Identification of novel DNA methylation markers in colorectal cancer using MIRA-based microarrays

HAI LI[1], YONG DU[2], DONG ZHANG[1], LI-NA WANG[1], CHUN YANG[1], BIN LIU[2], WEI-JIE WANG[1], LEI SHI[1], WEI-GUO HONG[3], LIANG ZHANG[3] and YIN-XUE YANG[1]

[1]Department of Colorectal Surgery and [2]Surgical Research Laboratory, General Hospital of Ningxia Medical University, Yinchuan, Ningxia 750004; [3]BioChain (Beijing) Science and Technology Inc., Beijing 100176, P.R. China

Received January 31, 2012; Accepted March 12, 2012

DOI: 10.3892/or.2012.1779

Abstract. To identify novel hypermethylated genes in colorectal cancer (CRC) and to test their potential application in CRC early diagnosis, a genome-wide screening of 57,723 CpG dinucleotides covering 4,010 genes was performed using MIRA-based microarrays in paired DNA samples extracted from 3 fresh frozen CRC tissues and their matching non-cancer tissues from 3 CRC patients undergoing curative surgery. Candidate hypermethylated genes screened by MIRA-based microarrays were further validated in independent CRC samples. A total of 297 CpG dinucleotides covering 211 genes were found to be hypermethylated in CRC tissues. From these 211 candidate methylated genes, three novel hypermethylated genes with more than four probes positive were picked up for validation. Direct bisulfite sequencing revealed that methylations occurred at multiple CpG sites of these three genes in cancer tissues, especially for PHOX2B and FGF12. Combined bisulfite restriction analysis showed that these three genes were methylated in cancer samples but not in non-cancer samples. We also compared the methylation levels of these three novel hypermethylated genes with those of vimentin and SEPT9, well-known hypermethylated genes in CRC, and found that methylated PHOX2B, FGF12 and GAD2 were better than methylated vimentin and SEPT9 in differentiating CRC cancer tissue from non-cancer tissue. Significant enrichment analysis of GO terms of the hypermethylated genes showed that a high proportion of hypermethylated genes in cancer tissues are involved in the regulation of transcription. In conclusion, we found a set of novel hypermethylated genes in CRC, which may have potential to be used as biomarkers for the early diagnosis of CRC.

Correspondence to: Dr Yin-Xue Yang, Department of Colorectal Surgery, General Hospital of Ningxia Medical University, Yinchuan, Ningxia 750004, P.R. China
E-mail: nyfy4442@126.com

Abbreviations: MIRA, methylated-CpG island recovery assay, CRC, colorectal cancer; COBRA, combined bisulfite restriction analysis

Key words: DNA methylation analysis, methyl-CpG binding protein, microarrays, CpG islands, colorectal cancer

Introduction

Colorectal cancer (CRC) is the third most common cancer in men and women, accounting for 11% of all cancer-related deaths (1). In the Chinese population, the morbidity and mortality of colorectal cancer has risen in recent years. Colorectal cancer is best treatable if diagnosed early. Surgery alone can be curative in more than 90% of patients with stage I CRC but less than 50% in most subgroups of stage III CRC (2). Therefore, early detection of CRC is of paramount importance and the key to ultimately curing the vast majority of patients.

DNA methylation is heritable chemical modification induced by enzyme that changes the appearance and structure of DNA without altering its sequence. Aberrant DNA methylation of tumor suppressor genes, a more frequent mechanism of deregulation of gene expression of tumor suppressor genes in the development of cancer than well-known genetic variations (3), may have a potential as cancer biomarkers to achieve high sensitivity and specificity for cancer screening. Many efforts have been invested to utilize such promoter methylation as a biomarker for potential early detection of CRC. As one of the most quickly evolving areas in CRC biomarker research, the importance of methylation biomarker in CRC screening is becoming increasingly recognized. Recently, many methylated genes have been demonstrated to be potential biomarkers for the detection of CRC, including methylated SEPT9 (4) in the plasma, methylated vimentin (5) and methylated TFPI2 (6) in stool. For examples, Lofton-Day *et al* (4) developed an assay for the detection of circulating methylated SEPT9 DNA in plasma, and evaluated its performance as a colorectal cancer biomarker in prospectively collected case-control samples. ColoSure™, detecting methylation status of vimentin in stool with Methylation-Specific PCR (MSP) technology, has been launched for CRC screening service commercially. Despite great progression in CRC methylation biomarker applications, more efforts are still needed to improve the sensitivity and specificity of methylation biomarker. Considering that tumorigenesis is multi-gene, multi-step and multi-pathway process, improvement in the sensitivity and specificity for

Table I. Primer pairs for COBRA validation of significantly hypermethylated genes.

