Genomics 98 (2011) 280–287

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome-wide profiling of DNA methylation in human cancer cells

Katsumi Ogoshi [a,1], Shin-ichi Hashimoto [a,b,2], Yoichiro Nakatani [b,2], Wei Qu [b,2], Kenshiro Oshima [b,2], Katsushi Tokunaga [c], Sumio Sugano [d,3], Masahira Hattori [b,2], Shinichi Morishita [b,2], Kouji Matsushima [a,*,1]

[a] Department of Molecular Preventive Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561 Japan
[c] Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[d] Department of Medical Genome, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561 Japan

ARTICLE INFO

Article history:
Received 11 January 2011
Accepted 8 July 2011
Available online 29 July 2011

Keywords:
DNA methylation
Second generation sequencing technology
Differentially methylated regions
Colon cancer

ABSTRACT

Global changes in DNA methylation correlate with altered gene expression and genomic instability in cancer. We have developed a methylation-specific digital sequencing (MSDS) method that can assess DNA methylation on a genomic scale. MSDS is a simple, low-cost method that combines the use of methylation-sensitive restriction enzymes with second generation sequencing technology. DNA methylation in two colon cancer cell lines, HT29 and HCT116, was measured using MSDS. When methylation levels were compared between the two cell lines, many differentially methylated regions (DMRs) were identified in CpG island shore regions (located within 2 kb of a CpG island), gene body regions and intergenic regions. The number of DMRs in the vicinity of gene transcription start sites correlated with the level of expression of TACC1, CLDN1, and PLEKHC1 (FERMT2) genes, which have been linked to carcinogenesis. The MSDS method has the potential to provide novel insight into the functional complexity of the human genome.

1. Introduction

DNA methylation is an important component of the epigenetic regulation of gene expression in eukaryotic cells. Methylation of CpG dinucleotides within transcriptional regulatory sequences often results in reduced expression or silencing of adjacent genes [1–3]. Global hypomethylation has been associated with chromosomal instability, while hypermethylation of the promoters of certain tumor suppressor genes is an important event in malignant transformation [4–8].

There are three predominant methods for investigating DNA methylation. The first method is an antibody-based, affinity purification procedure that captures methylated DNA [9–11]. The second method is bisulfite sequencing, which converts unmethylated cytosines to uracils that are subsequently recognized as thymines [12,13]. The third method makes use of methylation-sensitive restriction enzymes, which only cleave DNA when their recognition sites are unmethylated [14]. These three approaches can be combined with electrophoresis, tag based sequencing [15], microarray analysis [9,11,16–21] or high-throughput sequencing [22–25].

High-throughput sequencing can report the sequence of millions of DNA fragments in parallel. This technology has revolutionized the analysis of genome-wide histone modifications [26] and the re-sequencing of the human genome [27], as well as the analysis of transcription start sites (TSSs) [28], full-length RNAs [29], small RNAs [30], single nucleotide polymorphisms (SNPs) [31], and transcription factor combination domains [32].

The development of new technology has driven new discoveries in the area of DNA methylation. Meissner et al. showed that alterations in histone modification can reduce DNA methylation [22]. Irizarry et al. demonstrated that most tissue-specific differentially methylated regions (T-DMRs) and cancer-specific DMRs (C-DMRs) localize to CpG island shore regions [21,33]. Moreover, it has been established that DNA methylation in both the promoter and the gene body can affect gene expression [11,16,34,35]. Evolutionarily conserved regions (ECRs) within the gene body have been reported to act as alternative promoters under the control of DNA methylation [36]. On the other hand, DNA methylation occurs not only on CpG dinucleotides but also in non-CG contexts (mCHG and mCHH, where H = A, C or T) in embryonic stem cells [37].

Each method of methylation detection has its advantages and limitations. The use of antibodies allows determination of the methyl cytosine density within a region of interest. The use of methylation-sensitive restriction enzymes is only useful for the analysis of

* Corresponding author. Fax: +81 3 5684 2297.
E-mail addresses: oisho@m.u-tokyo.ac.jp (K. Ogoshi), hashimot@m.u-tokyo.ac.jp (S. Hashimoto), nakatani@cb.k.u-tokyo.ac.jp (Y. Nakatani), quwei@cb.k.u-tokyo.ac.jp (W. Qu), oshima@cb.k.u-tokyo.ac.jp (K. Oshima), tokunaga@m.u-tokyo.ac.jp (K. Tokunaga), ssugano@k.u-tokyo.ac.jp (S. Sugano), hattori@k.u-tokyo.ac.jp (M. Hattori), moris@cb.k.u-tokyo.ac.jp (S. Morishita), koujim@m.u-tokyo.ac.jp (K. Matsushima).
[1] Fax: +81 3 5841 3393.
[2] Fax: +81 4 7136 3977.
[3] Fax: +81 4 7136 4080.

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.07.003

methylation of an enzyme's specific recognition sites. Analysis of the global DNA methylation profile of the genome at the single-base level requires bisulfite sequencing. Whole genome bisulfite sequencing has been accomplished in Arabidopsis and humans by using second generation sequencing technology [23,24,37]. However, the widespread application of this method is currently limited due to the vast quantity of data it generates and its huge cost. Here, we report a simple, low-cost method to detect global DNA methylation, which combines the use of methylation-sensitive restriction enzymes together with second generation sequencing technology.

2. Methods

2.1. Cell culture and sample preparation

Tumor samples examined in the present study comprised colon cancer cell lines HT29 and HCT116 obtained from the ATCC. Cells were cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum. Genomic DNA was isolated using a QIAamp DNA mini kit (QIAGEN) according to the manufacturer's protocol.

