ORIGINAL ARTICLE

In silico Identification of *SFRP1* as a Hypermethylated Gene in Colorectal Cancers

Jongbum Kim, Sangsoo Kim*

Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea

Aberrant DNA methylation, as an epigenetic marker of cancer, influences tumor development and progression. We downloaded publicly available DNA methylation and gene expression datasets of matched cancer and normal pairs from the Cancer Genome Atlas Data Portal and performed a systematic computational analysis. This study has three aims to screen genes that show hypermethylation and downregulated patterns in colorectal cancers, to identify differentially methylated regions in one of these genes, *SFRP1*, and to test whether the SFRP genes affect survival or not. Our results show that 31 hypermethylated genes had a negative correlation with gene expression. Among them, *SFRP1* had a differentially methylated pattern at each methylation site. We also show that SFRP1 may be a potential biomarker for colorectal cancer survival.

Keywords: colorectal cancer, DNA Methylation, frizzled related protein-1, survival analysis

Introduction

The Encyclopedia of DNA Elements (ENCODE) project, launched in 2003 by the US National Human Genome Research Institute (NHGRI), has accelerated the study of various aspects of epigenetics [1], among which DNA methylation has been one of the most actively studied areas [2]. DNA methylation refers to the chemical conversion of a cytosine nucleotide in DNA to 5-methylcytosine by DNA methyltransferase. This phenomenon usually takes place at the genomic region where a cytosine is immediately followed by a guanine and where the dinucleotide is enriched more than expected from the G + C content of the genome. Such CpG islands (CGIs) are typically found in the promoter region of a gene, regulating its expression [3].

Tumor suppressor genes (TSGs), including DNA repair genes that repair DNA mismatch during DNA replication, maintain the normal activities of cells, preventing cancer development. On the other hand, their hypermethylation may repress their expression, influencing the development and growth of cancer cells. Moreover, TSG expression may cause gene silencing through the hypermethylation of CGIs, and such a paradigm opens up a new avenue of cancer diagnosis and treatment based on DNA methylation profiles [4].

In cancer cells, genes that play important roles in cell growth and differentiation, such as TSG, DNA repair genes, and apoptosis-related genes, show downregulation in expression due to hypermethylation in their promoter regions. For example, the expression of p16, a well-known TSG, is downregulated due to its promoter hypermethylation [5], and DNA repair genes, such as *BRCA1* and *hMLH1*, are also hypermethylated, repressing their gene expression [6].

It appears that hypermethylation profiles in CGIs may be useful targets for cancer treatment. Some DNA methylation inhibitors have been used as anticancer agents, paving a new avenue for the development of anticancer therapeutics [7]. Gene expression profiles have been used in cancer diagnosis and prognosis, but mRNA levels usually fluctuate temporally or are greatly influenced by environmental cues, resulting in unstable diagnostic sensitivity. On the other hand, DNA hypermethylation can show high specificity and sensitivity in cancer diagnosis [8]. Furthermore, its molecular diagnosis is gaining more and more acceptance due to its availability through real-time and quantitative blood testing [9]. In this study, we surveyed colon cancer data to look for genes that show hypermethylation, and investigated their effects on

patient survival.

Methods

Data

The Cancer Genome Atlas (TCGA), launched in 2009, aimed to analyze the genomic features through sequencing of about 30 different types of cancers. We downloaded the microarray datasets on gene expression and DNA methylation of colorectal cancer and normal samples from the following URL: https://tcga-data.nci.nih.gov/tcga/dataAccess Matrix.htm (the "Data Type" field was selected as "Clinical," "DNA Methylation," and "Expression-Genes"). There were 12 matched cancer/normal pairs of both data types, and these were defined as the discovery set. The microarray platforms were UNC_AgilentG4502A_07_3 and Illumina Infinium Human DNA Methylation 27 (Illumina, San Diego, CA,

USA) for gene expression and DNA methylation, respectively. For the methylation data, the so-called M-value was calculated as the $\log_2$ ratio of the intensities of the methylated probe versus unmethylated one, and the beta-value was the ratio of the methylated probe intensity and the overall intensity, either methylated or unmethylated [9].

For the clinical outcome analysis, we downloaded an expanded dataset from the URL, totaling 524 samples (202 samples on the Illumina Infinium Human DNA Methylation27 platform and 322 samples on the Illumina Infinium Human DNA Methylation450 platform), and 419 samples with clinical information were used in the survival analysis. For these samples, we downloaded a text file on tumor stage, last contact days, and vital status.

Table 1. Differentially hypermethylated and downregulated genes

Methyl probes[a]	Symbol[b]	Chr	MapInfo	Correlation	p-value[c]
cg03382304	*JAM2* (1e)	21	25934047	−0.944	0.00000
cg19246110	*ZNF671* (5u)	19	62930740	−0.934	0.00000
cg15839448	*SFRP1* (1e)	8	41285687	−0.923	0.00000
cg05436658	*PRKCB1* (1e)	16	23755069	−0.922	0.00000
cg06668300	*MAL* (1i)	2	95055482	−0.919	0.00000
cg01580681	*HAND2* (2e)	4	174686591	−0.888	0.00000
cg18239753	*KHDRBS2* (5u)	6	63053922	−0.884	0.00000
cg23113963	*SCNN1B* (p)	16	23220794	−0.879	0.00000
cg06744574	*BEND5* (1e)	1	49014946	−0.854	0.00000
cg21902544	*CBLN2* (5u)	18	68362495	−0.846	0.00000
cg09619146	*CPXM2* (1e)	10	125641024	−0.837	0.00000
cg00903242	*CHL1* (p)	3	213161	−0.835	0.00000
cg08045570	*FOXF2* (?)	6	1335501	−0.807	0.00000
cg16604516	*FBLN2* (1i)	3	13565419	−0.802	0.00000
cg12880658	*CDO1* (5u)	5	115180285	−0.789	0.00000
cg19774122	*LAMA2* (?)	6	129246315	−0.762	0.00002
cg00768439	*NAP1L3* (p)	X	92815367	−0.754	0.00002
cg02082342	*VSIG2* (p)	11	124127583	−0.742	0.00003
cg11656547	*MAMDC2* (5u)	9	71848567	−0.730	0.00005
cg04600618	*RSPH9* (1e)	6	43720958	−0.728	0.00006
cg00238052	*TCEAL2* (p)	X	101266947	−0.725	0.00006
cg23207990	*SFRP2* (p)	4	154929985	−0.707	0.00011
cg25422943	*PCDH9* (p)	13	66702734	−0.693	0.00017
cg14070647	*RSPO2* (5u)	8	109164440	−0.677	0.00028
cg22821324	*GPM6A* (1i)	4	177160433	−0.622	0.00117
cg00929855	*HSPA1A* (5u)	6	31891343	−0.570	0.00364
cg23043245	*PACSIN1* (1i)	6	34542054	−0.560	0.00442
cg06638451	*FAM107A* (5u)	3	58538309	−0.533	0.00736
cg13266631	*NEFL* (p)	8	24870923	−0.448	0.02810
cg09979256	*RSPO3* (5u)	6	127481797	−0.429	0.03651
cg19461621	*COLEC12* (p)	18	490979	−0.423	0.03957