Gene symbol	RefSeq	Primer (5'→3')
FGF12	NM_021032	F: TTCGTGCGGTTGGTGGTT R: TATCCAAAATATACCTACACATACATACAT
GAD2	NM_000818	F: AAATAGTTTTAGGTTCGTTTTATAGAAGG R: CCAAACACCCGAATCTTAACAAC
PHOX2B	NM_003924	F: GGCGTTTGTGTTAGAAAGGAT R: AAACCAACTCCTCCCRAATA
SEPT9	NM_001311493	F: GCGGTCGTAGTAGTTAGTTTAGTAT R: CGCCGAAAACRCTTCCTC
Vimentin	NM_003380	F: CGGCGGGGTTTAGTTTTTTGT R: ACGATAACRCRAACTAACTCCC

early detection of cancer cannot count on only using status of single gene. Recently, a new strategy that uses a panel of methylation markers to discriminate tumor from non-tumor tissue has emerged.

Currently, several high throughput tools have been developed for mapping DNA methylation on a genomic scale, which offer great convenience in discovering novel methylation biomarkers for this new strategy. Most of these tools should translate DNA methylation patterns into DNA sequence information or enrich methylated DNA fragments by using bisulfite conversion, methylation-sensitive restriction endonucleases, DNA demethylation regents or antibody against 5-methylcytosine. Bisulfite-based methods are limited by considerable bioinformatics challenges for the genome mainly consists of three DNA bases after sodium bisulfite conversion of cytosine. Methods based on methylation-sensitive restriction endonucleases are limited by the occurrence of the respective sites within the target sequence. DNA demethylation regent-based approaches, which use microarray to detect altered gene expression induced by DNA methyltransferase inhibitor *in vitro* cells, are effective but can only be used with cell lines, which DNA methylation profiles may be different from that observed in primary tumor tissue (7). Antibody based tools are also limited because it requires ssDNA for recognition, which is sometimes difficult to achieve in CpG-rich DNA regions.

Recently, the methylated-CpG island recovery assay (MIRA) in combination with microarray has been developed to define DNA methylation patterns across the human genome. This method makes use of the high affinity of methyl binding domain (MBD) protein for double-stranded CpG-methlation DNA and has several advantages over existing techniques for analyzing DNA methylation patterns. It does not depend on the occurrence of specific methylation-sensitive restriction sites within the target sequence, does not require ssDNA for recognition and has a low false positive rate in cataloguing methylated CpG islands on a genome-wide basis (8). It has been widely used to detect genome-wide DNA methylation, including in lung cancer (8,9), childhood acute lymphoblastic leukemia (10), breast cancer (11) and normal human B-lymphocytes (12). In this study, we conducted a genome-wide differential analysis of the methylation status of colorectal cancer in small samples using MIRA microarray to screen novel methylation biomarkers, and further validate those methylation biomarkers in independent samples.

Materials and methods

Subjects/patients. Primary cancer and matching adjacent non-cancer tissue samples from a cohort of 3 CRC patients undergoing curative surgery at our hospital were obtained for microarray screening. All cancers were clinically and histologically diagnosed as CRC. All samples were immediately frozen in liquid nitrogen and stored at -80˚C until nucleic acids extraction. The study was approved by the Ethics Committee at our hospital, and all patients gave their informed written consent. Patient information was obtained from medical records.

MIRA microarray analysis of DNA methylation. The MIRA microarray analysis was performed largely as previously reported (13). Briefly, DNA from fresh frozen tissues was cut using *Mse*I digestion (5'-TTAA). The fragmented DNA samples were then purified using MicroDNA Purification kit (Beijing CoWin Biotech Co., Beijing, China) to filter out segments <100 bp. The purified DNA fragments were used to enrich methylated DNA using Methylated Human DNA Enrichment kit (cat. # K5011425, BioChain Institute, Inc., Hayward, CA) according to the manufacturer's instructions. This kit uses MBD beads which were generated with an affinity binding of the recombinant MBD protein and Ni-beads through the His tag. MBD provides more sensitivity and specificity than antibody in binding methylated DNA and is capable of enriching methylated DNA fragments more efficiently than antibody (13). The MIRA-enriched DNA was purified and amplified using GenomePlex Whole Genome Amplification kit (Sigma), instead of conventional PCR amplification with linker primer. This modification led to better liner amplification, less DNA amplification bias and procedure omission of adding linker. Amplicons from the total input DNA without MBD enrichment and methylation-enriched DNA were purified and labeled with cy3-dCTP and cy5-dCTP with Klenow enzyme, respectively. Then the fluorescent dye labeled DNA

