Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression

Fei Song*, Joseph F. Smith*, Makoto T. Kimura†, Arlene D. Morrow*, Tomoki Matsuyma‡, Hiroki Nagase†, and William A. Held§

Departments of *Molecular and Cellular Biology and †Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263; ‡Plant Functions Laboratory, Computational Science Division, Institute of Physical and Chemical Research, Wako, Saitama 351-0198, Japan

Edited by Stanley M. Gartler, University of Washington, Seattle, WA, and approved January 12, 2005 (received for review November 15, 2004)

Early studies proposed that DNA methylation could have a role in regulating gene expression during development [Riggs, A.D. (1975) Cytogenet. Cell Genet. 14, 9–25]. However, some studies of DNA methylation in known tissue-specific genes during development do not support a major role for DNA methylation. In the results presented here, tissue-specific differentially methylated regions (TDMs) were first identified, and then expression of genes associated with these regions correlated with methylation status. Restriction landmark genomic scanning (RLGS) was used in conjunction with virtual RLGS to identify 150 TDMs [Matsuyama, T., Kimura, M.T., Koike, K., Ate, T., Nakao, T., Asami, T., Ebisuzaki, T., Held, W., A., Yoshida, S., & Nagase, H. (2003) Nucleic Acids Res. 31, 4490–4496]. Analysis of 14 TDMs by methylation-specific PCR and by bisulfite genomic sequencing confirms that the regions identified by RLGS are differentially methylated in a tissue-specific manner. The results indicate that 5% or more of the CpG islands are TDMs, disputing the general notion that all CpG islands are unmethylated. Some of the TDMs are within 5′ promoter CpG islands of genes, which exhibit a tissue-specific expression pattern that is consistent with methylation status and a role in tissue differentiation.

DNA methylation | epigenetic | mouse | tissue differentiation

There is a history of inverse correlation between DNA methylation of CpG island promoter regions and gene expression. Almost all CpG islands on the inactive X chromosome are methylated, and monoallelic methylation of imprinted genes is associated with monoallelic gene expression (1–3). Moreover, programmed changes in DNA methylation are essential features of development, with disruption frequently resulting in aberrant development (4). Although early studies proposed that DNA methylation could have a role in regulating development (5, 6), more recent studies of DNA methylation in known tissue-specific genes during development did not support a major role for DNA methylation. Studies by Warnecke and Clark (7) found that the tissue-specific expression of the skeletal α-actin gene in the adult mouse does not correlate with the methylation status of the promoter. Walsh and Bestor (8) investigated the 5′ methylation status of seven tissue-specific genes and found no correlation with tissue-specific expression. In the study presented here, rather than examining the methylation status of known tissue-specific genes, tissue-specific differentially methylated regions (TDMs) were first identified, and then genes located near the TDMs were analyzed for tissue-specific expression.

Restriction landmark genomic scanning (RLGS) is a method for the two-dimensional display of end-labeled DNA restriction fragments (9–11) and can be used to scan for genomic DNA methylation (11). Because the NoI recognition site contains two CpGs and the great majority of NoI sites are within CpG islands, RLGS (with NoI as the restriction landmark) displays CpG islands and adjacent regions. If a NoI site is methylated, it will not be digested and will not be end-labeled, resulting in the absence of the spot in the RLGS profile. Global analysis of genomic DNA methylation of different tissues by using RLGS indicates that there are a relatively large number of differences in profiles suggesting TDMs (12–14). Numerous differences in the RLGS profiles of ES cells and differentiated tissues, such as kidney and brain, have been identified (13), but the DNA sequence of these genomic regions was not established, and it is not known whether these methylation differences are associated with tissue-specific gene expression.

In the results presented here, RLGS was used in conjunction with virtual RLGS to identify 150 TDMs (9, 10). VI-RLGS is a new computational software application that utilizes the mouse genome sequence information to generate a virtual RLGS image (10). Our results indicate that many CpG islands are differentially methylated in a tissue-specific fashion, which could be confirmed by methylation-specific PCR (MSP) and bisulfite genomic sequencing. Some of these TDMs are within 5′ promoter CpG islands of genes that are expressed only if the region is unmethylated.

