Diverse DNA Methylation Statuses at Alternative Promoters of Human Genes in Various Tissues

Jieun Cheong,1,2 Yoichi Yamada,3 Riu Yamashita,4 Takuma Irie,1 Akinori Kanai,1 Hiroyuki Wakaguri,1 Kenta Nakai,1,4 Takashi Ito,5 Izumu Saito,1,2 Sumio Sugano,1 and Yutaka Suzuki1,*

Department of Medical Genome Sciences, Graduate School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan1, Graduate School of Medicine, the University of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo 113-0033, Japan2, Division of Electrical Engineering and Computer Science, Graduate School of Natural Science and Technology, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan3, Human Genome Center, Institute of Medical Science, the University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, 108-8639 Tokyo, Japan4 and Department of Computational Biology, Graduate School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan5

(Received 24 May 2006; revised 22 August 2006; published online 26 October 2006)

Abstract

We characterized the DNA methylation status at 144 tissue-biased and 37 non-tissue-biased alternative promoters of 61 human genes in five normal tissues. Analysis of the collected data revealed that (i) DNA methylation status differed greatly among alternative promoters belonging to the same gene; (ii) DNA methylation status differed between tissues for the majority of the individual promoters, and (iii) 80–90% of CpG-island-containing promoters were not methylated on either allele throughout the five tissues examined. Furthermore, although the statistical significance was not as clear as for the above features, we also found that (iv) the DNA methylation patterns of tissue-biased promoters changed more drastically than those of non-tissue-biased promoters; (v) tissue-biased promoters tended to be less methylated than their respective alternative promoters in the tissues where they were preferentially expressed, and (vi) the ‘null’ methylation pattern of a given promoter was enriched in the tissues where the transcription was most active. These findings together indicate that there are dynamic physiological changes of DNA methylation. DNA methylation appears to play a significant role in differential usage of alternative promoters and may be related to functional diversification between CpG-island-containing promoters and CpG-island-less promoters.

Key words: DNA methylation status; CpG islands

1. Introduction

Rapidly accumulating large-scale data produced from high-throughput analyses of the genome and transcriptome have recently revealed that the use of alternative promoters is prevalent in human genes.1–4 The alternative use of promoters consisting of different modules of transcriptional regulatory elements enables diversified transcriptional regulation within a single locus.5 It is speculated that multi-faceted use of genes may serve as a molecular basis for the complexity of highly elaborated systems in humans, such as the immune system and nervous system, despite the limited number of total genes.6,7 However, in spite of the growing interest in alternative promoters, very little is known about how and on what occasions those promoters are differentially regulated.

We have been analyzing the transcription start sites of mRNAs in humans by collective analyses of cDNAs that were isolated from full-length-enriched cDNA libraries constructed by our cap-targeted method,
By analyzing 1.8 million oligo-cap cDNAs, we recently demonstrated that about half of human protein-coding genes [defined as RefSeq genes (http://www.ncbi.nlm.nih.gov/RefSeq/)] have at least two distinct putative alternative promoters (PAPs) which are separated by >500 bp. On average, there were 3.1 alternative promoters per gene. In 17% of cases, tissue-biased utilization of the alternative promoters was observed. The richest tissue sources of such tissue-biased alternative promoters were testis and brain.

Interestingly, the identified alternative promoters generally consisted of a combination of one promoter which was located in a CpG island (CGI-containing promoter)\(^1\) and other promoter(s) which were not associated with CGI islands (CGI-less promoters). Although about 70% of genes were found to possess at least one CGI-containing promoter, it was rare that a particular gene had more than a single CGI-containing promoter among its alternative promoters. Even when the total number of promoters increased, the number of CGI-containing promoters did not increase proportionally.\(^1\) This biased distribution with regard to the presence of CpG islands led us to hypothesize that the transcriptional regulation mediated by DNA methylation might play some role(s) in the differential regulation of the alternative promoters, since the generation and maintenance of the CpG islands have been shown to be closely associated with the methylation of DNAs.