fragments (500 ng each) was mixed and co-hybridized to Agilent human CpG island microarray slide. This microarray allows interrogation of 57,723 CpG dinucleotides, covering 4,010 genes. Following hybridization, scanning and washing were performed on Agilent's Microarray Platform. The array data were extracted with Agilent Feature Extraction software. Probes with an intensity <300 were removed after global mean normalization. Considering that microarray experiments were performed with three independent biological replicates, probes with above 1.2-fold changes between MIRA/input signaling ratio from cancer vs. that from non-cancer and with q-value <0.05 using SAM software analysis were scored as positive.

Differentially methylated genes were further analyzed based on a significant enrichment of GO terms using hyper-geometric distribution in the R language package software.

Validation of hypermethylated genes. To further validate the reliability of methylation microarray analysis, three hyper-methylated genes with more than four probes positive in cancer tissues identified by methylation microarray analysis, which have yet not discovered methylation in CRC, were picked up and validated the methylation status using direct bisulfite sequencing. First, DNA was sodium bisulfite modified by using the DNA methylation detection kit (BioChain) as according to the manufacturer's instructions. Sodium bisulfite-modified DNA was then used for subsequent PCR amplification in 50 μl reaction volumes containing 1X PCR buffer (Mg^{2+} Plus), 200 μM of each dNTP, 0.5 μM of forward primer, 0.5 μM of reverse primer (primer sequences are provided in Table I) and 0.5 unit of Taq HS (Takara, Dalian, China). Amplification was performed with one cycle of 94˚C for 5 min, 25 cycles of 94˚C for 30 sec, 55˚C for 30 sec and 72˚C for 30 sec, followed by one cycle of 72˚C for 6 min. The forward amplification primers were used as the sequencing primers.

Validation was also performed for the three selected novel genes and other two interest genes, whose hypermethylation in CRC have been well-established, using combined bisulfite restriction analysis (COBRA). Sodium bisulfite conversion and PCR amplification were carried out using the same initial PCR reaction as described above for direct bisulfite sequencing analysis. After amplification, the PCR products were purified and digested with *Bst*UI (CG↓CG) restriction enzyme (New England Biolabs, Ipswich, MA, USA) overnight at 37˚C, followed by visualization on a 2% agarose gel supplemented with ethidium bromide.

Results

MIRA microarray analysis of DNA methylation. In this study, the global DNA methylation patterns in 3 colorectal cancers and corresponding neighboring non-cancer tissues were analyzed with the MIRA microarray. There were a total of 57,723 CpG dinucleotides covering 4,010 genes (average 14.4 CpG loci per gene) that were studied for methylation status per sample. A total of 297 CpGs with >1.2-fold changes in MIRA/input signaling ratios between cancer and non-cancer tissue, which corresponds to 211 differently methylated genes, were picked out by using two-class paired SAM analysis. The microarray data have been deposited in NCBI's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and are accessible through GEO series accession number GSE33545. Hierarchical cluster display of these hypermethylated 211 genes in the 3 paired specimens is shown in Fig. 1.

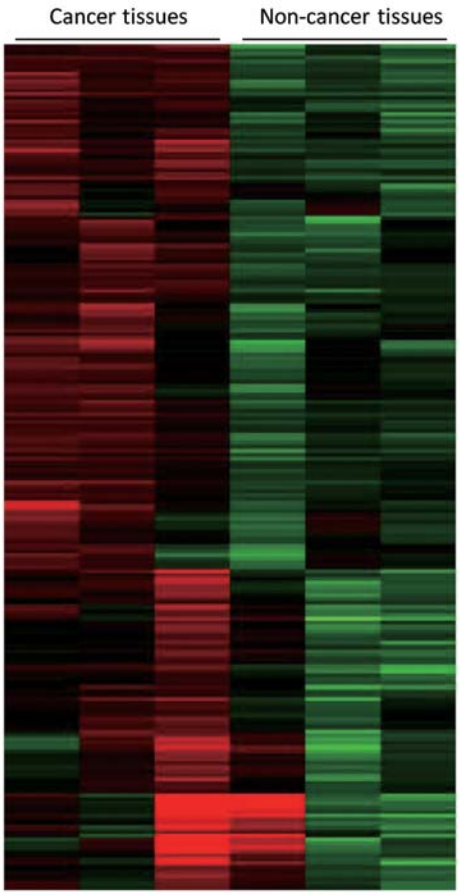

Figure 1. Clustering display of 211 hypermethylated genes in 3 colorectal cancer tissues and corresponding non-cancer tissues. Red indicates hyper-methylation and green indicates hypomethylation.