Methods

DNA and RNA Preparations. The DNA for RLGS was isolated from tissues of 12-week-old C57BL/6J male mice by using the protocols described in refs. 9 and 15. TRIzol was used to extract both DNA for MSP or bisulfite sequencing and RNA from the same samples (16). The RNA was quantified by a spectrophotometer and aliquots were checked for integrity by electrophoresis in denaturing agarose gels (17).

RLGS. RLGS was performed according to published protocols (9, 10, 15). VI-RLGS is a new computational software that displays a virtual RLGS image of mouse genome sequence information (10). The correspondence between the virtual and real RLGS spots for some of the loci were confirmed by PCR amplification of spot DNA eluted from RLGS gels by using primers derived from the sequences predicted by VI-RLGS (10). The identity between real and virtual spot sequences was further confirmed by sequence analysis of the PCR product.

Bisulfite Genomic Sequencing of TDMs. The bisulfite treatment of DNA and sequencing after PCR was done as described in ref. 10. The bisulfite PCR primers were designed by using the web program METHPRIMER (18). The primers were chosen to include the NoI landmark within the amplification product and also as many CpG dinucleotides in the product as possible. The melting temperatures of the primer pairs were constrained to be between 50°C and 60°C. The bisulfite PCR primers were synthesized by Integrated DNA Technologies (Coralville, IA). The PCR were carried out in either a Biometra T1 or an MJ Research PTC-100 thermocycler. The eluted PCR bands were sequenced at the...
Roswell Park Cancer Institute core facility’s Applied Biosystems model 3100 sequencer. To compare the direct bisulfite sequencing results with those obtained by cloning the PCR products (see Fig. 4 and Fig. 6, which is published as supporting information on the PNAS web site), the PCR products amplified from bisulfite-treated tissue DNA were cloned into pGEM-T easy vector (Promega) according to the manufacturer’s protocol. Plasmid DNA was purified with the DNA Purification system (Promega). Individual plasmids were then sequenced by using the Applied Biosystems model 3100 sequencer (Applied Biosystems).

MSP. MSP was performed as described in ref. 19 by using the various bisulfite-treated tissue DNAs. The methylated and unmethylated primers were designed by using METHPRIMER (18). The PCR (40 cycles) was performed by using the same conditions as for bisulfite genomic sequencing.

Quantitative, Real Time RT-PCR. Tissue-specific cDNAs were made by using the iScript cDNA kit from Bio-Rad according to the manufacturer’s protocol. One microliter of each tissue cDNA was used per quantitative PCR. PCR primers were designed to bridge the exon–intron boundaries within the gene of interest to exclude possible contamination by genomic DNA (except for Hspa1l). The SYBR Green primers were designed by the web program PRIMER3 (20) and purchased from Integrated DNA Technologies. A test RT-PCR was performed to check for a single PCR product before running the quantitative PCRs. The Assays on Demand primers sets were purchased from Applied Biosystems and used for the TaqMan assay. The quantitative PCRs were run on the Bio-Rad MyiQ Cycler for SYBR Green or iCycler for TaqMan according to the manufacturer’s recommendations for each probe. The appropriate master mixes were used for each application. The resulting PCR cycle time (Ct) values were collected by using the software provided for the iCycler or MyiQ systems, and the data were then analyzed in
Results and Discussion