2.2. Generation of methylation-specific digital sequencing (MSDS) libraries

Genomic DNA (1 μg) was digested sequentially with the methylation-sensitive enzymes, SacII, EagI and BssHII (Table 1). DNA was first digested by SacII (NEB) at 37 °C for 3 h, then by EagI (NEB) at 37 °C for 3 h, and finally by BssHII (NEB) at 50 °C for 3 h. Digested samples were then purified by phenol-chloroform extraction and the fragments were ligated to biotinylated linkers containing synthetic restriction sites using T4 DNA ligase (Invitrogen), as follows: (BssHII-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATG-3′, BssHII-B, 5′-pho-CGCGCATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′, EagI-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATC-3′, EagI-B, 5′-pho-GGCCGATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′, SacII-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATCCGC-3′, SacII-B, 5′-pho-GGATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′). The ligated DNA was then fragmented randomly using an ultrasonic device and 110–160 bp fragments were isolated by PAGE. Non-methylated ends were enriched using streptavidin beads. The ends of the DNA were repaired using the End-it DNA End repair kit (Epicentre) so that they were blunt-ended. Repaired fragments were then ligated to the following blunt-ended linkers using T4 DNA ligase: (blunt-linker-A, 5′-CTGCCCCGGGTTCCTCATTCTCT-3′, blunt-linker-B, 5′-AGAGAATGAGGAACCCGGGGCAGTT-3′). The ligated DNA was amplified by PCR using linker-specific primers (primer-A, 5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′, primer-B, 5′-CTGCCCCGGGTTCCTCATTCTCT-3′) and sequenced in 35 bp reads on the Life Technologies' SOLiD sequencer (Fig. 1A).

2.3. Data analysis

Sequencing data (30/35 bp per sequencing read) was analyzed using Corona Light software. Human genome sequence and mapping information (Mar. 2006, hg18) was downloaded from the University of California Santa Cruz Genome Bioinformatics Site and used to construct a virtual BssHII, EagI and SacII tag library on the basis of the genome sequence. Genes neighboring each restriction enzyme site were also identified in order to determine the effect of methylation on their expression.

Each enzyme site contributes two tags to the sequence library: a plus-strand library tag and a minus-strand library tag. The methylation level of an enzyme site can be inferred by either of the two sequence tag counts; we selected the tag that generated the greater number of counts. When the distance between two successive restriction enzyme sites is less than about 100 bp, the tags cannot be sequenced, because the tags are deleted by the PAGE selection. CpGs localized in close proximity to one another are often co-methylated [37], such that the level of methylation of one enzyme site can usually be inferred from the level of methylation of other sites localized within the same area, when the distance between sites is too small. In situations where the distance between these enzymes sites was <100 bp, the sites were considered together as a group, and we again selected the tag that generated the greatest number of counts within the group. If the tag sequence matched a site within a repeated sequence or at more than two sites within the whole human genome, the tag was excluded. Excluding such sequences, 86,897 sites were analyzed (Table 1).

2.4. Bisulfite sequencing

Bisulfite modification of genomic DNA was performed using the EpiTect Bisulfite Kit (Qiagen). Primers were designed using Methyl Primer Express software (Life Technologies). Bisulfite-treated DNA was amplified by PCR. The PCR products were cloned into the pCR2.1-TOPO Vector, then transformed into One Shot TOP10 Competent Cells (Invitrogen). At least 24 clones were sequenced using an ABI3730 Sequencer (Life Technologies). The data was analyzed using the Quantification Tool for Methylation Analysis (Riken Institute of Physical and Chemical Research).

2.5. Gene expression data

Gene expression data for the HT29 cells was generated as described previously [28], while for the HCT116 cells, this data was obtained using 5'SOLiD technology [28]. Analysis of gene expression data was based on the 5′ SOLiD method [28].

Table 1
The number of restriction sites for the methylation-sensitive enzymes BssHII, EagI and SacII, in silico data.

Enzyme	Recognition site		No. of total sites in human genome	No. of genes containing the recognition site within the promoter region	No. of CpG island containing the recognition site
BssHII	G^CGCGC	6 bp	72,899	8787	13,363
EagI	C^GGCCG	6 bp	90,190	9089	14,023
SacII	CCGC^GG	6 bp	66,312	9793	15,317
BssHII + EagI + SacII			229,401	13,978	21,164

Defined site data

	No. of defined sites within human genome	No. of genes containing the defined site within the promoter region	No. of CpG island containing the defined site	No. of CpG island shore containing the defined site
MSDS method	86,897	11,217	19,592	8522

Promoter region is defined as the region 500 bp either side of the TSS.

