MOLECULAR AND CELLULAR BIOLOGY, Aug. 1997, p. 4322–4329 Vol. 17, No. 8
0270-7306/97/$04.00+0

A 5′ 2-Kilobase-Pair Region of the Imprinted Mouse *H19* Gene Exhibits Exclusive Paternal Methylation throughout Development

KIMBERLY D. TREMBLAY, KRISTEN L. DURAN, AND MARISA S. BARTOLOMEI*

Howard Hughes Medical Institute and Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Received 25 March 1997/Returned for modification 24 April 1997/Accepted 29 April 1997

The imprinted mouse *H19* gene is hypermethylated on the inactive paternal allele in somatic tissues and sperm. Previous observations from a limited analysis have suggested that methylation of a few CpG dinucleotides in the region upstream from the start of transcription may be the mark that confers parental identity to the *H19* alleles. Here we exploit bisulfite mutagenesis coupled with genomic sequencing to derive the methylation status of 68 CpGs that reside in a 4-kb region 5′ to the start of transcription. This method reveals a 2-kb region positioned between 2 and 4 kb upstream from the start of transcription that is strikingly differentially methylated in midgestation embryos. At least 12 of the cytosine residues in this region are exclusively methylated on the paternal allele in blastocysts. In contrast, a 350-bp promoter-proximal region is less differentially methylated in midgestation embryos and, like most of the genome, is largely devoid of methylation on both alleles in blastocysts. We also demonstrate exclusive expression of the maternal *H19* allele in the embryos that exhibit paternal methylation of the upstream 2-kb region. These data suggest that the 2-kb differentially methylated region acts as a key regulatory domain for imprinted *H19* expression.

The maternal and paternal genomes are not functionally equivalent in mammals (8, 27), and as a result, both parental genomes are required for normal development (17, 30). This requirement is likely due to the products of imprinted genes that are expressed exclusively from a single parental allele. For example, deletion of the paternal allele of the imprinted insulin-like growth factor type 2 (*Igf2*) gene results in mice that are smaller than their wild-type littermates, while deletion of the maternal *Igf2* allele has no phenotypic consequences (6, 7). While it is not clear how many genes in the mammalian genome are imprinted, 18 imprinted genes have already been defined in mice and humans (1).

The presence of imprinted genes poses a unique problem for the mammalian embryo. The embryo must be able to distinguish between the maternal and paternal alleles of these genes in order that the appropriate allele is expressed. It is widely assumed that the parental alleles are distinguished or marked prior to fertilization, presumably during gametogenesis, while they are in separate compartments. It has also been proposed that the mark must be stable throughout development and reversible such that both alleles are appropriately reset during gametogenesis. The best candidate for the allele-specific mark that satisfies these criteria is the methylation of the cytosine residue in CpG dinucleotides, which is associated with gene repression of autosomal genes as well as those genes on the inactive X chromosome (23, 34). Consistent with the proposed role of DNA methylation is that all imprinted genes that have been analyzed exhibit allele-specific methylation (21).

One widely studied imprinted gene is the maternally expressed mouse *H19* gene (3). This gene, which encodes an RNA that is highly expressed in embryonic tissues of endodermal and mesodermal origin (5, 20), is hypermethylated on the inactive paternal allele (2, 4, 11). In somatic tissues and sperm, the paternal allele is methylated over a 7- to 9-kb region that includes the structural gene and a 4-kb region 5′ to the start of transcription (2). To provide further support for the idea that this methylation might be the mark that assigns parental identity to the alleles of *H19*, we previously demonstrated that at least two CpGs located in the 5′ portion of the upstream domain are differentially methylated in the gametes and that this difference is preserved during embryogenesis (32), including the period of genome-wide demethylation that occurs during preimplantation development (18, 24). Thus, methylation of these cytosine residues could act as the parental-specific imprinting mark. In contrast, CpG dinucleotides located 5′ to the promoter and within the body of the *H19* gene were not differentially methylated in preimplantation embryos and could not be acting to distinguish the alleles (4, 32). Although only a few CpG dinucleotides have been assayed for methylation differences, these studies suggest that there might be a larger imprinting domain that is located in the 5′ portion of the upstream region.

In the present study, we have addressed this question in a systematic fashion by determining the methylation status of the maternal and paternal alleles in the region upstream of the *H19* structural gene at different developmental stages. We defined the methylation status of 68 CpG dinucleotides using a bisulfite mutagenesis protocol and found that the region between −2 and −4 kb relative to the start of transcription exhibits marked hypermethylation of the paternal allele. In contrast, a 350-bp region proximal to the start of transcription is less differentially methylated and is more likely to function in the maintenance of the imprint. Furthermore, the 5′ border of the differentially methylated domain appears to be located around −4 kb relative to the start of transcription because the 5′ CpG dinucleotides analyzed by bisulfite mutagenesis are not as differentially methylated as the remainder of the 2-kb upstream domain. Additionally, the region 5′ to this domain exhibits a reduced number of CpG dinucleotides, and those dinucleotides that are found within methylation-sensitive restriction endonuclease sites are methylated on both alleles.

* Corresponding author. Phone: (215) 898-9063. Fax: (215) 573-6434. E-mail: bartolom@mail.med.upenn.edu.

Since we also demonstrate exclusive expression of the maternal allele in preimplantation and postimplantation embryos, we propose that the 2-kb region located between −2 and −4 kb relative to the transcription start site determines the imprinted expression of the mouse *H19* gene.

MATERIALS AND METHODS

Mice. C57BL/6J (B) and *Mus musculus castaneus* (*M. m. castaneus*) mice were purchased from the Jackson Laboratory. To facilitate the isolation of F_1 hybrids, a strain of mice was derived which served as the source of the *M. m. castaneus H19* allele [B6(CAST-*H19*)], which is designated C in this report (32). Natural matings were used to generate timed embryos with the day after conception considered day 0.5. Blastocysts were isolated as previously described (32). Note that for F_1 hybrid embryos (e.g., B × C) the maternal parent is designated first.

DNA and RNA analysis. DNA was isolated as previously described (2). Ten micrograms of genomic DNA was digested with the indicated restriction enzymes and separated by electrophoresis on a 1% agarose gel. The DNA was transferred to nitrocellulose filters (28), and the filters were hybridized to radiolabeled probes prepared by nick translation (22). The filters were washed (33) and exposed to Kodak XAR-5 X-ray film. RNase protection experiments were performed as previously described (3).

Bisulfite modification. One to ten micrograms of embryonic DNA or DNA from 300 to 430 B × C blastocysts was digested overnight with a restriction enzyme which cuts outside the region of interest. The DNA was phenol:chloroform:isoamyl alcohol extracted (25:24:1), ethanol precipitated, and resuspended in 100 μl of TE (Tris-EDTA) buffer (30 μl of TE for blastocysts). The DNA was then subject to denaturation with a final concentration of 0.3 M NaOH for 20 min at 42°C. Freshly prepared 10 mM hydroquinone and 3.6 M sodium bisulfite were added to final concentrations of 0.5 mM and 3.1 M, respectively. The reaction mix was overlaid with mineral oil and incubated in the dark at 55°C for 16 to 20 h (10, 12). After incubation, the DNA was purified with the Geneclean II kit (Intermountain Scientific Corporation). The purified sample was resuspended in 100 μl of TE (25 μl for the blastocysts) and then denatured in 0.3 M NaOH at 37°C for 15 min. After neutralization with ammonium acetate to a concentration of 3 M, the DNA was ethanol precipitated and resuspended in 100 μl of TE. The treated DNA was stored in the dark at −20°C. A treated sample was used as quickly as possible and was never used for more than 1 month.