[a]Illumina human methylation probes; [b]The genic region of the probe is given within parentheses as promoter (p), 5'-UTR (5u), first exon (1e), and unknown (?); [c]Linear regression analysis.
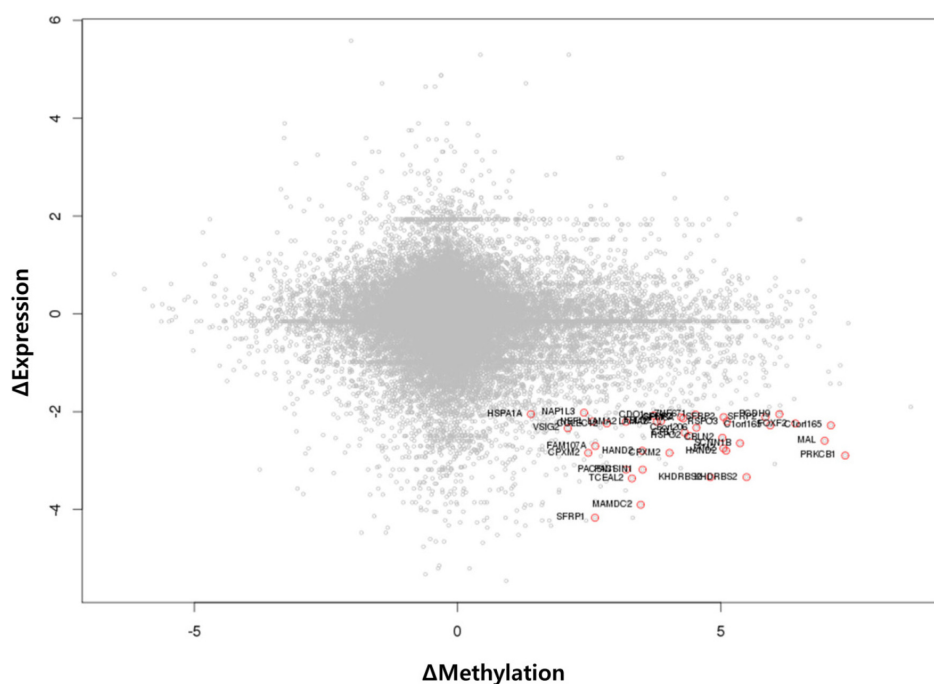
Fig. 1. Scatterplot of differential methylation versus differential expression. A total of 27,578 methylation probes are plotted in grey, while those probes showing downregulated expression by 4-fold or more (Δ expression < -2) and methylation value (M value) change of $+1$ or more are plotted as red points.

Identification of hypermethylated and downregulated genes

The discovery dataset was used for gene selection. Illumina Infinium Human DNA Methylation27 level 2 data were screened for differentially methylated regions (DMRs), where the median M of the tumor samples was greater than $+1$ and that of the matched normal samples was less than -1 [9]. For each gene, the mean mRNA expression level in tumor samples was compared to that in normal samples. Genes showing at least 4-fold down-regulation were selected. DAVID (http://david.abcc.ncifcrf.gov/) was used for the functional annotation of the gene list.

Survival analysis

Survival analysis was undertaken with the survival package in R (version 2.13.0). Survival time was represented by the days since cancer diagnosis in the hospital until the last contact date. The survival time of the patients who were still alive at the last contact date were treated as censored. Kaplan-Meier survival plot and Cox regression analysis were also performed with the survival package. In order to partition patients into two groups according to methylation level, the cutoff point was determined by the maximally selected rank statistics method, available in the R maxstat package.
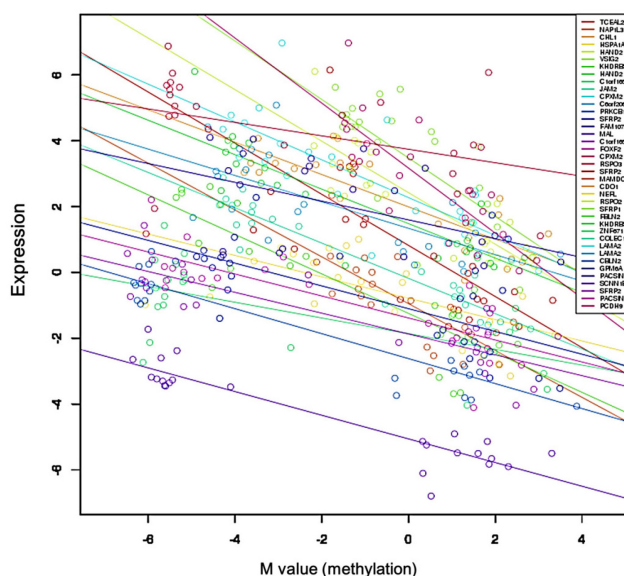

Fig. 2. Negative correlation between gene expression and DNA methylation in 31 selected genes. Each circle represents a sample.

