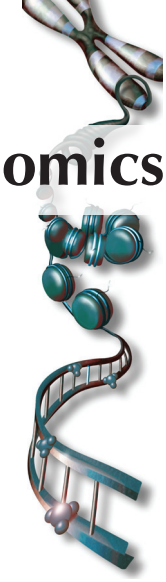
Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics



Epigenome-wide analysis of DNA methylation reveals a rectal cancer-specific epigenomic signature

Aim: The aim of the present study is to address a genome-wide search for novel methylation biomarkers in the rectal cancer (RC), as only scarce information on methylation profile is available. **Materials & methods:** We analyzed methylation status in 25 pairs of RC and adjacent healthy mucosa using the Illumina Human Methylation 450 BeadChip. **Results:** We found significantly aberrant methylation in 33 genes. After validation of our results by pyrosequencing, we found a good agreement with our findings. The *BPIL3* and *HBBP1* genes resulted hypomethylated in RC, whereas *TIFPI2*, *ADHFE1*, *FLI1* and *TLX1* were hypermethylated. An external validation by TCGA datasets confirmed the results. **Conclusion:** Our study, with external validation, has demonstrated the feasibility of using specific methylated DNA signatures for developing biomarkers in RC.

First draft submitted: 20 April 2016; Accepted for publication: 28 June 2016; Published online: 16 August 2016

Keywords: DNA methylation • Illumina Human Methylation 450 BeadChip • rectal cancer

Veronika Vymetalkova*[,1,2], Pavel Vodicka[1,2,3], Barbara Pardini[4], Fabio Rosa[4], Miroslav Levy[5], Michaela Schneiderova[6], Vaclav Liska[3,7], Ludmila Vodickova[1,2,3], Torbjörn K Nilsson[8] & Sanja A Farkas[9]

[1]Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic
[2]Institute of Biology & Medical Genetics, 1st Medical Faculty, Charles University, Prague, Czech Republic
[3]Biomedical Centre, Faculty of Medicine in Pilsen, Charles University, Czech Republic
[4]Human Genetics Foundation, (HuGeF), Torino, Italy
[5]Department of Surgery, 1st Faculty of Medicine, Charles University & Thomayer Hospital, Prague, Czech Republic
[6]Department of Surgery, General University Hospital, Prague, Czech Republic
[7]Department of Surgery, Teaching Hospital & Medical School in Pilsen, Charles University, Pilsen, Czech Republic
[8]Department of Medical Biosciences/Clinical Chemistry, Umeå University, Umeå, Sweden
[9]Department of Laboratory Medicine, Örebro University; Örebro, Sweden
*Author for correspondence:
Tel.: +420 296 2251
vpolakova@biomed.cas.cz

Background

Colorectal cancer (CRC) is the third most common cause of cancer and the second leading cause of cancer death in Europe and the USA [1]. From a clinical point of view, malignancies in the colon (CC) and the rectum (RC, comprising approximately 33%) represent two distinct entities that require different treatment strategies. The distinction between the CC and RC is largely anatomical but it impacts both surgical and radiotherapeutic management with often different prognoses [2]. In contrast with CC, which has a low incidence of local recurrence and longer survival time, patients with RC have a higher incidence of recurrence requiring the addition of pelvic radiation therapy (chemoradiation) [3,4]. As a consequence, the clinical management of patients with RC differs significantly from that of the CC in terms of surgical technique, the more frequent use of radiotherapy and method of chemotherapy administration [5].

There are some examples of studies that tried to clarify whether established CRC risk factors may or may not be risk factors for CC or RC separately [6]. For example, physical inactivity and body mass index have been associated with CC cancer but not with rectal cancer [7]. However, for RC only, very limited data are available, since existing studies usually failed to separate these entities.

From a molecular point of view, the prevalence of *K-ras* mutations and mutation patterns in the *TP53* gene in RC differs from those seen in CC [8]. For all these reasons, RC and CC should preferably be analyzed separately to reduce the attenuation of risk estimates for RC in the studies.

At present, it is generally assumed that CRC arises as a consequence of an accumulation of genetic and epigenetic alterations, which transforms colonic epithelial cells into adenocarcinoma cells. DNA methylation is an epigenetic event that alters gene expression and that may lead to cancer and other human dis-

Future Medicine part of fsg

eases. Global DNA hypomethylation is a type of altered DNA methylation which often occurs in repetitive elements of the genome such as long interspersed repeat sequences (LINE-1). Global DNA hypomethylation is associated with genomic instability and chromosome abnormalities [9,10]. Gene-specific methylation occurs at specific regions of the gene such as gene promoters and can either silence or activate the gene. It is generally accepted dogma that CpG island hypermethylation leads to transcriptional gene silencing [11].

The epigenetic changes associated with CRC, especially aberrant CpG island methylation in the promoter regions of tumor suppressor genes, have been tested in several studies [2]. In general, up to 10% of CpG islands in cancer epigenomes may be aberrantly methylated, which can lead to the silencing of thousands of gene promoters in the average cancer [12]. Moreover, CRC-associated aberrant methylation is not exclusively limited to CpG islands but may comprise 'CpG island shores' or areas that are less dense in CpG dinucleotides within 2 kb upstream of a CpG island [13,14]. Methylation of the CpG island shores may also be associated with the transcriptional inactivation and expression of splice variants [15].

Table 1. Patient's clinical characteristics.

Characteristic	n (total = 32)
Gender:	
– Male	20
– Female	12
Age mean:	65.8
– Male	68.4
– Female	63.1
CIN status:	
– MSS	32
– MSI	0
Tumor differentiation:	
– Low	10
– Moderate	19
– High	3
TNM stage:	
– Stage I	8
– Stage II	10
– Stage III	8
– Stage IV	6
Therapy:	
– Neoadjuvant therapy	5
– Adjuvant therapy	9

CIN: Chromosomal instability; MSI: Microsatellite instable; MSS: Microsatellite stable.

In the last decade, technologies for analyses on genome scale have progressed, and new tools have been implemented to characterize the full spectrum of molecular heterogeneity in many types of cancer cells [16]. With respect to CRC, genome-wide methylation changes have been identified in the past few years [13,15,17].

To our knowledge, no published studies focused on epigenetic diversity in RC by using a state-of-the-art high-density methylation array. In the present study, CpG-level methylation of tumor and matched adjacent tissues from RC patients were analyzed using Infinium HumanMethylation450K BeadChips. This enabled us to characterize differentially methylated regions involved in RC pathogenesis and identify novel candidate genes not previously associated with aberrant methylation in RC.

Materials & methods

Clinical samples

The study comprised 64 paired samples (tumor tissue and adjacent nonmalignant rectal mucosa [ANMRM]) from 32 patients with RC. We included only those patients that at the time of the collection, did not receive any adjuvant therapy yet. Clinical characteristics of patients are presented in Table 1. The Ethics committees of the Institute for Clinical and Experimental Medicine, the Thomayer Hospital, Prague (C.j. 786/09(G0-04-09), the General University Hospital, Prague (C.j. 12/11 Grant GAČR 1.LF UK) and the Teaching Hospital and Medical School, Pilsen (for project IGA MZCR NT14329) approved the study. All patients signed informed consent. The Ethical Review Board, Uppsala, Sweden approved the Swedish participation in the study.

DNA methylation array

The 64 RC and ANMRM specimens were homogenized by MagNA Lyser (Hoffmann-La Roche, Prague, Czech Republic) and the genomic DNA was extracted using the AllPrep DNA/RNA Isolation Kit (Qiagen, Prague, Czech Republic) protocol according to the manufacturer's instructions. DNA samples were analyzed with the Illumina Infinium HumanMethylation450K BeadChip according to standard laboratory procedures obtained from Illumina, described in details by Sandoval *et al.* [18]. The BeadChips were read by an iScan scanner and the data collection was performed in the GenomeStudio software (version 1.0).

Data processing

CpG sites with bead count less than three in more than 5% of the samples were removed, together with sites where more than 1% of the samples had a detection

p-value > 0.05 were filtered away from the raw data. The data were quantile normalized before the calculation of β-values. Beta mixture quantile dilation was used to eliminate the probe type bias in the Illumina Infinium technology. The combination of Quantile normalization together with beta mixture quantile dilation was recently suggested as the most effective normalization strategy when dealing with Illumina HumanMethylation450 BeadChip data [17]. Genetic variations may affect probe hybridization; therefore, probe filtering for single nucleotide polymorphisms was applied according to the Supplementary Table 1 from Nordlund *et al.* [19]. After filtering, 434,749 CpG sites remained for further analysis. Probes on X and Y chromosomes were removed prior to calculating the Δβ values. The gene body regional analysis included all the CpG sites annotated to the same RefSeq gene (within the boundaries of 5′ and 3′ UTRs) and a new averaged β-value was calculated. When a CpG site was annotated to more than one gene, it was used in the average calculation for all present genes. A CpG locus was considered differentially methylated if the Δβ-value (between tumor and ANMRM tissue samples) ≥ |0.3| and the adjusted p-value < 0.05. A cut-off value of |0.2| represents the detection limit of differential methylation with 99% CI [20].

Array validation with pyrosequencing

Six CpG sites targeted by the array were validated using the Pyrosequencing assay technology. CpG sites of interest were located on *TIFPI2*, *HBBP1*, *ADHFE1*, *BPIL3*, *FLI1* and *TLX1* genes, and 12 sample pairs previously analyzed with the Illumina Infinium HumanMethylation450K BeadChip were analyzed for validation. Five hundred nanograms of DNA were used for the bisulfite treatment using the EZ DNA Methylation Gold kit according to the instructions and eluted in 14 µl elution buffer (Zymo Research, Freiburg, Germany). PCR and sequencing primers were designed using Pyromark assay design software 2.0 (Qiagen). Fifty microliter PCR reaction was performed with the HotStarTaq DNA Polymerase Kit (Qiagen), containing 0.15 µmol/l of each primer, 1.25 units of Taq polymerase, 1.5–2.5 mM $MgCl_2$ and 0.1 mM each of dGTP, dATP, dTTP, dCTP and approximately 40 ng of bisulfite-treated DNA was added as a template. The PCR primers, annealing temperatures and amplicon sizes are shown in Supplementary Table 1. All primers were purchased from [21]. The PCR program was as follows: initial denaturation step of 5 min at 95°C, followed by 50 cycles of 45-s denaturation at 94°C, 45 s of annealing at 54 or 56°C with an extension of 45 s at 72°C and one cycle for 7 min at 72°C. After PCR, the samples were prepared for pyrosequencing using the Vacuum Prep Workstation (Qiagen): 37 µl of the amplicon, 3 µl Streptavidin Sepharose HP Beads (Amersham Biosciences, UK) and 40 µl binding buffer (10 mmol/l Tris-HCl, 2 mol/l NaCl, 1 mmol/l EDTA, 0.1% Tween-20, Milli-Q (18.2 MΩ × cm) water, pH 7.6) were mixed and used in the Vacuum Prep workstation. The biotinylated amplicons were immobilized onto the Streptavidin sepharose beads and then passed through one denaturation and two washing steps using the Vacuum Prep Workstation according to a standard protocol. The amplicons were subsequently transferred to a plate containing sequencing primers (0.4 µmol/l) in 40 µl annealing buffer (20 mmol/l Tris-Acetate, 2 mmol/l Magnesium acetate, pH 7.6). Sequencing was performed using a Pyromark Gold Q96 Reagent Kit and a PSQ 96ID system (Qiagen). The nucleotide addition order was optimized by the Pyro Q-CpG software version 1.0.9 (Qiagen). Results were automatically analyzed using the same software.

