Genomics 96 (2010) 67–72

Contents lists available at ScienceDirect

Genomics

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

Presence of 5-methylcytosine in CpNpG trinucleotides in the human genome

Juna Lee [a,b], Se Jin Jang [c,*], Nicole Benoit [b], Mohammad O. Hoque [b], Joseph A. Califano [b], Barry Trink [b], David Sidransky [b,d], Li Mao [e], Chulso Moon [a,b,d,f,*]

[a] Graduate Program in Human Genetics, Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[b] The Head and Neck Cancer Research Division, Department of Otolaryngology—Head and Neck Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[c] Department of Pathology, Asan Medical Center, Seoul, Republic of Korea
[d] Sidney Kimmel Comprehensive Cancer Center, Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[e] Molecular Biology Laboratory, Department of Thoracic/Head and Neck Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA
[f] Cleo Craig Cancer Research Program, Lawton OK, 73505, USA

ARTICLE INFO

Article history:
Received 18 May 2009
Accepted 23 March 2010
Available online 9 April 2010

Keywords:
Methylation
CpNpG
DNA methyltransferase
DNMT1
DNMT3a

ABSTRACT

While the methylation machinery of mammalian cells has been shown to be capable of both maintenance and *de novo* methylation at CpNpG sites, CpNpG methylation in the human genome has not been demonstrated. Here, we report the first observation of 5-methylcytosines in CpNpG triplets in the human genome. We identify the existence of CpNpG methylation in a number of genes which contain trinucleotide repeat regions, including the androgen receptor (*AR*). We further analyzed DNA extracted from primary tissue samples and found the same pattern of CpNpG methylation. To confirm our results, we performed Southern blot analysis by analyzing the cleavage sites of restriction enzymes within exon 1 of the *AR* gene and found direct evidence of the presence of 5mCs in CpNpG triplets in the human genome. Our results also suggest that this methylation pattern may be due to the human DNA methyltransferases DNMT1 and DNMT3A. Although the functional significance needs to be tested further, the discovery of inheritable CpNpG methylation in the human genome may have important implications in our understanding of gene regulation and of the development of various diseases, including cancer.

Introduction

Methylation of cytosine residues in prokaryotic and eukaryotic genomes is an important epigenetic mechanism in gene regulation [1–4]. It is generally believed that cytosine methylation occurs exclusively in CpG dinucleotides in mammalian genomes [4–6]. In contrast, in plant genomes, 5-methylcytosines (5mC) in CpNpG triplets is as frequent as in CpG dinucleotides, where N can be any base and the methylcytosine occurs at a symmetrical sequence [7]. The existence of CpNpG methylation in mammalian genomes *in vivo* remains contentious. While data derived from nearest neighbor analysis suggests that a considerable proportion of 5mC was not located exclusively in CpG dinucleotides, the data from genomic sequencing of mammalian DNA has not uncovered abundant methylation at non-CpG sites [8–13]. These consistencies have been partially answered by Clark et al. who demonstrated that the methylation machinery of mammalian cells has been found to be capable of both maintenance and *de novo* methylation at CpNpG sites [14]. However, they failed to demonstrated CpNpG methylation in mammalian genomes although such methylation was identified within a stably integrated segment of plasmid DNA. In this report, using bisulfite sequencing and Southern blot analysis, we provide direct evidence demonstrating the presence of 5mCs in CpNpG triplets in the human genome.

Results

Presence of CpNpG methylation in cell lines

To determine if CpNpG methylation is present in the human genome, we screened 29 genes; 22 were encoded in the nuclear genome while 7 were encoded in the mitochondrial genome (Table 1). Fifteen of these genes contained a CpNpG trinucleotide repeat region. Using the bisulfite sequencing approach, we analyzed the methylation status of these genes in a number of cancer cells lines of different origins. The large array of cell lines was used to determine if CpNpG methylation was tissue-specific, a phenomenon that is seen with some CpG methylation. Genomic DNA from the 35 cancer cell lines were treated with sodium bisulfite which converted unmethylated cytosine residues to uracils while methylated cytosine residues resisted such treatment. Sodium bisulfite-modified DNA was amplified using PCR primers designed according to sequences not affected by methylation. Of these 29 genes, 6 demonstrated extensive CpNpG

* Corresponding authors. C. Moon is to be contacted at Department of Otolaryngology, The Johns Hopkins University SOM, Cancer Research Building II, 5M3, 1550 Orleans Street, Baltimore, MD 21205, USA. Fax: +1 410 614 1411.
E-mail addresses: jangsejin@amc.seoul.kr (S.J. Jang), cmoon5@jhmi.edu (C. Moon).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.03.013

Table 1
Summary of genes examined, primer sequences for bisulfite sequencing, and CpNpG methylation status. MT: mitochondrial genome. M: CpNpG methylation present. U: no CpNpG methylation present.