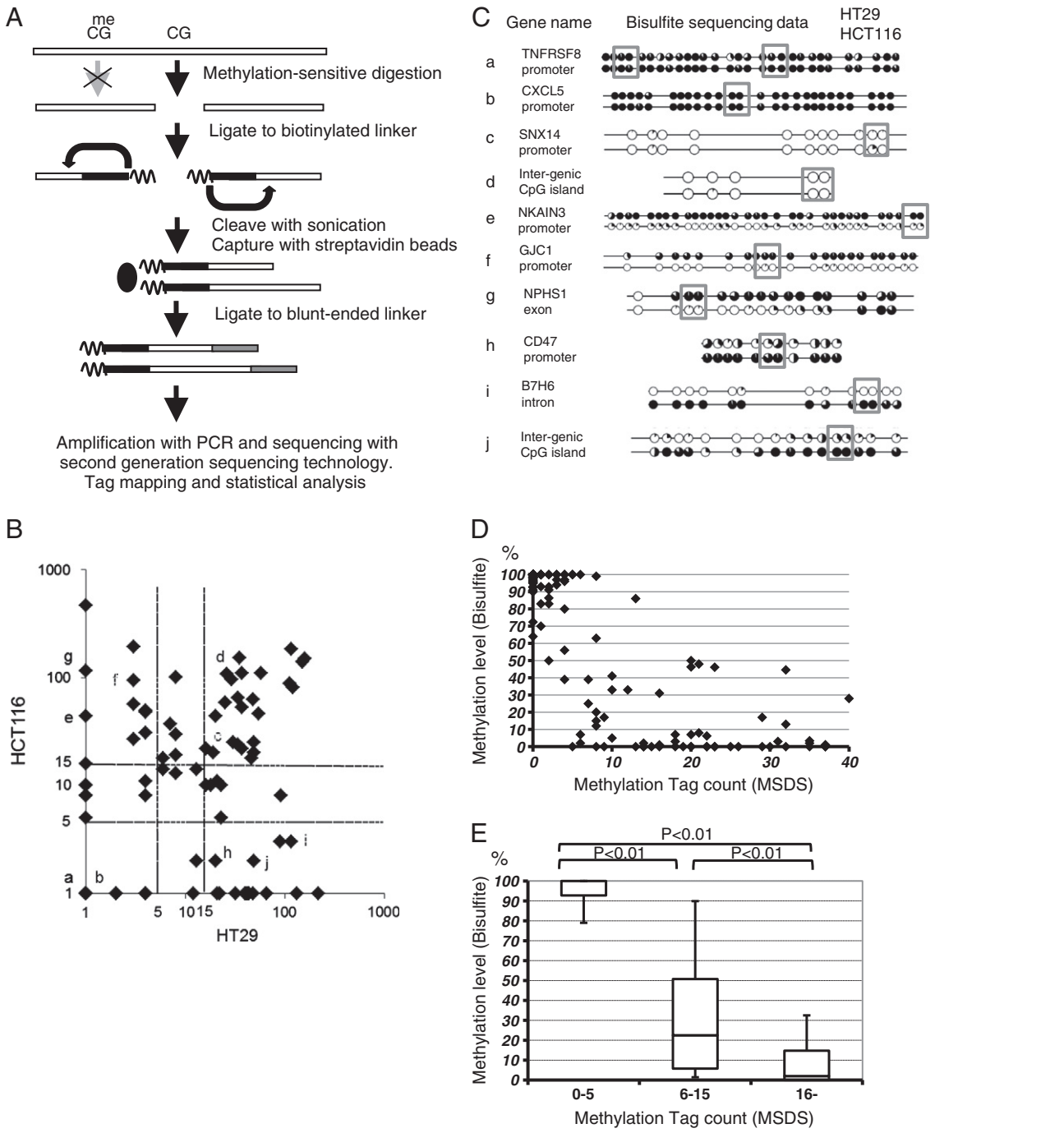

Fig. 1. Flow chart schematic of the methylation-specific digital sequencing (MSDS) method and the correlation between methylation level as assessed by bisulfite sequencing and the MSDS methylation tag count. A. Genomic DNA was digested with methylation-sensitive restriction enzymes and ligated to biotinylated linkers. The ligated sample was then cleaved by sonication and the fragments were captured with streptavidin beads. The captured DNA fragments were ligated to blunt-end linkers, amplified by PCR, and sequenced on the Life Technologies' SOLiD sequencer. B. Scatter diagram of MSDS tag counts corresponding to 100 regions selected at random that were also subjected to bisulfite sequencing. Sample labels a–d and e-j indicate sets of genes whose pattern of methylation was similar (a–d) or different (e–j) between HT29 and HCT116 cells. C. DNA methylation of genes a-j as analyzed by bisulfite sequencing. Gray squares represent CpG sites within restriction enzyme sites. Circles represent potential methylation sites (CpG) where the shading intensity indicates the frequency at which the site was found to be methylated amongst the clones analyzed (0–100%). a. TNFRSF8 (tumor necrosis factor receptor superfamily, member 8) gene promoter; b. CXCL5 (chemokine (C–X–C motif) ligand 5) gene promoter; c. SNX14 (sorting nexin 14) gene promoter; d. intergenic region within CpG island; e. NKAIN3 (Na^+/K^+ transporting ATPase interacting 3) gene; f. GJC1 (gap junction protein, gamma 1, 45kDa) gene promoter; g. NPHS1 (nephrosis 1, congenital, Finnish type (nephrin)) gene exon; h. CD47 (CD47 molecule) gene promoter; i. B7H6 (B7 homolog 6) gene intron; j. intergenic region within CpG island. D. and E. Correlation between methylation level as assessed by bisulfite sequencing and methylation tag count as assessed by MSDS. Horizontal bars represent methylation as determined by bisulfite sequencing. In E, boxes represent the quartiles and whiskers mark the 5th and 95th percentiles.

3. Results

3.1. MSDS analysis

We have developed a simple method, which we call MSDS, to detect global DNA methylation. The method makes use of three methylation-sensitive restriction enzymes (BssHII, EagI, and SacII) and second generation sequencing technology. There are two CpG dinucleotides within the recognition sequence of each enzyme (GCGCGC, CGGCCG, and CCGCGG, respectively). Short sequence tag fragments derived from the sites cleaved by these restriction enzymes can be mapped to unique sites within the genome. In silico

analyses indicate that there are 72,899, 90,190 and 66,312 recognition sites for BssHII, EagI and SacII within the human genome, respectively (Table 1). These restriction sites cover 64% (13,978) of unique gene promoters (within 500 bp of the TSS) and 75% (21,164) of the CpG islands within the genome. This combination of restriction enzymes allowed the most comprehensive coverage of CpG islands without using a greater number of enzymes. After excluding certain uninformative sites as described in Section 2.3, we analyzed a total of 86,897 sites. These defined sites cover 51% (11,217) of unique gene promoters (Supplemental Table 1), 69% (19,592) of CpG islands, and 8522 CpG island shores within the genome (Table 1).

Two colon cancer cell lines (HT29 and HCT116) were utilized to confirm the utility of this method. Using the SOLiD platform, 2,397,132 and 2,971,226 tags were mapped to the human genome in HT29 and HCT116 cells, respectively. After excluding uninformative sites as noted above, 1,493,950 tags (62.3% of total tags) in HT29 cells and 1,886,310 tags (63.5% of total tags) in HCT116 cells were matched to the remaining informative sites (Table 2, Supplementary Fig. 1). There were 35,309 (40.6%), and 48,141 (55.4%) zero tag count or un-hit sites in HT29 and HCT116 cells, respectively (Table 3).