External validation in TCGA rectal tumor & normal samples

Results from methylation profiling on RC patients were compared with an open access dataset of RC individuals from the cancer genome atlas (TCGA) project. Level 3 genomic data from RC publicly available in TCGA were employed as validation test set. TCGA DNA methylation data were generated using the Illumina Infinium Meth450K platform and presented as β values, with 0 indicating 0% DNA methylation and β values of 1 indicating 100% DNA methylation [2]. Methylation data on 485,577 CpG sites from 98 rectal tumors and seven ANMRM rectal tissue samples were obtained from the TCGA Data Portal. To validate our results, we also investigated RNA expression data for RC patients (RNAseq. Level 2 data available at the time of manuscript preparation) by the Wanderer tool [22], an interactive viewer to explore DNA methylation and gene expression data in TCGA. This set comprises 91 RC patients and nine ANMRM. The overlap of tissues for methylation and expression data was the following: 89 RC patients and two ANMRM. The unadjusted p-values < 0.05 were considered as statistically significant, since these analyses were hypothesis driven.

Gene functional classification

We used the functional annotation tool available in the Database for Annotation, Visualization and Integrated Discovery to identify the gene ontology (GO) terms that were over-represented in the list of hyper- and hypomethylated genes. To narrow the number of genes in the list, we selected genes with differentially methylated CpG sites in the 5′UTR and upstream

gene region: in total, 806 hypermethylated genes and 302 hypomethylated genes were included.

Statistical analysis

All statistical analyses on data measured with the DNA methylation array were performed using the R-package environment [23]. Color balance adjustment and background correction were performed using the methylumi package available from the Bioconductor project [24]. Filtering was performed using the 'pfilter' function with default settings from the R-package wateRmelon, also available from the Bioconductor project. The data were quantile normalized, using the 'nanet' method from the same R-package, prior to calculating the p-values for difference in DNA methylation using the empirical Bayes' moderated t-statistic [25].

The Benjamini–Hochberg method was used to adjust the p-values for multiple testing [26]. A simple t-test was used for TCGA 3 level methylation data.

Results

Data preprocessing

We have analyzed 64 tissue samples using the Illumina Methylation array (485,577 CpG sites investigated). Two samples failed to meet the criteria of the probe call rate and were excluded from further analysis along with their paired specimen. Due to the preoperative radiotherapy, rectal tumors may contain fairly variable number of cancer cells. For this reason, we initially performed a paired chromosomal analysis displaying DNA methylation changes $\Delta\beta \geq |0.3|$ and excluded the sample pairs 5, 11, 12, 18 and 29 with no changes in DNA methylation of this magnitude, Supplementary Figure 1. The inclusion of patients with $\Delta\beta \leq |0.3|$ is in fact only likely to dilute the solid findings, not adding new, extra findings to the results. The exclusion of five sample pairs (<10% of patients) with $\Delta\beta$-value $\leq |0.3|$ was, therefore, mandatory according to the design. In total, seven sample pairs were removed.

Differentially methylated genes

In the principal component analysis based on the differential DNA methylation of the CpG loci there was a marked separation among RC and ANMRM (Figure 1). Applying the criterion $\Delta\beta \geq 0.3$, we found 5929 CpG sites differentially methylated in RC with the majority of them (4350) located within a gene (Figure 2A & B). Of these last, 3527 were hypermethylated and 823 hypomethylated. These CpG sites mapped to 1192 different genes, mostly on chromosomes 1, 4, 6–8 and 13 (Figure 2A & B). Hypermethylation was predominantly observed in CpG islands. On the other hand, hypomethylation was predominantly found in intergenic regions, so called open sea (Figure 2C & D). We used Database for Annotation, Visualization and Integrated Discovery to identify GO terms enriched in our list of differentially methylated genes. The hypomethylated genes were involved in processes such as receptor and membrane activity and the hypermethylated genes in multicellular organism development, neurogenesis or regulation of cellular processes (Table 2).

When performing an unguided gene regional analysis (including CpG sites from 5′ and 3′UTRs), we found 33 genes differentially methylated in RC. The specimens were divided into two clusters representing on the left, the ANMRM tissues and on the right, the RC (Figure 3): only two patients, RC9 and RC10 were misclassified.

We identified those genes with the highest quantitative differences in methylation between RC and ANMRM. The unguided analysis was based primarily on the analyzed $\Delta\beta$-values and CpG sites with an adjusted p-value < 0.05. The differential methylation for the top ten hypo- and hypermethylated CpG sites and genes ranged between -0.47 and -0.53, and from 0.55 to 0.62, respectively (Table 3). The majority of these CpG sites were located within gene body regions. Particularly, the most hypermethylated CpG sites in *ADHFE1*, *TFPI2* and *FLI1* genes exhibited a large number of significantly aberrantly methylated CpG sites in cancer tissue: 19 CpG sites out of 23 analyzed covering *ADHFE1* gene; 23 out of 25 covering *TFPI2* and 39 out of 47 in *FLI1*. RC tissues displayed significant hypermethylation in the upstream of the 5′ region of *ADHFE1* (Figure 4) while hypomethylation was found in the gene body. Several CpG sites in this gene (cg01588438, cg20912169, cg09383816 and cg20040765) pointed to regions with considerable difference between tumor and healthy tissues ($\Delta\beta$ fractions of -0.26 and 0.57–0.62).

RC tissues also displayed significant CpGs hypermethylation in the 5′upstream region of the *FLI1* gene; in contrast CpG sites belonging to the so called S_Shore displayed hypomethylation (from cg01681098 till the end of the gene; Figure 4).

Other genes with a similar switch pattern were *ZNF804B* and *ZNF793*, whereas *ZNF385B* or *RASSF2* showed the opposite direction of the switch (Figure 4).

The 3′UTR region of *GPR85* in RC tissues showed a significant hypomethylation while the gene body and 5′UTR region had significant hypermethylation. Interestingly, the region 1500kb upstream of a transcription start site (TSS) again showed hypomethylation of RC (Figure 4).

Validation of the methylation array data by pyrosequencing

The differential methylation of six selected genes was validated by pyrosequencing in a subset of tissues already analyzed with the Illumina BeadChip

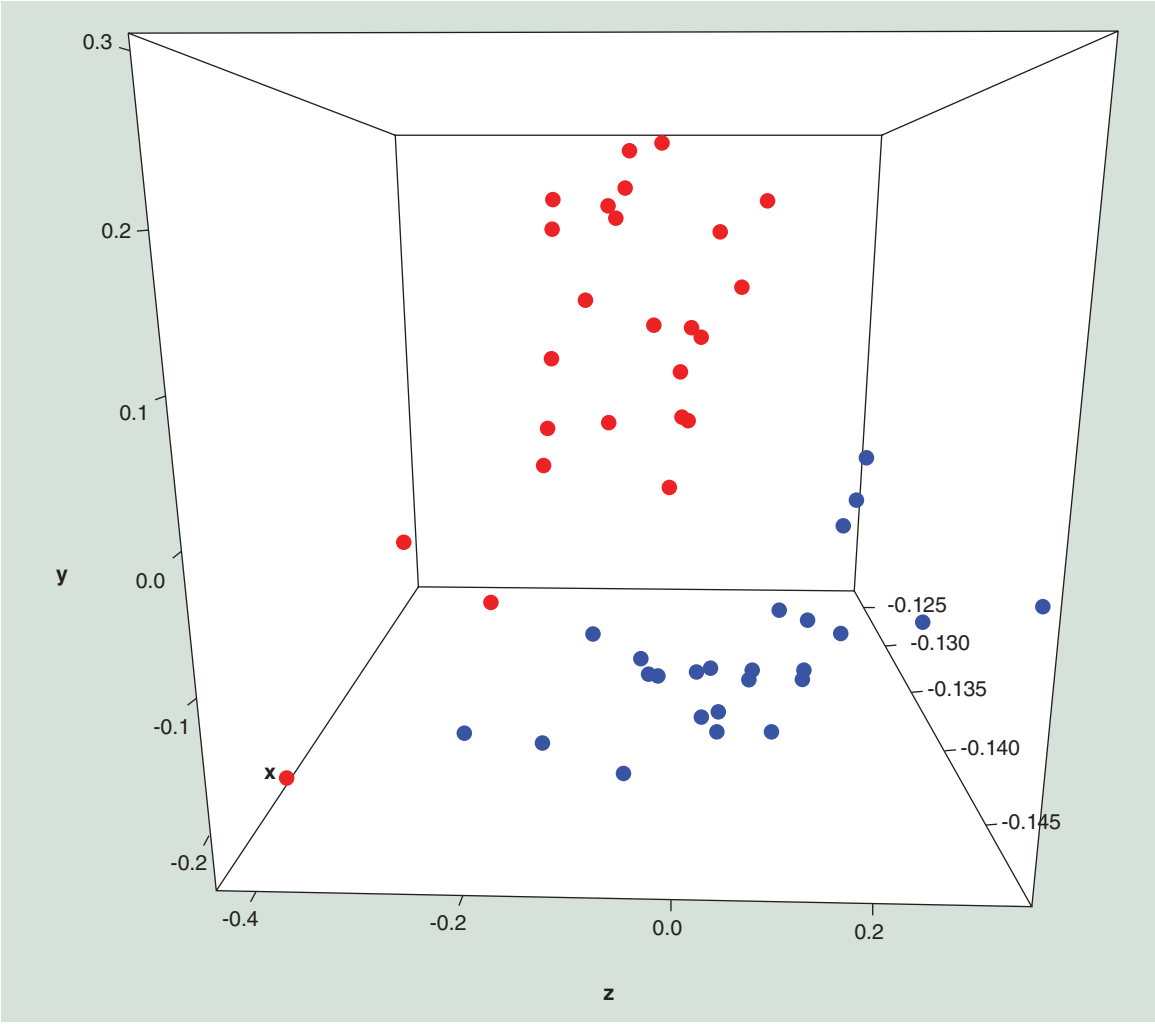


Figure 1. Principal component analysis composed of CpG sites after filtering for cancer. The tumor tissue (red) and adjacent nonmalignant (blue) rectal mucosa specimens differ in DNA methylation.

arrays. Results were in agreement with the array data (Supplementary Figure 2). RC tissues displayed significant hypomethylation in the *BPIL3* and *HBBP1* genes, whereas in *TIFPI2*, *ADHFE1*, *FLI1* and *TLX1* tumors resulted as hypermethylated.