Results and Discussion

Hypermethylated and downregulated genes in colorectal cancer

We used the discovery set (12 pairs of colorectal tumor and matched normal samples) to survey genes showing correlating differential DNA methylation and differential gene expression. All methylation probes (n = 27,578) were

Table 2. Differentially hypermethylated and downregulated genes

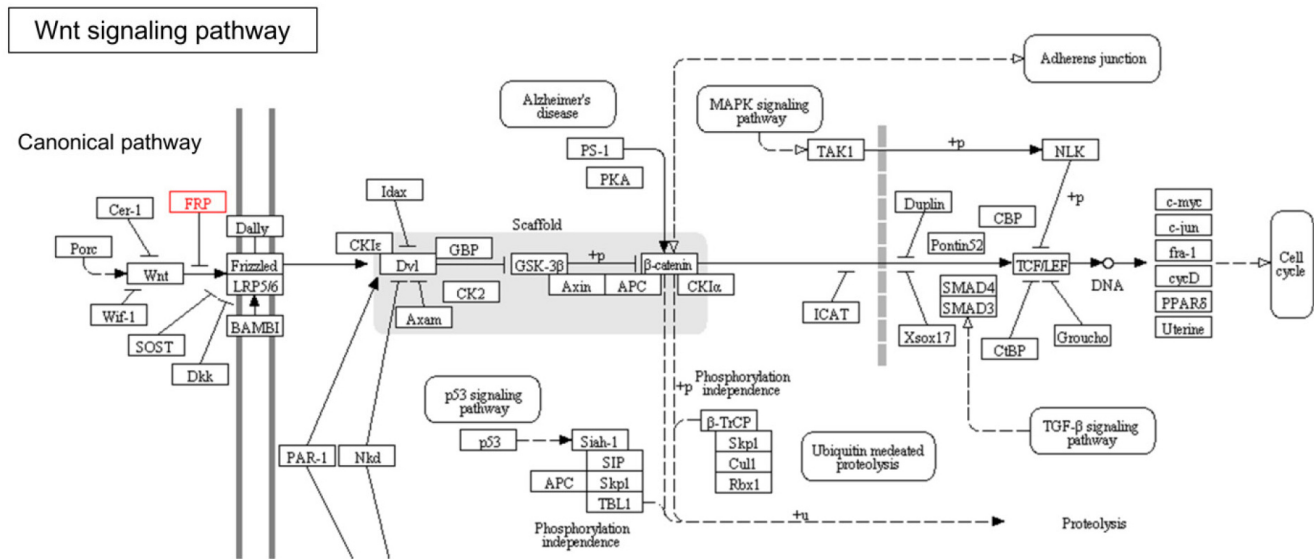| Term	Gene	p-value	Fisher exact
Wnt receptor signaling pathway	*SFRP1, SFRP2, RSPO2, RSPO3*	1.6E-3	9.6E-5
Cell adhesion	*CPXM2, CHL1, JAM2, LAMA2, PCDH9*	3.3E-2	8.2E-3
Biological adhesion	*CPXM2, CHL1, JAM2, LAMA2, PCDH9*	3.3E-2	8.2E-3
Somitogenesis	*SFRP1, SFRP2*	5.9E-2	1.8E-3
Segmentation	*SFRP1, SFRP2*	8.2E-2	3.5E-3

Fig. 3. Wnt signaling pathway from Kyoto Encyclopedia of Genes and Genomes (KEGG) (SFRP genes are in red box).
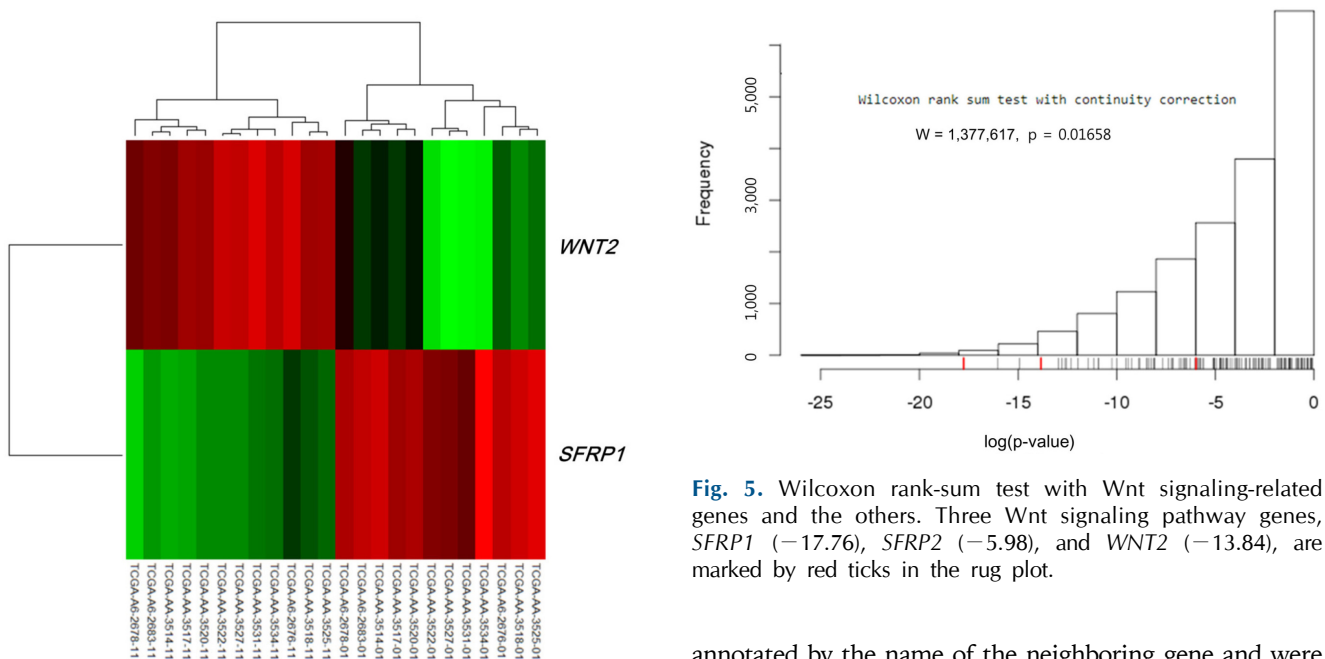

Fig. 4. Heat map of *WNT2* and *SFRP1* in the discovery dataset. The 12 samples on the left are normal, while the 12 samples on the right are tumors.