DNA methylation consists of the addition of a methyl group to the cytosine of the dinucleotide CpG by DNA methyltransferase enzymes in mammals.\(^1\) Such methylation generally represses the gene expression of the downstream gene, either by direct interference with the interaction between transcription factors and their binding sites or by altering the chromatin structure of the surrounding genomic region. The regulation of DNA methylation and demethylation is thought to be crucially important during development, and many studies have shown that disruption of proper DNA methylation has profound effects on carcinogenesis.\(^12,\)\(^13\) DNA methylation is thought to be one of the major causative factors for the generation of CpG islands. Methylated cytosine can be converted to thymine by deamination, and thus, in the absence of selection pressure, CpG dinucleotides are rapidly erased from genomic sequences. Actually, CpG dinucleotides are markedly underrepresented throughout the human genome. This reduction of CpG does not occur in genome regions that are devoid of methylation. Rather, CpG dinucleotides are clustered as CpG ‘islands’, which are known to be especially abundant around promoter regions, and are thought to play an important role in proper transcriptional regulation.\(^14,\)\(^16\)

In this study, in order to clarify how the differential regulation of alternative promoters is realized under various cellular circumstances, we analyzed the DNA methylation status of alternative promoters and compared the findings with the tissue-usage-bias of the promoters. Among our collection of alternative promoters, we selected genes with alternative promoters whose usage was biased between tissues, since differential promoter usage was most marked in this population. We characterized the DNA methylation status of the genomic regions proximal to the transcriptional start sites (−1 kb to +200 bp; transcriptional start site: 0) for 181 alternative promoters in 61 genes in five normal tissues in total. Here we describe our first study of the tissue-profile of DNA methylation in a wide variety of normal tissues.

2. Materials and methods

2.1. Primer design

PCR primers were designed using Primer3 (http://fokker.wi.mit.edu/cgi-bin/primer3/primer3, www/cgi). For primers for HpaII-McrBC PCRs, the parameters were set as: primer length: 20–25 bp; melting temperature: 50–60°C; target length: 1 kb; and for primers for bisulfite genomic sequencing, the parameters were set as: 20–30 bp; melting temperature: 50–70°C; target length: 300–600 bp. The RT–PCR primers were localized inside of the first exons that were unique to the respective alternative promoters using the following parameters: primer length: 20–25 bp; melting temperature: 50–60°C; target length: 50–200 bp. The sequence of each of the designed primers is shown in Supplementary Table S1. Also note that, since in almost all cases, each of the alternative promoters contained multiple transcriptional start sites, the sites used to excise the upstream putative promoters were the sites where the largest number of the 5’-ends of the oligo-cap cDNAs were mapped. The human genome information was in accord with UCSC hg17 (http://genome.ucsc.edu/) except for the locations of CpG islands (see below).

2.2. HpaII-McrBC PCR assay

Normal human genomic DNAs were obtained from BioChain (BioChain; USA). Each genomic DNA sample was incubated in 1 M NaCl solution overnight at 37°C. The genomic DNAs were purified by phenol-chloroform extraction and ethanol precipitation and dissolved in distilled water. Aliquots (5–15 µg) were digested with 60 U of HpaII (Takara; Japan), MspI (Takara; Japan) or McrBC (New England Biolabs; USA) overnight at 37°C in 100 µl of reaction mixture following the instructions of the manufacturer. To assure complete digestion, the
digestion step was repeated once more. Control samples were treated in the same way without the addition of the restriction enzyme. After incubation, digested samples were purified by phenol–chloroform extraction and ethanol precipitation and dissolved in distilled water.

For each digestion, the amount of undigested DNA was measured by Quantitative (Real Time) PCR using 7900HT (ABI; USA). For PCR, 1.0 µl (20–50 ng) of digested genomic DNA was used in a 20 µl reaction solution containing: 10 µl of SYBR GREEN PCR Master mix (ABI; USA): 0.125 pmol of each primer: 4 µl of distilled water. The PCR conditions were 50°C for 2 min, 95°C for 10 min, 5 cycles of: 95°C for 30 s, 58°C for 30 s, 60°C for 1 min, 5 cycles of: 95°C for 30 s, 56°C for 30 s, 60°C for 1 min, 30 cycles of: 95°C for 30 s, 55°C for 30 s, 60°C for 1 min, and 60°C for 5 min. The integrity of the PCR products (lack of primer-dimers or non-specific amplicons) was confirmed by agarose gel electrophoresis. The integrity was also assured by measuring the monomial of the dissociation curve of the PCR products using Dissociation Curve Analysis in SDS Analytic Tools (ABI; USA).

