RESEARCH COMMUNICATION

Genome-wide Analysis of Aberrant DNA Methylation for Identification of Potential Biomarkers in Colorectal Cancer Patients

Wei-Jia Fang[1], Yi Zheng[1], Li-Ming Wu[2], Qing-Hong Ke[2], Hong Shen[3], Ying Yuan[3], Shu-Sen Zheng[2]*

Abstract

Background: Colorectal cancer is one of the leading causes of mortality worldwide. Genome wide analysis studies have identified sequence mutations causing loss-of-function that are associated with disease occurrence and severity. Epigenetic modifications, such DNA methylation, have also been implicated in many cancers but have yet to be examined in the East Asian population of colorectal cancer patients. **Methods:** Biopsies of tumors and matched non-cancerous tissue types were obtained and genomic DNA was isolated and subjected to the bisulphite conversion method for comparative DNA methylation analysis on the Illumina Infinium HumanMethylation27 BeadChip. **Results:** Totals of 258 and 74 genes were found to be hyper- and hypo-methylated as compared to the individual's matched control tissue. Interestingly, three genes that exhibited hypermethylation in their promoter regions, CMTM2, ECRG4, and SH3GL3, were shown to be significantly associated with colorectal cancer in previous studies. Using heatmap cluster analysis, eight hypermethylated and 10 hypomethylated genes were identified as significantly differentially methylated genes in the tumour tissues. **Conclusions:** Genome-wide methylation profiling facilitates rapid and simultaneous analysis of cancerous cells which may help to identify methylation markers with high sensitivity and specificity for diagnosis and prognosis. Our results show the promise of the microarray technology in identification of potential methylation biomarkers for colorectal cancers.

Keywords: Genome-wide analysis - colorectal cancer - DNA methylation - bisulphite conversion method

Introduction

     Colorectal cancer is one of the leading causes of mortality worldwide. The etiology of this disease has proven complex and includes environmental and genetic components; as such, it remains to be fully understood. Sequence mutations causing loss-of-function have been associated with disease onset and severity. The recent advances in methods to examine epigenetic modifications, such as DNA methylation, have led to interest in determining the genome-wide epigenetic profiles characterizing many different disease states.

     Methylation of cytosine-guanine (CpG) dinucleotides has been identified as one of the most important epigenetic alterations underlying human carcinogenesis. More epigenetically altered genes than genetcally altered genes have been found to exist in any given tumour (Khare, 2012). For example, 11 genes on average are mutated in human cancers (Tobias et al., 2006) but about 200 genes were found to be hypermethylated in the representative human colon adenocarcinoma cell line SW48 (Michael et al., 2005). This finding implied that detection of epigenetic changes in colorectal tumours may be a helpful strategy to improve diagnosis and prognosis of this deadly disease.

     Information on aberrant methylation of genes that correlate to a particular disease or disease state can be obtained from samples sourced from any body fluid or from stool (Asif and Jean, 2004). Various efficient technological platforms have been developed in recent years for high-throughput genome-wide analysis of DNA methylation; one of these is the Infinium HumanMethylation27 BeadChip by Illumina. This platform has been successfully applied to detect aberrantly methylated genes in diabetes mellitus (Christopher et al., 2010), and ovary (Christina et al., 2010) and breast cancers (Van et al., 2010).

     Epidemiological evidence for the association among ethnicity, ancestry, heritage, geographical origin, and occurrence of colorectal cancer has also become recently available. Unfortunately, no genome-wide level epigenetic studies have been performed with the precise focus on colorectal cancer in East Asia. Thus, we obtained cancerous and type matched non-cancerous tissues from East Asian colorectal cancer patients to investigate

*[1]Department of Medical Oncology, [2]Key Laboratory of Combined Multi-organ Transplantation, Ministry of Public Health, Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, [3]Department of Medical Oncology, The Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China *For correspondence: shusenzheng@zju.edu.cn*

and characterize the differential methylation profile corresponding to the disease.

Materials and Methods

Participants and tissue samples

Fresh tumor tissues and type matched non-cancerous (normal) tissues were obtained by surgical biopsy from three patients (Cases 1, 2, and 3) on May 28, 2010 in our hospital. The non-cancerous tissues from the same tissue type (colon region) were obtained as negative controls and run in parallel for comparative analysis to identify differential gene expression in the experimental (cancerous) tissue. Each patient gave voluntary informed consent and the study was carried out with approval from the Institutional Ethics Committee.