Fig. 5. Wilcoxon rank-sum test with Wnt signaling-related genes and the others. Three Wnt signaling pathway genes, *SFRP1* (−17.76), *SFRP2* (−5.98), and *WNT2* (−13.84), are marked by red ticks in the rug plot.

annotated by the name of the neighboring gene and were matched to the gene expression matrix. In order to look for the probes that were specifically hypermethylated only in tumors, we used the M-value of the methylation data; the
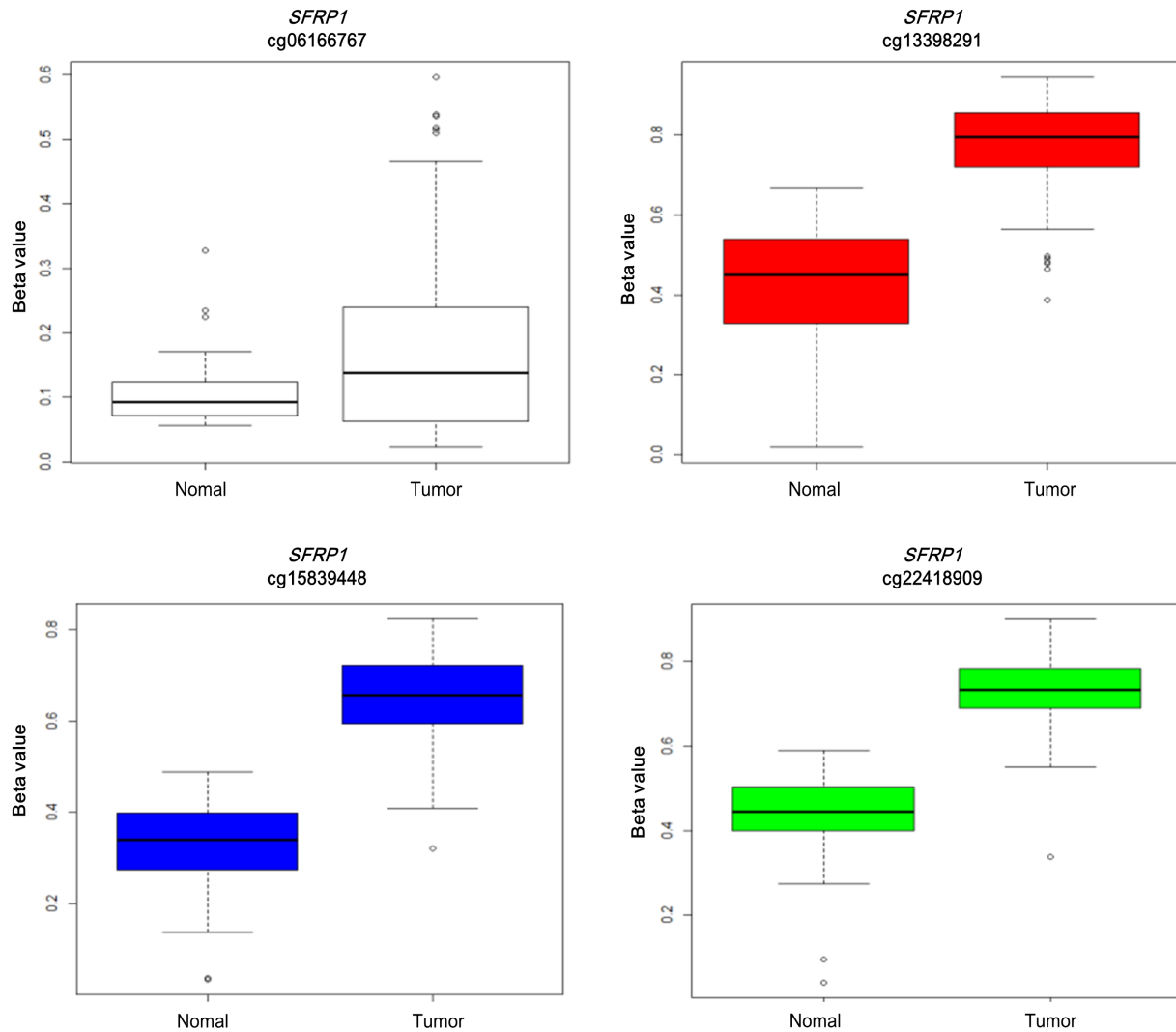
Fig. 6. Differential methylation in four *SFRP1* probes (tumor, 166; normal, 37).

Table 3. Differentially hypermethylated and downexpressed genes

Name	Symbol	Chr	MapInfo	Transcription start site (TSS)	Distance to TSS	Gene strand	p-value[a]
cg06166767	*SFRP1*	8	41287005	41286137	868	-	0.1956
cg13398291	*SFRP1*	8	41285326	41286137	811	-	0.0002
cg15839448	*SFRP1*	8	41285687	41286137	450	-	0.1E-9
cg22418909	*SFRP1*	8	41285895	41286137	242	-	0.607

[a]Linear regression analysis.

median M of tumors (n = 12) was greater than 1, while the matched normal samples (n = 12) had a median M of less than −1. There were 634 such probes. We also identified 707 genes showing more than 4-fold downregulation in tumor samples compared to their matched normal samples. There were 31 genes that were shared by those two lists (Table 1). Fig. 1 shows the scatterplot of the mean differen-