External validation of the DNA methylation & mRNA expression

The most hypo- and hyper-methylated CpGs identified in our study, as well as 33 genes differentially methylated from the unguided hierarchical clustering analysis (see heatmap, Figure 3) were investigated also in the TCGA datasets (98 rectal tumors and seven ANMRM rectal tissue) as an external validation. There was a good agreement in differences in DNA methylation between RC and ANMRM tissues among our and the TCGA datasets (Table 3 & Supplementary Table 2).

For RNA-seq, the TCGA dataset available on Wanderer tool were for 91 rectal tumors and nine ANMRM rectal tissues. When we overlapped methylation data and RNA-seq in Wanderer, the data in common were only for 89 rectal cancers and two ANMRM tissues.

Generally, the genes with hypermethylated CpG sites evinced lower expression levels in tumors when compared with their ANMRM (particularly *ADHFE1*, *FLI1*, *NPY* and *ITGA4*). Higher mRNA levels in tumors were observed for *PRKAR1B*, *SND1* and *TFPI2*. For the genes with hypomethylated CpG sites, only *CYP27A1* showed increased expression levels. On the other side, decreased expression levels were observed for *ITGBL1* and *MYBPC3* genes. The majority of these CpG sites were localized in the gene body or 5′UTR region, none of them were located in the promoter region of genes.

For those genes observed differentially methylated in the heat map, a general lower expression of all tested genes was observed in RC tissues when compared with their ANMRM counterparts (from p = 0.02

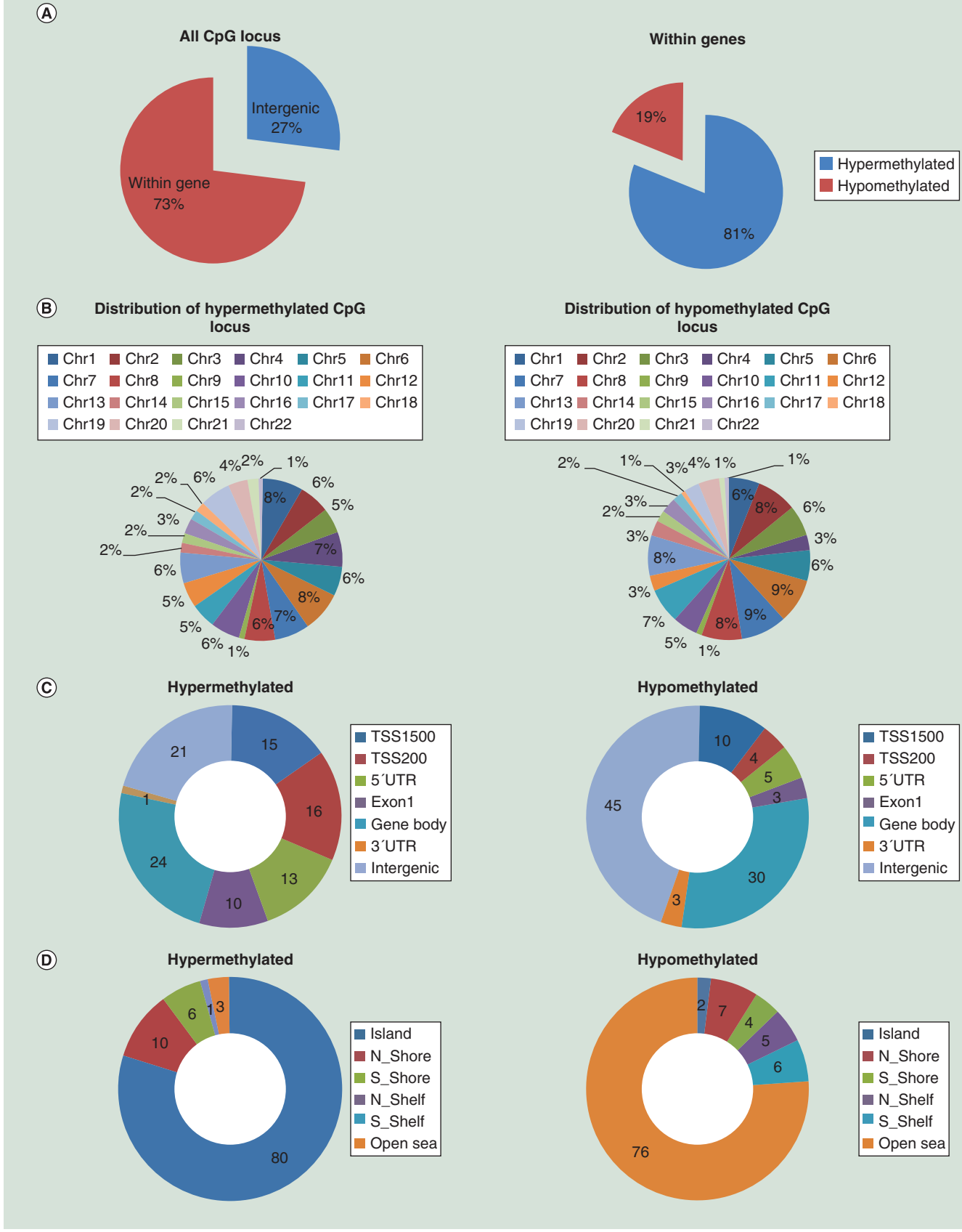

A
All CpG locus
Intergenic 27%
Within gene 73%
Within genes
19%
81%
Hypermethylated
Hypomethylated
B
Distribution of hypermethylated CpG locus
Distribution of hypomethylated CpG locus
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6
Chr7 Chr8 Chr9 Chr10 Chr11 Chr12
Chr13 Chr14 Chr15 Chr16 Chr17 Chr18
Chr19 Chr20 Chr21 Chr22
C
Hypermethylated
Hypomethylated
TSS1500
TSS200
5´UTR
Exon1
Gene body
3´UTR
Intergenic
15 16 13 10 24 1 21
10 4 5 3 30 3 45
D
Hypermethylated
Hypomethylated
Island
N_Shore
S_Shore
N_Shelf
S_Shelf
Open sea
80 10 6 1 3
2 7 4 5 6 76

Figure 2. Genomic distribution of differentially methylated regions (see facing page). (A) Regional distribution of differentially methylated CpG sites in rectal cancer patients. **(B)** Chromosomal distribution of differentially methylated CpG sites in rectal cancer patients. **(C)** The distribution of the hypermethylated and hypomethylated CpG sites over seven gene categories: TSS1500, TSS200, 5′UTR, first exon, gene body, 3′UTR and intergenic regions. **(D)** The distribution of the hypermethylated and hypomethylated CpG sites over CpG islands, shores, shelves and open sea regions. For categorization, the CpG counts were normalized by the number of CpGs in the same category represented on the 450K array. The percentage of normalized CpG counts is indicated in the bars.

to $p < 10^{-7}$). The only exception was for the *TLX1* gene, where significantly higher expression levels were observed in RC in comparison to ANMRM ($p < 10^{-6}$; Supplementary Figure 3). Particularly in the gene body region of *TLX1*, tumor hypermethylation was noticed in our study.

Using the TCGA database, we searched for correlations between methylation and RNA-seq expression data for the most hypo- and hyper-methylated CpGs identified in our study. RNA-seq expression data were available for a subset of RC patients in the TCGA project (89 tumor tissues and two ANMRM). Because of this limited number of ANMRM in TCGA dataset, we were not able to investigate whether changes in DNA methylation observed by us were correlated with respective transcript levels.

Discussion

The design of the present study was guided by the insight of the different histological and molecular features among CC and RC tissues [28]. We focused on RC, which has in general been under-represented in epigenetic CRC studies. In the present study, RC tissues were characterized by differential methylation of a limited subset of genes showing altered methylation throughout all regions of the gene. In addition, the absolute values of the differences (Δβ) were quite large, usually >0.50. These two features highlight that this limited set of genes constitutes a strong epigenetic signature for RC.

From the epidemiological and therapeutic point of view, CC and rectal tumors are considered as different entities [4]. They are functionally different and exposed to fecal matter for different time laps. RC, in particular, is exposed to stool in a more concentrated and direct way compared with CC. Also, as undigested matters pass through the CC, they are coated with alkaline mucus. The different levels of pH in proximal and distal locations may influence susceptibility to environmental factors [29]. In addition, the peculiar microenvironment of the RC could have modulatory effects on tumor behavior in addition to promoter methylation and could also obscure any methylation differences [30].

The recently published data from the TCGA project suggested that the overall patterns of changes in methylation, mRNA and miRNA are indistinguishable between CC and RC [2]. Methylation data obtained from the clinical rectal paired samples studied in the present work may be adequate for use in comparative analyses in other RC methylation studies. The different epigenetic landscapes between adjacent nonmalignant mucosal tissues from the right CC, the left CC and the RC may contribute to determining which genes will show up with the largest Δβ when comparing the tumor and the ANMRM tissues from these different locations. It is the delta value, not the absolute one in the tumor tissues that determine which genes are players in the cancer process.

In our study, we found common RC-specific methylation patterns consisting of 5929 CpGs that were significantly different from those of their healthy counterparts. These CpG sites mapped 1192 different genes, mostly at chromosomes 1, 4, 6–8 and 13. GO and pathway analysis showed a significant enrichment of the genes containing hypermethylated CpG sites that were related to developmental and regulation activities such as regulation of metabolic or biosynthetic processes, of transcription or gene expression. By contrast, the hypomethylated CpG sites were related to signal transduction and receptor activity, suggesting that quite diverse cellular processes may be influenced by methylation events and participate in the development of pathological processes in RC.

There is scarce information on the role of the proposed gene signature in RC pathogenesis. The TCGA portal presents only few healthy tissues available for RC. The present study reports for the first time an equal representation of the tumor tissue and its healthy counterpart making the observed results solid. Moreover, TCGA cases are from an American mixed population. It is known that diet and other lifestyle risk factors may modify global and gene-specific DNA methylation [31–33]. There is evidence also of a differential DNA methylation in various ethnicities and by gender which may be connected with differences in the dietary habits and in socioeconomic conditions [31–33]. Interestingly, the Czech group of cases derives from a relatively small country characterized by a homogeneous population (all Caucasians) with individuals sharing a very similar diet. This is an important issue since the observed differential methylation in tumor tissues is, therefore, independent from important bias such as diet and ethnicity.