2.3. Bisulfite genomic sequencing

Five micrograms of the normal human genomic DNAs (BioChain; USA) were treated in a 600 µl sodium bisulfite reaction solution containing: 10 mM hydroquinone, 8 M urea and 3.6 M bisulfite (Sigma; USA) at pH 5.0 (pH adjustment with 10 M NaOH) at 55°C overnight, according to the standard procedure.\(^{17,18}\) Fifty nanograms of the bisulfite-treated DNA was used for PCR amplification in a 10 µl reaction mixture containing: 10 nM Tris–HCl at pH 7.5, 50 mM KCl, 1.5 mM MgCl\(_2\), 1 mM DTT, 10 nM 2-mercaptoethanol, each dNTP at 0.25 mM, each primer at 0.25 µM and 2.5 U of Ex-Taq DNA polymerase (TaKaRa; Japan). The PCR conditions were: 95°C for 3 min, 5 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, 5 cycles of 95°C for 30 s, 52°C for 30 s, 72°C for 1 min, 35 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 1 min and a final incubation at 72°C for 5 min. The amplified PCR products were purified using the Wizard DNA Clean-up system (Promega; USA) and were TA-cloned into pCR2.1 vector (Invitrogen; USA). Ten individual clones were sequenced from both ends using a 3730 ABI sequencer according to the standard protocol (ABI; USA).

2.4. Data processing

Bias in the usage of the alternative promoters between tissues (Table 1) and statistical significance of the enrichments (Tables 3 and 4 and Fig. 3B) were evaluated by calculating hyper-geometric distributions using the following equation:

\[
\sum_{x=i}^{N_p} \binom{N_p}{x} \cdot \binom{N(1-p)}{n-x} \cdot \binom{n}{N}^{-1},
\]

where \(N\) = total number of data, \(p\) = expected frequency of the feature of concern estimated from the overall data (number of feature-associated data/\(N\)), \(n\) = total number of data belonging to the subgroup, and \(i\) = number of the data which belong to the subgroup and which are associated with the feature of concern within the subgroup. For example, for Table 1: \(N\) = total number of cDNAs belonging to the locus, \(p\) = expected frequency of the cDNAs associated with a given tissue within the locus (number of associated cDNAs/\(N\)), \(n\) = total number of cDNAs belonging to the PAP of concern, and \(i\) = number of the cDNAs associated with a given tissue within the PAP.

CpG islands were defined as a genomic region that was 200 bp or greater in length with a C+G content >50% and a CpG percentage that exceeded by 60% that expected in random sequences.\(^{11}\)

As for the difference in DNA methylation status between tissues or between PAPs belonging to the same genes, ‘Changed’ was defined as meaning that there were at least two different statuses of DNA methylation in the group. ‘Not Changed’ was defined as meaning that there was only one status of DNA methylation in the group. For defining ‘Not Changed’, we required that the data from at least two PAPs were not ‘n.d.’ for analyzing mutual differences between PAPs, and we required at least three PAPs for analyzing tissue profiles.

2.5. Semi-quantitative (real-time) RT–PCR assays

Total RNAs (50 µg) of five normal human tissues were obtained from BioChain (BioChain; USA). We
synthesized first strand cDNAs with 20 pmol of dT primer (17 mer of dT). The quality and quantity of the obtained RNA were assessed using BioAnalyzer (Agilent; USA). Real-time PCR was performed using the reagents described above. The PCR conditions were: 50°C for 2 min, 95°C for 10 min, and 30 cycles of 95°C for 30 s, 60°C for 1 min. The results from three independent experiments were averaged. The PCR products were also analyzed by agarose gel electrophoresis to assess the integrity (single-band-ness). The relative expression level was evaluated as the deviation of the Ct value (constant of threshold) from the Ct values averaged for the five tissues using the equation:

\[ \text{Expression(relative)} = 2^{\frac{C_t}{C_{t\text{average}}}} \]

3. Results

3.1. Selection of alternative promoters

In our previous study, we generated a collection of 30 707 PAPs in 7674 RefSeq genes. Among them, we identified 1731 PAPs in 1287 genes which showed biased usage between tissues or cell types [tissue-biased PAPs (TB-PAPs); \( P < 0.01 \); statistical significance was evaluated by calculating hyper-geometric distributions using the formula shown in Materials and Methods]. We further selected the genes that had at least two TB-PAPs containing restriction enzyme sites essential for the HpaII-McrBC PCR assays in the region from −1 kb to +200 bp of the transcription start sites. As a result, we obtained 61 genes that had 144 PAPs. Also, we selected an additional 37 PAPs belonging to the same genes but showing no tissue bias (non-TB-PAPs). Thus, we selected a total of 181 PAPs in 61 genes as targets for methylation analysis (Table 1). For a schematic representation of the selection flow chart, see Supplementary Figure S1. For detailed information about each of the promoters, see Supplementary Table S2. Also, the numbering of the PAPs corresponds to that used in our database, DBTSS (http://dbtss.hgc.jp/), so that visual inspection can be performed for each of the PAPs.