tial methylation versus mean differential expression of these 31 genes (red points) on the background of 27,578 methylation probes (grey points). At the sample level, these 31 genes showed a negative correlation between methylation and expression (Fig. 2).
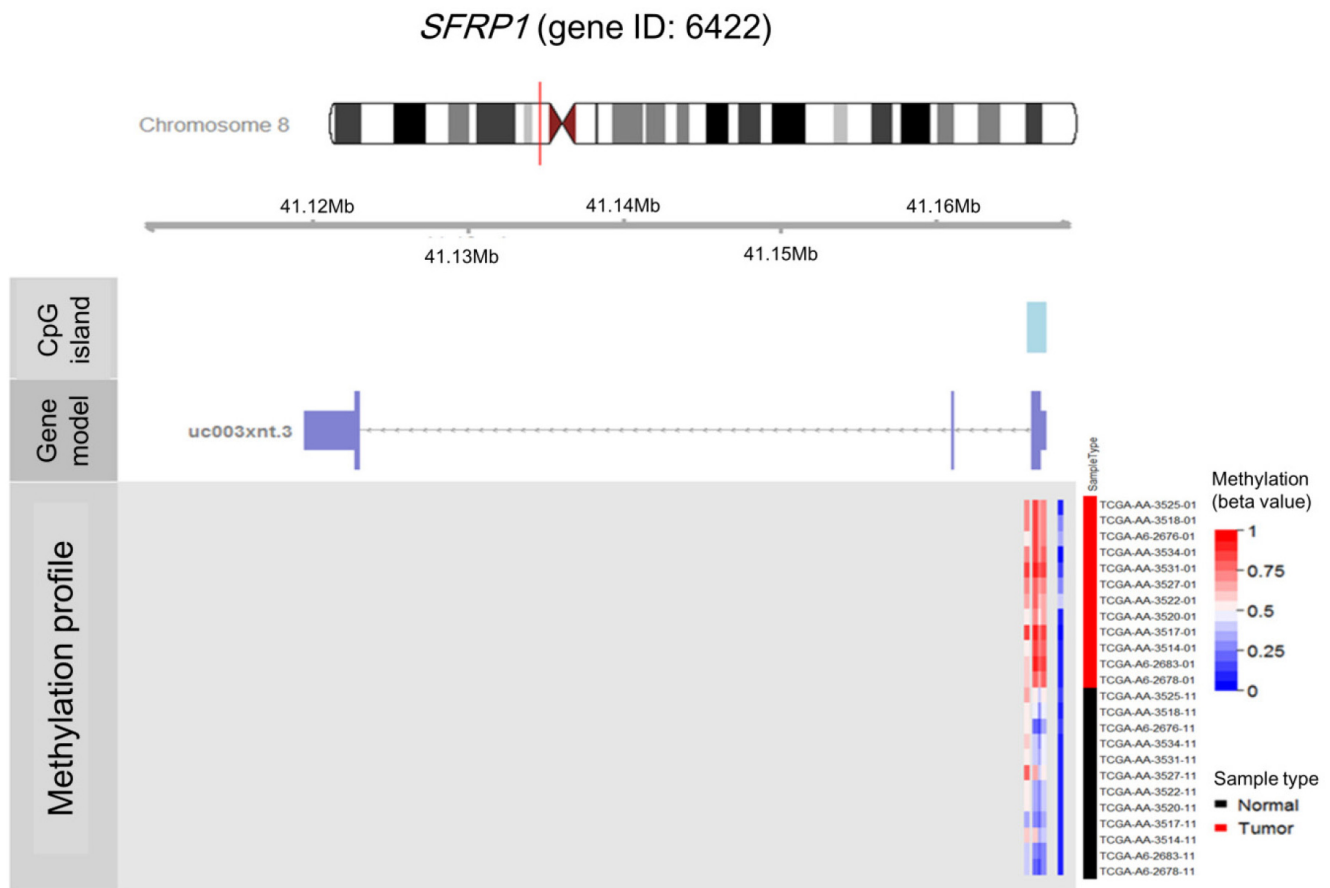
Fig. 7. Methylation heat map of four *SFRP1* probes. *SFRP1* is marked by a vertical red line in the ideogram of chromosome 8, while tracks of CpG islands and exon-intron structures are shown. The heat map in methylation profile track shows the methylation beta-values of the four probes (cg06166767, cg13398291, cg15839448, and cg22418909 from left to right). Each row of the heat map represents one of the samples, which are ordered as shown by the The Cancer Genome Atlas (TCGA) sample codes. The last two digits of the TCGA sample codes represent sample types: "01" for tumor sample and "11" for matched normal.

Gene ontology and pathway analysis

Gene ontology analysis using DAVID revealed that the 31 genes were enriched in the Wnt receptor signaling pathway (Fisher exact = 9.6E-5); 4 genes—*RSPO2, RSPO3, SFRP1,* and *SFRP2*—belonged to this pathway (Table 2). Site of Frizzled proteins (SFRP) genes interact directly with Wnt, the ligand of the Wnt signaling pathway Frizzled (Fz) receptor, inhibiting the binding of Wnt to Fz receptor [10].

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway diagram (http://www.genome.jp/kegg/) also confirmed that Wnt and SFRP regulate the Wnt signaling pathway at the upstream level (Fig. 3).

Wnt signaling pathway-related genes in colorectal cancer

Both the *SFRP1* and *SFRP2* genes are known to show DNA hypermethylation in colorectal cancers; their relatively low expression causes upregulation of Wnt signaling and thus tumor cell proliferation [11]. Wnt2, a member of the Wnt gene family, and *SFRP1* showed drastically opposite gene expression patterns in the discovery dataset (Fig. 4). Given the premise that downregulation of *SFRP1* and *SFRP2* due to DNA hypermethylation deteriorates the inhibition of the Wnt ligand signaling pathway, we surveyed the gene expression profile of the pathway genes, including SFRP. The significance of the differential expression of each gene was measured by t test using the discovery dataset. The p-value distributions were compared between the genes involved in Wnt signaling (n = 151) and those that were not involved in Wnt signaling (n = 17,658) using Wilcoxon rank-sum test. The genes involved in Wnt signaling showed more significant differential expression than the others (p = 0.01658). *SFRP1* and *WNT2* were among the top ranked genes of the pathway (Fig. 5).