In order to identify functions associated with these differentially methylated CpGs and genes, we performed gene ontology (GO) analysis using hyper-geometric distribution in the R language package software. The results showed a high proportion of hypermethylated genes in cancer tissues are involved in regulation of transcription, P<0.01 (Table II).

Validation of differentially methylated genes. To validate the reliability of methylation microarray analysis, three novel hypermethylated genes from the 14 hypermethylated genes (Table III) with more than four probes positive, identified by methylation microarray analysis were picked up for validation at a single-base-pair level in CRC cancer tissues and adjacent non-cancer tissues using bisulfite sequencing. Widespread methylations for those three genes were observed across the entire fragments analyzed in cancer tissues, especially for PHOX2B and FGF12. Fig. 2 shows that CpG methylation occurred at nucleotide +59, +65, +67, +74 and +102 in colorectal cancer tissue but not in adjacent non-cancer tissue.

To further validate the reliability of methylation microarray analysis, PHOX2B, FGF12 and GAD2 were further subjected to COBRA validation. All of them were confirmed to be methylated in cancer samples but not in non-cancer samples.

Table II. Significant enrichment analysis of GO terms of the differentially methylated genes using the R language package software.

Pathways	Genes
Regulation of transcription, DNA-dependent	ZNF263, OLIG2, DLX4, EBF1, ESRRB, FOXC1, TLX3, VSX1, MNX1, TLX1, HOXA5, HOXA9, IRF4, MEIS1, MEOX2, MLLT3, NFIX, NFKB1, NR4A2, OTX2, PAX6, PAX9, PITX2, POLR2A, POU4F2, SIM1, SIX1, SOX1, SUPT4H1, TFAP2B, WT1, ZNF135, PHOX2B, LHX2 and NR1D1
Positive regulation of transcription	TOPORS, EBF1, NUP62, AATF, GLI3, IRF4, MYF6, NFKB1, BLM, TBX5, TGFBR1 and WT1
Positive regulation of transcription from RNA polymerase II promoter	CDKN2A, FOXC1, AATF, GLI3, HIPK2, TLX1, HOXA5, PDX1, ASCL1, NEUROG3, PAX9, POU4F2, RB1, SALL1, SIX1, SOX9, SUPT4H1 and WT1

Table III. Hypermethylated genes with more than four probes positive.

Gene	Total probes	Changed probes	Ratio	Gene description
MNX1	33	7	2.24	Motor neuron and pancreas homeobox 1
PHOX2B	31	6	4.59	Paired-like homeobox 2b
EPHA7	24	5	4.34	EPH receptor A7
GAD2	31	5	3.77	Glutamate decarboxylase 2 (pancreatic islets and brain, 65 kDa)
OLIG2	82	5	3.41	Oligodendrocyte lineage transcription factor 2
FGF12	34	4	5.10	Fibroblast growth factor 12
GAD1	64	4	4.23	Glutamate decarboxylase 1 (brain, 67 kDa)
GDNF	41	4	4.08	Glial cell derived neurotrophic factor
GRIA2	17	4	4.43	Glutamate receptor, ionotropic, AMPA 2
PAX6	145	4	4.10	Paired box 6
TFAP2C	77	4	3.58	Transcription factor AP-2 γ (activating enhancer binding protein 2 γ)
TIAM1	36	4	3.73	T-cell lymphoma invasion and metastasis 1
WIT1	29	4	3.27	Wilms tumor upstream neighbor 1
WT1	52	4	3.59	Wilms tumor 1

Several probes are present on the microarray for each gene. Ratio, representing the average ratio of probe density of positive probe in cancer tissue to that in adjacent non-cancer tissue, was calculated by SAM software.