Although we selected only 144 (about 8%) TB-PAPs of 1731 for the following assays, we think that this promoter set retains the overall character of TB-PAPs. As described in our previous report, testis and brain were the richest source of the tissue-biased alternative promoters. This was also true for the present set (Table 1). Also, among the total set of 181 promoters, 59 (33%) contained CpG-islands and 122 (67%) were CpG-island-less (Table 1). This frequency was similar to that of the total 30 707 PAPs (CpG-island-containing: 34%, see Ref.1). We therefore consider the features of the current set of promoters to be representative of the overall features of alternative promoters at least in these respects.

3.2. Measurement of the DNA methylation status of promoters in five normal tissues

We examined the DNA methylation status using HpaII-McrBC PCR assays (HM-PCR assays). For each PAP, the DNA methylation status in five normal tissues, namely, brain, testis, spleen, stomach, and liver, was determined. According to the observed PCR band patterns, we classified the DNA methylation status into four groups: ‘complete’, ‘composite’, ‘null’ and ‘incomplete’ methylation (abbreviated CL, CP, NL and IC respectively), which are classification categories often employed in previous studies (Fig. 1A; and see Ref. 19). Typical examples of PCR band patterns assigned to each of the groups are shown in Fig. 1B.

We observed 138 (24%) ‘complete’, 264 (45%) ‘null’, 165 (28%) ‘composite’, and 16 (3%) ‘incomplete’ methylations among the 144 TB-PAPs (Table 2). For these 583 cases, the classification was entirely straightforward based on the PCR band patterns (the presence or absence of the PCR bands; see Fig. 1B and the legend). Moreover, we used real-time PCR for the quantification of the band amplification. Thus, we could assess the extent of the presence or absence of the bands in a quantitative manner (for the quantitative data, see Supplementary Figure S2). In 137 (15% of a total of 905) cases, the band patterns were ambiguous (or a mixture) between the two categories, and thus they were classified as ‘not determined (n.d.)’. (For further discussion of this issue, see Supplementary Figure S4). For the 37 non-TB-PAPs, we classified 37 (24%), 64 (42%), 47 (31%), 5 (3%) and 32 (18%) of total 181 promoters as ‘complete’, ‘null’, ‘composite’, ‘incomplete’ and ‘n.d.’, respectively (Table 2). It was interesting that ‘composite’, which is the category in which one of the two alleles is completely methylated and the other is null methylated, was observed far more frequently than ‘incomplete’, which is the category in which both alleles are partially methylated. This suggests that the gene expression statuses from two alleles might differ for a significant population of genes, although it is also possible that the presence of mixed populations of multiple cell types within the used tissues accounted for this observation (Note that the ‘composite’ pattern can raise from cell-type difference, not from allelic difference. If both alleles are completely methylated in certain cell types but completely unmethylated in other cell types in the same tissue, this can give a composite pattern. For extensive analysis regarding this issue, see Supplementary Figure S5). This general tendency was also observed for both tissue-biased and non-tissue-biased PAPs. Altogether, we obtained data explicitly representing the DNA methylation status of
736 proximal promoters. This dataset is the largest so far collected for the DNA methylation status at promoters in normal tissues.

We further confirmed the results obtained from the HM-PCR assays by bisulfite genomic sequencing assays. Since the bisulfite genomic sequencing assays could not be scaled up to high-throughput assays, we selected 5, 4, 15, 11 cases as representative cases randomly selected from the cases categorized as ‘complete’, ‘composite’, ‘null’ and ‘incomplete’, respectively, by HM-PCR assays (Fig. 1A; results from 35 PAPs are presented in Supplementary Table S3).
cases, the classifications obtained from the two assays were consistent with each other. This strong agreement between the two assays provides support for our conclusion real-time PCR was useful for quantitatively discriminating distinct DNA methylation patterns in the HM-PCR assay.