***H19* bisulfite-specific PCR primers.** The primers used to generate products from bisulfite-altered DNA are specific for the top strand of mutagenized DNA. The sequences of the primers, with the nucleotide position of the first base indicated in parentheses (GenBank accession number U19619), are as follows: BMsp1t3, 5′-ATAGATTAGATTTGAGGGGAAGAGT-3′ (763); BMsp1t4, 5′-GATTGATAAAGGTATGTTAATTTG-3′ (845); BMsp1t5, 5′-AATTCTTATACCTCCTAAATACTC-3′ (1301); BMsp1t6, 5′-CCAAAAATAAACATAATCTCCTTAC-3′ (1328); BMsp2t1, 5′-GAGTATTTAGGAGGTATAAGAATT-3′ (1278); BMsp2t2, 5′-GTAAGGAGATTATGTTTATTTTTGG-3′ (1304); BHha1t4, 5′-CCTCATTAATCCCATAACTAT-3′ (1726); BHha1t3, 5′-ATCAAAAACTAACATAAACCCCT-3′ (1751); BHha2t1, 5′-ATAGTTATGGGTTTTATGAGG-3′ (1706); BHha2t2, 5′-AGGGGTTTATGTTAGTTTTTGATAA-3′ (1729); BMsp3t, 5′-ACACCCAAAACTTAATATAAAATTCC-3′ (2131); BMsp3t2, 5′-CCTCTTCAATTAATTTTAACT-3′ (2153); BMsp4t1, 5′-GGAATTTTATATTAAGTTTTGGGTGT-3′ (2106); BMsp4t2, 5′-AGTTAAAATTAATTGAAGAGG-3′ (2133); BHha4t3, 5′-ATTCCAACCTCTTATAAACCATAT-3′ (2538); BHha4t2, 5′-AACCCCCTCCAAAAACTCAAAT-3′ (2560); BHha5t2, 5′-TTGTGAGTGGAAAGATTAAATTGTTTGG-3′ (2355); BHha5t, 5′-TAGAGATAGTTAAAGTTAAGGTTTGTTTATG-3′ (2415); BHha5t3, 5′-ATACACACATCTTACCACCCCTATAAATCCC-3′ (2748); BHha6t, 5′-GGGATTTATAGGGGTGGTAAGATGTGTGTAT-3′ (2778); BHha7t2, 5′-ACCTAAAATACTCAAACTTTATCACAAC-3′ (3192); BHha7t, 5′-TAAAATATCACAAATACCTAATCCCT-3′ (3225); BMsp6t, 5′-GGTTGAGGATTTGTTAAGGTGTTATTG-3′ (4343); BMsp6t2, 5′-GAGTGGTTATGATTGGTTAGTTTTTGAG-3′ (4395); and BMsp7t, 5′-TAATAACTAATTTAAACACTCCTCACC-3′ (4777).

PCR amplification, cloning, and sequencing. Since much of the DNA was degraded as a result of the harsh mutagenesis conditions, two rounds of PCR were performed with fully or partially nested primer pairs. Each 25-μl reaction mixture contained 0.1 to 4.0 μl of bisulfite-treated DNA, 12.5 ng of each primer, 1.5 mM NaCl, 100 μM (each) deoxynucleoside triphosphate, and 0.5 U of Ampli*Taq* DNA polymerase (Perkin-Elmer). The primer sets used for the first and second rounds of PCR at each region are listed in Table 1. First-round PCR was performed under the following conditions: 4 min at 94°C, 2 min at 55°C, and 2 min at 72°C for two cycles. Thirty-five cycles of PCR were then performed for 1 min at 94°C, 2 min at 55°C, and 2 min at 72°C. For the second round of PCR, 0.01 to 1.0 μl of the first-round sample was used. The conditions for second-round PCR were the same as those for the first round except that the first two cycles were omitted. The PCR products were cloned with the TA Cloning kit (Invitrogen), and double-stranded sequencing was performed on positive clones with the Sequenase kit (U.S. Biochemical). To determine the parental origin of the PCR products, sequence polymorphisms were identified between the parental strains (Table 1). The number of CpGs assayed by each set of primers is indicated.

TABLE 1. Primer sets and types and locations of polymorphisms

PCR round 1/round 2 primer set	No. of CpGs assayed	B→C polymorphism	Position of polymorphism (nt)[a]
5′BMsp1t3-3′BMsp1t6/5′BMsp1t4-3′BMsp1t5	6	T→G	964
5′BMsp2t1-3′BHha1t3/5′BMsp2t2-3′BHha1t4	14	G→A	1566
5′BHha2t1-3′BMsp3t2/5′BHha2t2-3′BMsp3t	6	G→A	2016
5′BMsp4t1-3′BHha4t2/5′BMsp4t2-3′BHha4t3	10[b]	A→G	2420
5′BHha5t2-3′BHha5t3/5′BHha5t-3′BHha5t3	9[b]	A→T	2569
5′BHha6t-3′BHha7t/5′BHha6t-3′BHha7t2	18	G→A	2891
5′BMsp6t-3′BMsp7t/5′BMsp6t2-3′BMsp7t	9	C→G	4432

[a] nt, nucleotide.
[b] Four CpGs overlap between the two sets of primers.

To obtain the bisulfite data, the presence of a band in the cytosine lane was recorded as a methylated cytosine. To ensure the thoroughness of each bisulfite treatment, a YAC clone which harbors the murine *H19* locus, and is free of any methylated cytosine residues, was treated. After each mutagenesis, several strands were sequenced, and the mutagenesis reaction was considered complete only if greater than 99% of the cytosines were converted to thymines. A total of 13 bisulfite mutagenesis reactions were performed on the DNA from midgestation embryos. The data obtained from each set of primers were compiled from at least two separate mutagenesis reactions and several PCRs. No bias was observed between the 11.75 days postcoitum (dpc) B × C and the 13.5 dpc C × B midgestation samples. For each bar in the graphs of the midgestation embryo and sperm samples shown in Fig. 2, 3, 5, and 6, 8 to 32 clones were sequenced. For the blastocyst DNA, two mutagenesis reactions were performed, and 7 to 21 clones were sequenced for each allele.

RESULTS

Bisulfite mutagenesis of genomic DNA from midgestation embryos. Using genomic DNA analysis, we previously demonstrated that a 4- to 6-kb region located upstream of the *H19* promoter was hypermethylated on the inactive paternal allele in adult somatic tissues and sperm (2). When this region was tested for allelic methylation differences in oocytes and preimplantation embryos, it was found that three CpG dinucleotides contained within methylation-sensitive restriction sites were preferentially methylated on the paternal allele (32). Two of the sites, *Hha*I site 1 and *Hha*I site 5 (Fig. 1), were methylated exclusively in sperm and on the paternal allele in four-cell embryos, morulae, and blastocysts. More recently, we have determined that *Hpa*II site 1 and *Hha*I site 2 are exclusively methylated on the paternal allele in blastocysts (data not shown). Because these sites are differentially methylated at all times assayed thus far, they are candidates for the mark that designates parental identity.