Identification of TDMs. RLGS profiles were performed for several different mouse tissues by using two restriction enzyme combinations (NotI–PstI–PvuII and NotI–PvuII–PstI, the same restriction enzymes but in a different order of digestion) that survey different C57BL/6J mice (Fig. 1 shows a portion of RLGS profiles from several tissues from C57BL/6J mice (12 weeks of age). In Fig. 1b Left, there are several spots present in testis that are absent in kidney and liver (Pvu6, Pvu7, and Pvu80), indicating that these regions are unmethylated in testis but methylated in kidney and liver. In contrast, Pvu56 is present in kidney and liver (unmethylated) but absent (methylated) in testis. Pvu56 was present only in kidney, but at less than full intensity, which indicates partial methylation. Additional TDMs are shown in Fig. 1b Right. RLGS analyses of tissues from different C57BL/6J mice indicate that the tissue-specific RLGS profiles are consistent (see Fig. 7, which is published as supporting information on the PNAS web site). Vi-RLGS (10) identified 150 unique RLGS loci/spots that were present (unmethylated) in some tissues but absent (methylated) in others (Fig. 2; for a larger version of Fig. 2 see Fig. 8, which is published as supporting in formation on the PNAS web site). There was a large group (43 loci) that was unmethylated only in testis, but the majority (86 loci) were unmethylated in two to five of the tissues analyzed. BLAT analysis (University of California, Santa Cruz, Genome Bioinformatics database available at http://genome.ucsc.edu) of genomic sequences determined by Vi-RLGS was used to identify genomic regions and genes associated with the TDMs. Approximately two-thirds of the loci (100) were located within CpG islands. Although 56% of the TDMs are located in apparent 5' promoter regions, there are many that are located in exons, introns, or intergenic regions. The entire listing of TDMs identified by Vi-RLGS can be found in Table 4, which is published as supporting information on the PNAS web site.

Table 1. TDM loci confirmed by MSP or bisulfite sequencing

<table>
<thead>
<tr>
<th>Loci Chr</th>
<th>Location, MB</th>
<th>CpG island</th>
<th>NotI site location</th>
<th>Gene</th>
<th>Function</th>
<th>Liv</th>
<th>Kid</th>
<th>Brain</th>
<th>Col</th>
<th>Mus</th>
<th>Tes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PvuII</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>114.98</td>
<td>No</td>
<td>Exon</td>
<td>mRNA AK037416</td>
<td>Mct transporter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>10.22</td>
<td>Yes</td>
<td>3' exon</td>
<td>Ref5 D43000K24Rik</td>
<td>Ring Zn finger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>47.12</td>
<td>Yes</td>
<td>5' promoter</td>
<td>Ref5 C330046L10Rik</td>
<td>Ubiquitin hydrolase</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>58.76</td>
<td>Yes</td>
<td>5' promoter</td>
<td>Est BE134335</td>
<td>DNA J domain</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>14</td>
<td>47.79</td>
<td>Yes</td>
<td>5' promoter</td>
<td>Ref5 4930548G07Rik</td>
<td>Chromatin binding</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
<td>16.79</td>
<td>Yes</td>
<td>5' promoter</td>
<td>mRNA 6130401J40Rik</td>
<td>Ubiquitin-conj. E2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>66</td>
<td>11</td>
<td>43.4</td>
<td>Yes</td>
<td>3' exon</td>
<td>RefSeq Adra1b</td>
<td>G protein receptor</td>
<td>0.1</td>
<td>2</td>
<td>ND</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

| PstI     |             |           |                   |                                          |                          |     |     |       |     |     |     |
| 3        | 13          | 111.46    | Yes               | 5' promoter                              | Ref5 Ddx4                 | RNA helicase              | 0   | 0   | 0     | 0   | 2   |     |
| 6        | 17          | 34.63     | Yes               | 3' exon                                  | Ref5 HSPLike              | Heat shock                | 0   | 0   | 0     | 0   | 2   |     |
| 32       | 12          | 67.21     | Yes               | 3' exon                                  | Ref5 Dact1                | Dsh signaling             | 0.5 | 0   | 0     | 0   | 2   |     |
| 33       | 4           | ND        | No                | Exons                                    | mRNA BC059845             | Unknown                   | 0.5 | 0   | 0     | 0   | 2   |     |
| 61       | 6           | 89.64     | Yes               | 5' promoter                              | mRNA GATA-2               | Transcription             | 0   | 2   | 2     | 2   | 2   | 2   |
| 21       | 19          | 6.3       | No                | Intron                                    | Predicted gene            | Unknown                   | 2   | 0   | 0     | 0   | 0   | 0   |
| 46       | 17          | 17.6      | No                | Intron                                    | Predicted gene            | Unknown                   | 0   | 2   | 0     | 0   | 2   | 2   |