Case 1 samples yielded poor genomic DNA upon extraction, and were excluded from further analysis. Cases 2 and 3 yielded sufficient high-quality DNA and were carried through subsequent analysis. These two patients were males, 44 and 76 years old, and diagnosed with rectal cancer. Their pathologic characteristics were moderate-poor differentiation adenocarcinoma with TNM classification of T2N1M0 (IIIA).

DNA isolation and bisulphite conversion

Genomic DNA from Cases 2 and 3 was isolated by using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA). Bisulphite conversion was performed by means of the EZ DNA Methylation kit (Zymo Research Corp, Orange, CA, USA). The protocol used had been adapted by Illumina (San Diego, CA, USA) to improve bisulphite conversion efficiency, and includes a cyclic denaturation step during the conversion reaction. Successful bisulphate conversion is dependent upon 5-methyl cytosine being resistant to deamination by bisulphite treatment. Therefore, only unmethylated cytosines are deaminated to uracils (which are then converted into thymines following subsequent PCR), while 5-methylcytosines remain unchanged during the reaction that is carried out on single-stranded DNA. We used a total of 500 ng DNA for each bisulphite conversion reaction.

Processing of HumanMethylation27 BeadChips

The Infinium HumanMethylation27 BeadChip protocol was composed of six steps: whole-genome amplification; fragmentation; hybridization; washing; counterstaining; and, scanning. The protocol was carried out exactly as recommended by the manufacturer, and which can be found at http://www.illumina.com/downloads/ InfMethylation_AppNote.pdf. A total of 500ng–1ug bisulphate-converted starting DNA was used as the starting sample. Considering the 28 million CpG sites currently known throughout the haploid human genome, Illumina designed the Infinium methylation probes for a set of 27,578 CpG sites located in promoter regions (up to 1kb upstream or 500 bp down stream of transcription start sites (TSS)). Of these, 27,324 correspond to 14,475 consensus coding sequences (CCDS), including about 1,000 cancer-associated genes. Furthermore, probes were preferentially selected to occur within CpG islands by using the NCBI "relaxed" definition of a CpG island: CpG islands identified bioinformatically as having a CpG content of >50% and an observed/expected ratio of >0.6 (http://www.ncbi.nlm.nih.gov/projects/mapview/static/ humansearch.html#comp).

Statistical analysis

The scanner data and image output files, including chromosome methylation status and heatmap analysis, were managed and generated with the GenomeStudio software Methylation module v1.0. All reagents were provided by Illumina. The beta value (β) was used to estimate the methylation level of the CpG locus for calculating the ratio of intensities between methylated and unmethylated alleles ($1>AVG$-$\beta>0$: represents fully to non-methylated). An AVG-$\beta >0.7$ was considered to be significantly hypermethylated in the locus, as compared to other loci in the same sample. For a locus covered by multiple probes, the DiffScores across probes were averaged ($-374.344<$ DiffScore <374.344; control = 0). Absolute value >100 was considered to be significantly hyper- or hypomethylated in the sample, as compared to other samples. The heatmap analysis compared the data obtained from an experimental (cancer) and control (non-cancerous, type-matched) tissue from two colorectal cancer patients, and linear dependence was determined bioinformatically using the average linkage method and Pearson's correlation metric.

Results

Methylation profile of arrays

Aberrant methylation of genes was observed in the two tumour tissues analyzed, as compared to the individual's normal tissues. The degree of methylation status of the differentially methylated genes and their frequency were allocated into a histogram (Figure 1). DiffScore is a value that Illumina software used to illustrate the difference of two groups of data. It is directional p-value, closed related to p-value (See Illumina GenomeStudio document
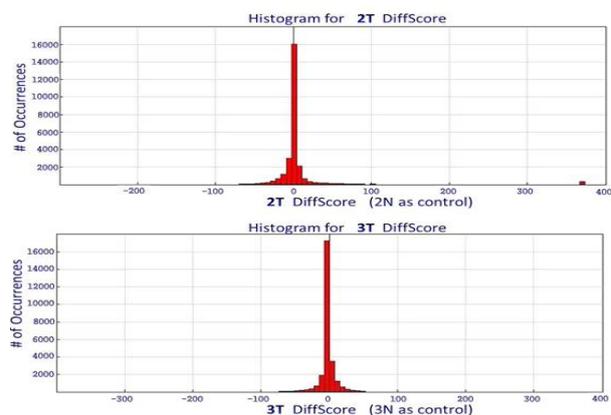