3.2. Correlation between methylation level (bisulfite sequencing) and MSDS tag counts

To validate the accuracy of the MSDS method, MSDS tag counts were compared with the results of direct DNA methylation analysis, which were generated by bisulfite sequencing at more than 100 random CpG sites (Fig. 1B). In both cell lines, CpG sites that generated the highest number of tags corresponded to those that also exhibited a low level of methylation by bisulfite sequencing analysis, such as those within the SNX14 (sorting nexin 14) gene promoter. CpG sites that generated relatively fewer tags in both cell lines corresponded to those that were highly methylated in both cell lines according to the bisulfite sequence analysis, such as those within the TNFRSF8 (tumor necrosis factor receptor superfamily, member 8) gene promoter. Some sites appeared to be differentially methylated between the two cell lines. CpG sites that gave rise to many tags in HT29 cells alone were poorly methylated only in these cells, such as those within the promoter of the CD47 (CD47 molecule) gene. Similarly, CpG sites that gave rise to many tags in HCT116 cells alone were poorly methylated only in these cells, such as those within the promoter of the NKAIN3 (Na^+/K^+ transporting ATPase interacting 3) gene (Fig. 1C). Thus, the methylation levels determined by bisulfite sequencing correlated inversely with the MSDS tag count in both cell lines. Based on these results (Fig. 1D), we classified the genes into three arbitrary groups, based on the number of tag counts at each restriction enzyme site**.** Sites represented by ≥16 tags were considered to be poorly methylated (<20% methylated), whereas sites represented by ≤5 tags were considered to be highly methylated (>80% methylated), and sites represented by 6–15 tags were considered to be methylated at an intermediate level (21–79% methylated) (Fig. 1E).

Table 2
Summary of the results of sequence analysis for methylation tags, MSDS data.

Enzyme	No. of tags mapped to the human genome (% of total tags)		No. of defined tags	
	HT29	HCT116	HT29	HCT116
BssHII	648,117 (27.0%)	889,793 (29.9%)	415,643	563,851
EagI	1,077,327 (44.9%)	1,141,883 (38.4%)	660,181	740,139
SacII	671,688 (28.0%)	939,550 (31.6%)	418,126	582,320
BssHII + EagI + SacII	2,397,132	2,971,226	1,493,950	1,886,310

Table 3
Level of DNA methylation at CpG sites in HT29 and HCT116 cells.

Methylation level	No. of CpG sites			
	HT29	%	HCT116	%
Low	27,478	31.60%	25,240	29.00%
Middle	10,112	11.60%	5485	6.30%
High	49,307	56.70%	56,172	64.60%
0 tag count sites	35,309	40.63%	48,141	55.40%
Total CpG sites	86,897	100%	86,897	100%

Low level corresponds to 0–20% methylation, by bisulfite sequencing analysis, and corresponds to a tag count of greater than 16, whereas the high level corresponds to 80%~methylation and is correlated with tag counts between zero and five.

3.3. Genome-wide DNA methylation

The results of the genome-wide DNA methylation analysis are presented in Table 3. 56.7% (49,307) and 64.6% (56,172) of unique CpG sites were highly methylated, and 31.6% (27,478) and 29.0% (25,240) were poorly methylated in HT29 and HCT116 cells, respectively, indicating that a majority of CpG sites were either highly or poorly methylated, which is consistent with what has been reported previously [37,38]. As CpGs in close proximity are often co-methylated [37], we considered the tag with the greatest number of counts within such promoter regions as representative of the methylation level of those regions. Some gene promoters overlap CpG islands, whereas others do not. 48% of the promoters evaluated in both cell lines were poorly methylated (Fig. 2B-1); those that did not overlap CpG islands were methylated to a greater extent than those that contained a CpG island within the promoter region ($P<0.01$) (Figs. 2B-2, 3). Within gene body regions, 63% of enzyme sites were highly methylated in both cell lines (Figs. 2C-1) and the methylation level of CpG islands in such regions was lower than in other regions of the gene body ($P<0.01$) (Figs. 2C-2, 3). In promoter and gene body, the extent of methylation within CpG islands was lower than that in regions outside of the CpG islands ($P<0.01$). In addition, there was a difference in the extent of methylation between CpG islands within the promoter and CpG islands within the gene body. Methylation within CpG islands has previously been reported to be dependent upon the position of the island within the promoter region or within the gene body [36].

3.4. Comparison of MSDS methylation with gene expression

There is a strong relationship between gene expression and DNA methylation. Therefore, we analyzed gene expression in the HT29 and HCT116 cell lines and compared the results to the methylation data from the same cell lines. The genes were divided into three groups: high expression (≥100 copies), low expression (≤1 copy), and moderate expression (1<copies <100) [28]; and the level of expression of the genes within these groups was compared to the level of methylation. Highly and moderately expressed genes exhibited a low level of methylation within the promoter and a high level of methylation within the gene body. Genes with a low expression level exhibited a moderate level of methylation within both the gene promoter and the gene body. These results are consistent with those of previous reports [35,36] (Fig. 3).

3.5. Differentially methylated regions between HT29 and HCT116

Next, we compared the extent of methylation of the HT29 and HCT116 cells. Of the sites measured, 77% were methylated to a similar extent in the two cell lines: 52% of the sites measured were poorly methylated, and 23% of the sites measured were highly methylated (Figs. 2A-1). DMRs between the two cell lines were defined as enzyme sites that were highly methylated in one cell line and poorly methylated in the other cell line. Such sites made up 9% of the sites