3.3. Variable methylation patterns between alternative promoters and between tissues

At first glance, we were surprised by the variable pattern of DNA methylation at the alternative promoters. Fig. 2 illustrates the case of the SEPT9 gene as an extreme example. The SEPT9 gene was identified as a causative gene of an autosomal dominant hereditary disease, neuralgic amyotrophy, and is thought to play roles in cell proliferation. This gene possessed at least 20 alternative promoters, represented by an array of 20 mutually exclusive first exon variants (Fig. 2A). Of these, the DNA methylation status was determined for 14 in this study (Fig. 2B). A surprising aspect of the data is that no two PAPs in Fig. 2B share the same methylation pattern variation among the five tissues. For example, PAP-ID1 and 2 are separated by only 800 bases. Yet, PAP-ID1 showed the CP-CP-CL-CP-NL pattern, while PAP-ID2 showed the NL-NL-NL-NL-NL pattern. PAP-ID10 showed the CP-CL-CL-CP-CP pattern, which was similar to the pattern of PAP-ID11, namely, CP-CL-CP-CP-CP. However, these two promoters are separated by 4 kb. These observations suggest that the DNA methylation pattern may be unique for each PAP even within the same gene. This could allow very fine tuning of the gene expression. Only PAP-ID5 and PAP-ID13 showed a uniform ‘composite’ patterns.
methylation pattern among the five tissues (although surprisingly, PAP-ID5 and PAP-ID13 are separated by 100 kb). PAP-ID2 also showed a uniform ‘null’ methylation pattern. For the remaining 11 PAPs (79% of the 14 PAPs), the methylation pattern clearly varied among the five tissues. Thus, variable methylation patterns among the tissues may be more common than a uniform pattern.

In order to obtain a more generalized picture, we compared the methylation patterns between PAPs belonging to the same gene. Although we assayed 181 PAPs for 61 genes, there was about 20% ‘n.d.’ data. Thus, there were 48 genes for which we had data for at least three tissues at two PAPs per gene. Only five of these 48 genes showed a perfect match of PAP methylation patterns in all five tissues (Table 3). What we observed for SEPT9 was true for all the genes we analyzed. It is noteworthy that the methylation pattern was uniformly ‘null’ for the matched PAPs. There were 10 PAPs for five genes that showed a perfect match. They were ‘null’ methylated in all five tissues (Table 3). The number of genes with a perfect match would only be seven even if a mismatch in one tissue were allowed. There were two additional genes possessing five PAPs that had one mismatch (Table 3). Two of them showed a ‘composite’ pattern in liver, while the rest were all ‘null’ in the other tissues (Table 3).

Indeed, the DNA methylation status of the individual PAPs differed greatly depending on the tissue. As shown in Table 3, the DNA methylation status of 93 of 154 (60%) PAPs varied among the five tissues. This confirms that variable methylation patterns among the tissues may be more common than a uniform pattern. We also noted that the TB-PAPs seemed to have more highly variable methylation statuses than non-TB-PAPs. TB-PAPs had variable methylation patterns in 77 of 124 cases (62%), while only 16 of 30 (53%) non-TB-PAPs had such patterns, indicating slightly greater variability in the former, although the statistical significance was not very high ($P < 0.14$; evaluated by calculating hyper-geometric distributions following the equation shown in Materials and Methods), possibly due to the limited size of the current dataset.

### Table 2. DNA methylation status of the TB-PAPs.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>Testis</th>
<th>Brain</th>
<th>Spleen</th>
<th>Stomach</th>
<th>Liver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>31 (28%)</td>
<td>33 (27%)</td>
<td>42 (34%)</td>
<td>12 (11%)</td>
<td>20 (17%)</td>
</tr>
<tr>
<td>Null</td>
<td>47 (42%)</td>
<td>48 (39%)</td>
<td>50 (41%)</td>
<td>52 (46%)</td>
<td>67 (58%)</td>
</tr>
<tr>
<td>Composite</td>
<td>30 (27%)</td>
<td>39 (32%)</td>
<td>27 (22%)</td>
<td>44 (39%)</td>
<td>25 (22%)</td>
</tr>
<tr>
<td>Incomplete</td>
<td>4 (3%)</td>
<td>2 (2%)</td>
<td>3 (3%)</td>
<td>4 (4%)</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>Subtotal</td>
<td>112 (100%)</td>
<td>122 (100%)</td>
<td>122 (100%)</td>
<td>112 (100%)</td>
<td>115 (100%)</td>
</tr>
<tr>
<td>n.d.</td>
<td>32</td>
<td>22</td>
<td>22</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>144</td>
<td>144</td>
<td>144</td>
<td>144</td>
<td>144</td>
</tr>
</tbody>
</table>

The numbers of PAPs whose DNA methylation status was categorized as the corresponding patterns in the indicated tissue are shown. Frequencies of the observed patterns were calculated relative to the subtotal, which excluded “n.d.”, and are shown in parentheses.