Figure 1. DiffScore of Tumour Samples. The difference of methylation status between the tumour (T) and normal (N) tissues in Case 2 (2) and Case 3 (3) samples. In 2T, 258 gene loci had a DiffScore >100, which lead to a small column that lies across the 100 DiffScore (x-axis) and a more obvious column near the 400 DiffScore. In 3T, only 74 gene loci had a DiffScore >100, which lead to almost no visible columns on or to the right of the 100 DiffScore zero point on the horizontal ordinate
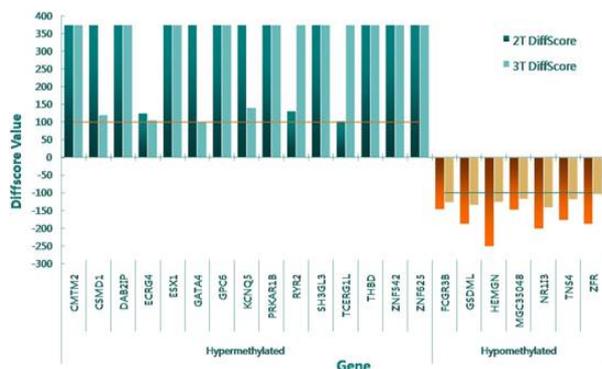
Figure 2. Aberrant Methylation Genes in Both Samples.
Fifteen hypermethylated (blue) and seven hypomethylated (orange) genes that were significant and concordant in both tumour tissues examined are shown on the vertical ordinate
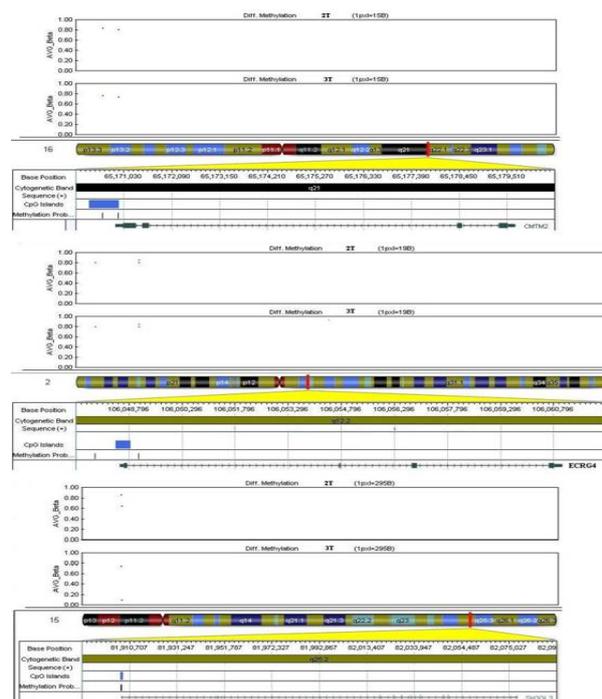

Figure 3. Chromosome Analysis of Three Genes.
Hypermethylated loci were found in promoter regions for each of these genes (Top: CMTM2, Middle: ECRG4, Bottom: SH3GL3) in both tumor tissues examined. At least one loci in 2T and 3T have AVG-β values >0.7, signifying significant hypermethylation as compared to other loci in the same sample

for detail. In Case 2, 258 gene loci were found to have a DiffScore >100. Among these loci, 178 were considered to be significantly hypermethylated while 80 were significantly hypomethylated. In contrast, Case 3 had only 74 gene loci with a DiffScore >100 and 30 genes were hypermethylated while 44 were hypomethylated. We use the DiffScore >100 here as its correspondant p-value is less 0.01. With this score we could be work with a proper number of genes or methylation sites that are significant enough in their methylation status.