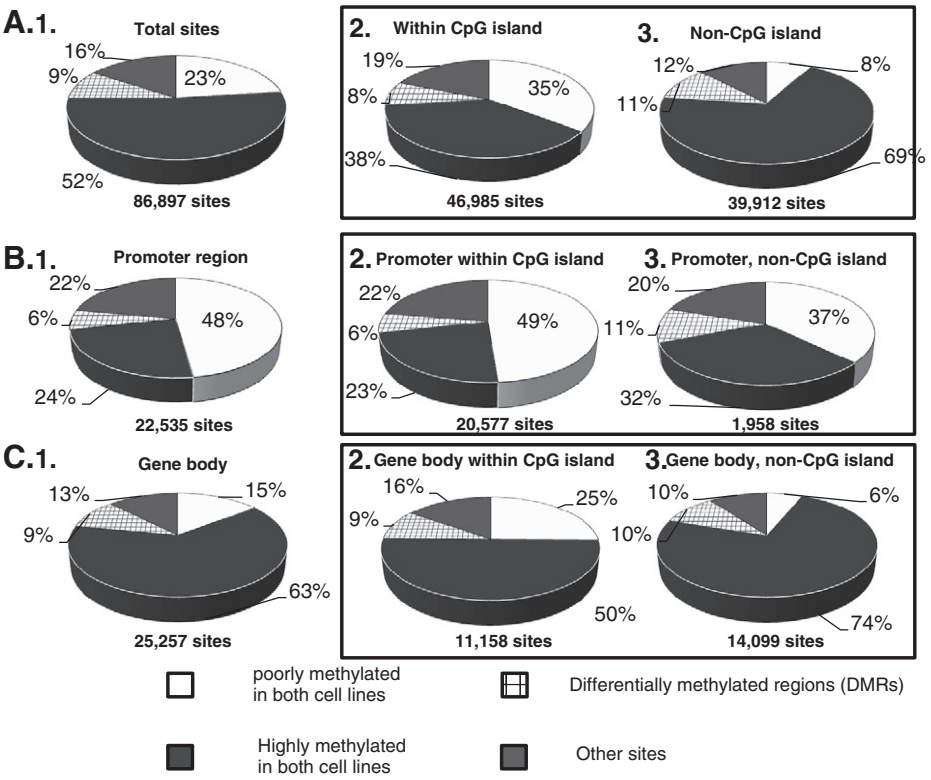

Fig. 2. DNA methylation in the vicinity of the TSSs and gene bodies. The number under each panel represents the total number of CpG sites for each pie chart. A. Pie charts represent the distribution of restriction enzyme sites, which are divided into four categories: poorly methylated in both cell lines, highly methylated in both cell lines, DMRs and other sites. B. The chart on the left represents methylation within 500 bp of the TSS. The charts on the right represent methylation within or outside of the CpG islands that are within 500 bp of the TSS. C. The chart on the left represents methylation within the gene body. The charts on the right represent methylation within or outside of CpG islands in the gene body.

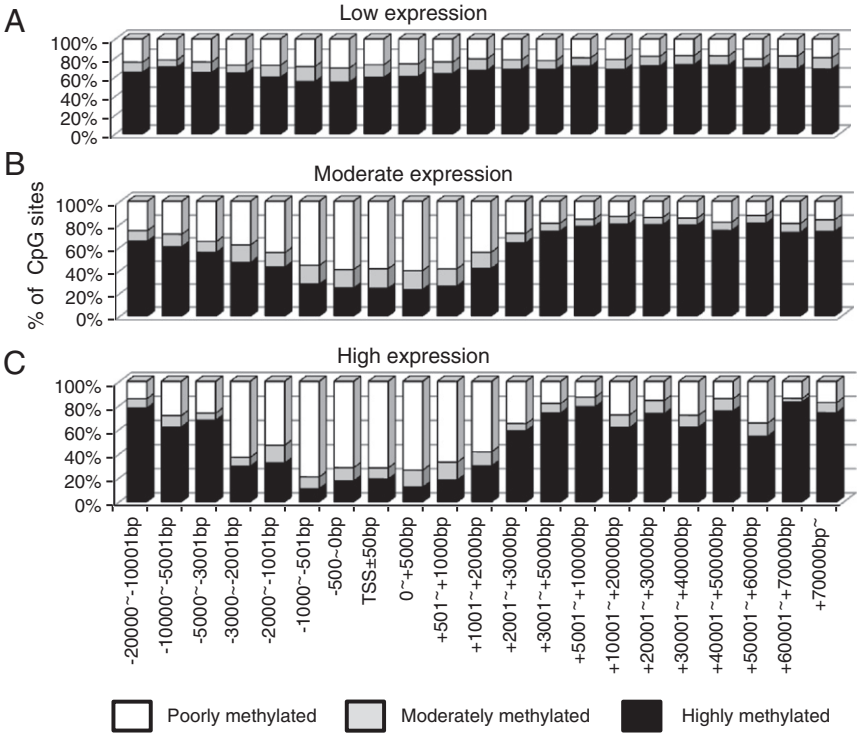

Fig. 3. Correlation of the extent of methylation assessed by MSDS relative to the distance from the TSS with the level of gene expression assessed by 5′-end SOLiD transcriptome analysis. Genes were categorized into three groups based on the 5′-end SOLiD transcriptome data: low expression, ≤1copy; high expression, ≥100 copy; moderate expression, 1<copies<100.

measured. Two percent of the sites measured were highly methylated in HT29 cells and poorly methylated in HCT116 cells, while 7% of the sites measured were highly methylated in HCT116 cells and poorly methylated in HT29 cells. As shown in Table 4, many DMRs were observed not in the gene promoter region, but instead in the gene body, intergenic or CpG island shore regions.

Finally, we investigated the correlation between the extent of methylation in the vicinity of the TSS and the expression of the associated gene. We evaluated the expression of several genes localized to DMRs within 1000 bp of the TSS and whose expression level differed at least five-fold between HT29 and HCT116 cells. Of these genes, we analyzed those linked to carcinogenesis. These included SMPD3 (sphingomyelin phosphodiesterase 3, neutral membrane (neutral sphingomyelinase II)), VASH1 (vasohibin 1), and TACC1 (transforming acidic coiled-coil containing protein 1), which are involved in the cell cycle; RIPK3 (receptor-interacting serine-threonine kinase 3), and TNFSF9 (tumor necrosis factor (ligand) superfamily, member 9), which are involved in apoptosis; and CLDN1 (claudin 1), NELL1 (NEL-like 1 (chicken)), ROBO1 (roundabout, axon guidance receptor, homolog 1 (Drosophila)), ENG (endoglin), PLEKHC1 (pleckstrin homology domain containing, family C) (FERMT2 (fermitin family member 2)), which are involved in cell adhesion. In these genes, the DMRs localize to several positions within the gene, in some cases only upstream of the TSS (e.g. VASH1 and TACC1); in some cases only downstream of the TSS (e.g. SMPD3, RIPL3, TNFSF9, CLDN1, NELL1 and ENG); and in some cases, both upstream and downstream of the TSS (e.g. ROBO1 and PLEKHC1 (FERMT2)) (Fig. 4). In TNFSF9, NELL1 and ENG genes, CpG sites closer to the TSS than the DMR were methylated to a similar level in both cell lines (Fig. 4). This result demonstrates that these DMRs are not localized solely in the vicinity of the TSS.