3.4. **Relationship between tissue-bias and DNA methylation status of alternative promoters**

We further examined the relationship between the tissue-preferential usage of the individual PAPs and the tissue-profiles of their DNA methylation status. We compared the DNA methylation status of the TB-PAPs between their ‘preferred’ tissues and other tissues. We focused this analysis on the testis-prefering PAPs and brain-prefering PAPs, as the largest number of TB-PAPs was obtained from these tissues (also see Table 1). We found that the DNA methylation status at the ‘testis-prefering’ PAPs was considerably biased to ‘null’, and ‘complete’ was extremely rare ($P < 0.003$; evaluated by calculating hyper-geometric distributions of ‘null’ and ‘composite’ against their frequencies in the total population; Table 4). Such bias toward the ‘null’
Figure 2. DNA Methylation Status at the PAPs of the SEPT9 gene. (A) Schematic representation of the structure of the SEPT9 gene. Fourteen PAPs which were selected from 20 and were subjected to the analyses are marked with asterisks. (B) Observed patterns of DNA methylation at each of the PAPs in each of the tissues. In the 2nd column CGI-containing PAPs are represented as 1, and CGI-less PAPs as 0. In the 3rd column, tissue-preferences of the PAPs are shown.
state was not seen in other tissues for testis-preferring PAPs or for non-TB-PAPs in the testis. These observations suggest that unique regulation may be exerted for TB-PAPs in the testis. We observed no such bias for the other PAPs of the genes that had testis-preferring PAPs with ‘null’ methylation in the testis (Table 4).

When the same analysis was performed for brain-preferring PAPs, we did not see any bias toward ‘null’. However, interestingly, in brain, the DNA methylation patterns in accompanying PAPs of the same gene were skewed. As shown in Table 4, even when the brain-preferring PAPs were ‘compositely’ methylated, their counterpart PAPs rarely had ‘null’ status. Although further extensive data collection will be necessary, these observations indicate that there may be several distinct methylation mechanisms for tissue-biased gene expression depending on the tissue type. Nevertheless, it seemed that actively used PAPs are generally less extensively methylated than their accompanying PAPs.

Table 3. Changes in the DNA methylation status between PAPs and between tissues, the numbers of TB-PAPs and non-TB-PAPs whose DNA methylation status was Changed (upper line) or Not-Changed (lower line) between tissues

<table>
<thead>
<tr>
<th>RefSeq# or Promoter ID</th>
<th>CpGisland</th>
<th>Biased expression</th>
<th>Testis</th>
<th>Brain</th>
<th>Spleen</th>
<th>Stomach</th>
<th>Liver</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002065</td>
<td>1</td>
<td>Testis</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_002394</td>
<td>0</td>
<td>Lung</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_004046</td>
<td>0</td>
<td>Lung</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_005104</td>
<td>0</td>
<td>Brain</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_006759</td>
<td>1</td>
<td>Other</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_008916</td>
<td>0</td>
<td>Other</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
</tr>
<tr>
<td>NM_018093</td>
<td>0</td>
<td>Testis</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>CP</td>
</tr>
<tr>
<td>NM_018093</td>
<td>0</td>
<td>Testis</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>NL</td>
<td>CP</td>
</tr>
</tbody>
</table>

(A) The exceptional cases in which the DNA methylation patterns of a PAP perfectly matched with those of the accompanying PAPs in all five tissues

(B) The cases in which the DNA methylation patterns of a PAP matched with those of the accompanying PAPs in all five tissues, when one mismatch was allowed

3.5. Relationship between DNA methylation status and mRNA expression level

In order to examine the relationship between the DNA methylation status and the level of mRNA expression, we measured the relative amount of mRNAs transcribed from each of the alternative promoters in the five tissues. For this purpose, we used real-time RT–PCR assays. Based on our full-length cDNA information, RT–PCR primers were designed at unique regions inside of the first exons of the transcript variants corresponding to each of the PAPs which showed different DNA methylation statuses between tissues. Ninety-three such primer pairs were considered, and in 49 cases, positive signals from either of the tissues were observed. (For an example, see Fig. 3A.) Fig. 3B shows the distribution
of the relative expression level calculated separately for each of the DNA methylation patterns. Although the difference in the distributions of the relative expression levels between ‘null’ and ‘composite’ populations was not highly statistically significant as a whole, we still observed that ‘null’ or relatively light methylation patterns of PAPs were enriched (P < 0.02; hypergeometric distributions calculated similarly as in Section 3.4.) in the tissues in which the mRNA expression was the strongest (relative expression level/C21/C8; also see Fig. 3A, Supplementary Table S2). We also observed that ‘complete’ DNA methylation was the only pattern in which the mRNA expression was the weakest (relative expression level/C20/C1/C8). It is highly possible that the changes in the DNA methylation status are responsible for the modulations of the gene expression in these cases.