Interestingly, these initial analyses indicated that the distal upstream sites exhibited more stringent methylation differences than those sites located adjacent to the promoter or within the body of the gene (4, 32). However, because the initial analyses utilized a PCR-based assay which requires that CpG dinucleotides reside within recognition sequences assayable by methylation-sensitive restriction endonucleases (i.e., *Hha*I and *Hpa*II), only a small percentage of the CpG dinucleotides in the upstream region and transcription unit were analyzed (Fig. 1). To determine whether the upstream sites were part of a larger differentially methylated region, we employed the bisulfite mutagenesis procedure that assays the methylation status of all CpG dinucleotides (10, 12). In this method, sodium bisulfite treatment of DNA converts cytosine residues to uracil but leaves 5-methylcytosine unchanged. Following mutagenesis, the DNA is PCR amplified, cloned, and sequenced. The cytosine residues that were methylated in the original

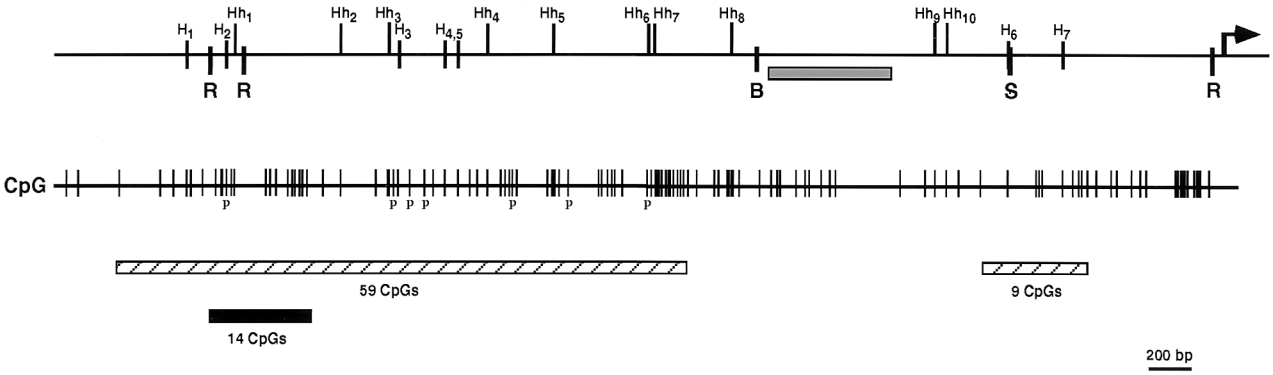

FIG. 1. Location of CpG dinucleotides in the upstream region of the *H19* gene. A 4.6-kb region upstream of the transcription start site (arrow) is depicted on the top line. The *Hha*I (Hh, vertical lines above the gene line) and *Hpa*II (H, vertical lines through the gene line) methylation-sensitive restriction endonuclease sites are indicated on the top line. Other restriction endonuclease sites include *Eco*RI (R), *Bam*HI (B), and *Sac*I (S). The location of the G-rich repetitive element is designated by the shaded box. The bottom line shows the location of all CpG dinucleotides found in the upstream region, as indicated by the sequence designated GenBank no. U19619 (32). Those CpG dinucleotides which are polymorphic in the strains used for the bisulfite mutagenesis assay (C57BL/6J and *M. m. castaneus*) are indicated by the letter P. These dinucleotides were not included in the bisulfite data. The two hatched boxes correspond to the regions analyzed by bisulfite mutagenesis of sperm and midgestation embryonic DNA (the number of CpGs is shown underneath the boxes). Due to the repetitive nature of the DNA between these hatched boxes, the methylation status of the corresponding CpG dinucleotides could not be determined by the bisulfite mutagenesis procedure. The regions designated by the black box (14 CpGs) and the promoter-proximal hatched box were analyzed by bisulfite mutagenesis of blastocyst DNA.

sample will appear as cytosine on the sequencing gel while cytosines that were unmethylated in the original sample will appear as thymine.

The bisulfite mutagenesis method provides a more informative methylation analysis for several reasons. First, as stated above, it directly assays all CpGs within the region defined by the PCR primers. Second, each sequence represents a single strand of DNA that was present in the original sample and therefore describes the methylation profile of a single chromosome, as demonstrated in Fig. 4. Third, unlike the PCR-based methylation-sensitive restriction enzyme assay, this technique is quantitative because it determines the number of DNA strands that are methylated at a given residue. Finally, the parental origin of each strand of DNA can be unambiguously assigned based on sequence polymorphisms between the two strains of mice used to generate the embryonic DNA samples.

To determine whether a methylation imprint was located 5′ to the start of transcription we analyzed two distinct regions in sperm and midgestation F_1 hybrid embryos, one that is highly likely to harbor a methylation imprint that is inherited from sperm and preserved throughout development, and one that is less likely to harbor such an imprint (Fig. 1). The 5′-most region is 2.2 kb in length and contains 59 assayable CpGs, 4 of which are potential candidates for the imprinting mark. The 3′ region is 350 bp in length and harbors nine CpGs, two of which are not candidates for the imprinting mark. Results of the analysis of the upstream region are presented in Fig. 2. The maternal allele was hypomethylated relative to the paternal allele along most of the 2.2-kb region (Fig. 2A). Of the 59 CpG dinucleotides assayed, only 22 exhibited any cytosine methylation. With the exception of the four cytosine residues at the 5′ end of this domain, the highest percentage of maternal clones that were methylated at any one of the cytosines was 25%. Furthermore, methylation of one cytosine residue within a maternally inherited clone was independent of the methylation status of adjacent CpG residues. Conversely, the paternal allele was hypermethylated at every cytosine residue within the entire 2.2-kb region (Fig. 2B). Except for the seven CpG dinucleotides at the 5′ end, many of the cytosines were methylated on every strand analyzed, and most were methylated on greater than 80% of the strands. Sperm DNA was also highly