Table 2. Gene ontology enrichment of hyper- and hypomethylated CpG sites in the 5′UTR and upstream gene regions.

GO terms	Number of genes (%)	adj. p-value[†]
Hypomethylated genes		
Cluster 1, enrichment score 11.7:		
– GO:0004930~G-protein-coupled receptor activity	44 (15)	<0.001
– GO:0004888~transmembrane receptor activity	51 (18)	<0.001
– GO:0007186~G-protein-coupled receptor protein signaling pathway	48 (17)	<0.001
Cluster 2, enrichment score 11.1:		
– GO:0004984~olfactory receptor activity	29 (10)	<0.001
– GO:0007606~sensory perception of chemical stimulus	31 (11)	<0.001
– GO:0007608~sensory perception of smell	29 (10)	<0.001
Cluster 3, enrichment score 8.8:		
– GO:0004888~transmembrane receptor activity	51 (18)	<0.001
– GO:0004872~receptor activity	57 (20)	<0.001
– GO:0004871~signal transducer activity	63 (22)	<0.001
– GO:0060089~molecular transducer activity	63 (22)	<0.001
Hypermethylated genes		
Cluster 1, enrichment score 37.5:		
– GO:0007275~multicellular organismal development	279 (33)	<0.001
– GO:0048731~system development	242 (29)	<0.001
– GO:0048856~anatomical structure development	254 (30)	<0.001
– GO:0032502~developmental process	291 (35)	<0.001
Cluster 2, enrichment score 28.7:		
– GO:0022008~neurogenesis	103 (12)	<0.001
– GO:0048699~generation of neurons	95 (11)	<0.001
– GO:0030182~neuron differentiation	83 (10)	<0.001
Cluster 3, enrichment score 14.6:		
– GO:0031326~regulation of cellular biosynthetic process	232 (28)	<0.001
– GO:0009889~regulation of biosynthetic process	233 (28)	<0.001
– GO:0019219~regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	223 (27)	<0.001
– GO:0051171~regulation of nitrogen compound metabolic process	224 (27)	<0.001
– GO:0080090~regulation of primary metabolic process	243 (29)	<0.001
– GO:0031323~regulation of cellular metabolic process	250 (30)	<0.001
– GO:0019222~regulation of metabolic process	257 (31)	<0.001
– GO:0045449~regulation of transcription	199 (24)	<0.001
– GO:0010556~regulation of macromolecule biosynthetic process	209 (25)	<0.001
– GO:0010468~regulation of gene expression	206 (25)	<0.001
– GO:0060255~regulation of macromolecule metabolic process	224 (27)	<0.001

[†]p-values were adjusted using the Benjamini–Hochberg method.
adj: Adjusted; GO: Gene ontology.

For two CpGs from our panel (cg1588438 and cg9383816), located in the TSS200 region of *ADHFE1*, we observed the same methylation pattern previously described in CC cancer patients only [15]. *ADHFE1* encodes iron-containing alcohol dehydrogenase, an enzyme responsible for the oxidation of 4-hydroxybutyrate in mammals [34]. The hypermethylation of its promoter in CRC has been previously reported [15,35,36].

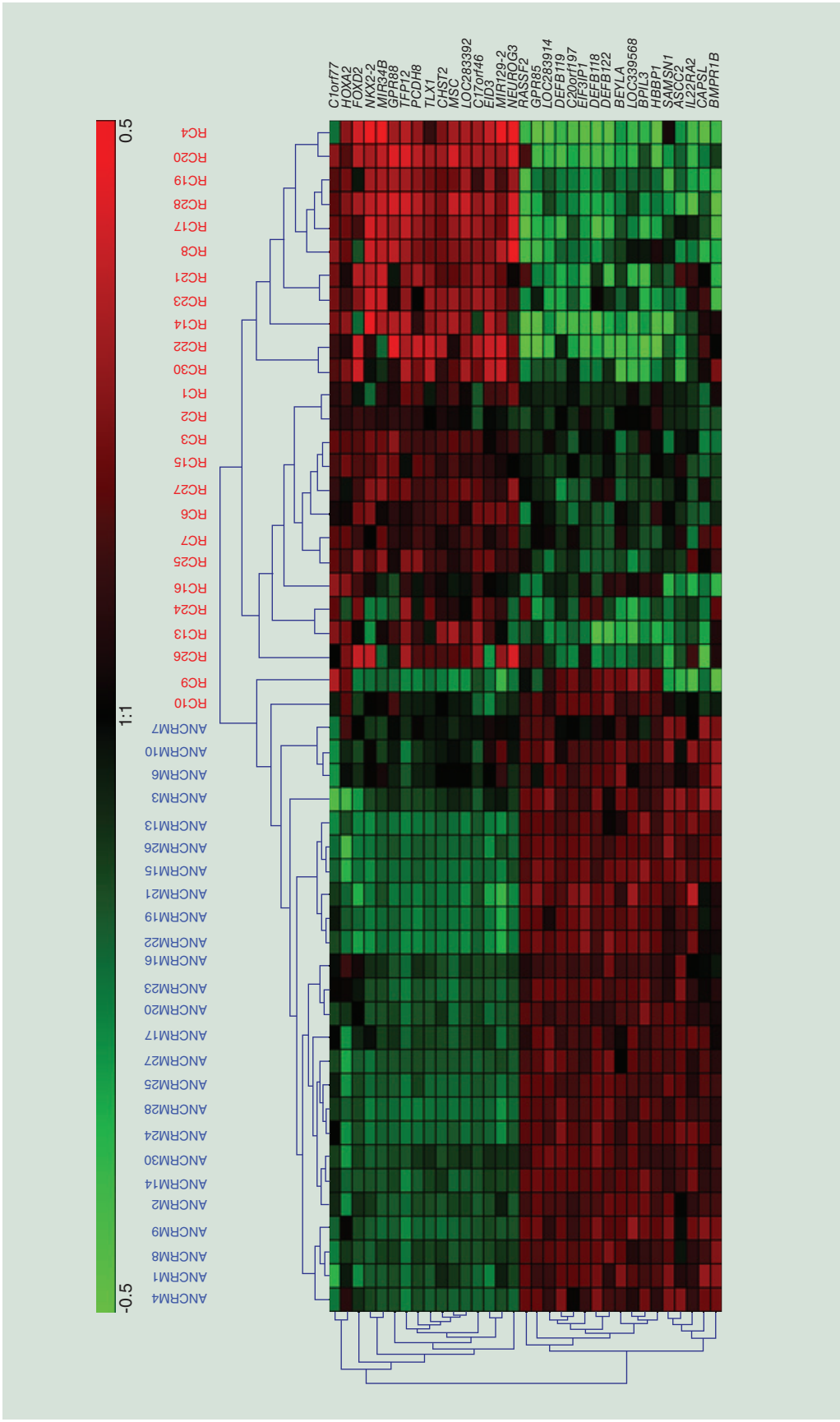
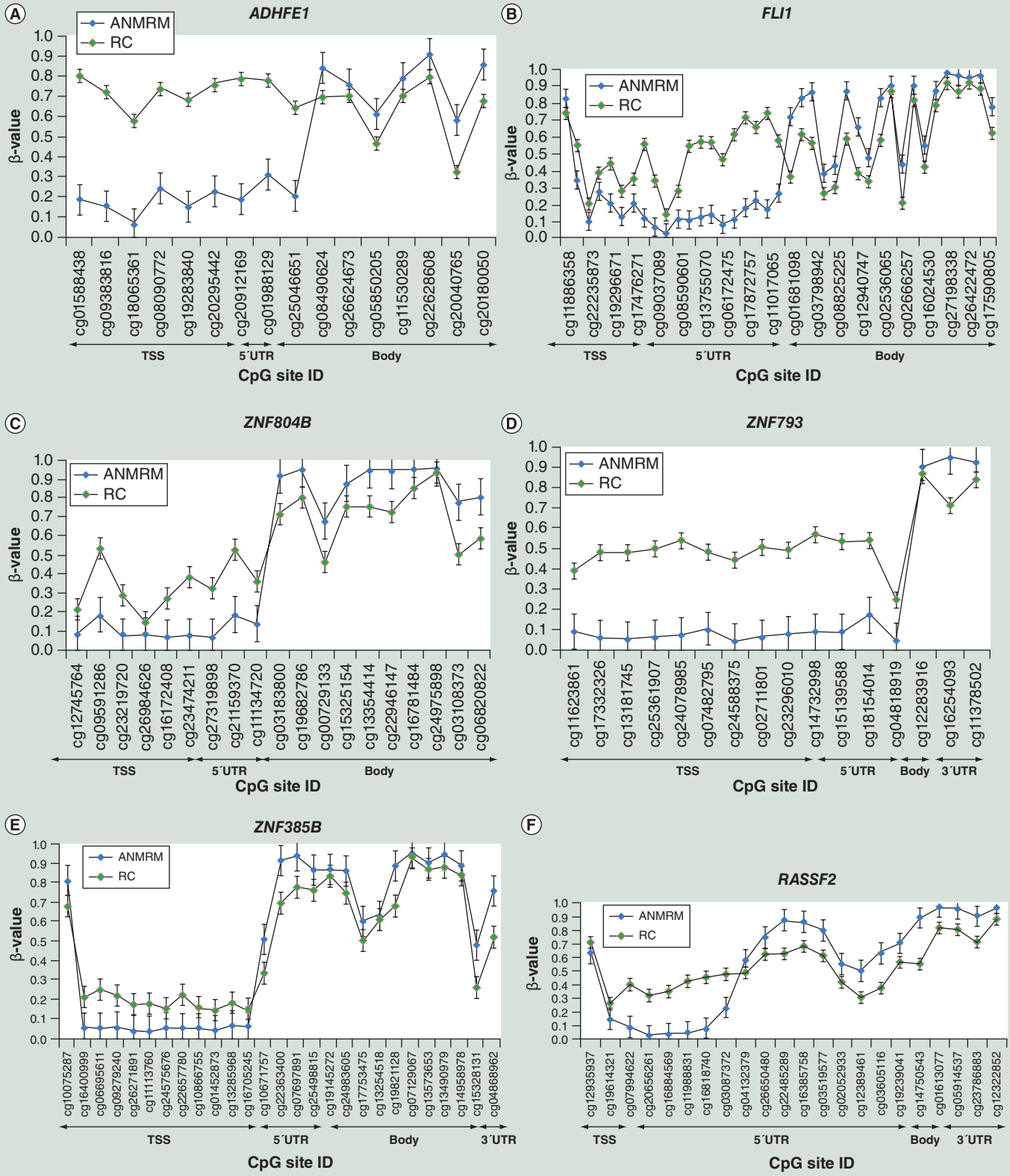


Figure 3. Heatmap showing differentially methylated genes in rectal mucosa (red letters) and adjacent nonmalignant rectal mucosa (blue letters).


