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science 342, 744–747 (2013).
22. McCleary, G. et al. Identification of genetic variants that affect histone modifications in human cells. Science 342, 747–749 (2013).
23. Stanler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory locations. Nature 480, 490–495 (2011).
24. Lienert, F. et al. Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. 43, 1091–1097 (2011).
Beatrice B. Barnabas2, Gerard G. Bouffard2, Shelise Y. Brooks2, Holly Coleman2, Lyudmila Dekhtyar2, Xiaoabin Guan2, Joel Han2, Shi-ling Ho2, Richelle Legaspi2, Quino L. Maduro2, Catherine A. Masiello2, Jennifer C. McDowell2, Cassandra Montemayor2, Morgan Park2, Nancy L. Riebow2, Karen Schandler2, Chanthra Scharer2, Brian Schmidt2, Christina Sison2, Sirintorn Stantripop2, Pamela J. Thomas2, Meghana Vemulpal2 & Alice C. Young2

Acknowledgments
This work was supported, in part, by the Intramural Research Program of the National Institute of Environmental Health Sciences, NIH (ES103146 to P.A.W.). The authors express gratitude to Drs. Richard Woychik, Trevor Archer, Angela Ting, and Paula Vertino for useful comments throughout the course of this work.

Author contributions
Study design: S.A., J.R.B., J.E.F., B.A.M., R.R.T., J.W.T., J.C.M., and P.A.W. Data generation: T.S., M.T., C.G.D., and N.I.S.C. comparative sequencing program. Data analysis: S.A.G., J.W.T., B.B., J.R.B., A.B., Y.D., C.G.D., J.F.F., J.L., B.A.M., R.R.T., X.X., P.B., D.C.F., J.C.M., and P.A.W. Data visualization: S.A.G., F.D., Y.D., T.W., and X.X. Manuscript preparation: All authors participated in drafting and editing of the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-08067-z.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2019