Gene	Locus	Forward primer (5′→3′)	Reverse primer (5′→3′)	Methylation status
AIB1 (NCOA3)	20q12	AGGGTTTTTTTAATGTTTAAATGGT	ATCCTGCCAAAACCCATCCATACT	M
AR	X	AAGATTTATTGAGGAGTTTTTTAGAAT	AATAACCTATAAAACCTCTACAATAAA	M
ASH1 (ASCL1)	12q22–q23	GAAAGTTTTGTTAAGATGGAGAG	CAAAACCCAAATTAACCAAC	M
BHLHB5 (BHLHE22)	8q13	GAGGGGTGTTTTAATGTTTATT	CCTTTTACTCTTTAAATTTCTTACT	U
BRCA1	17q21	AGTTTATGGGAAGTAGTTATGTATTTTAGG	CTAAAATAATATTCCTAAAATACCTTTACC	U
DMPK	19q13.2–q13.3	TTGTTTTGAGTGTGGGTTTT	AACCCAACTTACTACCTTCC	U
DRPLA (ATN1)	12p13.31	GTGTGGAGTTAGGGTTTTT	AAACCATTAAAACCTACTTAACTATAA	U
E2F4	16q22.1	GGGGGTTATTATTGTAGTGA	CCTTAATAAACTCAAAAAAA	U
GAGE7B (GAGE12I)	X	GTGGGATTTTTTTATTTTTAAT	ACTACCTCACAACTCCCTAAC	U
H19	11p15.5	TGTTTTTTGAGGGGAGATAGTGGTTTGGGA	AATATCCCTCATAAACTC	U
HD (HTT)	4p16.3	GATTTTGGAAAAGTTGATGAAGGT	AATATCCCTCATAAACTC	U
NOTCH4	6p21.3	GGATTGGGGTTTGGAGAA	TAAAAATCTCCTCCATCCA	U
POLG	15q25	AAAAGAAGTTAAGTTGGAGTTT	ACATCTCCCCTCCTTACC	U
RAI1	17p11.2	TTATTTTTATAAGAGTTGTA	TAAAAAATTTCCTAAACATA	U
SCA1 (ATXN1)	6p23	AGTTTATTGGGTTTTTTTAA	ACCCCCAAATAATAAACCCC	M
SCA2 (ATXN2)	12q24	TGTATGGGTTTTTTATTATG	CCTCATATTTACATAAATTCCATCAA	U
SCA3 (ATXN3)	14q24.3–q31	TGATAGGTTATTTTGTGATGAAA	CTAAATCACTCCCAAATACTCC	M
SCA7 (ATXN7)	3p21.1–p12	GGTTGTGGATGATGTTAGGG	TAAAAACCTCAACCCACAAA	U
SCA8 (C10orf2)	13q21	AGTATGAGGAAGTATGGAAAAA	TCCCAATTCCTTAACTAAACC	M
SCA12 (PPP2R2B)	5q31–q33	AGGGGAGGAGGTTGGAAGAG	ACTCACCCTCACACCCACAC	U
SCA17 (TBP)	6q27	GTTTATTTTGGAAGAGTAATAAAGG	AATACCTAACCTAAAATTCCCTA	U
ATP6 (MT-ATP6)	MT	GGGTGTAGTGATTATAGGTT	AACTCTTCCAATTAAATACA	U
COX1 (MT-CO1)	MT	GGGAATTATTTTTATTTTGGAG	AAAAACCAATTAATATCATAACTCA	U
COX2 (MT-CO2)	MT	TTGGTTATTAATGGTATTGAATTT	AACATAAAACTATAATTTACTCCACA	U
COX3 (MT-CO3)	MT	GATTTTTGGTTTAGTTATGTGA	AAACCACATCTACAAAATACCA	U
CYTB (MT-CYB)	MT	TTATTTTGAGGGGTTATAGTAAT	CAAAAACTCCTCCTAATTTAT	U
ND1 (MT-ND1)	MT	AATTGTAATGGTATTTTTAATGTTT	TCTAATATTAAAACCTAAAACTAATTC	U
ND5 (MT-ND5)	MT	ATTTATGTGTTTAGATTAAGAAGTT	TTTTAAATAATCCTCCTATTTTTC	U