A
ADHFE1
ANMRM
RC
β-value
cg01588438
cg09383816
cg18065361
cg08090772
cg19283840
cg20295442
cg20912169
cg01988129
cg25046651
cg08490624
cg26624673
cg05850205
cg11530289
cg22628608
cg20040765
cg20180050
TSS
5´UTR
Body
CpG site ID
B
FLI1
ANMRM
RC
β-value
cg11886358
cg22235873
cg19296671
cg17476271
cg09037089
cg08590601
cg13755070
cg06172475
cg17872757
cg11017065
cg01681098
cg03798942
cg08825225
cg12940747
cg02536065
cg02666257
cg16024530
cg27198338
cg26422472
cg17590805
TSS
5´UTR
Body
CpG site ID
C
ZNF804B
ANMRM
RC
β-value
cg12745764
cg09591286
cg23219720
cg26984626
cg16172408
cg23474211
cg27319898
cg21159370
cg11134720
cg03183800
cg19682786
cg00729133
cg15325154
cg13354414
cg22946147
cg16781484
cg24975898
cg03108373
cg06820822
TSS
5´UTR
Body
CpG site ID
D
ZNF793
ANMRM
RC
β-value
cg11623861
cg17332326
cg13181745
cg25361907
cg24078985
cg07482795
cg24588375
cg02711801
cg23296010
cg14732998
cg15139588
cg18154014
cg04818919
cg12283916
cg16254093
cg11378502
TSS
5´UTR
Body
3´UTR
CpG site ID
E
ZNF385B
ANMRM
RC
β-value
cg10075287
cg16400999
cg06695611
cg09279240
cg26271891
cg11113760
cg24575676
cg22657780
cg10866755
cg01452873
cg13285968
cg16705245
cg10671757
cg22363400
cg07697891
cg25498815
cg19145272
cg24983605
cg17753475
cg13254518
cg19821128
cg07129067
cg13573653
cg13490979
cg14958978
cg15328131
cg04868962
TSS
5´UTR
Body
3´UTR
CpG site ID
F
RASSF2
ANMRM
RC
β-value
cg12935937
cg19614321
cg07994622
cg20656261
cg16884569
cg11988831
cg16818740
cg03087372
cg04132379
cg26650480
cg22485289
cg16385758
cg03519577
cg02052933
cg12389461
cg03605116
cg19239041
cg14750543
cg01613077
cg05914537
cg23786883
cg12322852
TSS
5´UTR
Body
3´UTR
CpG site ID

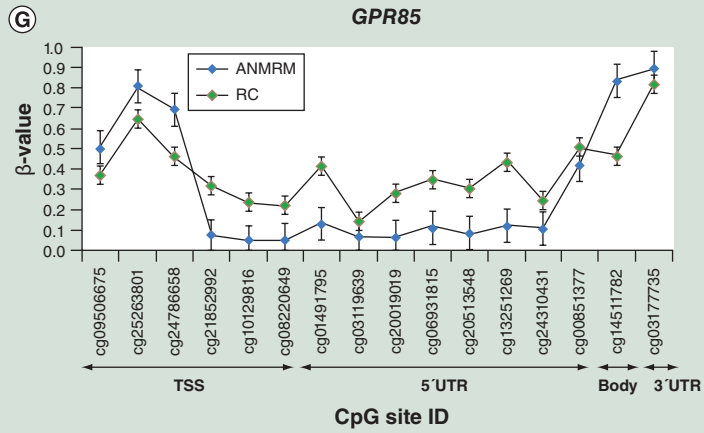


Figure 4. The different DNA methylation profile across the whole gene, covering TSS1500, TSS200, 5′UTR, 1st exon, gene body and 3′UTR regions (see facing page). The DNA methylation profile of the genes: **(A)** *ADHFE1*; **(B)** *FLI1*; **(C)** *ZNF804B*; **(D)** *ZNF793*; **(E)** *ZNF385B*; **(F)** *RASSF2* and **(G)** *GPR85*. The y-axes show the absolute methylated fraction (β-value) of each CpG site. The x-axes show the CpG ID coordinates.

Additionally, a previous study reported an inverse correlation similar to our study between *ADHFE1* methylation and its expression [36]. Overall, the results on *ADHFE1* methylation suggest a role of this gene in both CC and RC pathogenesis but the mechanism involved remains unclear. Kim *et al.* [37] showed that *ADHFE1* transcripts exhibit differentiation-dependent expression during *in vivo* brown and white adipogenesis. In another study, *ADHFE1* was related to bacterial γ-hydroxybutyrate dehydrogenase and resulted having a conserved NAD-binding site [38].

The methylation of *TFPI2* and its loss of expression is a frequent event in human cancers [39,40], including CRC [35,41]. Recently, DNA methylation differences assessed by a targeted DNA microarray in RC tumors also identified *TFPI2* as a potential methylation biomarker in RC [42]. This potential tumor suppressor gene, a Kunitz-type serine proteinase inhibitor, protects the extracellular matrix of cancer cells from degradation and inhibits *in vitro* colony formation and proliferation [43]. A loss of the *TFPI2* function could predispose cells toward a proinvasive program, consistent with an important role of this protein in later stages of carcinogenesis. *TFPI2* belongs to the group of embryonic cell Polycomb group (*PcG*)-marked genes which participate in the 3D structure of nuclear DNA [44] and may target genes with a characteristic 'bivalent' promoter chromatin structure containing both active and repressive histone modifications. Such PcG-marked genes may be predisposed to methylation [45] and may thus be good targets for investigation as early diagnostic biomarkers. This phenomenon was confirmed by Glockner *et al.* [45] by the detection of *TFPI2* methylation in stool DNA.

Concerning *FLI1*, an inverse correlation between the hypermethylation and gene expression was previously observed [36]. This gene has also been found hypermethylated in CRC [46,47]. The *FLI1* gene is a transcriptional activator playing a role in gene expression regulation, which is also expected to be important in cancer development. All these data were from CRC studies and the similarity with our present results may indicate that RC and CC share common features in their pathogenesis. Interestingly, this particular gene resulted as strongly hypermethylated as well as in our previous study on HER2+ breast cancer [14].

A majority of the differentially methylated genes identified by us were not known to be CRC or RC related. These included several genes with large differences in β-values: *PRKAR1B*, *TRBJ2-6*, *HOXA2*, *NKX2-2*, *PCDH8*, *TLX1* and *MIR129-2*. Interestingly, almost all of them presented altered methylation profiles in several cancers, therefore their functional role in RC should be the subject of further research.

Aberrant methylation of the *HOXA2* gene was recently observed in patients with non-small-cell lung carcinoma [48], nasopharyngeal carcinoma [49] and malignant and benign biliary tissues [50]. For *NKX2-2*, a hypomethylation was observed in glioblastoma multiforme [51] and in breast cancer [52] but not in our previous study [14] where we instead found *NKX2-4* and *NKX2-6* to be hypermethylated. *PCDH8* methylation is a frequent event in clear cell renal cell carcinoma [53], bladder [54], breast [14] and gastric cancers [55]. Tommasi *et al.* [56] and Lindqvist *et al.* [14] found a hypermethylated *TLX1* in breast cancer. Finally, altered methylation of *MIR129-2*, a

Table 3. Ten most statistically significantly hypo- or hypermethylated CpG sites and genes according to the β-value identified in this study.

Gene name	Present study				TCGA data				Chr	CpG ID (Location)
	Mean β-value				Mean β-value					
	adj. p-value	Δβ	RC	Normal mucosa	adj. p-value	Δβ	RC	Normal mucosa		
Hypomethylated CpG sites										
MAP3K5	2.1×10^{-15}	-0.53	0.19	0.72	1.4×10^{-16}	-0.45	0.25	0.70	6	7,474,842 Gene body
ITGBL1	3.2×10^{-14}	-0.51	0.24	0.73	2.3×10^{-14}	-0.50	0.16	0.66	13	11,838,152 Gene body
MYBPC3	7.5×10^{-12}	-0.50	0.46	0.96	2.5×10^{-67}	-0.64	0.32	0.98	11	14,642,259 Gene body
adj. to *TRBVB*	2.2×10^{-13}	-0.48	0.30	0.78	1.8×10^{-35}	-0.42	0.23	0.65	7	1,703,205
MAP2K2	1.7×10^{-13}	-0.48	0.31	0.79	6.5×10^{-22}	-0.53	0.21	0.74	19	14,573,876 Gene body
adj. to *Mafb*	2.9×10^{-15}	-0.48	0.15	0.64	6.2×10^{-11}	-0.47	0.12	0.59	16	3,498,081
adj. to *ILα*	5.1×10^{-14}	-0.48	0.36	0.84	5.1×10^{-10}	-0.56	0.24	0.80	2	9,841,889
HPVC1	2.4×10^{-15}	-0.48	0.33	0.81	1.3×10^{-13}	-0.49	0.24	0.73	7	23,860,325 TSS1500
MAP3K5	2.9×10^{-13}	-0.48	0.39	0.87	NA	NA	NA	NA	6	26,680,608 Gene body
SPAG4L	1.4×10^{-11}	-0.47	0.33	0.80	3.1×10^{-35}	-0.59	0.29	0.88	20	2,510,802 1st Exon, 5′UTR
CYP27A1	3.4×10^{-13}	-0.47	0.24	0.71						2,930,667 Gene Body
Hypermethylated CpG sites										
ADHFE1	9.6×10^{-16}	0.62	0.81	0.19	1.2×10^{-16}	0.66	0.75	0.09	8	1,588,438 TSS200
TFPI2	8.8×10^{-16}	0.60	0.75	0.15	NA	NA	NA	NA	7	16,934,178 TSS200
ADHFE1	1.9×10^{-13}	0.60	0.79	0.19	8.8×10^{-38}	0.68	0.72	0.04	8	20,912,169 5′UTR, 1st, Exon
PRKAR1B	9.6×10^{-14}	0.59	0.83	0.23	NA	NA	NA	NA	7	18,601,167 5′UTR, TSS200
TRBJ2–6	3.7×10^{-12}	0.59	0.82	0.23	2.4×10^{-09}	0.61	0.64	0.13	7	9,493,063 CpG Island
PRKAR1B	7.3×10^{-14}	0.58	0.80	0.27	NA	NA	NA	NA	7	13,895,235 5′UTR, TSS200
adj. to *SOX-1*	8.4×10^{-12}	0.57	0.76	0.18	3.7×10^{-16}	0.61	0.69	0.08	13	25,570,913 CpG Island
ADHFE1	2.1×10^{-14}	0.57	0.72	0.16	1.0×10^{-31}	0.62	0.69	0.07	8	9,383,816 TSS200
TFPI2	1.2×10^{-14}	0.56	0.69	0.13	5.3×10^{-07}	0.55	0.70	0.15	7	20,230,721 Gene body
FLI1	4.0×10^{-10}	0.56	0.74	0.18	2.6×10^{-23}	0.63	0.70	0.07	11	11,017,065 Gene body,5′UTR
NPY	9.9×10^{-14}	0.56	0.74	0.21	NA	NA	NA	NA	7	16,964,348 TSS200
OPLAH	1.1×10^{-12}	0.56	0.69	0.16	NA	NA	NA	NA	8	17,698,295 Gene body
ITGA4	9.9×10^{-14}	0.56	0.62	0.09	NA	NA	NA	NA	2	6,952,671 5′UTR;1stExon
SND1	7.3×10^{-15}	0.55	0.67	0.14	NA	NA	NA	NA	7	9,296,001 Gene body

Δβ, delta beta is the value of the differential methylation. Negative Δβ-values reflect hypomethylated status while positive Δβ an hypermethylated one.
The adjusted p-value was considered significant when < 0.05 [27].
adj.: Adjusted; Adj. to: Adjacent to gene from 5′ side; Body: Intragenic CpG sites; CGI: CpG island; CGI shore: Regions 2000 bp away from the CpG island;
Chr: Chromosome number; CpG ID (Location): The coordinate of the CpG location according to the human genome build 37; NA: Data missing; RC: Rectal cancer.

tumor suppressive miRNA frequently methylated in lymphoid but not myeloid malignancies, leads to its reversible silencing [57]. Recently, a strong hypermethylation was observed in HER2+ breast cancer [14].