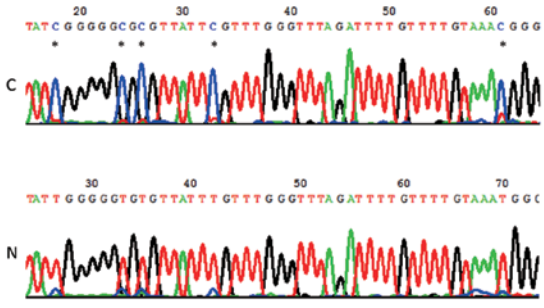

Figure 2. Direct bisulfite sequencing picture of FGF12. Methylated CpGs (*) were observed at nucleotide +59(18), +65(24), +67(26), +74(33) and +102(61) in colorectal cancer tissue (C) but not in adjacent non-cancer tissue (N).

We also compared the methylation levels of these three novel hypermethylated genes with those of vimentin and SEPT9, well-known hypermethylated genes in CRC. The results showed that methylated PHOX2B, FGF12 and GAD2 were better than methylated vimentin and SEPT9 in differentiating CRC cancer tissue from non-cancer tissue (Fig. 3).

Discussion

Two methylated DNA enrichment approaches have been developed to identify methylated genes in genome-wide scale. Antibody-based approach enriches single-stranded DNA while the MBD-based approach captures double-stranded DNA. A comparative study between these two methylated DNA enrichment approaches have shown that each technique operates in a different domain of the CpG density landscape and overall degree of enrichment is higher for MBD-based approach, especially for CpG-dense methylated DNA, suggesting the use of MBD-based approach may favor interrogation of CpG-rich

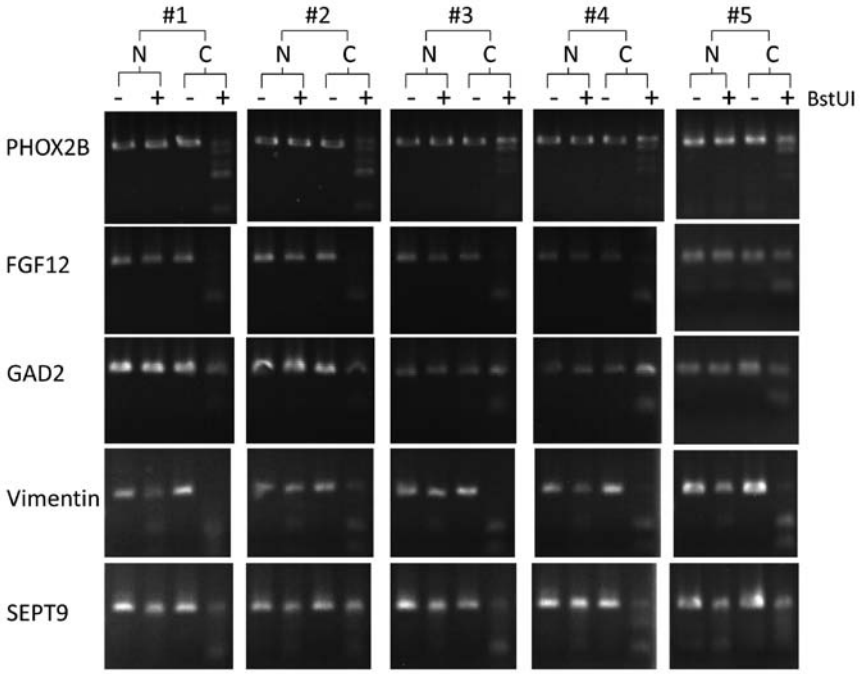

Figure 3. Results from combined bisulfite restriction analysis for PHOX2B, FGF12, GAD2, vimentin and SEPT9 in five colorectal cancer tissues (C) and paired adjacent non-cancer tissues (N). In each case, undigested product (-) is shown next to digested product (+).

regions (14). In this study, MIRA-based microarray analysis was performed in 3 paired DNA samples from CRC tissues and adjacent non-cancer tissues to look for novel DNA methylation in CRC. A total of 297 CpG dinucleotides in CRC covering 211 genes were found. Three novel hypermethylated genes with more than four probes positive identified by methylation microarray analysis were picked up for further validation and all of them were confirmed to be methylated in cancer samples but not in non-cancer samples using combined bisulfite restriction analysis. Significant enrichment analysis of GO terms of the hypermethylated genes showed that these hypermethylated genes were involved in many important pathways, including regulation of transcription.