4. Discussion

Diverse methods for the measurement of DNA methylation have been developed [9,11,16–24]. Moreover, second generation sequencing technology has facilitated genome-wide analyses of DNA methylation. However, this technology generates too much data and is too expensive to be accessible to most researchers in the field. Therefore, we have developed a method for detecting global DNA methylation that makes use of methylation-sensitive restriction enzymes combined with second generation sequencing technology. The methylation level can be determined by monitoring the number of times a particular recognition site is sequenced within the genomic pool. This method optimizes data handling by limiting the data generated to only the restriction enzyme sites.

Table 4
Differentially methylated regions between the HT29 and HCT116 cell lines.

Gene position	No. of sites in each gene position	No. of differentially methylated sites	*P* value
Tags mapped to genome	86,897	7903	
CpG island	46,985	3682	1.67E−44
CpG island shore[a]	9895	1219	2.12E−32
Other region	30,017	3002	1.47E−11
Gene body			
−5000~2001 bp	2216	240	0.00399
−2000~1001 bp	1642	187	0.001
−1000~501 bp	2503	264	0.01
−500~1 bp	9178	605	1.16E−18
TSS ± 50 bp	3562	187	3.65E−16
1st exon	9168	651	2.21E−12
Intron	23,164	2111	0.908
Exon	4191	314	0.0002172
Last exon	2090	180	0.437
Intergenic	32,730	3348	1.55E−19

[a] CpG island shore; 2000 bp within CpG island.

Changes in DNA methylation profiles have been observed during normal development and tumorigenesis. However, differences in the genome-wide methylation pattern between related cancer cell lines derived from related tumors have not been examined previously. We found many sites, localized predominantly to CpG island shore and intergenic regions, which are differentially methylated in HT29 and HCT116 cells. Differences between the HT29 and HCT116 cell lines may have arisen in the process of transforming from normal tissue to cancer tissue, or may be due to differences in the cells before they became cancerous. Regions of the genome that are differentially methylated may provide insight into the variety of regions of the genome where differential gene expression is linked to the development of cancer. By searching regions of the genome that are differentially methylated between cell lines from similar tissue, it may be possible to reveal differences in the initiation of cancer and/or in the types of cells that cause the malignant transformation: from a basal cell or layer cell, for example.

The following genes have been reported to exhibit tumor suppressor activity and to be regulated by methylation within their promoter regions: SMPD3 [39], VASH1 [40], RIPK3 [41], TNFSF9 [42], NELL1 [43,44], ROBO1 [45–47], ENG [43,48]. In this study, we identified additional genes characterized by a link between DNA methylation and expression: TACC1, CLDN1, and PLEKHC1 (FERMT2). These genes may also be important in carcinogenesis.

Analysis of DNA methylation using microarrays suffers from a bias towards gene promoter regions and CpG islands. The microarray method can only provide relative data, and therefore requires a reference standard. Methods based on restriction enzymes possess inferior sensitivity to bisulfite sequencing, but do allow a quantitative comparison of methylation patterns. Bisulfite sequencing allows analysis of DNA methylation at the single nucleotide level, but at a very high cost for a genome the size of the human genome [37].

Established methods that make use of methylation-sensitive restriction enzymes include methylation-specific digital karyotyping (MSDK) [15], modified methylation-specific digital karyotyping (MMSDK) [25], methyl-sensitive cut counting (MSCC) [35] and others. The MMSDK method is similar to the MSDS method that we have developed. Both methods are based on digital karyotyping technology. The process of digestion of genomic DNA with a methylation-sensitive restriction enzyme, and the use of a fragmenting enzyme with which to generate short sequence tags is essential to both methods. The key difference between MSDS and MMSDK is the location of the short sequence read using second generation sequencing technology. The MMSDK method targets the nearest NlaIII sites to the methylation-sensitive enzymes, while the MSDS method targets the methylation-sensitive enzyme sites themselves. This difference simplifies the construct preparation process. It results in a shortening of the duration of the procedure and reduces loss of the sample. Furthermore, it requires less PCR cycles, and therefore minimizes PCR bias. During data analysis, the directly identifiable measurement positions make it easy to perform direct sequencing of the cleavage site. However, because direct sequencing is employed, when a restriction enzyme site exists within a repeated sequence, these tags are considered unreliable and excluded. Accordingly, MSDS method is not able to measure repeated sequences. The difference between MSCC and MSDS is the combination of methylation-sensitive enzymes used. MSCC uses the HpaII enzyme, while MSDS uses three enzymes (BssHII, EagI and SacII). Methods that make use of methylation-sensitive restriction enzymes are limited to profiling the methylation of these enzyme's recognition sites, suggesting that use of enzymes that cleave at a greater number of sites is advantageous. The HpaII and MspI enzymes are four base encoding enzymes, and the recognition site of each enzyme contains one CpG dinucleotide within an identical recognition sequence, CCGG. HpaII cleaves only at unmethylated sites, whereas MspI cleaves both methylated and unmethylated sites. In silico analyses reveal that there are approximately two million recognition sites for these enzymes