For the other genes identified in the present study (*FOXD2*, *GPR85*, *GPR88*, *CHST2*, *NEUROG3*, *DEF118*, *DEF118*, *DEF122*, *EIF3IP1*, *BEYLA*, *BPIL3*, *HBB1*, *ASCC2*, *IL22RA2*, *CAPSLB* and *HPR1B*) no data regarding methylation and cancer were available, and further cell biological studies of their functional roles could be potentially rewarding.

Crossvalidation against TCGA methylation dataset showed the reliability and reproducibility of the methylation differences identified in our array study and confirmed the high discriminative and noncoincidental potential of the selected biomarkers.

The majority of differentially methylated CpG sites in our study were located in the gene body region. Intragenic DNA methylation may also affect the transcription from alternate promoters or the transcription of noncoding RNAs [58,59]. CpGs methylation in gene exons is a major cause of C to T transition mutations, leading to cancer causing mutations in somatic cells [60]. The functional role of intragenic DNA methylation needs further validation by expression analyses since the available data are conflicting [19]. For example, promoter methylation is inversely correlated with expression, whereas methylation in the gene body is positively correlated with expression [58]. Thus, in mammals, it is the initiation of transcription but not transcription elongation that seems to be sensitive to DNA methylation silencing. This could be one of the reasons for lack of any inverse correlation of hypermethylated genes with their expression levels. However, the TCGA dataset contained very few paired ANMRM samples to go with the tumor tissues, which detracts from its usefulness to validate methylation/expression correlations.

During early stages of CRC or RC, epigenetic alterations appear to exceed the frequency of genetic mutations, suggesting their greater potential for the next generation of diagnostic biomarkers for the detection of increased risk of cancer transformation.

Our data may further contribute in understanding the role of aberrant methylation and other molecular mechanisms in RC pathogenesis. Collaborative efforts will ultimately result in the employment of epigenetically based approaches to be commonly used to guide RC prevention and treatment. Limiting a study to a well-defined anatomical location such as the RC may reduce the noise levels in the array data studies and thereby increase the rate of successful identification of novel epigenetic biomarkers.

In conclusion, our large and sufficiently powered clinical study with independent external validation has demonstrated the feasibility of using specific methylated DNA signatures for developing putative diagnostic biomarkers in RC.

Future perspective

Our data contribute to improved understanding of the role of gene-specific aberrant methylation in rectal cancer pathogenesis. Treating rectal cancer as an independent entity may improve discovery of biomarkers used for early detection and prognosis. In the future, new biomarker genes will be established and their association with patients' survival will be addressed. The current study contributes to the establishment of such new biomarkers, with the identification of *BPIL3*, *HBBP1*, *TIFPI2*, *ADHFE1*, *FLI1* and *TLX1* genes.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/epi-2016-0044

Acknowledgements

The authors are very thankful to E Van Emburgh for his technical support.

Financial & competing interests disclosure

Methylation profiling was performed by the SNP&SEQ Technology Platform in Uppsala. The platform is part of Science for Life Laboratory at Uppsala University and supported as a national infrastructure by the Swedish Research Council. This work was supported by the Internal Grant Agency of the Czech Ministry of Health (NT 13424, NT/14329-3), Czech Science Foundation (15-27580A; GA15-08239S), COST LD14050 and by the European Regional Development Fund (project number CZ.1.05/2.1.00/03.0076), Lions Cancer Foundation and Nyckelfonden, Örebro läns landstin. This study was also supported by the National Sustainability Program I (NPU I) Nr. LO1503 provided by the Ministry of Education Youth and Sports of the Czech Republic. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Rectal cancer (RC) comprises about a third of all colorectal cancer cases and the location of RC makes it difficult to perform operative resection; therefore irradiation is used as an alternative treatment.
- The DNA methylation profile of 32 pairs of RC and adjacent nonmalignant rectal mucosa's showed that majority of the CpG sites are hypermethylated in RC.
- The *BPIL3* and *HBBP1* genes were hypomethylated in rectal cancer, whereas *TIFPI2*, *ADHFE1*, *FLI1* and *TLX1* genes were hypermethylated.
- Cross-validation against TCGA methylation dataset showed the reliability and reproducibility of the methylation differences identified in our array study and confirmed the high discriminative and noncoincidental potential of the selected biomarkers.

References

1 Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin. Colon Rectal Surg.* 22(4), 191–197 (2009).

2 Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407), 330–337 (2012).

3 Wolpin BM, Meyerhardt JA, Mamon HJ, Mayer RJ. Adjuvant treatment of colorectal cancer. *CA Cancer J. Clin.* 57(3), 168–185 (2007).

4 Minsky BD. Counterpoint: long-course chemoradiation is preferable in the neoadjuvant treatment of rectal cancer. *Semin. Radiat. Oncol.* 21(3), 228–233 (2011).

5 Zlobec I, Vuong T, Hayashi S *et al.* A simple and reproducible scoring system for EGFR in colorectal cancer: application to prognosis and prediction of response to preoperative brachytherapy. *Br. J. Cancer* 96(5), 793–800 (2007).

6 Wei EK, Giovannucci E, Wu K *et al.* Comparison of risk factors for colon and rectal cancer. *Int. J. Cancer* 108(3), 433–442 (2004).

7 Colditz GA, Cannuscio CC, Frazier AL. Physical activity and reduced risk of colon cancer: implications for prevention. *Cancer Causes Control* 8(4), 649–667 (1997).

8 Servomaa K, Kiuru A, Kosma VM, Hirvikoski P, Rytomaa T. *p53* and *K-ras* gene mutations in carcinoma of the rectum among Finnish women. *Mol. Pathol.* 53(1), 24–30 (2000).

9 Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 31(1), 27–36 (2010).

10 Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* 1775(1), 138–162 (2007).

11 Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 21(35), 5427–5440 (2002).

12 Pfeifer GP, Rauch TA. DNA methylation patterns in lung carcinomas. *Semin. Cancer Biol.* 19(3), 181–187 (2009).

13 Farkas SA, Vymetalkova V, Vodickova L, Vodicka P, Nilsson TK. DNA methylation changes in genes frequently mutated in sporadic colorectal cancer and in the DNA repair and Wnt/beta-catenin signaling pathway genes. *Epigenomics* 6(2), 179–191 (2014).

14 Lindqvist BM, Wingren S, Motlagh PB, Nilsson TK. Whole genome DNA methylation signature of HER2-positive breast cancer. *Epigenetics* 9(8), 1149–1162 (2014).

15 Naumov VA, Generozov EV, Zaharjevskaya NB *et al.* Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. *Epigenetics* 8(9), 921–934 (2013).

16 The Cancer Genome Atlas Research Network (2012). http://cancergenome.nih.gov

17 Yousefi P, Huen K, Aguilar Schall R *et al.* Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. *Epigenetics* 8(11), 1141–1152 (2013).

18 Sandoval J, Heyn H, Moran S *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6), 692–702 (2011).

19 Nordlund J, Backlin Cl, Wahlberg P *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* 14(9), r105 (2013).

20 Bibikova M, Barnes B, Tsan C *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* 98(4), 288–295 (2011).

21 Biomers.net. www.biomers.net

22 Wanderer. http://maplab.imppc.org/wanderer/#

23 Wang X, Kang DD, Shen K *et al.* An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics* 28(19), 2534–2536 (2012).

24 Bioconductor. www.bioconductor.org

25 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article3 (2004).

26 Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B (Methodological)* 57(1), 289–300 (1995).

27 The Cancer Genome Atlas: TCGA Data Portal. https://tcga-data.nci.nih.gov/tcga/

28 Jass JR. Gastrointestinal polyposes: clinical, pathological and molecular features. *Gastroenterol. Clin. North Am.* 36(4), 927–946, viii (2007).

29 Iacopetta B. Are there two sides to colorectal cancer? *Int. J. Cancer* 101(5), 403–408 (2002).

30 Beggs AD, Jones A, El-Bahrawy M, Abulafi M, Hodgson SV, Tomlinson IP. Whole-genome methylation analysis of benign and malignant colorectal tumours. *J. Pathol.* 229(5), 697–704 (2013).

31 Zhang FF, Cardarelli R, Carroll J *et al.* Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics* 6(5), 623–629 (2011).

32 Zhang FF, Santella RM, Wolff M, Kappil MA, Markowitz SB, Morabia A. White blood cell global methylation and IL-6 promoter methylation in association with diet and lifestyle risk factors in a cancer-free population. *Epigenetics* 7(6), 606–614 (2012).

33 Delgado-Cruzata L, Zhang W, Mcdonald JA *et al.* Dietary modifications, weight loss, and changes in metabolic markers affect global DNA methylation in Hispanic, African–American, and Afro-Caribbean breast cancer survivors. *J. Nutr.* 145(4), 783–790 (2015).

34 Kardon T, Noel G, Vertommen D, Schaftingen EV. Identification of the gene encoding hydroxyacid-oxoacid transhydrogenase, an enzyme that metabolizes 4-hydroxybutyrate. *FEBS Lett.* 580(9), 2347–2350 (2006).

35 Kim YH, Lee HC, Kim SY *et al.* Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations. *Ann. Surg. Oncol.* 18(8), 2338–2347 (2007).