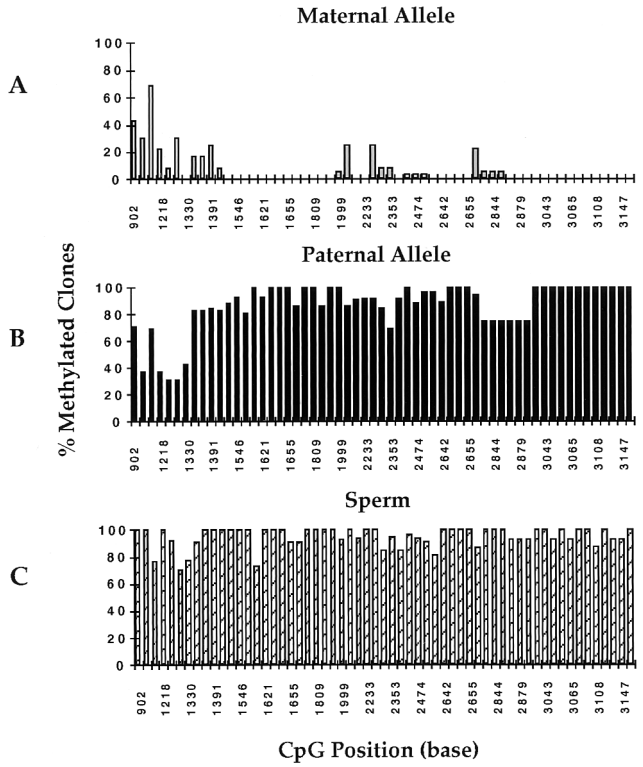

FIG. 2. Methylation profile of 59 CpG dinucleotides as assayed by the bisulfite mutagenesis sequencing assay. The 2.2-kb region (indicated by the hatched box labeled 59 CpGs in Fig. 1) is approximately 2 kb upstream from the *H19* transcriptional start site. The first assayed cytosine is located at nucleotide 902, and the last cytosine is located at nucleotide 3164 (GenBank no. U19619; the nucleotide position of every fourth CpG is shown). Each bar along the horizontal axis represents a CpG dinucleotide, and the numbers beneath the bars indicate the nucleotide positions of the cytosine residues. The height of each bar indicates the percentage of clones that were methylated. The DNA used to obtain data for the maternal (A) and the paternal alleles (B) was derived from two reciprocally crossed F_1 hybrid embryos: 11.75 dpc (B × C) F_1 embryos and 13.5 dpc (C × B) F_1 embryos, respectively. The sperm DNA was derived from B6(CAST-*H19*) adults (C). *Hpa*I site 1 (Hpa1) is located at position 1218, Hha1 is at position 1397, Hha2 is at position 1809, Hha4 is at position 2583, Hha5 is at position 2653, and Hha7 is at position 3034.

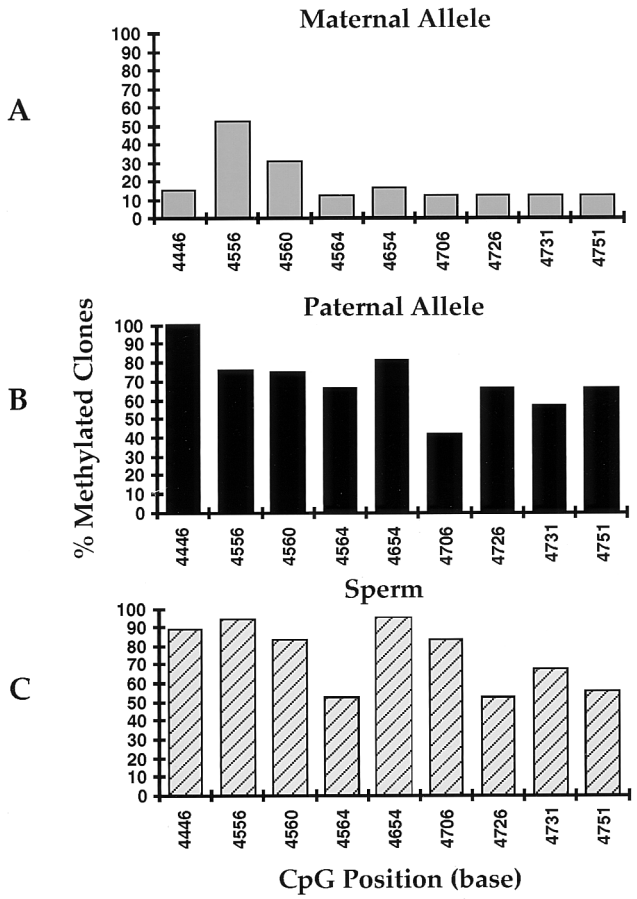

FIG. 3. Methylation profile of nine CpG dinucleotides as assayed by the bisulfite mutagenesis sequencing assay. This region is 328 bp in length, spans from position 4423 to 4751 (GenBank no. U19619), and is approximately 500 bp 5′ to the *H19* transcriptional start site (Fig. 1). There are nine assayable CpG dinucleotides in this region. The primary sequence data are shown in Fig. 4. The methods and DNA samples used and the presentation of data are the same as those described in the legend for Fig. 2. *Hpa*II sites 6 and 7 are located at positions 4446 and 4654, respectively.

methylated over the entire region and showed a very similar methylation profile to the paternal allele derived from midgestation embryos (Fig. 2C). These results indicate that the paternal methylation observed for this region in the midgestation embryos was likely to be inherited from the sperm.

Interestingly, the seven cytosine residues located at the 5′ end of the 2.2-kb region were not as differentially methylated as the remainder of the upstream region. It is also noteworthy that *Hpa*II site 1 is less differentially methylated in the midgestation embryos than in the blastocysts (data not shown). Thus, it is likely that the 5′ border of differential methylation resides adjacent to the 5′ *Eco*RI site indicated in Fig. 1.

The methylation profile of the 350-bp promoter-proximal region is presented in Fig. 3 and 4. In contrast to the upstream region, this region was not as differentially methylated. Although the maternal allele was hypomethylated relative to the paternal allele, each cytosine residue was methylated on a subset of the maternal allele clones (Fig. 3A and 4). The paternal allele was hypermethylated at each site, but only one site was found to be methylated on all of the paternally derived clones (Fig. 3B and 4). Similar results were observed in DNA derived from sperm (Fig. 3C and 4). Furthermore, as demonstrated by the methylation profiles from individual chromosomes (Fig. 4), the methylation of the promoter-proximal region on both alleles varied considerably between cells. In contrast, the upstream DNA methylation profiles of individual chromosomes were much less variable (data not shown).

The bisulfite mutagenesis assay indicated that the most differential region of methylation was contained within the 2-kb region approximately 2 kb upstream from the *H19* transcriptional start site. Consequently, this region is the most promising for harboring the imprinting mark.

Bisulfite mutagenesis of blastocyst DNA. A comparison of the results obtained with the PCR assay of preimplantation embryos and the bisulfite mutagenesis of midgestation embryos suggested that, in most cases, the methylation statuses of the *Hpa*II and *Hha*I sites were similar at these developmental periods and indicated that the methylation inherited from the sperm was largely maintained in the preimplantation embryo and preserved throughout embryogenesis. To confirm this, we performed the bisulfite mutagenesis on blastocysts and assayed a subset of the sites that were analyzed in the midgestation embryos.