DMR of *SFRP1* gene in colorectal cancer

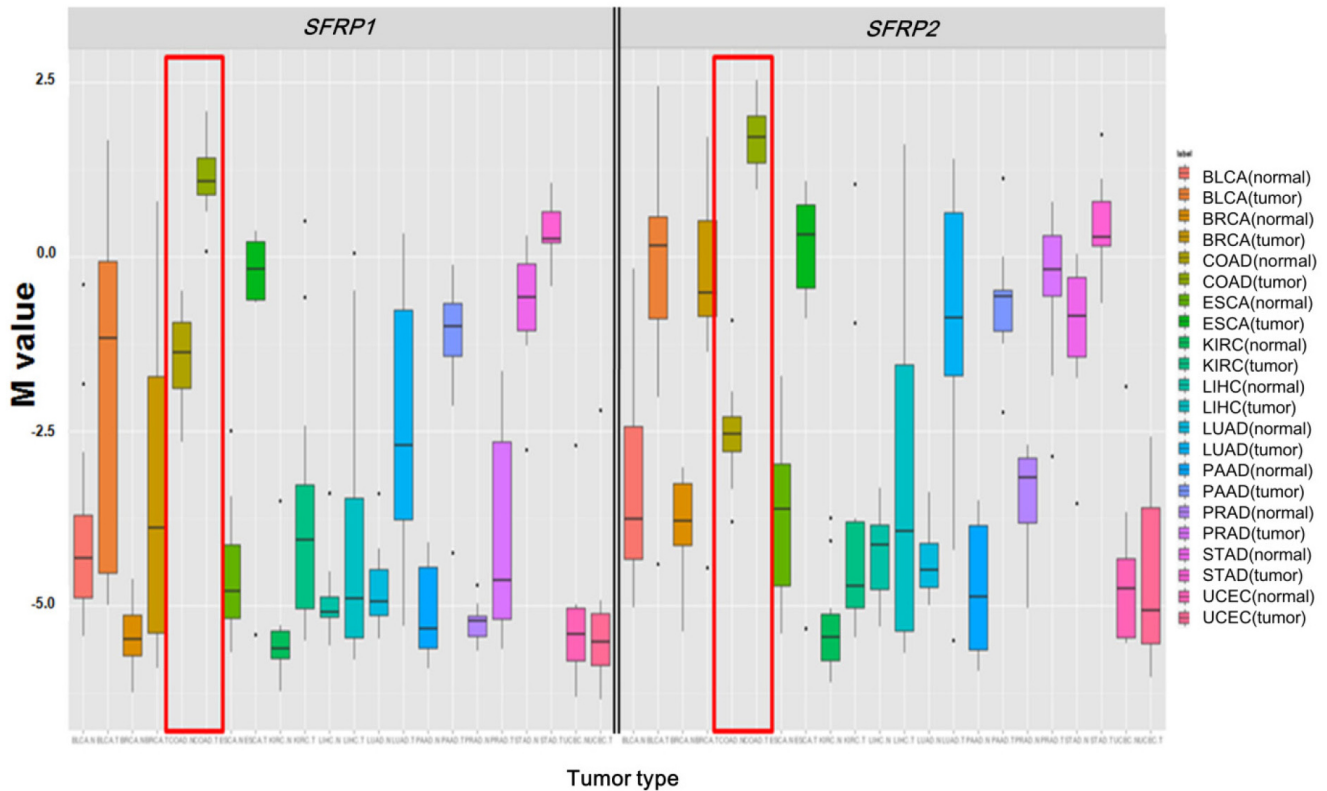Illumina Human DNA Methylation27 and 450 are

Fig. 8. Comparison of *SFRP1* (left panel) and *SFRP2* (right panel) methylation among 11 cancer types. Each cancer type is shown by the boxplot of normal samples, followed by that of tumor samples. Colorectal cancers are enclosed by red boxes.
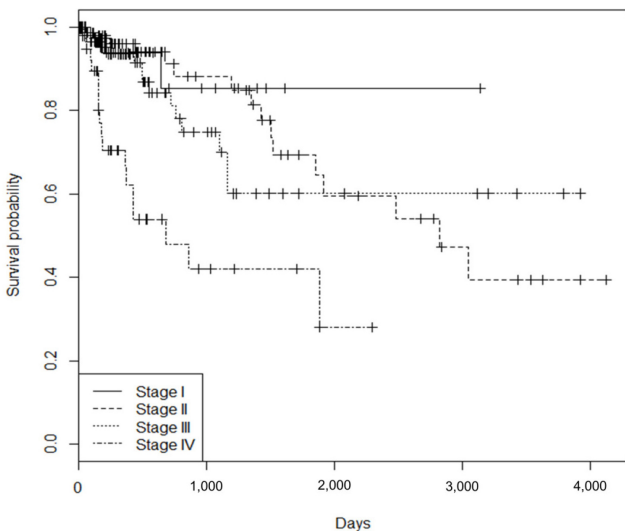

Fig. 9. Survival curves by colorectal cancer stage.

Fig. 10. SFRP gene cutoff points by maximally selected rank statistic. The cutoff points are marked by vertical dotted lines at 0.598 and 0.713 for *SFRP1* (A) and *SFRP2* (B), respectively.