DNA methylation analysis
    When the two datasets of Case 2 and Case 3 genes with DiffScores >100 were compared we found that 15 of the hypermethylated and seven of the hypomethylated genes were similar between the two (Figure 2).
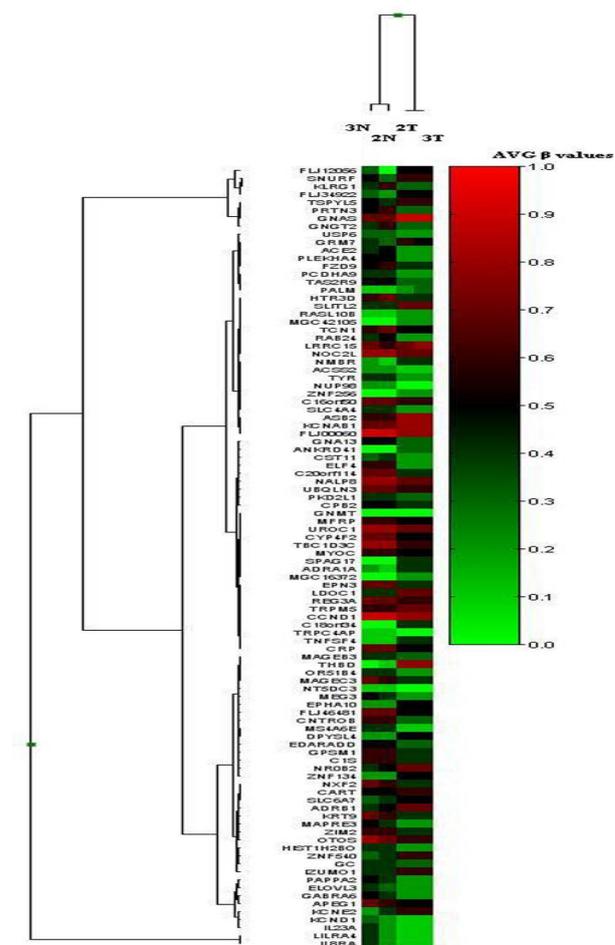

Figure 4. Heatmap Cluster of Two Samples. Hyper- or hypomethylation levels in the tumour and normal tissues in Case 2 and 3 samples. Methylation levels are indicated by a gradient map representing the hypo- to hypermethylated status (green to red) of different genes according to AVG-β estimation. (N), normal tissue; (T), tumour tissue; genes with (AVG-β) >0.7 were considered to be significantly hypermethylated. For example, THBD gene was significantly hypermethylated in tumor as compared to normal tissue, as indicated by the gradient between these two colours being significantly different according to the design of the heatmap cluster method: normal tissue side (bright green) and tumour tissue side (bright red)

Chromosome promoter region methylation status analysis
    Among the 15 hypermethylated genes that were common between the two samples were the colorectal cancer-associated genes CMTM2, ECRG4, and SH3GL3. All three of these genes contained hypermethylated loci in their promoter regions (Figure 3). Moreover, the AVG-β values for the three genes were >0.7. Detailed information is shown in Table 1. In contrast, the seven hypomethylated genes that were shared between the two samples had no hypomethylated loci that corresponded to their promoter regions.

Heatmap cluster
    In order to further verify the similarities observed in methylation status of the Case 2 and 3 tissues, we used Genomestudio software to separate the genes associated with the tumor and the normal tissues into two parts, according to their AVG-β values. AVG-β is a value Illumina software use it to describe the methylation

Table 1. Details of Three Hypermethylated Genes

Gene	Transcription initiation locus	Coding initiation locus	Target ID	CpG locus	2T AVG-β values	3T AVG-β values
CMTM2	65,170,852	65,171,012	cg01683883	65,170,554	0.8215514	0.7548316
ECRG4	106,048,545	106,048,653	cg11024597	106,047,843	0.7981651	0.7825059
SH3GL3	81,907,287	81,950,590	cg22946150	81,907,111	0.8596994	0.7393214

proportion in the sample. The big the AveBeta, the stronger the methylation of CpG site (see Illumina document for detail). As shown in Figure 4, the heatmap that was generated displayed different methylation status between the tumour and normal tissues. Eight significantly hypermethylated genes were found in the tumour tissues: THBD, ADRB1, KCH1D1, SLITL2, LDOC1, NRDB2, ZNF540, and IZUMO1. Meanwhile, ten significantly hypomethylated genes were found in the tumour: KRT9, NXF2, HTR3D, ELF4 C20orf114, EPN3, CRP, MAGEC3, CNTROB, and IIMD.

Discussion

Recent methodologies based on sodium bisulphite conversion of cytosine have allowed for accurate, whole genome investigations of the methylation status of human tissues using sequencing strategies (Maria and Jerd, 2010). Restriction landmark genomic scanning was one of the earliest of such methods to be introduced and successfully used to determine the global methylation status of 1,184 unselected CpG islands (Joseph et al., 2000). Later, Pei et al. used the GoldenGate microarrays from Illumina to study 1,505 genomic loci; they found 202 loci from 132 genes were differentially methylated in colorectal cancer samples obtained from Western Australian patients (Pei et al., 2010).