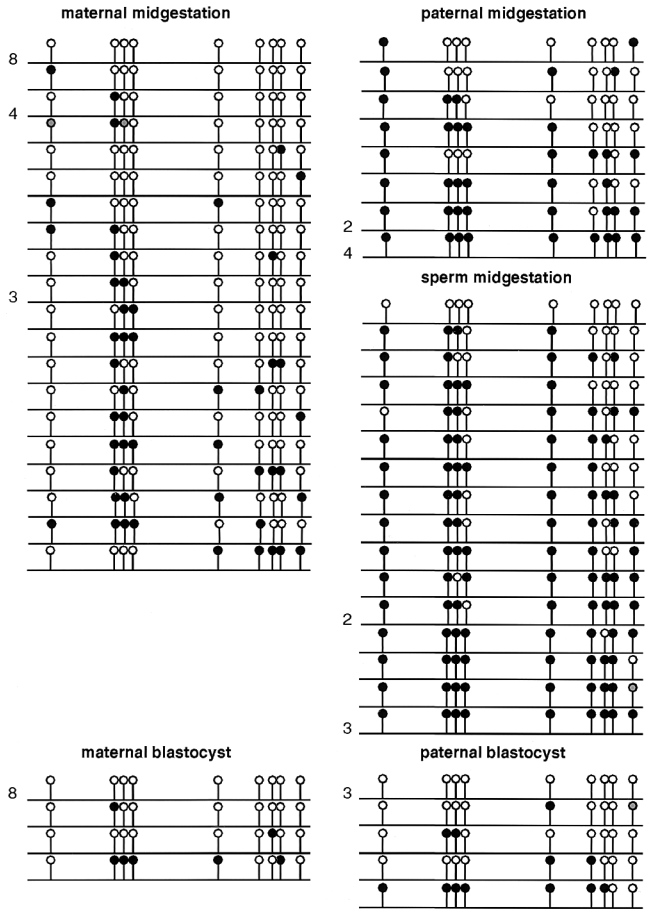

FIG. 4. Methylation status of individual strands of DNA in the promoter-proximal region. These data are summarized in the bar graphs in Fig. 3 and 6. The heterogeneity of the methylation states of the individual strands is evident in this representation of the primary data. Each line corresponds to an individual strand of DNA. Nine CpG dinucleotides are represented (shown 5′ to 3′) by the circles above the line. A filled circle corresponds to a methylated cytosine, and an open circle corresponds to an unmethylated cytosine. Those cytosines that could not be read on the sequencing gels are shaded. In cases where a given profile was found multiple times, the number of times which it was found is indicated to the left of the line. Maternal alleles are indicated on the left of the figure, and paternal alleles and sperm are indicated on the right.

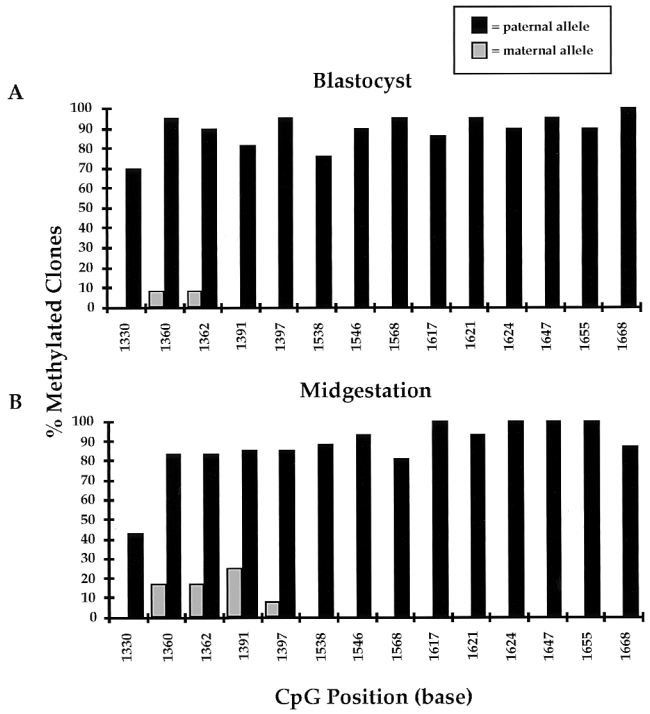

FIG. 5. Methylation profile of 14 CpG dinucleotides as assayed by the bisulfite mutagenesis procedure. The location of this region is indicated by the black box in Fig. 1. The methylation profile of DNA derived from blastocysts (A) and the methylation profile of maternally and paternally derived clones from bisulfite-treated midgestation DNA (B) are shown. The data in panel B are also shown in Fig. 2 but have been added here for ease of comparison with the data in panel A. See Fig. 2 and 3 for details.

Two 300- to 400-bp regions (Fig. 1, 14 CpG and 9 CpG boxes) were analyzed as described above. The most distal region assayed contains 14 CpG dinucleotides (Fig. 5). For 12 of the 14 CpG dinucleotides, a methylated cytosine was never found on the maternal clones. In striking contrast, the paternal allele was hypermethylated at every cytosine residue. The exclusive paternal methylation of the CpG corresponding to *Hha*I site 1 (Fig. 5A, CpG position 1397) also confirms the results obtained with the PCR assay of this same site (32). Moreover, the differences between the maternal and paternal alleles were slightly more pronounced than those observed in midgestation embryos (Fig. 5B).

In contrast to the distal region, the bisulfite mutagenesis assay revealed that the promoter-proximal region exhibited little methylation on either allele in blastocysts (Fig. 4 and 6). The reduced methylation of the CpG corresponding to *Hpa*II site 7 (Fig. 6A, CpG position 4654) is consistent with the earlier PCR assay of blastocyst DNA (32). The bisulfite results were not consistent, however, with the earlier PCR analysis of *Hpa*II site 6 (Fig. 6A, CpG position 4446) which showed that the paternal allele was preferentially methylated in blastocysts (32). These discrepant results may indicate that the methylation of the proximal region is labile. The CpG methylation on the individual strands of DNA corresponding to each parental allele shown in Fig. 4 is consistent with this proposal. Taken together, the bisulfite and PCR assays of gamete, blastocyst, and midgestation DNA indicate that the preferential methylation on the paternal allele in the 2-kb upstream region is inherited from the sperm and is preserved throughout development. Thus, this differential methylation could serve as the signal that designates the genomic imprint. The gametic methylation observed in the promoter-proximal region, on the other hand, is erased during preimplantation development and is reestablished later.

An upstream region is hypermethylated on both alleles. The slight differential methylation of the distal CpG dinucleotides analyzed by bisulfite mutagenesis suggests that the 5′ border of differential methylation resides around the 5′ *Eco*RI sites (Fig. 1 and 2). To determine if these sites do indeed represent the 5′ border of paternal-specific methylation, an additional 6 kb of DNA 5′ to the sequences depicted in Fig. 1 was cloned and analyzed. Digestion with the methylation-sensitive restriction enzymes *Hha*I and *Hpa*II revealed only four *Hpa*II sites. Preliminary sequence analysis of a portion of this region also indicated the presence of very few CpG dinucleotides. This reduced number of CpG dinucleotides is in striking contrast to the CpG-rich region found further downstream.

The methylation status of three of the four upstream *Hpa*II sites was determined. These sites are found within a 3.4-kb *Bam*HI-*Hin*dIII fragment located approximately at −7 to −10.4 kb relative to the start of transcription. To determine the allelic methylation pattern, it was necessary to distinguish the maternal and paternal alleles. Because no polymorphism had been defined in this region, we took advantage of a strain of mice that had been generated to study the function of the mouse *H19* gene (14). These mice carried a deletion mutation spanning the structural gene and 10 kb of 5′ flanking sequence. Through the use of animals heterozygous for the deletion, the methylation status of the intact allele was determined. Previous analyses indicated that the deleted allele did not affect the methylation pattern of the undeleted allele (data not shown).