methylation (Table 1, Fig. 1a). As seen in the sequencing results for the androgen receptor (*AR*) gene, one of the two alleles from each cell line was unmethylated in most of the CpG sites and in all CpNpG sites in the genomic region (Figs. 1a and b). The other allele showed extensive CpG methylation consistent with the nature of an inactivated X chromosome. In addition, we observed several 5mCs in the CpNpG sites in the latter allele, including three consecutive CTG sites and two CAG sites (Fig. 1a). Interestingly, we observed this methylation phenomenon only in genes with trinucleotide repeat regions (Table 1, Figs. 1a and b). Furthermore, we found that this pattern of methylation was associated with the trinucleotide repeat region; sequencing of the other regions did not demonstrate 5mCs in CpNpG sites (data not shown). It should be noted that the sequencing did not demonstrate tissue specificity; while CpNpG methylation was not observed in every cell line examined, it was seen in cell lines of differing origins. For example, methylation of the trinucleotide region of the *AR* gene was observed in both lung cancer cell lines and prostate cancer cell lines (data not shown).

Presence of CpNpG methylation in tissues

Because it is believed that DNA methylation patterns can be significantly altered in cell culture conditions and in particular because DNA methylation density can be increased, we analyzed constitutional DNA extracted from 30 head and neck tumor samples, 30 lung tumor samples, and 8 normal lung tissue samples. Similar to our examination of CpNpG methylation in cancer cell lines, a large number of primary tissue samples were sequenced to determine if there was a difference in CpNpG methylation between normal and tumor samples. We observed CpNpG methylation in tissue samples as well (Table 1, Fig. 1b). However, the sequencing did not demonstrate a difference in CpNpG methylation between normal tissue samples and tumor samples. The result indicates that the 5mCs in CpNpG sites are found not only in culture conditions but are also found naturally in the human genome. These data, therefore, provide direct evidence of the presence of 5mCs in CpNpG triplets in the human genome.

Confirmation of CpNpG methylation by Southern blot analysis

To confirm our results, we performed Southern blot analysis. As shown with the *AR* in Fig. 2a, a roughly 3.8-kilobase (kb) fragment including the relevant methylation region could be generated by AccI (GT↓MKAC) digestion. Four PstI sites (CTGCA↓G) were identified within the 3.8-kb fragment (located at 34 bp, 856 bp, 1125 bp, and 1580 bp distant to the proximal AccI site) (Fig. 2a). Because the restriction enzyme PstI is methylation-sensitive (it cannot cleave the site if the 5′ cytosine is a 5mC), we predicted that the PstI site located at 856 bp distant to the proximal AccI site could not be cleaved due to CTG methylation. Genomic DNA samples were digested by AccI alone or AccI plus PstI, followed by Southern blot analysis using a 572-bp DNA probe located between the first two PstI sites (Fig. 2a). When DNA was digested by AccI alone, only a single 3.8-lb band could be detected in all four specimens as predicted (Fig. 2b). When DNA was digested by AccI and PstI, two bands, 0.8-kb and 1.1-kb, were observed in each of the specimens (Fig. 2b). This pattern was consistent with our prediction that the second PstI site could not be cleaved by PstI in one of the two alleles because of the presence of a 5′ 5mC (Fig. 1b). The lack of any band at the 3.8-kb and 1.5-kb positions indicated that the most distal PstI site was completely unmethylated in both alleles, and that enzymatic cleavage was complete.

CpNpG methylation may be mediated by DNMT1 and DNMT3a

There are four members of the human DNA methyltransferase (DNMT) family: DNMT1, DNMT2, DNTM3A, and DNMT3B. While the methylation machinery of mammalian cells has been shown to be capable of both maintenance and *de novo* methylation of CpNpG sites, it has not been conclusively determined which DNA methyltransferase is responsible [14]. To determine which DNMT is associated with