Here, we describe our use of DNA methylation profiling of colorectal cancer tissues from two East Asian male patients by means of the recently developed Illumina Infinium® HumanMethylation27 BeadChip. The 27K Infinium assay enabled direct investigation of 27,578 individual cytosines at the CpG loci within proximal promoter regions, encompassed within 1.5 kb upstream and 1 kb downstream of the transcription start sites of 14,475 consensus genes.

Dr. Manel Esteller published a list of 11 genes that are silenced by CpG island promoter hypermethylation in colorectal cancer tissues: RARβ2, CRBP1, FAT, SFRP1, DKK1, WIF1, COX2, GATA4, GATA5, TPEF/HPP1, and WRN (Manel, 2007). The methylation of these genes in our two tumour samples in our study yielded exactly opposite results for these genes, with the sole exception of GATA4. In addition, the general methylation patterns between our two samples were largely discordant; Case 2 had more hyper- or hypomethylated genes than did tissues from Case 3. Nevertheless, there were some aberrantly methylated genes that were concordant between the two: 15 hypermethylated and seven hypomethylated. This anomalous phenomenon merits further investigation. However, this may simply imply the distinct characteristics of DNA methylomes in individual colorectal cancer patients.

FDR (False Discovery Rate) is a measure used in a lot of high throughput studies to describe the possibility of false positives. We could use its concept to calculate our risk of running into false positives. Out of about 27000 CpG sites on the Illumina chip, we think that 258 in case 2 and 74 in case 3 are significantly hypermethylated genes, while 80 in case 2 and 44 in case 3 are hypomethylated. Among them, 15 are common genes in hypermethylated while 7 are common in hypomethylated groups. With the FDR concept, the false positive rate should be less than 1%. Therefore we believe the resulted genes in both hypermethylated and hypomethylated genes should very likely related to the disease. More confirmation studies are planned.

Three genes with hypermethylated loci in the promoter region were found among the aberrantly methylated genes in both samples. They are the CKLF-like MARVEL transmembrane domain-containing family (CMTM), the oesophageal cancer-related gene 4 (ECRG4), and the SH3-containing Grb2-like 3 gene (SH3GL3). CMTM represents a novel family of proteins linking classical chemokines and the transmembrane 4 superfamily(Henan et al., 2010). CMTM5 exhibits tumour suppressor activities, but frequently undergoes epigenetic inactivation in carcinoma cell lines (Luning et al., 2007). CMTM5-v1 inhibits the growth of cervical cancer (Luning et al., 2009) and pancreatic cancer cell lines (Xiaohuan et al., 2009) by inducing apoptosis. ECRG4 is a newly discovered gene expressed in oesophageal squamous cells, colorectal carcinomas, and gliomas. It acts to suppress tumour cell migration and invasion (Silke et al., 2009; Linwei et al., 2010). It is considered to be an independent prognostic factor indicating poor survival in patients with oesophageal squamous cell carcinoma (Yoichiro et al.,2007), and it is a molecular marker for the prediction of biochemical, local, and systemic recurrence of prostate cancer (Donkena et al., 2009, Lin-Wei et al., 2009). SH3GL3 shows an overall high level of expression in the human brain and testis. Its role in the prognosis of cervical lymph node metastasis in oral cavity cancer has been investigated (Norihiro et al., 2003; Sue et al., 2007) .

In stark contrast to hypermethylation, promoter hypomethylation can lead to increased expression of oncogenes (Igor and Frederick, 2009; Wang et al., 2012). Interestingly, only genes with hypomethylated loci outside the promoter region were found in our study. TNS4 was one of these genes. Studies by others have shown that tumours with high TNS4 mRNA expression exhibit aggressive cancerous behaviour (Katsuya et al., 2008). Therefore, having no hypomethylated loci in the promoter region is not the sole criterion for epigenomic activity in oncogenes.

Since single tumour biomarkers may not adequately reflect the phenotype of cells, panels of genomic DNA hypermethylation biomarkers are increasingly being

used in medical practice. In the current study, a heatmap was generated to distinguish the differences (by colour gradient) among aberrant methylated genes in the tumour milieu. In the heatmap, genome-wide methylation information on normal or tumour tissues from the two samples was clustered based on common characteristics. In other words, if genes with obvious differences in methylation status were detected in a separate normal or tumour sample, they would be removed from the heatmap result, no matter how significantly hyper- or hypomethylated they were. Thus, the heatmap cluster analysis was used to precisely explore the different methylation profile of genes in the tumour tissues, as compared to the normal tissues. We identified a panel of several differentially methylated genes (8 hyper- and 10 hypomethylated) in the colorectal tumour samples; future analysis will yield more such genes that may fine-tune this biomarker panel and correlate to different disease stages and prognoses.