36 Oster B, Thorsen K, Lamy P *et al.* Identification and validation of highly frequent CpG island hypermethylation

in colorectal adenomas and carcinomas. *Int. J. Cancer* 129(12), 2855–2866 (2011).

37 Kim JY, Tillison KS, Zhou S, Lee JH, Smas CM. Differentiation-dependent expression of Adhfe1 in adipogenesis. *Arch. Biochem. Biophys.* 464(1), 100–111 (2007).

38 Lyon RC, Johnston SM, Panopoulos A, Alzeer S, McGarvie G, Ellis EM. Enzymes involved in the metabolism of gamma-hydroxybutyrate in SH-SY5Y cells: identification of an iron-dependent alcohol dehydrogenase ADHFe1. *Chem. Biol. Interact.* 178(1–3), 283–287 (2009).

39 Hibi K, Goto T, Kitamura YH *et al.* Methylation of the *TFPI2* gene is frequently detected in advanced gastric carcinoma. *Anticancer Res.* 30(10), 4131–4133 (2010).

40 Wang S, Xiao X, Zhou X *et al. TFPI-2* is a putative tumor suppressor gene frequently inactivated by promoter hypermethylation in nasopharyngeal carcinoma. *BMC Cancer* 10, 617 (2010).

41 Ashktorab H, Rahi H, Wansley D *et al.* Toward a comprehensive and systematic methylome signature in colorectal cancers. *Epigenetics* 8(8), 807–815 (2013).

42 Exner R, Pulverer W, Diem M *et al.* Potential of DNA methylation in rectal cancer as diagnostic and prognostic biomarkers. *Br. J. Cancer* 113(7), 1035–1045 (2015).

43 Wong CM, Ng YL, Lee JM *et al.* Tissue factor pathway inhibitor-2 as a frequently silenced tumor suppressor gene in hepatocellular carcinoma. *Hepatology* 45(5), 1129–1138 (2007).

44 Sparmann A, Van Lohuizen M. Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* 6(11), 846–856 (2006).

45 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(11), 4691–4699 (2009).

46 Moon JW, Lee SK, Lee JO *et al.* Identification of novel hypermethylated genes and demethylating effect of vincristine in colorectal cancer. *J. Exp. Clin. Cancer Res.* 33, 4 (2014).

47 Lin PC, Lin JK, Lin CH *et al.* Clinical relevance of plasma DNA methylation in colorectal cancer patients identified by using a genome-wide high-resolution array. *Ann. Surg. Oncol.* 22(Suppl. 3), 1419–1427 (2015).

48 Heller G, Babinsky VN, Ziegler B *et al.* Genome-wide CpG island methylation analyses in non-small cell lung cancer patients. *Carcinogenesis* 34(3), 513–521 (2013).

49 Li HP, Peng CC, Chung IC *et al.* Aberrantly hypermethylated Homeobox A2 derepresses metalloproteinase-9 through TBP and promotes invasion in Nasopharyngeal carcinoma. *Oncotarget* 4(11), 2154–2165 (2013).

50 Shu Y, Wang B, Wang J, Wang JM, Zou SQ. Identification of methylation profile of *HOX* genes in extrahepatic cholangiocarcinoma. *World J. Gastroenterol.* 17(29), 3407–3419 (2011).

51 Chiang JH, Cheng WS, Hood L, Tian Q. An epigenetic biomarker panel for glioblastoma multiforme personalized medicine through DNA methylation analysis of human embryonic stem cell-like signature. *OMICS* 18(5), 310–323 (2014).

52 Kamalakaran S, Varadan V, Giercksky Russnes HE *et al.* DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol. Oncol.* 5(1), 77–92 (2011).

53 Lin YL, Wang YL, Fu XL, Ma JG. Aberrant methylation of *PCDH8* is a potential prognostic biomarker for patients with clear cell renal cell carcinoma. *Med. Sci. Monit.* 20, 2380–2385 (2014).

54 Lin YlL, Ma JH, Luo XL, Guan TY, Li ZG. Clinical significance of protocadherin-8 (*PCDH8*) promoter methylation in bladder cancer. *J. Int. Med. Res.* 41(1), 48–54 (2014).

55 Zhang D, Zhao W, Liao X, Bi T, Li H, Che X. Frequent silencing of protocadherin 8 by promoter methylation, a candidate tumor suppressor for human gastric cancer. *Oncol. Rep.* 28(5), 1785–1791 (2012).

56 Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Res.* 11(1), R14 (2009).

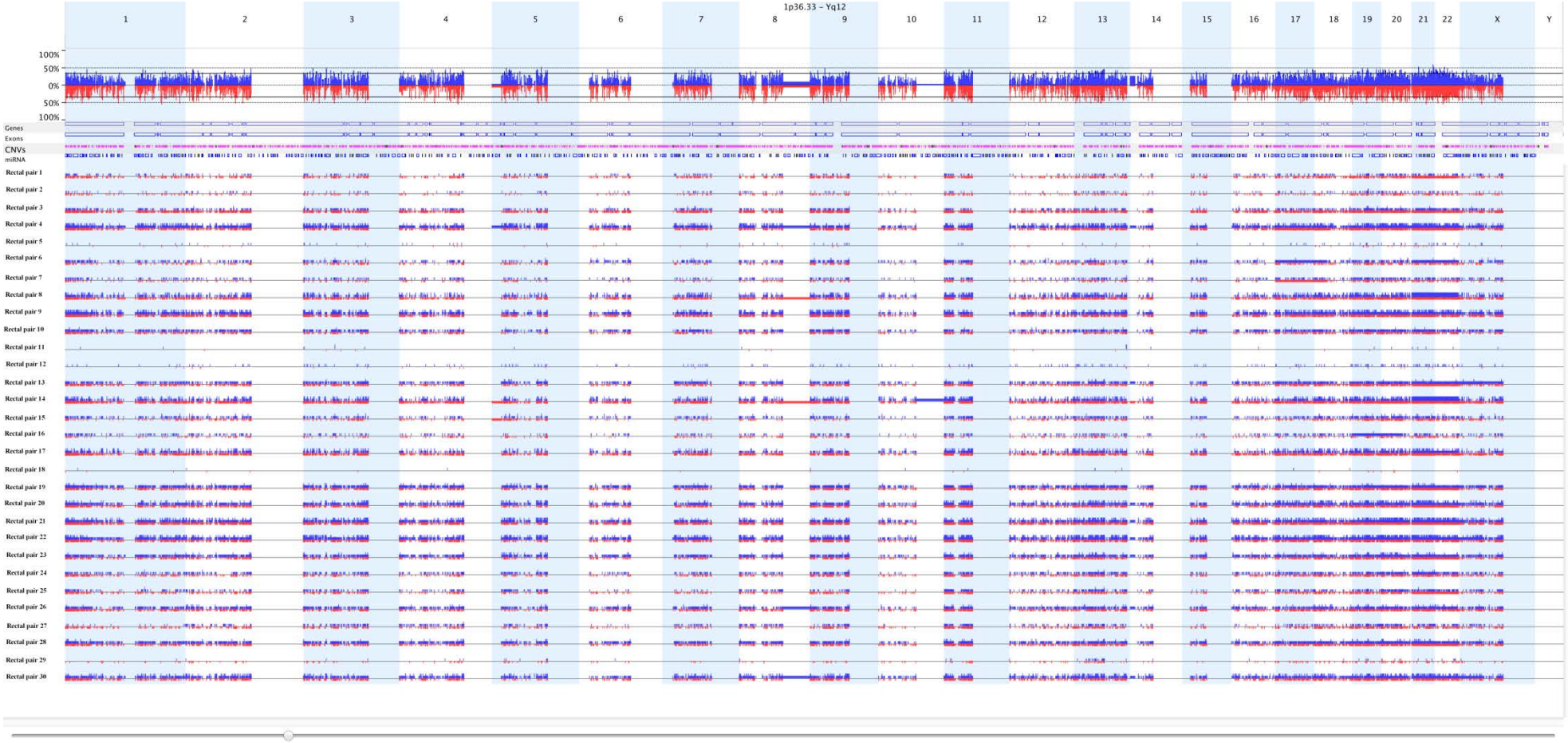
57 Wong KY, Yim RL, Kwong YL *et al.* Epigenetic inactivation of the MIR129-2 in hematological malignancies. *J. Hematol. Oncol.* 6, 16 (2013).

58 Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13(7), 484–492 (2012).

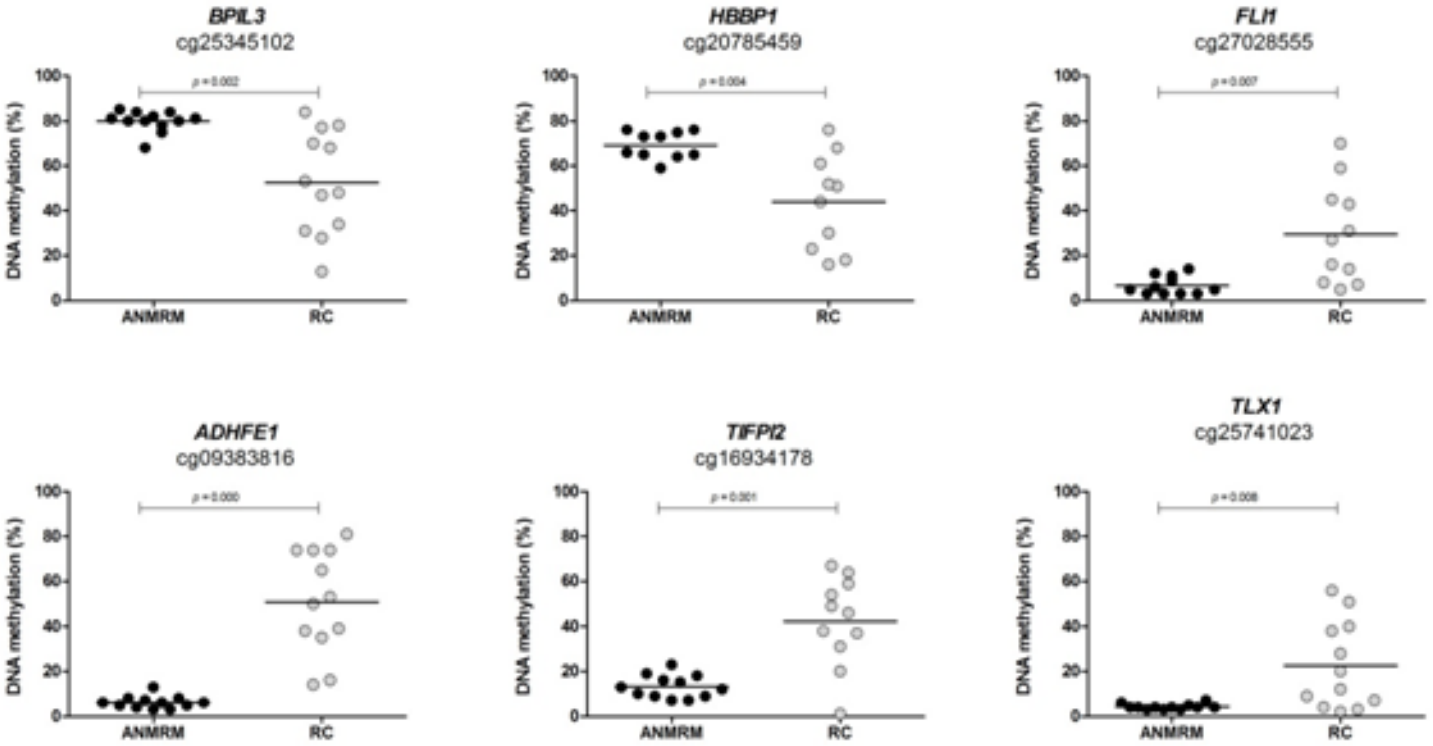
59 Kulis M, Queiros AC, Beekman R, Martin-Subero JI. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta* 1829(11), 1161–1174 (2013).