As shown in Fig. 7, the 3.4-kb *Bam*HI-*Hin*dIII fragment remained largely undigested for both the paternal and the

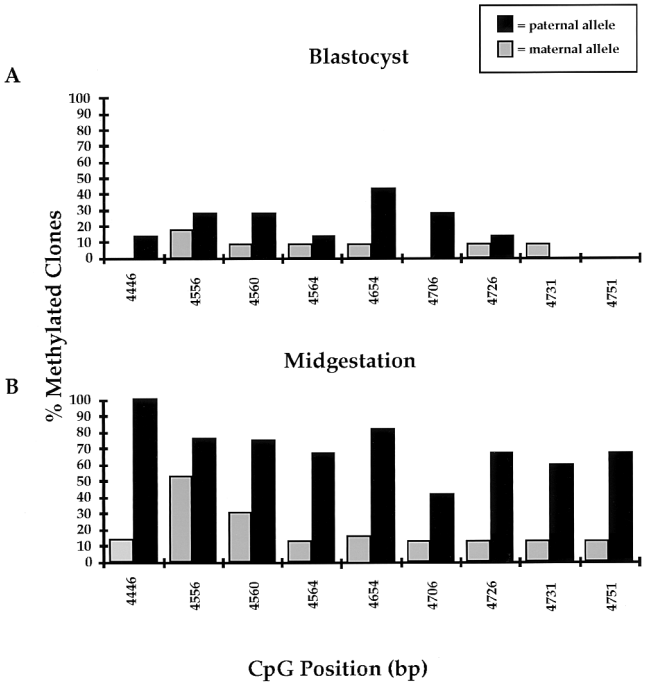

FIG. 6. Summary of the methylation profile of 9 CpG dinucleotides in the promoter-proximal region obtained by using the bisulfite mutagenesis assay. The location of this region is outlined by the hatched box (9 CpGs) in Fig. 1. The primary data are shown in Fig. 4. The methylation profiles of blastocyst DNA (A) and of DNA derived from midgestation embryos (B) are shown. The data in panel B are also shown in Fig. 3 but have been added here for ease of comparison with the data in panel A. See Fig. 2 and 3 for details.

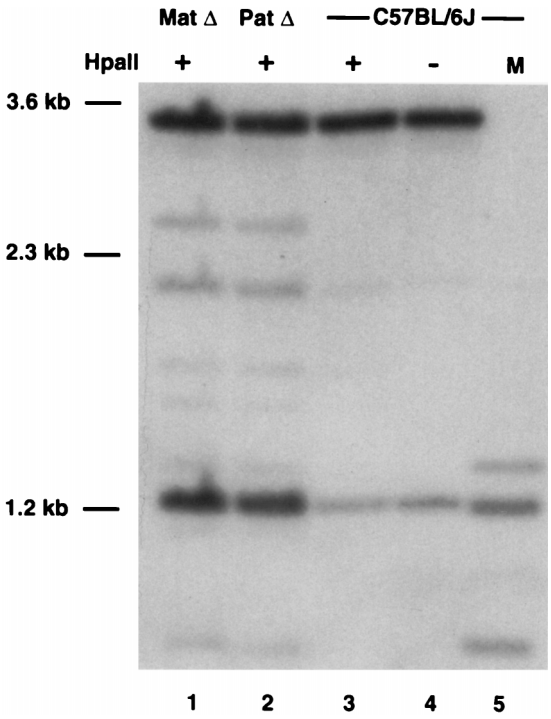

FIG. 7. Genomic DNA methylation analysis of the upstream 3.4-kb *Bam*HI-*Hin*dIII fragment. This region of the DNA is located at approximately −7 to −10.4 kb relative to the start of transcription and contains three methylation-sensitive *Hpa*II sites. Mat Δ and Pat Δ refer to DNA from the spleens of 7-day-old mice that have either the maternal or the paternal *H19* allele deleted, respectively, with the result that all hybridizing fragments except the one at approximately 1.2 kb correspond to the undeleted allele (14). The 1.2-kb band is the *Bam*HI-*Hin*dIII fragment located 5′ to the 3.4-kb fragment which is intact on the deletion allele. The probe is a 4.6-kb *Hin*dIII fragment. All DNA samples are digested with *Bam*HI and *Hin*dIII. The DNA in lanes 1 to 3 was additionally digested with *Hpa*II, and the DNA in lane 5 was additionally digested with *Msp*I. The 3.4-kb *Bam*HI-*Hin*dIII fragment is largely undigested in all samples, leading to the conclusion that both alleles are hypermethylated. M, *Msp*I.

maternal alleles (lanes 1 and 2, respectively), indicating that both alleles were hypermethylated. To confirm that this lack of digestion was not a function of the source of the DNA used in the analysis, the methylation status of neonatal liver DNA from C57BL/6J mice was also tested, and no significant amount of product corresponding to unmethylated DNA was detected (Fig. 7, lane 3). While we cannot distinguish the parental alleles of *H19* in this sample, the lack of digestion strongly indicates that both alleles of *H19* are not differentially methylated in a tissue where the gene is exclusively expressed from the maternal allele (3) and that this region is not likely to be important for the imprinted expression of *H19*. Based on this and the bisulfite experiments, we conclude that the 5′ end of the differentially methylated region resides around −4 kb.

Expression analysis in midgestation embryos. To establish that the differentially methylated 2-kb upstream region is the mark that confers the parental imprint to the *H19* gene, it is necessary to show that *H19* is exclusively expressed from the maternal allele in tissues in which the differential methylation of the upstream 2-kb region was detected. Since we previously demonstrated that the *H19* gene was exclusively expressed from the maternal allele in blastocysts and 7.5-day embryos (32), it was also necessary to determine if *H19* was monoallelically expressed in the midgestation embryos that were subjected to the bisulfite mutagenesis. As shown in Fig. 8 (lanes 4 to 7), exclusive expression of the maternal allele was detected in embryos from 9.75 to 16.5 days. Furthermore, placenta and yolk sac from 11.75-day embryos also exhibited this same expression pattern (Fig. 8, lanes 8 and 9, respectively). Thus, the blastocyst and midgestation embryos that we have used to investigate the methylation status of the parental alleles display imprinted expression. These results strengthen the hypothesized causative relationship between paternal methylation of the 2-kb upstream region and repression of the paternal allele.

DISCUSSION

The mechanism by which the parental alleles of imprinted genes are distinguished so that their appropriate pattern of expression is established and maintained in the developing embryo is poorly understood. Recent evidence suggests that the methylation of the cytosine residue in CpG dinucleotides is involved in the imprinting process and considerable effort has focused on understanding the extent to which methylation determines the imprinted expression of the *H19* gene. The inactive, paternal allele of *H19* is hypermethylated in both mice and humans (2, 4, 11, 13, 35). In the mouse, the paternal allele is hypermethylated in somatic tissues and sperm over a 7- to 9-kb region that includes 4 kb of 5′ flanking sequence and the structural gene (2). Although this differential methylation is quite striking, in order for methylation to be acting as the mark that distinguishes the parental alleles, it is still necessary to show that the differential methylation present in the somatic tissues is inherited from the gametes and is preserved throughout embryogenesis. Using a PCR-based assay, we identified four CpGs that are differentially methylated in gametes and preimplantation embryos (reference 32 and data not shown), supporting our assertion that differential methylation could serve as the mark that assigns parental identity to the alleles of the *H19* gene.