Acknowledgements

References

Asif R, Jean PJ (2004). CpG island methylation in gastroenterologic neoplasia: a maturing field. *Gastroenterology*, **127**, 1578-88.

Claudia G, Erica L, Letizia L, et al (1997). A novel SH3-containing human gene family preferentially expressed in the central nervous system. *Genomics*, **41**, 427-34.

Christopher GB, Andrew ET, Vardhman KR, Alexander PM, Stephan B (2010). Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics*, **3**, 33.

Christina T, Marianne E, Andrew F, et al (2010). Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods*, **52**, 248-54.

Donkena KV, Mathias E, Dirk VdB, et al (2009). Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. *Cancer Invest*, **27**, 549-60.

Henan L, Xiaohuan G, Luning S, et al (2010). CMTM5-v1, a four-transmembrane protein, presents a secreted form released via a vesicle-mediated secretory pathway. *BMB Rep*, **43**, 182-7.

Igor PP, Frederick AB(2009). DNA hypomethylation in the origin and pathogenesis of human diseases. *Cell Mol Life Sci*, **66**, 2249-61.

Joseph FC, Michael CF, Dominic JS, et al (2000). Aberrant CpG-island methylation has non-random and tumour-type–specific patterns. *Nat Genet*, **25**, 132-8.

Katsuya S, Koshi M, Fumiaki T, et al (2008). Prognostic relevance of Tensin4 expression in human gastric cancer. *Ann Surg Oncol*, **15**, 2606-13.

Khare S, Verma M (2012). Epigenetics of colon cancer. *Methods Mol Biol*, **863**, 177-85.

Luning S, Yan C, Hong YL, et al (2007). CMTM5 exhibits tumor suppressor activities and is frequently silenced by methylation in carcinoma cell lines. *Clin Cancer Res*, **13**, 5756-62.

Luning S, Xiaohuan G, Markus P, et al (2009). CMTM5-v1 induces apoptosis in cervical carcinoma cells. *Biochem Biophys Res Commun*, **379**, 866-71.

Li LW, Yu XY, Yang Y, et al (2009). Expression of esophageal cancer related gene 4 (ECRG4), a novel tumor suppressor gene, in esophageal cancer and its inhibitory effect on the tumor growth in vitro and in vivo. *Int J Cancer*, **125**, 1505-13.

Li L, Zhang C, Li X, Lu S, Zhou Y (2010). The candidate tumor suppressor gene ECRG4 inhibits cancer cells migration and invasion in esophageal carcinoma. *J Exp Clin Canc Res*, **29**, 133.

Michael W, Jonathan JD, David W, et al (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, **37**, 853-62.

Manel E (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature*, **8**, 286-98.

Maria AH, Gerd PP (2010). Methods for genome-wide analysis of DNA methylation in intestinal tumors. *Mutat Res*, **693**, 77-83.

Pei WA, Marie L, Natalia L, et al (2010). Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer*, **10**, 227.

Su TN, Shogo H, Hitoshi T, et al (2007). Identification of a predictive gene expression signature of cervical lymph node metastasis in oral squamous cell carcinoma. *Cancer Sci*, **98**, 740-6.

Silke G, Valeska F, Thilo T, et al (2009). ECRG4 is a candidate tumor suppressor gene frequently hypermethylated in colorectal carcinoma and glioma. *BMC Cancer*, **9**, 447.

Tobias S, Siân J, Laura DW, et al (2006). The consensus coding sequences of human breast and colorectal cancers. *Science*, **314**, 268-74.

Van dAI, Yu W, Suo L, et al (2010). Array-based DNA methylation profiling for breast cancer subtype discrimination. *PloS One*, **5**, e12616.

Wang TP, Hsu SH, Feng HC, Huang RF (2012). Folate deprivation enhances invasiveness of human colon cancer cells mediated by activation of sonic hedgehog signaling through promoter hypomethylation and cross action with transcription nuclear factor-kappa B pathway. *Carcinogenesis*, Apr 26.

Xiaohuan G, Ting L, Yu W, et al (2009). CMTM5 induces apoptosis of pancreatic cancer cells and has synergistic effects with TNF-a. *Biochem Biophys Res Commun*, **387**, 139-42.

Yoichiro M, Hideyuki I, Yoshiyuki K, et al (2007). Expression of ECRG4 is an independent prognostic factor for poor survival in patients with esophageal squamous cell carcinoma. *Oncol Rep*, **18**, 981-5.