microarray platforms that can measure the methylation values of promoter regions, with 27,578 and 331,182 probes, respectively. On the former platform, there are four probes in the promoter region of *SFRP1*, and their methylation status varies probe to probe, while the latter has 33 additional probes for *SFRP1*. We examined which probes showed
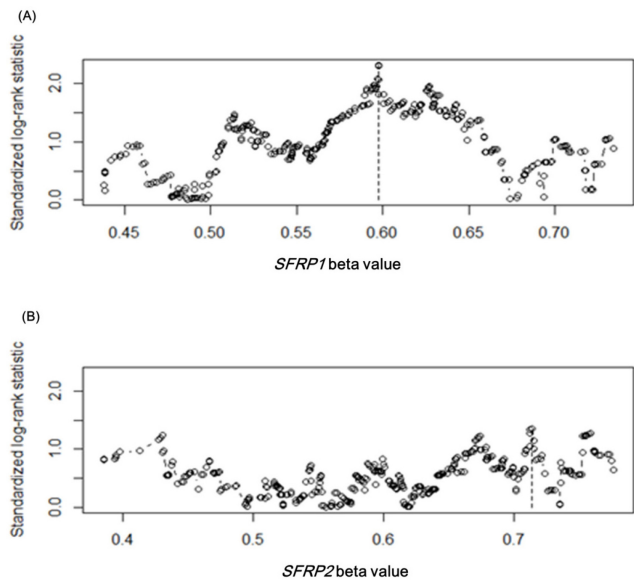
hypermethylation and correlated with expression change using the expanded dataset, for which both data types were available (n = 203). Among the four *SFRP1* probes that were
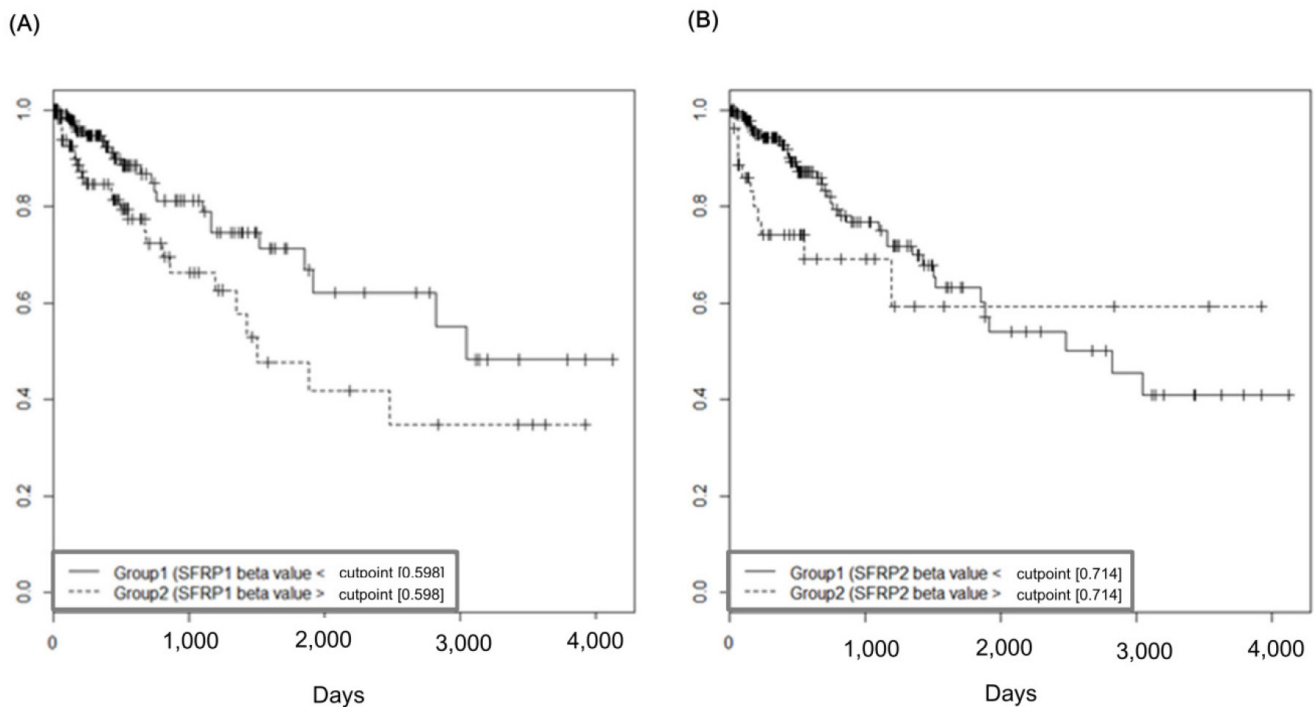
Fig. 11. Survival curves of *SFRP1* (A) and *SFRP2* (B) by methylation beta-values. X-axis represents survival time in days.

Table 4. Hazard ratio of colorectal patients by tumor stage and SFRP methylation status

	No. of patients (%)	HR (95% CI)	p-value[a]
Tumor stage			
II	167 (39.48)	1.536 (0.448–5.275)	0.490
III	123 (29.07)	1.828 (0.528–6.326)	0.340
IV	60 (14.18)	6.083 (1.778–20.812)	0.004
SFRP1 methylation higher group	186 (43.97)	1.93 (1.119–3.329)	0.018
SFRP2 methylation higher group	91 (21.51)	1.703 (0.892–3.25)	0.107

HR, hazard ratio; CI, confidence interval.
[a]Cox Proportional-Hazards Regression.

shared between the two platforms, one of them (cg06166767, chr8:41287005) did not show hypermethylation, while the other three showed similar hypermethylation (Fig. 6). Table 3 shows the correlation between differential methylation and differential expression, which was measured using linear regression, revealing a significant correlation of cg15839448 (chr8:41285687) (p = 0.1E-9). R/Bioconductor package methy Analysis was used to depict the chromosomal location in relation to CGI, along with the methylation beta-value, as a heat map (Fig. 7), confirming the variation among probes. Aberrant methylation patterns of SFRP have been reported not only in colorectal cancers but also in gastric cancers [12], breast cancers [10], and pancreatic cancers [13]. From TCGA, we downloaded 12 matched pairs for each cancer of the following organs: bladder, breast, esophageal, kidney, liver, lung, pancreas, prostate, stomach, and uterine corpus endometrium. Except for liver and endometrial cancers, the DNA methylation levels of both *SFRP1* and *SFRP2* were consistently higher in cancers than in the normal counterparts (Fig. 8). It should be noted that in colorectal cancers, the mean methylation M-value of the tumor samples was greater than 0. While stomach cancers (STAD) also showed a mean M of greater than 0, the difference between the tumor and normal sample was not as large as in colorectal cancers. In all other cancers, the mean M was below 0, implying marginal hypermethylation at most.