Shown is the chromosomal location (University of California, Santa Cruz, Bioinformatics Database, October 2003 assembly available at http://genome.ucsc.edu), whether the TDM was located within a designated CpG island, and the location of the TDM NotI site relative to a nearby gene. Gene and function information were also taken from links to the University of California, Santa Cruz, Genome Bioinformatics database. The RLGS spot intensity for each tissue is given, with 2 corresponding to a diploid single copy sequence, 1 indicating haploid spot intensity, and 0 indicating that the spot is absent. Values in bold indicate that the sequence is unmethylated (0.5–2 intensity). Values of <2 suggest that methylation is heterogeneous with the tissue. A complete listing of all TDM loci identified by Vi-RLGS can be found in Table 4, which is published as supporting information on the PNAS web site.

Estimates of the Number of NotI Sites and CpG Islands That Are Differentially Methylated in the Mouse Genome. The mouse genome contains 6,057 NotI sites with 5,175 in nonrepetitive DNA (21). Because most of the NotI sites within repetitive DNA are methylated and not “visible,” RLGS can potentially resolve 5,175 NotI sites. The two restriction enzyme combinations used in this study display ~3,200 RLGS spots. Although each NotI site produces two fragments, there is a <1% chance that both will be in the analyzed fragments (22). Also, there were no overlaps between NotI–PstI–PvuII and NotI–PvuII–PstI spot sequences identified by Vi-RLGS. If one simply presumes that each RLGS spot corresponds to a single NotI site and that it was possible to identify 171 TDMs (150 identified by vi-RLGS) of 3,200 spots (171/3,200 = 5.3%), there would be 274 (5,175 × 0.053) NotI sites that are TDMs and 184 (274 × 0.67) that are in CpG islands that contain NotI sites [100 of 150 (67%) TDMs were in CpG islands]. There are estimated to be 15,500 CpG islands in the mouse genome (23), with 3,478 (22.4%) containing NotI sites (21). Therefore, one can estimate that 836 CpG islands contain TDMs (184/0.22), which suggests that ~5.6% (836/15,500) of the CpG islands are TDMs. Because only six tissues were surveyed at only one developmental time point, it is quite likely that this calculation is an underestimate.

Confirmation of TDMs by MSP and Bisulfite Genomic Sequencing. Fourteen loci were analyzed more extensively by MSP (19) and/or bisulfite genomic sequencing (24) to confirm the RLGS results and determine the density of methylation in the TDM (Figs. 3 and 4, Table 1, and Figs. 6 and 9, which are published
CpG2 is unmethylated in kidney and brain. See text for further explanation.

Pst6 (CpG2). CpG1 is strongly methylated in all tissues except testis. In contrast, Pst21 was methylated in all tissues except liver. Pst46 (Pvu42 and Pst6), and methylated only in liver. Pst46 (Pvu42 and Pst6), and methylated only in liver. Pst61 was methylated in testis and colon. These results indicate that tissue-specific CpG island (CpG2) 470 bp downstream of Pst6 is unmethylated in kidney and brain, whereas Pst6 (CpG1) is almost completely methylated in all tissues except testis. Similar bisulfite sequencing results were found for Pvu66 CpG island, except testis. Similar bisulfite sequencing results were found for Pvu66 CpG island, except testis. Similar bisulfite sequencing results were found for Pvu66 CpG island, except testis. Similar bisulfite sequencing results were found for Pvu66 CpG island, except testis.

MSP analysis of CpG island regions just upstream or downstream of TDMs indicates both similarities and differences in methylation status. For example, MSP analysis indicates a small fraction of the cells in colon. As a positive control for unmethylated primers, bacterial artificial chromosome genomic clones containing the genomic regions of interest (RP23-180J7, -28SK3, -59D1, -430C5, -288H11, -467M14, -381B4, -278A5, and -446I5) were mixed, bisulfite treated, and used as a template for MSP (Cont). (b) MSP analysis of the methylation status of Pst6 TDM and an adjacent CpG island. Genomic DNA was treated with sodium bisulfite and then amplified with primers specific for methylated and unmethylated sequences within the Pst6 TDM and an adjacent CpG island. The results are in general agreement, although the cloned sequences exhibit considerable heterogeneity in methylation. Fig. 10, which is published as supporting information on the PNAS web site, shows more detail as a larger version of this figure.