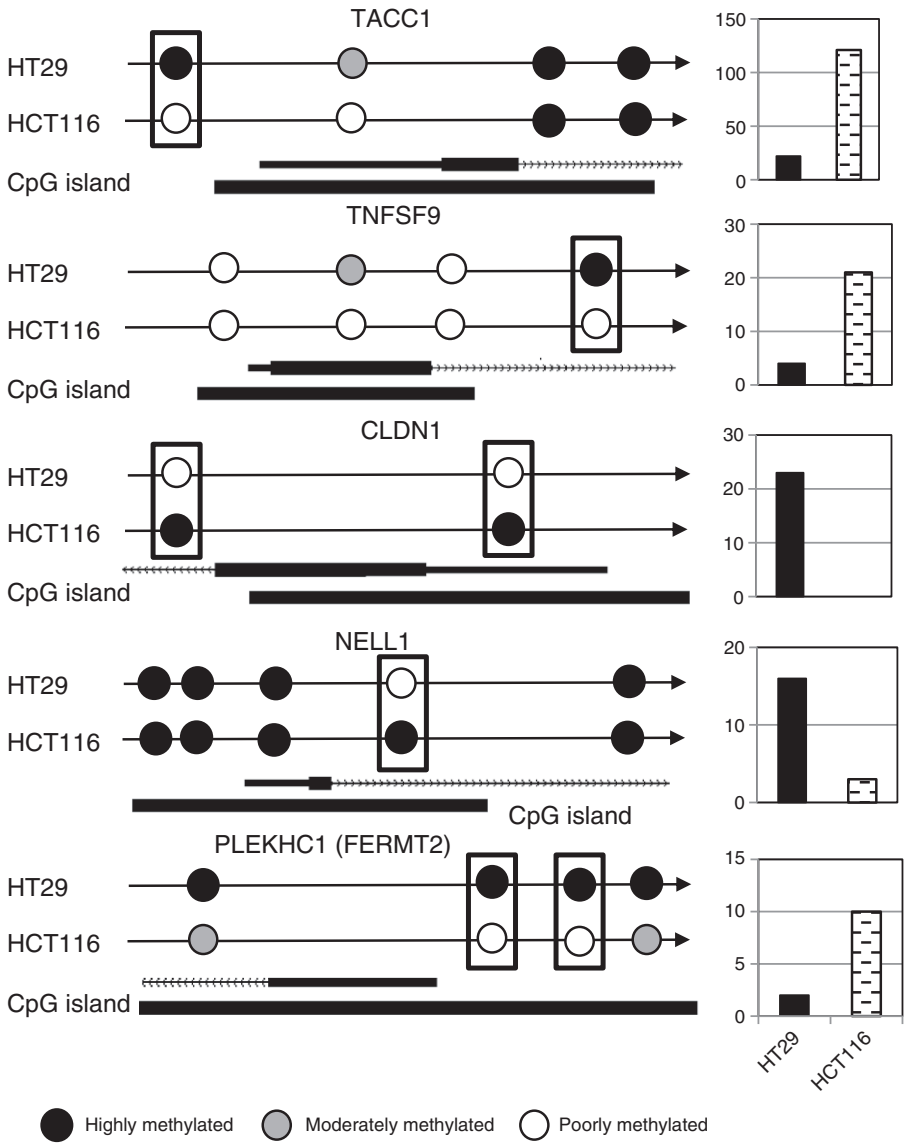

Fig. 4. Correlation between the extent of methylation within DMRs and the level of expression of the associated gene. We selected five genes containing DMRs in close proximity to the TSSs: TACC1, TNFSF9, CLDN1, NELL1, and PLEKHC1 (FERMT2). The panel to the left shows relative MSDS tag counts (circles) at the relevant sites within each gene in each cell line. A black box around the two circles represents the position of the DMRs that were selected. The panel on the right shows the relative expression of each gene in each cell line.

within the human genome. The MSDS method reported here makes data analysis simpler because it results in analysis of fewer sites and therefore fewer tags.

Within the data obtained via the MSDS methods, zero tag count or un-hit sites made up about 40–50% of the uninformative sites. These sites may be fully methylated or just false negative. The use of enzymes that recognize six-base target sequences obviates the need for a comparison target, as is the case when evaluating MspI against HpaII. To verify this, we performed bisulfite sequencing at 44 CpG sites that gave rise to a MSDS tag count of zero. We found that 41 of the 44 sites were highly methylated. This finding suggests that the bias due to PCR or cleavage efficiency of the restriction enzymes was small, and that the number of sequence tag counts was enough to cover the total restriction enzyme sites. At the same time, this false negative rate is considered low. Under the MSDS method, we determined the cut-off values for poorly, moderately and highly methylated sites, based on the results of bisulfite sequencing. The depth of the MSDS sequence affects the cut-off values, and deeper sequencing coverage should improve accuracy, as is the case with the MSCC method [35]. Quantitative information was based on the comparison between bisulfite sequencing data and MSDS tag counts in Fig. 1E. The reliability of these data might be affected by imperfect cleavage by BssHII, EagI and SacII enzymes or a bias derived from the PCR amplification used in MSDS. However, the cutting efficiency of these enzymes was verified, and only a limited number of PCR cycles were used to make the construct.

The MSDS method was able to detect the difference between highly methylated and poorly methylated sites. In the process of the carcinogenesis, DNA methylation is changed in many ways. MSDS was unable to detect minor changes (10–20%) in DNA methylation. However, the MSDS method was able to detect large changes in DNA methylation (e.g. a change from 0–20% methylation to 80–100% methylation).

The MSDS method is a robust means of detecting global DNA methylation. This method has the potential to provide novel insight into the functional complexity of the human genome, and may also serve as a basis for the diagnosis of diseases such as cancer.

5. Accession number

MSDS tags have been deposited in the NCBI Short Read Archive under the project accession SRA028119. Expression tags for HT29 and

HCT116 cell lines have been deposited in NCBI Short Read Archive under the project accessions SRA002659 and SRA028119, respectively.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.07.003.

Acknowledgments

This work was partly supported by a Grant-in Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by Core Research for Evolutional Science and Technology (CREST) funding from the Japan Science and Technology Agency (JST). We thank M. Sarashi, H. Hanaoka, O. Takahashi, K. Azuma, W. Suda, E. Iioka, K. Furuya, M. Ogawa, S. Tomura, F. Shand, Y. Okajima, K. Harukawa, and S. Hatanaka for technical assistance.