Survival analysis: SFRP methylation in association with clinical factors

It is well known that the survival rate of colorectal cancer
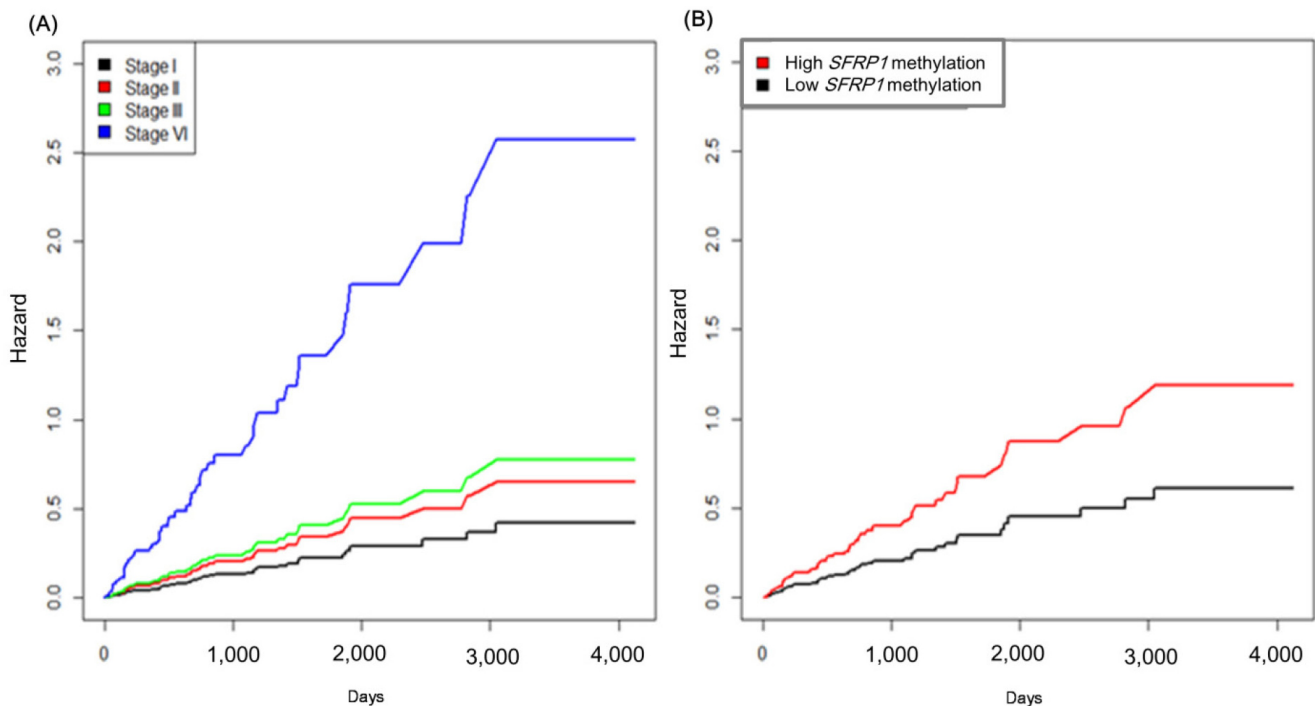
Fig. 12. Cumulative hazard curve of colorectal cancer by stage (A) and *SFRP1* methylation (B) values. X-axis represents survival time in days.

is negatively correlated with the tumor stage: the higher the stage, the poorer the survival. We confirmed this trend in the expanded dataset (n = 419). Stages IIA and IIB were merged into stage II, and similarly, stages IIIA, IIIB, and IIIC were merged into stage III. The Kaplan-Meier survival curves showed, as expected, lower survival for higher stages (Fig. 9). The Cox regression analyses also indicated a significant association of stage IV with survival rate (mean hazard ratio 6.083) (95% confidence interval [CI], 1.778 to 20.812; p = 0.004) (Table 4).

Upon confirming the anticipated survival trends by stage in the expanded dataset, we tested whether the methylation level of *SFRP1* correlated with the survival rate. Among the four methylation probes, cg15839448 was chosen, as its methylation level showed the most consistent correlation with the expression level of *SFRP1*. In order to separate the samples into two groups that showed a survival difference according to *SFRP1* methylation beta-values, we used maximally selected rank statistics, implemented in R (the maxstat package), yielding the cutoff point (beta-value = 0.598) (Fig. 10A). Similarly, cg23207990 was chosen for *SFRP2*, yielding the cutoff point (beta-value = 0.713) (Fig. 10B).

The group with highly methylated *SFRP1* showed poorer survival than the other group (p = 0.0181) (Fig. 11A). On the other hand, *SFRP2* methylation did not exhibit a significant relationship (p = 0.107) (Fig. 11B). The corresponding

hazard ratios from the Cox regression analysis were 1.93 (95% CI, 1.119 to 3.329) and 1.703 (95% CI, 0.892 to 3.25) for *SFRP1* and *SFRP2*, respectively (Table 4). As higher tumor stage shows a higher risk rate in the cumulative hazard curve (Fig. 12A), higher *SFRP1* methylation also increases the risk rate (Fig. 12B).

In conclusion, hypermethylation of SFRP genes has been linked to the downregulation of their expression in colorectal cancers [11]. Here, we confirm it using independent TCGA datasets. The evidence was stronger for *SFRP1* but only marginal for *SFRP2*. We also demonstrate that promoter hypermethylation of *SFRP1* is linked to poor survival of colorectal cancer patients.

Acknowledgments

References

1. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, *et al*. Identification and

analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799-816.

2. Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;358: 1148-1159.

3. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature* 1986;321:209-213.

4. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* 2010;19:698-711.

5. Foster SA, Wong DJ, Barrett MT, Galloway DA. Inactivation of p16 in human mammary epithelial cells by CpG island methylation. *Mol Cell Biol* 1998;18:1793-1801.

6. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet* 2001;10:687-692.

7. Issa JP. DNA methylation as a therapeutic target in cancer. *Clin Cancer Res* 2007;13:1634-1637.

8. Levenson VV, Melnikov AA. DNA methylation as clinically useful biomarkers-light at the end of the tunnel. *Pharmaceuticals (Basel)* 2012;5:94-113.

9. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, *et al*. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;11:587.

10. Veeck J, Niederacher D, An H, Klopocki E, Wiesmann F, Betz B, *et al*. Aberrant methylation of the Wnt antagonist *SFRP1* in breast cancer is associated with unfavourable prognosis. *Oncogene* 2006;25:3479-3488.

11. Qi J, Zhu YQ, Luo J, Tao WH. Hypermethylation and expression regulation of secreted frizzled-related protein genes in colorectal tumor. *World J Gastroenterol* 2006;12:7113-7117.

12. Kinoshita T, Nomoto S, Kodera Y, Koike M, Fujiwara M, Nakao A. Decreased expression and aberrant hypermethylation of the SFRP genes in human gastric cancer. *Hepatogastroenterology* 2011;58:1051-1056.

13. Bu XM, Zhao CH, Zhang N, Gao F, Lin S, Dai XW. Hypermethylation and aberrant expression of secreted frizzled-related protein genes in pancreatic cancer. *World J Gastroenterol* 2008;14:3421-3424.