Fig. 4. Bisulfite genomic sequencing of TDM Pst3. (a) Sequence traces obtained from the PCR products from bisulfite-treated DNA by using primers that would amplify a portion of the Pst3 locus. Because the PCR products were not cloned, the trace represents an approximation of the “average” methylation status at each CpG residue. The normal sequence and bisulfite sequence are shown. CpG sites are highlighted. (b) Diagrammatic representation of Pst3 within the 5′ promoter region of Ddx4 (DEAD-box protein 4) on chromosome 13. The CpG island includes the 5′ promoter, exon 1, and a portion of the first intron. The arrow indicates the direction of transcription. The methylation status of the CpGs (black circles, fully methylated; blue circles, partial methylation; white circles, unmethylated) within a 450-bp region of Pst3 is shown along with the location of the restriction landmark NotI site. Note the dense hypermethylation in liver, kidney, colon, muscle, and brain and the hypomethylation in testis. Fig. 6 shows the bisulfite sequence of a portion of the Pst3 region that was determined from cloned PCR products. The results are in general agreement, although the cloned sequences exhibit considerable heterogeneity in methylation. Fig. 10, which is published as supporting information on the PNAS web site, shows more detail as a larger version of this figure.

sequence, found in the 5′ end of the Ddx4 gene, is methylated in kidney but not in testis, consistent with the RLGs results (Fig. 4a and Table 1). The more extensive bisulfite sequencing results shown diagrammatically in Fig. 4b (see also Fig. 6) indicate that almost the entire CpG island is densely methylated in all tissues except testis. Similar bisulfite sequencing analysis results were found for Pvu2, located in exon 3 (National Center for Biotechnology Information accession no. AK037416) of a monocarboxylate transporter-like gene (Fig. 9), whereas for the Pvu66 CpG island in the 3′ end of the adrenergic receptor a1b (Adra1b), methylation was demonstrated in the liver but not kidney and testis (data not shown). These results indicate that tissue-specific CpG methylation (TDMS) may be associated with both the 5′ and 3′ regions of genes as well as with internal exons.

MSP analysis of CpG island regions just upstream or downstream of TDMs indicates both similarities and differences in methylation status. For example, MSP analysis indicates a small CpG island (CpG2) 470 bp downstream of Pst6 is unmethylated in kidney and brain, whereas Pst6 (CpG1) is almost completely methylated in these tissues (Fig. 3b). However, both CpG1 and CpG2 are methylated in colon and muscle and unmethylated in testis. These results indicate that TDMS appear to be localized...
to small genomic regions, possibly a single CpG island. However, additional work is necessary to establish the precise boundaries of the TDMs.

**Tissue-Specific Expression of Genes Associated with TDMs.** Promoter CpG island DNA methylation has frequently been associated with silencing of gene expression in cancer (25) and in normal cells (14). The dense methylation of the CpG island promoter region of DEAD-box protein 4 (Ddx4) (Fig. 4), associated with the Pst3 TDM, would suggest that Ddx4 is silenced in most tissues except testis. Quantitative, real-time RT-PCR indicated a very high level of expression in testis (Fig. 5 and Table 2) and a low level expression in brain that was several hundred times less than testis. Furthermore, interrogation of the Genomics Institute of Novartis Research Foundation (GNF) gene expression database (available at http://symatlas.gnf.org/SymAtlas) (26) indicates that Ddx4 is expressed at very high levels in C57BL/6J testis and is either not expressed or expressed at much lower levels in other tissues. The targeted homozygous knockout of Ddx4 causes a deficiency in the proliferation and differentiation of mouse male germ cells resulting in sterility (27). Thus, there is a clear association between tissue-specific expression of Ddx4 and an unmethylated CpG island promoter region. Quantitative RT-PCR analysis and interrogation of the GNF database (26) of other genes associated with TDMs (Fig. 5 and Table 2) indicated that 5′ promoter methylation is frequently associated with silencing of gene expression. Although Pst6 was located in a 3′ exon, MSP indicated that a CpG island around 500 bp upstream of the promoter for Hspa1l was also unmethylated in testis and methylated in the other tissues (data not shown). Quantitative RT-PCR (Fig. 5 and Table 2) and the GNF database (26) (data not shown) indicated essentially exclusive expression of Hspa1l in testis. The gene (4930548G07Rik) associated with the promoter region of Pvu29 is expressed at a relatively high level in testis (Figs. 5 and Table 2), although the GNF database (26) indicates expression in embryonic tissues as well. Pst61, which is located in a promoter CpG island associated with Gata2 and is methylated only in liver (Figs. 2 and 3) had the lowest level of expression in liver [GNF database (26)] (Fig. 5 and Table 2). Pvu2 and Pvu66 are located within nonpromoter region exons, and there is no apparent association between DNA methylation and tissue-specific silencing of genes associated with these TDMs. In fact, methylation of Pvu66 in liver is coupled with a