3.6. Relationship between DNA methylation status and the presence of CpG islands

We further attempted to clarify the relationship between the DNA methylation status and the presence of CpG islands. We compared the DNA methylation status between the populations of CGI-containing PAPs and CGI-less PAPs and found a striking difference. In contrast to the so-far observed variable methylation patterns, we found that the CGI-containing PAPs were predominantly ‘null’ methylated throughout the five tissues (P < 1.0E–55; evaluated by standard Fisher’s test; Table 5). Consequently, the DNA methylation status was unchanged throughout the tissues, remaining constantly ‘null’. These tendencies held irrespective of whether the PAPs were TB-PAPs or non-TB-PAPs (P < 30E–9; evaluated by standard Fisher’s test; Table 5). In contrast, the DNA methylation in CGI-less PAPs was highly variable among tissues. When the CGI-containing PAPs were ‘null’ methylated, the pattern of the DNA methylation of the other PAPs of the same gene, which were usually CGI-less, was similar to the overall distribution patterns.

4. Discussion

In this study, we characterized the DNA methylation status at 181 PAPs in 61 genes which were identified
by our previous full-length cDNA analyses. This is the
first report describing the profiles of the DNA methyla-
tion of PAPs in a variety of tissues and may be the
largest collection of data on the methylation of human
promoters in general for normal tissues. In total, we
analyzed the methylation status of 8612 CpG dinucleo-
tides in 181 PAPs located in the proximal regions
of transcriptional start sites (also see Supplementary
Table S2). Although this number is small relative to the
total number of CpG dinucleotides existing within the
whole genome, this novel analysis of these CpG islands
should be significant, because they may have direct
biological consequences via modulating the transcription
levels of genes.

Many of the previous studies using DNA-chip-based
technologies were focused on comparing the DNA
methylation status between cancerous cells and normal
cells. From the data we collected here, we found:
(i) PAPs of the same gene tend to have different
methylation pattern variation depending on the tissue;
(ii) DNA methylation status differs among tissues for the
majority of the individual PAPs; (iii) CGI-containing
PAPs are an exception to observation, (ii) in that they
tend to have a uniform ‘null’ pattern throughout the
tissues examined.

The last result is somewhat of a confirmation of the
findings of our previous study, although it was not
previously tested on this scale. We made a comprehen-
sive analysis of CpG islands on human chromosome
21 and observed that 84% of CpG islands located within
promoter regions were ‘null’ methylated in leukocytes. The
general lack of methylation of promoter CpG islands,
regardless of whether the corresponding genes are
expressed or not, was also suggested by other
previous studies, although the numbers of the assayed
promoters and cell types were limited. In this study,
we assayed 181 PAPs in five normal tissues and observed
that most of the CGI-containing PAPs were
constantly ‘null’ methylated irrespective of the tissue
(Table 5). Based on this observation, we concluded
that methylation appears not to be the reason for the
biased expression of most of the CGI-containing
TB-PAPs.

In contrast to the case of the CGI-containing
TB-PAPs, we found surprisingly wide variation in the
methylation pattern among the tissues examined. We
observed that the majority of TB-PAPs showed alteration
of their DNA methylation status among tissues
(62%; Table 5). If we counted only CGI-less TB-PAPs,
the proportion went up to 80% (Table 5). We also found

Figure 3. Methylation Pattern and mRNA Expression in the Five Tissues. (A) Examples of the semi-quantitative (real-time) RT-PCR results.
The Ct values (constant of threshold; as of the default setting of ABI7900HT) and the relative expression levels are shown in the margin. The
results of a similar analysis for the GAPDH mRNA are shown in Supplementary Figure S3, although we did not use GAPDH as a normalization
standard. Ct: Constant of threshold as for default setting. Also note that relatively high expression levels are correlated with relatively light
methylation patterns in both cases, even if their expression patterns are non-specific. (B) Composition of the DNA methylation patterns in the
population of the PAPs with the indicated relative expression level. Based on the results of the semi-quantitative (real-time) RT-PCR assays,
the relative expression levels were calculated as the deviation from the averaged expression level in the five tissues [also see (A) and Materials
and Methods]. The relative expression levels observed for each of the PAPs in each of the tissues were correlated with their DNA-methylation
status and summed for the five tissues.
that TB-PAPs seemed to have more variation of their methylation patterns than non-TB-PAPs (Table 3). We examined only five tissues in this study. We expect that the proportion of PAPs that show a variable methylation pattern among the tissues may increase if the number of tissues assayed is increased. Since we assayed 181 PAPs that were located on various chromosomes, these results indicate that various regions in the human genome are differentially methylated among various tissues. This suggests that the methylation of promoters is more dynamically regulated from one tissue to another than previously anticipated.