Interestingly, the CpG dinucleotides that exhibit exclusive methylation of the paternal allele are located in the distal portion of the upstream region. While it is possible that this could be a sampling artifact, it is also possible that these dinucleotides are pointing to an imprinting control region. To test these possibilities, a technique was utilized which assayed

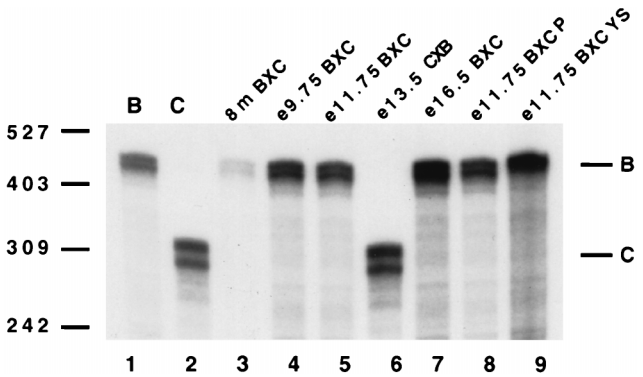

FIG. 8. Exclusive expression of the maternal allele of *H19* in midgestation embryos. The results of an RNase protection assay of total RNA from C57BL/6J (B), *M. m. castaneus* (C), and F_1 hybrid embryos are shown. Molecular size markers are indicated at the left and the locations of the B and C products are indicated at the right of the figure. Neonatal muscle RNA from the parental strains is assayed in lanes 1 and 2. Other RNA samples include 8-month-old muscle RNA from B × C mice (lane 3), RNA from the bodies of embryonic day 9.75 and day 11.75 B × C and day 13.5 C × B mice (lanes 4 to 6, respectively), liver RNA from embryonic day 16.5 mice (lane 7), and placenta and yolk sac RNA from embryonic day 11.75 mice (lanes 8 and 9, respectively). Three micrograms of RNA was assayed for all samples except yolk sac, for which 1 μg of RNA was assayed.

all CpG dinucleotides. The bisulfite mutagenesis and sequencing assay defined a 2-kb region located from −2 kb to −4 kb relative to the start of transcription that is highly differentially methylated in midgestation embryos as well as in blastocysts, strengthening the hypothesis that paternal methylation in this region is inherited from the sperm and maintained throughout development. Therefore, it is plausible that this 2-kb region could serve as an imprinting control region.

Transgenic experiments support a role for this upstream 2-kb differentially methylated region in the imprinting of *H19*. A 14-kb transgene consisting of the 4-kb upstream region, a 2-kb internally deleted *H19* structural gene, and an 8-kb 3′ flanking region, where the endodermal-specific enhancers are located, is imprinted similarly to the endogenous *H19* gene (2). That is, the transgene is hypomethylated and expressed when maternally transmitted, and it is hypermethylated and repressed when inherited from the father. These experiments suggest that the signals necessary to confer both parental-specific expression and methylation at an exogenous locus are present in this transgene. When the 2-kb region corresponding to the sequences exhibiting the most striking pattern of differential methylation at the endogenous locus are deleted, the transgene is expressed and hypomethylated regardless of parental origin (9). Although this experiment does not prove that the differential methylation is essential to imprinting, it does confirm that a signal necessary to silence the paternal allele is located in the 2-kb region.

While we did not subject all of the sequences adjacent to the 2.2-kb upstream region to the bisulfite mutagenesis assay, it is not likely that the highly differentially methylated domain extends much beyond this region. The distal seven cytosine residues analyzed in the upstream region were only slightly more methylated on the paternal allele than on the maternal allele in midgestation embryos. Attempts to examine CpG dinucleotides located 5′ to this region by using genomic DNA and sequencing analysis revealed a striking absence of CpG dinucleotides. Furthermore, those few dinucleotides that are contained within methylation-sensitive restriction endonuclease sites are methylated on both alleles. The region that is adjacent to the 3′ end of the 2.2-kb domain harbors a G-rich repetitive element that has not been amenable to PCR and bisulfite mutagenesis analysis. While the repetitive elements may be crucial to determining the imprinted pattern of many genes (19), we cannot determine if the few CpGs located in this region are differentially methylated. Immediately 3′ of the repetitive element, however, lies the 350-bp promoter-proximal region that was tested in the bisulfite mutagenesis assay. This region is hypomethylated on both alleles in blastocyst DNA but hypermethylated on the paternal allele during midgestation, indicating that the paternal methylation is acquired after implantation. Thus, the promoter-proximal region probably does not harbor elements that are key to conferring the parental imprint during early development. More likely, as with those sites assayed in the gene body (4), the reacquisition of methylation serves later to help silence the paternal allele when the transcription of *H19* is highest. This late methylation is analogous to that observed for the maternally expressed insulin-like growth factor type 2/cation-independent mannose-6-phosphate receptor gene (29). In this case the promoter is methylated on the inactive paternal allele after fertilization and coincides with the exclusive expression of the maternal allele.

Razin and colleagues have observed dynamic alterations in allelic methylation patterns in the promoter and body of the *H19* gene during preimplantation development (4, 26). As stated above, we too have found that the methylation of the promoter-proximal region is labile. These results are consistent with the hypothesis that methylation of the downstream region is not important for the initial imprint. In contrast, the paternal methylation in the upstream 2-kb region is inherited from the sperm and maintained during preimplantation and postimplantation development. Thus, the methylation in this region is stably propagated during development.

If the upstream 2-kb differentially methylated region acts as the primary imprint, *H19* must be monoallelically expressed from the maternal allele during all times in which the paternal methylation is observed. In fact, such is the case since we observe exclusive expression of the maternal allele when blastocysts and 7.5-day embryos are analyzed using a reverse transcription-PCR assay (32) and when embryos from 9.75 to 16.5 days are analyzed using an RNase protection assay (Fig. 8). Curiously, others have observed significant expression from the paternal allele (25, 31). This difference in expression could be attributed either to genotypic differences in the embryos or, alternatively, to the conditions under which these embryos were generated. All embryos used in the experiments described in this report were derived by natural matings and in vivo development while a variable amount of culturing was employed in the generation of embryos used by others (25, 31). Thus, culture may result in expression of the paternal allele. In fact, Sasaki and colleagues observed biallelic *H19* expression in embryos that were cultured until the late blastocyst stage but observed monoallelic expression in embryos that were derived by in vitro fertilization followed by embryo transfer at the two-cell stage (25). Indeed, under certain conditions, we too have observed biallelic expression in embryos that have been cultured from the two-cell to the blastocyst stage (32a). Therefore, culturing embryos may result either in the loss of the specific imprint or the failure of the embryo to recognize the imprint.

The characterization of *H19* gene methylation in the embryo adds credence to the hypothesis that methylation is the mark that specifies the allelic imprint. Other evidence implicating methylation as a key element for monoallelic *H19* expression comes from a strain of mice that are deficient for DNA methyltransferase, the hemimethylase responsible for the maintenance methylation activity present in the embryo (16). Consistent with a role for DNA methylation in regulating maternal monoallelic *H19* expression is the finding that the *H19* gene is unmethylated and biallelically expressed in embryos deficient in DNA methyltransferase activity (15). While these experiments cannot distinguish between a role for methylation in setting versus maintaining the imprint, they do establish that methylation is essential to monoallelic expression of *H19*.