![Graphical representation (SYBR Green RT-PCR) of tissue expression of genes associated with TDMs relative to expression in liver. Expression levels are shown in bold, and ΔCt values with the lowest mRNA level are shown in italics. A difference of one cycle is equivalent to an ~2-fold difference in mRNA. Also indicated is the expression pattern determined from the GNF gene expression database (http://symatlas.gnf.org/SymAtlas). Kid, Kidney; Liv, liver; Col, colon; Tes, testis; Mus, muscle.](http://www.pnas.org/cgi/doi/10.1073/pnas.0408436102)

### Table 2. Tissue expression pattern of genes associated with TDMs: quantitative RT-PCR analysis

<table>
<thead>
<tr>
<th>Locus</th>
<th>TDM location</th>
<th>Methyl. (tissue)</th>
<th>Gene</th>
<th>RT-PCR method</th>
<th>Brain</th>
<th>Kid</th>
<th>Liv</th>
<th>Col</th>
<th>Tes</th>
<th>Mus</th>
<th>GNF expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pst3</td>
<td>5′ promoter</td>
<td>U (testis)</td>
<td>Ddx4</td>
<td>SYBR-G</td>
<td>13.5 ± 0.6</td>
<td>14.6 ± 1.5</td>
<td>18.2 ± 1.7</td>
<td>14.6 ± 0.7</td>
<td>−1.5 ± 0.3</td>
<td>17.4 ± 1.4</td>
<td>Testis only/high expression</td>
</tr>
<tr>
<td>Pst6</td>
<td>3′ exon</td>
<td>U (testis)</td>
<td>Hsp1-like</td>
<td>SYBR-G</td>
<td>9.4 ± 1.4</td>
<td>7.6 ± 1.6</td>
<td>8.6 ± 1.2</td>
<td>5.9 ± 0.9</td>
<td>−0.9 ± 0.1</td>
<td>10.7 ± 0.5</td>
<td>Testis only/high expression</td>
</tr>
<tr>
<td>Pst6</td>
<td>3′ exon</td>
<td>U (testis)</td>
<td>Hsp1-like</td>
<td>TaqMan</td>
<td>8.9 ± 1.3</td>
<td>10 ± 0.3</td>
<td>12.8 ± 1.1</td>
<td>10.7 ± 0.9</td>
<td>−1.3 ± 1.1</td>
<td>6.7 ± 3.6</td>
<td>Testis only/high expression</td>
</tr>
<tr>
<td>Pvu29</td>
<td>5′ promoter</td>
<td>U (testis)</td>
<td>RIKEN cDNA</td>
<td>SYBR-G</td>
<td>7.2 ± 0.2</td>
<td>9.1 ± 0.4</td>
<td>8.8 ± 0.6</td>
<td>8.9 ± 0.5</td>
<td>2.7 ± 0.7</td>
<td>10.9 ± 0.3</td>
<td>Broad pattern/high expression in testis</td>
</tr>
<tr>
<td>Pst61</td>
<td>5′ promoter</td>
<td>M (liver)</td>
<td>Gata2</td>
<td>SYBR-G</td>
<td>9.7 ± 0.4</td>
<td>7.5 ± 1.1</td>
<td>12.4 ± 0.9</td>
<td>11.1 ± 1.1</td>
<td>5.6 ± 0.4</td>
<td>10.2 ± 0.1</td>
<td>Broad pattern/low expression in liver</td>
</tr>
<tr>
<td>Pst61</td>
<td>5′ promoter</td>
<td>M (liver)</td>
<td>Gata2</td>
<td>TaqMan</td>
<td>3.9 ± 1.5</td>
<td>1.6 ± 0.3</td>
<td>9.6 ± 1.8</td>
<td>6.4 ± 0.9</td>
<td>9.1 ± 0.6</td>
<td>7.6 ± 3.5</td>
<td>Broad pattern/low expression in liver</td>
</tr>
<tr>
<td>Pvu2</td>
<td>3′ exon</td>
<td>U (testis)</td>
<td>Mct-like</td>
<td>SYBR-G</td>
<td>8.4 ± 0.8</td>
<td>7.4 ± 0.4</td>
<td>8.1 ± 1.4</td>
<td>4.2 ± 0.9</td>
<td>7.5 ± 0.7</td>
<td>11.3 ± 1.6</td>
<td>ND</td>
</tr>
<tr>
<td>Pvu66</td>
<td>3′ exon</td>
<td>M (liver)</td>
<td>Adra1b</td>
<td>SYBR-G</td>
<td>8.1 ± 0.3</td>
<td>10.9 ± 1.1</td>
<td>6.9 ± 1.5</td>
<td>12.5 ± 2.5</td>
<td>11.8 ± 0.8</td>
<td>12.9 ± 5.7</td>
<td>Broad pattern/low expression in liver</td>
</tr>
</tbody>
</table>