Because it is rarely feasible to get sufficient affinity-purified DNA for microarray hybridization, DNA amplification before hybridization to microarrays is often needed either for MBD approach or antibody approach. In this study, we used whole genome amplification to amplify MIRA-captured DNA, replacing conventional PCR amplification with linker primer. This modification led to better liner amplification, less DNA amplification bias and more easy operation. In this study, however, we did not detect vimentin methylation, a widely used biomarker for CRC. This phenomenon might be explained by the finding that DNA amplification is often difficult for CpG-rich regions and results in depletion of GC-rich regions, and therefore reduces sensitivity and permits only a sub-fraction of the genome to be interrogated (14).

In this study, we identified three novel hypermethylated genes in colorectal cancer. FGF12 is a novel gene whose methylation status has not been reported in any human disease, nor its relation to malignant features. It encodes a member of the fibroblast growth factor (FGF) family, which are involved in a variety of biological processes, including tumor growth and invasion. Methylations of the other two genes have been reported in cancer. GAD2 has been reported to be methylated in follicular lymphoma (15) and PHOX2B in Wilms tumor (16) and neuroblastoma (16,17). However, this was first study in which that methylation of GAD2 and PHOX2B was identified in colorectal cancer. It has been known that GAD2 encodes one of several forms of glutamic acid decarboxylase, which are responsible for catalyzing the production of γ-aminobutyric acid from L-glutamic acid. PHOX2B encodes a member of the paired family of homeobox proteins localized to the nucleus. It remains unresolved whether hypermethylated genes are the etiological causes or the consequence of colorectal cancer. However, the novel methylated genes identified in this study, are likely to be potentially developed into an innovative strategy for disease detection, prognosis, and monitoring the treatment outcomes. Next, we will further assess the utility of the three novel hypermethylated genes in stool DNA as molecular markers for early detection of CRC.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J and Thun MJ: Cancer statistics, 2009. CA Cancer J Clin 59: 225-249, 2009.
2. Myers AP, Meyerhardt JA and Cantley LC: Getting knit-PI3Ky: PIK3CA mutation status to direct multimodality therapy? Clin Cancer Res 15: 6748-6750, 2009.
3. Schuebel KE, Chen W, Cope L, *et al*: Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. PLoS Genet 3: 1709-1723, 2007.
4. Lofton-Day C, Model F, Devos T, *et al*: DNA methylation biomarkers for blood-based colorectal cancer screening. Clin Chem 54: 414-423, 2008.
5. Itzkowitz SH, Jandorf L, Brand R, *et al*: Improved fecal DNA test for colorectal cancer screening. Clin Gastroenterol Hepatol 5: 111-117, 2007.
6. Glockner SC, Dhir M, Yi JM, *et al*: Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer. Cancer Res 69: 4691-4699, 2009.

7. Houshdaran S, Hawley S, Palmer C, *et al*: DNA methylation profiles of ovarian epithelial carcinoma tumors and cell lines. PLoS One 5: e9359, 2010.
8. Rauch T, Li H, Wu X and Pfeifer GP: MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res 66: 7939-7947, 2006.
9. Pfeifer GP and Rauch TA: DNA methylation patterns in lung carcinomas. Semin Cancer Biol 19: 181-187, 2009.
10. Dunwell T, Hesson L, Rauch TA, *et al*: A genome-wide screen identifies frequently methylated genes in haematological and epithelial cancers. Mol Cancer 9: 44, 2010.
11. Hill VK, Hesson LB, Dansranjavin T, *et al*: Identification of 5 novel genes methylated in breast and other epithelial cancers. Mol Cancer 9: 51, 2010.
12. Rauch TA, Wu X, Zhong X, Riggs AD and Pfeifer GP: A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci USA 106: 671-678, 2009.
13. Rauch TA and Pfeifer GP: The MIRA method for DNA methylation analysis. Methods Mol Biol 507: 65-75, 2009.
14. Robinson MD, Stirzaker C, Statham AL, *et al*: Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res 20: 1719-1729, 2010.
15. Bennett LB, Schnabel JL, Kelchen JM, *et al*: DNA hypermethylation accompanied by transcriptional repression in follicular lymphoma. Genes Chromosomes Cancer 48: 828-841, 2009.
16. Abbaszadeh F, Barker KT, McConville C, Scott RH and Rahman N: A new familial cancer syndrome including predisposition to Wilms tumor and neuroblastoma. Fam Cancer 9: 425-430, 2010.
17. de Pontual L, Trochet D, Bourdeaut F, *et al*: Methylation-associated PHOX2B gene silencing is a rare event in human neuroblastoma. Eur J Cancer 43: 2366-2372, 2007.