In conclusion, we have described a 2-kb domain which is 2 kb 5′ to the transcription start site and which is methylated in sperm and on the paternal allele in blastocysts and midgestation embryos. We propose that this domain is crucial to establishing the molecular imprint of the *H19* gene in the early embryo. Since the early expression of *H19* is exclusively maternal in origin, the modifications present at this time are sufficient to confer imprinted expression. Following implantation, the *H19* gene is transcribed at a higher level (20) and the differentially methylated domain extends downstream to include the promoter-proximal region and the transcription unit (2, 4, 11). As suggested by the biallelic *H19* expression in DNA methyltransferase mutant embryos, we believe that the widespread methylation of the paternal allele that occurs later in development helps to silence this allele during the periods of maximal transcription. Thus, activation of the maternal allele and repression of the paternal allele appear to involve multiple elements that are invoked at different times. To test the hy-

pothesis that the 2-kb region is a key domain for establishing the *H19* imprint, we are currently deleting this region from the endogenous locus using gene targeting in ES cells. Absence of the region should result in mice that are unable to silence the paternal allele of the *H19* gene.

ACKNOWLEDGMENTS

We thank Melanie Cranston and David Elson for technical assistance, Michael Stadnick for help in the initial phase of this work, Amy Hark for the spleens from the *H19* knockout mice, Andrea Webber for the *H19* YAC, and Laurie Zimmerman for secretarial support. We also thank Rebecca Oakey, Richard Schultz, Timothy Bestor, Michael Malim, and the members of the laboratory for critical reviews of the manuscript.

This work was supported by U.S. Public Service grant GM51279 and the Howard Hughes Medical Institute. K.D.T. was supported by U.S. Public Service training grant GM07229.

REFERENCES

1. **Barlow, D. P.** 1995. Gametic imprinting in mammals. Science **270:**1610–1613.
2. **Bartolomei, M. S., A. L. Webber, M. E. Brunkow, and S. M. Tilghman.** 1993. Epigenetic mechanisms underlying the imprinting of the mouse *H19* gene. Genes Dev. **7:**1663–1673.
3. **Bartolomei, M. S., S. Zemel, and S. M. Tilghman.** 1991. Parental imprinting of the mouse *H19* gene. Nature **351:**153–155.
4. **Brandeis, M., T. Kafri, M. Ariel, J. R. Chaillet, J. McCarrey, A. Razin, and H. Cedar.** 1993. The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. EMBO J. **12:**3669–3677.
5. **Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman.** 1990. The product of the *H19* gene may function as an RNA. Mol. Cell. Biol. **10:**28–36.
6. **DeChiara, T. M., A. Efstratiadis, and E. J. Robertson.** 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature **345:**78–80.
7. **DeChiara, T. M., E. J. Robertson, and A. Efstratiadis.** 1991. Parental imprinting of the mouse insulin-like growth factor II gene. Cell **64:**849–859.
8. **Efstratiadis, A.** 1994. Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Dev. **4:**265–280.
9. **Elson, D. A., and M. S. Bartolomei.** 1997. A 5′ differentially methylated sequence and the 3′ flanking region are necessary for *H19* transgene imprinting. Mol. Cell. Biol. **17:**309–317.
10. **Feil, R., J. Charlton, A. P. Bird, J. Walter, and W. Reik.** 1994. Methylation analysis on individual chromosomes: improved protocol for bisulfite genomic sequencing. Nucleic Acids Res. **22:**695–696.
11. **Ferguson-Smith, A. S., H. Sasaki, B. M. Cattanach, and M. A. Surani.** 1993. Parental-origin-specific epigenetic modification of the mouse *H19* gene. Nature **362:**751–754.
12. **Frommer, M., L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, and C. L. Paul.** 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. USA **89:**1827–1831.
13. **Jinno, Y., K. Sengoku, M. Nakao, K. Tamate, T. Miyamoto, T. Matsuzaka, J. S. Sutcliffe, T. Anan, N. Takuma, K. Nishiwaki, Y. Ikeda, T. Ishimaru, M. Ishikawa, and N. Niikawa.** 1996. Mouse/human sequence divergence in a region with a paternal-specific methylation imprint at the human *H19* locus. Hum. Mol. Genet. **5:**1155–1161.
14. **Leighton, P. A., R. S. Ingram, J. Eggenschwiler, A. Efstratiadis, and S. M. Tilghman.** 1995. Disruption of imprinting caused by deletion of the *H19* gene region in mice. Nature **375:**34–39.
15. **Li, E., C. Beard, and R. Jaenisch.** 1993. Role for DNA methylation in genomic imprinting. Nature **366:**362–365.
16. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:**915–926.
17. **McGrath, J., and D. Solter.** 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37:**179–183.
18. **Monk, M., M. Boubelik, and S. Lehnert.** 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development **99:**371–382.
19. **Neumann, B., P. Kubicka, and D. P. Barlow.** 1995. Characteristics of imprinted genes. Nat. Genet. **9:**12–13.
20. **Poirier, F., C.-T. J. Chan, P. M. Timmons, E. J. Robertson, M. J. Evans, and P. W. J. Rigby.** 1991. The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. Development **113:**1105–1114.
21. **Razin, A., and H. Cedar.** 1994. DNA methylation and genomic imprinting. Cell **77:**473–476.
22. **Rigby, P. W. J., M. Dieckmann, D. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:**237–251.
23. **Riggs, A. D., and G. P. Pfeifer.** 1992. X-chromosome inactivation and cell memory. Trends Genet. **8:**169–174.
24. **Sanford, J. P., H. J. Clark, V. M. Chapman, and J. Rossant.** 1987. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. **1:**1039–1046.
25. **Sasaki, H., A. C. Ferguson-Smith, A. S. W. Shum, S. C. Barton, and M. A. Surani.** 1995. Temporal and spatial regulation of *H19* imprinting in normal and uniparental mouse embryos. Development **121:**4195–4202.
26. **Shemer, R., Y. Birger, W. L. Dean, W. Reik, A. D. Riggs, and A. Razin.** 1996. Dynamic methylation adjustments and counting as part of imprinting mechanisms. Proc. Natl. Acad. Sci. USA **93:**6371–6376.
27. **Solter, D.** 1988. Differential imprinting and expression of maternal and paternal genomes. Annu. Rev. Genet. **22:**127–146.
28. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
29. **Stoger, R., P. Kubicka, C.-G. Liu, T. Kafri, A. Razin, H. Cedar, and D. P. Barlow.** 1993. Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. Cell **73:**61–71.
30. **Surani, M. A. H., S. C. Barton, and M. L. Norris.** 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308:**548–550.
31. **Szabo, P. E., and J. R. Mann.** 1995. Allele-specific expression and total expression levels of imprinted genes during early mouse development: implications for imprinting mechanisms. Genes Dev. **9:**3097–3108.
32. **Tremblay, K. D., J. R. Saam, R. S. Ingram, S. M. Tilghman, and M. S. Bartolomei.** 1995. A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene. Nat. Genet. **9:**407–413.

32a. **Tremblay, K. D., M. S. Bartolomei, R. Schultz, and A. Doherty.** Unpublished data.

33. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76:**3683–3687.
34. **Yeivin, A., and A. Razin.** 1993. Gene methylation patterns and expression, p. 523–568. *In* J. P. Jost and H. P. Saluz (ed.), DNA methylation: molecular biology and biological significance. Birkhauser Verlag, Basel, Switzerland.
35. **Zhang, Y., T. Shields, T. Crenshaw, Y. Hao, T. Moulton, and B. Tycko.** 1993. Imprinting of human H19; allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. Am. J. Hum. Genet. **53:**113–124.