Genes associated with TDMs were identified by BLAT by using the University of California, Santa Cruz, Genome Bioinformatics database (http://genome.ucsc.edu). The methylation (methyl.; U, unmethylated; M, methylated) state (inferred by RLGS, MSP, and/or bisulfite sequencing) and location of the TDM relative to the gene are indicated. Although Pst6 is located in a 3′ exon, a CpG island 500 bp upstream of the promoter for Hspa1l had a methylation status similar to Pst6 (data not shown). Quantitative, real-time PCR was performed by using SYBR Green (SYBR-G) and TaqMan probes (see Methods). Expression levels were standardized to GAPDH by calculating ΔCt (ΔCt = Gene Ct − GAPDH Ct). The lower the ΔCt, the higher the mRNA level. ΔCt values with the highest mRNA level are shown in bold, and ΔCt values with the lowest mRNA level are shown in italics. A difference of one cycle is equivalent to an ~2-fold difference in mRNA. Also indicated is the expression pattern determined from the GNF gene expression database (http://symatlas.gnf.org/SymAtlas). Kid, Kidney; Liv, liver; Col, colon; Tes, testis; Mus, muscle.
higher level of expression of Adra1b (Fig. 5 and Table 2). These results suggest that tissue-specific DNA methylation in CpG island promoter regions is associated with tissue-specific gene repression. Fig. 5 shows the tissue-specific expression level of each of the genes relative to expression in liver ($\Delta \Delta Ct = \Delta Ct_{(liver)} - \Delta Ct_{(tissue)}$), clearly indicating the tissue-specific expression of genes associated with Pst3, Pst6, and Pvu29.

TDMs. Based on RLGS analysis of six tissues, our results indicate that about 5% of the RLGS loci are TDMs. Previous RLGS analysis of mouse stem cells before and after differentiation, as well as somatic tissues (brain, kidney, placenta, and sperm) indicated that $\approx 16\%$ of the RLGS loci exhibit differences (13). However these studies did not identify the DNA demethylated in Dnmt3a and Dnmt3b (13). However these studies did not identify the DNA demethylated in Dnmt3a and Dnmt3b (13). It is presently not known whether any of these regions associated with tissue-specific gene expression. Strichman-Almashanu et al. (28) identified $\approx 40$ CpG islands that are normally methylated in the human genome. It is presently not known whether any of these regions associated with tissue-specific gene expression. Recently, Hattori et al. (29) identified 236 loci by virtual RLGS that are demethylated in Dnmt3a$^{-/}$ and Dnmt3b$^{-/}$ ES cells. Two of 40 loci that were further confirmed by methylation-sensitive real-time PCR correspond to TDM loci identified by virtual RLGS in our study (Pst13 and Pst42; see Table 4).