References

[1] T.H. Bestor, The DNA methyltransferases of mammals, Hum. Mol. Genet. 9 (2000) 2395–2402.
[2] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.
[3] W. Reik, Stability and flexibility of epigenetic gene regulation in mammalian development, Nature 447 (2007) 425–432.
[4] P.W. Laird, The power and the promise of DNA methylation markers, Nat. Rev. Cancer 3 (2003) 253–266.
[5] A.P. Feinberg, The epigenetics of cancer etiology, Semin. Cancer Biol. 14 (2004) 427–432.
[6] S.E. Ozanne, M. Constancia, Mechanisms of disease: the developmental origins of disease and the role of the epigenotype, Nat. Clin. Pract. Endocrinol. Metab. 3 (2007) 539–546.
[7] P.A. Jones, S.B. Baylin, The epigenomics of cancer, Cell 128 (2007) 683–692.
[8] M. Esteller, Epigenetics in cancer, N. Engl. J. Med. 358 (2008) 1148–1159.
[9] I. Keshet, et al., Evidence for an instructive mechanism of de novo methylation in cancer cells, Nat. Genet. 38 (2006) 149–153.
[10] S.H. Cross, J.A. Charlton, X. Nan, A.P. Bird, Purification of CpG islands using a methylated DNA binding column, Nat. Genet. 6 (1994) 236–244.
[11] T.A. Rauch, X. Wu, X. Zhong, A.D. Riggs, G.P. Pfeifer, A human B cell methylome at 100-base pair resolution, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 671–678.
[12] M. Frommer, et al., A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 1827–1831.
[13] S.J. Clark, J. Harrison, C.L. Paul, M. Frommer, High sensitivity mapping of methylated cytosines, Nucleic Acids Res. 22 (1994) 2990–2997.
[14] A.P. Bird, E.M. Southern, Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis, J. Mol. Biol. 118 (1978) 27–47.
[15] M. Hu, et al., Distinct epigenetic changes in the stromal cells of breast cancers, Nat. Genet. 37 (2005) 899–905.
[16] A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome, Science 315 (2007) 1141–1143.
[17] B. Khulan, et al., Comparative isoschizomer profiling of cytosine methylation: the HELP assay, Genome Res. 16 (2006) 1046–1055.
[18] X. Zhang, et al., Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis, Cell 126 (2006) 1189–1201.
[19] D. Zilberman, M. Gehring, R.K. Tran, T. Ballinger, S. Henikoff, Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription, Nat. Genet. 39 (2007) 61–69.
[20] M. Bibikova, et al., High-throughput DNA methylation profiling using universal bead arrays, Genome Res. 16 (2006) 383–393.
[21] R.A. Irizarry, et al., The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores, Nat. Genet. 41 (2009) 178–186.
[22] A. Meissner, et al., Genome-scale DNA methylation maps of pluripotent and differentiated cells, Nature 454 (2008) 766–770.
[23] R. Lister, et al., Highly integrated single-base resolution maps of the epigenome in Arabidopsis, Cell 133 (2008) 523–536.
[24] S.J. Cokus, et al., Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning, Nature 452 (2008) 215–219.
[25] J. Li, et al., An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines, BMC Genomics 10 (2009) 223.
[26] A. Barski, et al., High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.
[27] L.W. Hillier, et al., Whole-genome sequencing and variant discovery in *C. elegans*, Nat. Methods 5 (2008) 183–188.
[28] S. Hashimoto, et al., High-resolution analysis of the 5′-end transcriptome using a next generation DNA sequencer, PLoS One 4 (2009) e4108.
[29] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. 18 (2008) 1509–1517.
[30] M. Hafner, et al., Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing, Methods 44 (2008) 3–12.
[31] K.E. Holt, et al., High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi, Nat. Genet. 40 (2008) 987–993.
[32] J.D. McGhee, et al., ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult, Dev. Biol. 327 (2009) 551–565.
[33] S. Yagi, et al., DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression, Genome Res. 18 (2008) 1969–1978.
[34] Y.J. Shann, et al., Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines, Genome Res. 18 (2008) 791–801.
[35] M.P. Ball, et al., Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, Nat. Biotechnol. 27 (2009) 361–368.
[36] A.K. Maunakea, et al., Conserved role of intragenic DNA methylation in regulating alternative promoters, Nature 466 (2010) 253–257.
[37] R. Lister, et al., Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322.
[38] F. Eckhardt, et al., DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 38 (2006) 1378–1385.
[39] B. Demircan, et al., Comparative epigenomics of human and mouse mammary tumors, Genes Chromosomes Cancer 48 (2009) 83–97.
[40] C. Lu, et al., Regulation of tumor angiogenesis by EZH2, Cancer Cell 18 (2010) 185–197.
[41] M. Fukasawa, et al., Microarray analysis of promoter methylation in lung cancers, J. Hum. Genet. 51 (2006) 368–374.
[42] S. Yamashita, Y. Tsujino, K. Moriguchi, M. Tatematsu, T. Ushijima, Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2′-deoxycytidine treatment and oligonucleotide microarray, Cancer Sci. 97 (2006) 64–71.
[43] Y. Mori, et al., A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer, Gastroenterology 131 (2006) 797–808.
[44] Z. Jin, et al., Hypermethylation of the nel-like 1 gene is a common and early event and is associated with poor prognosis in early-stage esophageal adenocarcinoma, Oncogene 26 (2007) 6332–6340.
[45] A. Dallol, et al., Tumour specific promoter region methylation of the human homologue of the Drosophila Roundabout gene DUTT1 (ROBO1) in human cancers, Oncogene 21 (2002) 3020–3028.
[46] J. Xian, et al., Targeted disruption of the 3p12 gene, Dutt1/Robo1, predisposes mice to lung adenocarcinomas and lymphomas with methylation of the gene promoter, Cancer Res. 64 (2004) 6432–6437.
[47] S. Ghosh, et al., Alterations of ROBO1/DUTT1 and ROBO2 loci in early dysplastic lesions of head and neck: clinical and prognostic implications, Hum. Genet. 125 (2009) 189–198.
[48] V.C. Wong, et al., Identification of an invasion and tumor-suppressing gene, Endoglin (ENG), silenced by both epigenetic inactivation and allelic loss in esophageal squamous cell carcinoma, Int. J. Cancer 123 (2008) 2816–2823.