Among 48 genes for which we compared PAPs within the same gene, only about 10% had the same methylation pattern in all five tissues. All of these methylation patterns were ‘null’, as described in Table 3. Actually, CGI-containing PAPs accounted for these uniform ‘null’ patterns. Seven PAPs contained CpG-islands out of 10 PAPs of 5 genes that showed perfect matches of the methylation patterns (Table 3). We could not find any simple correlation of methylation pattern variation between PAPs within the same gene. The methylation of PAPs thus seems to be individually controlled even within the same gene.

We also searched for specific patterns of methylation according to the expression pattern or expression level of the mRNA transcribed from a PAP. We observed that high levels expression correlated with the ‘null’ or relatively light methylation pattern of PAPs, although the statistical significance was not as clear as in the above cases, possibly due to the limited size of the current dataset. There seems to be a tendency for CGI-containing PAPs to be ‘null’ methylated and to show high expression levels. Our previous findings suggested that most genes have only one CGI-containing PAP among all their PAPs.1 It is possible that this unique CGI-containing PAP is the major promoter of the gene, playing a key role among the group of alternative promoters.

Together, in this study, we showed that information about DNA methylation is as indispensable as the mere DNA sequences of the promoters for a thorough understanding of the regulation of the expression of the genes encoded by the human genome. Modifications of the histones, methylation of the DNA, and the chromatin structure are suggested to be related with each other.24 Thus, information about DNA methylation may in general be useful as an indicator of the chromatin structure. Advances in genome and transcriptome research have opened a way for decoding the gene expression regulation of the genomic DNA sequences. Actually, in yeast, a recent report demonstrated that it was possible to predict the level of gene expression directly from the genomic sequence with some accuracy.

---

**Table 5.** Changes in the DNA Methylation Status between CGI-containing and CGI-less Promoters (A) Numbers of observed patterns of DNA methylation status in CGI-containing PAPs and CGI-less PAPs are shown. The DNA methylation patterns observed in five tissues were summed. (B) Numbers of PAPs with “Changed” or “Not-Changed” DNA methylation status in five tissues are shown. (C) Numbers of DNA methylation patterns of CGI-less PAPs when their accompanying CGI-containing PAPs were “null” methylated.

<table>
<thead>
<tr>
<th></th>
<th>CGI-containing PAPs (%)</th>
<th>CGI-less PAPs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>5 (2%)</td>
</tr>
<tr>
<td></td>
<td>Null</td>
<td>202 (87%)</td>
</tr>
<tr>
<td></td>
<td>Composite</td>
<td>23 (10%)</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Total</td>
<td>232 (100%)</td>
<td>504 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CGI-containing PAPs (%)</th>
<th>CGI-less PAPs (%)</th>
<th>PAP Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TB-PAP</td>
<td>Non-TB-PAP</td>
<td>TB-PAP</td>
</tr>
<tr>
<td>Changed</td>
<td>10 (25%)</td>
<td>1 (12%)</td>
<td>67 (80%)</td>
</tr>
<tr>
<td>Not-Changed</td>
<td>30 (75%)</td>
<td>7 (88%)</td>
<td>17 (20%)</td>
</tr>
<tr>
<td>Total</td>
<td>40 (100%)</td>
<td>8 (100%)</td>
<td>84 (100%)</td>
</tr>
</tbody>
</table>

**Accompanying CGI-less PAPs when CGI-containing PAPs = Null**

|                  | Complete | 72 (34%) |
|                  | Null     | 49 (23%) |
|                  | Composite | 83 (40%) |
|                  | Incomplete | 6 (3%) |
| Total            | 210 (100%) |
using clustered microarray data as an education dataset. The sequence information of the DNA together with its higher-level chromatin structure will enable us to better understand how the code of the DNA directs the complex regulation of gene expression in human cells. Based on that knowledge, we should eventually be able to understand how the alternative utilization of the promoters plays roles in the diversification of gene functions.

Acknowledgements: We are grateful to K. Abe, K. Toya and K. Imamura for technical support. We are also thankful to E. Nakajima for critical reading of the manuscript. This work was supported by grants from the New Energy and Industrial Technology Development Organization (NEDO) project of the Ministry of Economy, Trade and Industry (METI) of Japan, the Japan Key Technology Center project of METI of Japan, and a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan.

Supplementary Data: Supplementary data are available online at http://www.dnaresearch.oxfordjournals.org

References