The results presented here indicate that the majority of the TDMs appear to be associated with 5’ promoter CpG islands and may have important roles in establishing or maintaining gene silencing during or after tissue differentiation. Other studies also support a role for DNA methylation in tissue-specific gene expression. Recent results indicate that the methylation of the maspin (SERPINB5) gene promoter contributes to the cell-type-specific expression pattern (30). Methylation of a critical CpG site within a signal transducer and activator of transcription-3-binding element of the glial fibrillary acidic protein promoter regulates expression in developing brain (31). Also, the promoter regions of a number of germ-line-specific genes, such as MAGE-type genes, are densely methylated in all tissues except testis, in which they are expressed (32). It is also possible that tissue-specific domains of active or inactive chromatin include several genes.

It is important that future studies establish the timing of DNA methylation at TDMs relative to alterations in gene expression during development and tissue differentiation. An unmethylated promoter region may be necessary but not sufficient for gene expression. Also, it is possible that silencing precedes DNA methylation. Although our results indicate that tissue-specific DNA methylation is associated with gene expression patterns, it is quite possible that there are other mechanistic consequences. In this context, it will be important to determine the significance of the nonpromoter TDMs. Almost half of the differentially methylated regions are not associated with 5’ promoter CpG islands and may have novel effects on genome structure and gene regulation. Preliminary analysis does not link the TDMs to any common repetitive sequence. Additional bioinformatics analysis of the sequence and genomic features of these regions should help to elucidate the critical roles of DNA methylation in development and disease.

We thank Michael Higgins and Dominic Smiraglia for critical comments on the manuscript. This research was supported by National Cancer Institute Grant CA102423 (to W.A.H.) and National Cancer Institute Core Center Grant CA16056 (to Roswell Park Cancer Institute).

Supplementary Figure 1

Mouse 1       Mouse 2

Testis

Brain

Kidney

Liver

Spot Present

Spot Reduced or Absent
NPP VuII loci

Liver 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Kidney 19 20 21 22 23 24 25 26
Brain ND
Colon
Muscle
Testis

unmethylated in testis only
unmethylated in liver only
unmethylated in brain only
unmethylated in kidney only
unmethylated in colon only
unmethylated in 2 tissues
unmethylated in 3 tissues
unmethylated in 4 tissues
unmethylated in 5 tissues

NPP stI loci

Liver 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Kidney 19 20 21 22 23 24 25 26
Brain
Colon
Muscle
Testis

unmethylated in testis only
unmethylated in liver only
unmethylated in brain only
unmethylated in kidney only
unmethylated in colon only
unmethylated in 2 tissues
unmethylated in 3 tissues
unmethylated in 4 tissues
unmethylated in 5 tissues

Methylated  Partial Methylation  Unmethylated

1 unmethylated in testis only
2 unmethylated in liver only
3 unmethylated in brain only
4 unmethylated in kidney only
5 unmethylated in colon only
6 unmethylated in 2 tissues
7 unmethylated in 3 tissues
8 unmethylated in 4 tissues
9 unmethylated in 5 tissues
Supplementary Figure 2

(a) Chromatograms showing the DNA sequences from liver and testis. The sequences are aligned with the Mct-like region.

(b) Schematic diagram illustrating the Mct-like region and the BGS PCR product. The Not I restriction site is indicated.
Kidney

Testis

Ex 18 Ex 2 Ex 1

53.9 Kb Pst 3 (362 bp)

CpG island (543 bp, 68 CpGs)

NotI site

Bisulfite Sequencing (450 bp, 50 CpGs)

Testis

Liver

Kidney

Colon

Muscle

Brain

- Methylated
- Partial methylated
- Unmethylated

Fig. 6 (131bp, 26 CpGs)