60 Rideout WM 3rd, Coetzee GA, Olumi AF, Jones PA. 5-methylcytosine as an endogenous mutagen in the human LDL receptor and *p53* genes. *Science* 249(4974), 1288–1290 (1990).

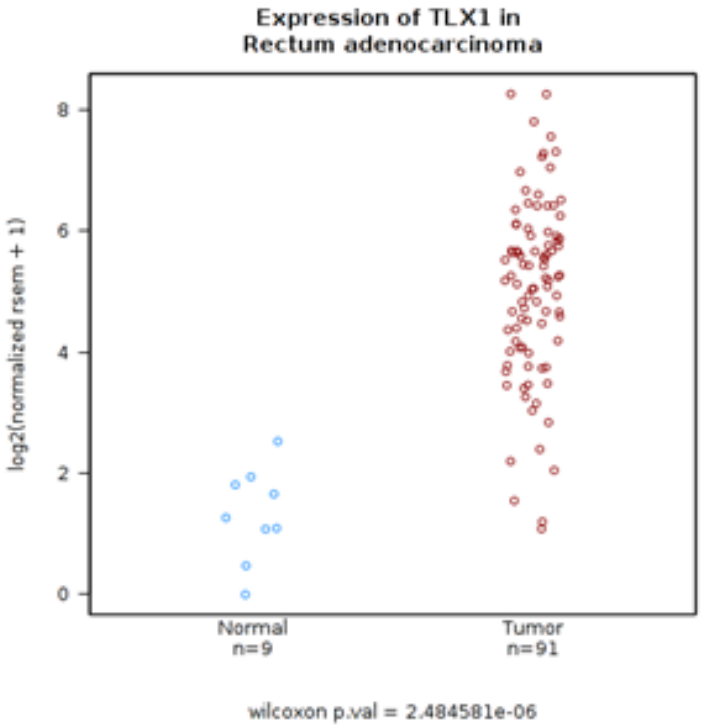
Supplementary Figure 1. Differential DNA methylation (tumour vs. ANMRM) across chromosomes. Blue bars show hypermethylation and the red bars hypomethylation. Rectal pairs 5, 11, 12, 18 and 29 were removed from further analysis due to complete lack of methylation differences.



Supplementary Figure 2. Genes validated with Pyrosequencing assay technology using the adjacent non-malignant rectal mucosa (ANMRM, n =10-12) and rectal cancer (RC, n= 10-12).



Supplementary Figure 3. Expression levels of *TLX1* gene from TCGA dataset (**https://tcga-data.nci.nih.gov/tcga/**)



Supplementary Table S1. Genes validated with Pyrosequencing assay technology.
The amplicon size, annealing temperature and Mg^{2+} concentration are shown for each PCR primer pair. The sequence to analyze and the target CpG site is underlined and colored red.

Gene CpG site ID	Primer type	Sequence (5'-3')	Amplicon size (bp)	Ta (ºC)	Mg^{2+} (mM)
TIFPI2 **cg16934178**	Fw	GGGGTTTATGGTGTAGGG	229	54	2.5
	Rev-biotin	TCCTCTCCCTCTTACACAATT			
	Seq primer	AGGTTTTTTGTTTTAGG			
	Seq to analyze	YGGTTYGG GTGTTYGTTT TATGYGGGGY GAGYGTTYGG TYGATTTTYG			
HBBP1 **cg20785459**	Fw	GTAAGGAAGTTTTGTAAGATGGATA TT	210	56	1.5
	Rev-biotin	ATCCTTTTCAATTATAACTTATTCCT ATTTATCA			
	Seq primer	GTGTGTAGTTTATTTTTTAAGGA			
	Seq to analyze	TTTYGTTAGT AGTYGTTAAT AATTTGAAT			
ADHFE1 **cg01588438**	Fw	TTGATAAGTAAGGAGATTTAAGGTA GAA	142	54	2.0
	Rev-biotin	AACTCAAAACCATTTTCCCAC			
	Seq primer	GGTAATTTTAAGGGTGGA			
	Seq to analyze	TGGTGYGAGY GTYGTTGGGG TAGTTGG			
BPIL3 **cg25345102**	Fw	AGGGGGTTTAATTTGTTTTGAGA	139	56	2.0
	Rev-biotin	ACCAATCAACAAACTACAAAATACC			
	Seq primer	TAGGGTTTGGTGGGT			
	Seq to analyze	TGYGTTTAAT TTAGGTTTTT TGGTATAG			
FLI1 **cg27028555**	Fw	GTTAGTTTTTTGGGGATTAGGAAG	117	56	2.0
	Rev-biotin	CCCTAAACCACCTATCC			
	Seq primer	TGTTTGGGAGTTAGTG			
	Seq to analyze	T TTGGGTYGTY GGGTTYGGGT AAGYGTTGGG			
TLX1 **cg25741023**	Fw-biotin	GTATAGTTAATGGAGAGATTTAGT	138	54	2.0
	Rev	TATTCTTCCCCTCTCTAACTTCTACT T			
	Seq primer	ATACAAAAAAACTTC			
	Seq to analyze	RCRATTTCRA CTAAATCTCT CCATTA			

Supplementary Table S2. Comparison of the CpG sites identified in our study by the unguided hierarchical clustering analysis and from the TCGA database **(https://tcga-data.nci.nih.gov/tcga/)**. The order of genes is reflected with the respect to those shown by heatmap.

	Present study				TCGA database					
	Mean β-value				Mean β-value					
Gene name	**adj. *p***	**Δβ**	**RC**	**Normal mucosa**	**adj. *p***	**Δβ**	**RC**	**Normal mucosa**	**Chr**	**CpG ID (Location)**
C1orf77	$4.1e^{-10}$	0.32	0.74	0.42	$5.4e^{-6}$	0.34	0.71	0.37	1	2482497 Gene body
HOXA2	$4.7e^{-9}$	0.39	0.83	0.44	$2.5e^{-6}$	0.53	0.84	0.31	7	2803819 Gene body
FOXD2	$3.3e^{-7}$	0.29	0.65	0.36	$2.5e^{-23}$	0.34	0.65	0.31	1	3440588 TSS1500
NKX2-2	$2.7e^{-7}$	0.41	0.59	0.17	$2.5e^{-29}$	0.50	0.56	0.06	20	22474464 Gene body
MIR34B	$2.9e^{-8}$	0.45	0.55	0.10	$2.1e^{-36}$	0.54	0.59	0.04	11	22879515 TSS200
GPR88	$1.8e^{-6}$	0.16	0.94	0.77	$2.5e^{-5}$	0.29	0.90	0.61	5	18421360 Gene body
TFPI2	$2.1e^{-14}$	0.50	0.58	0.08	$2.8e^{-15}$	0.52	0.61	0.09	7	12973591 Gene body
PCDH8	1.2e-8	0.35	0.55	0.21	$2.2e^{-25}$	0.39	0.49	0.10	13	287312 1st Exon
TLX1	5.2e-10	0.39	0.48	0.09	$2.5e^{-37}$	0.45	0.54	0.09	10	299972 Gene body
CHST2	$6.3e^{-12}$	0.43	0.67	0.23	$4.8e^{-6}$	0.51	0.69	0.18	3	17373442 Gene body
MSC	$5.3e^{-12}$	0.44	0.60	0.16	$2.2e^{-9}$	0.31	0.59	0.18	8	6269753 Gene body
LOC283922	0.30	0.004	0.05	0.05	$2.7e^{-3}$	0.01	0.09	0.08	16	27401698 Gene body
C17orf46	$1.3e^{-9}$	0.51	0.60	0.09	$3.4e^{-41}$	0.55	0.58	0.03	17	4992638 Gene body
EID3	$1.8e^{-5}$	0.25	0.54	0.29	NA	NA	NA	NA	12	5057777 1st Exon
MIR129-2	$8.1e^{-9}$	0.47	0.70	0.24	$8.5e^{-6}$	0.42	0.59	0.17	11	1939477 TSS200
NEUROG3	$2.6e^{-6}$	0.30	0.46	0.16	$4.8e^{-13}$	0.30	0.47	0.17	10	12938159 Gene body
RASSF2	$3.2e^{-7}$	-0.34	0.55	0.89	$9.9e^{-25}$	-0.38	0.45	0.83	20	14750543 Gene body
GPR85	$4.6e^{-10}$	-0.37	0.47	0.84	$2.6e^{-34}$	-0.49	0.35	0.84	7	14511782 Gene body
LOC283914	$2.3e^{-11}$	-0.35	0.46	0.81	NA	NA	NA	NA	16	269291631 Gene body
DEFB119	$2.5e^{-10}$	-0.34	0.43	0.77	$7.9e^{-41}$	-0.39	0.32	0.71	20	18462653 Gene body
C20orf197	$2.3e^{-12}$	-0.27	0.47	0.75	$2.1e^{-15}$	-0.31	0.41	0.72	20	9076077 TSS1500
EIF3IP1	$1.1e^{-10}$	-0.34	0.43	0.77	$1.4e^{-41}$	-0.51	0.35	0.86	7	14146669 Gene body
DEFB118	$4.1e^{-9}$	-0.37	0.51	0.88	$2.7e^{-28}$	-0.46	0.39	0.85	20	20312687 Gene body

DEFB122	$4.4e^{-10}$	-0.33	0.46	0.79	$2.0e^{-10}$	-0.46	0.32	0.78	20	1349088 Gene body
BEYLA	$2.2e^{-9}$	-0.32	0.41	0.74	NA	NA	NA	NA	8	12866122 Gene body
LOC339568	$5.4e^{-9}$	-0.35	0.37	0.72	$3.0e^{-13}$	-0.52	0.22	0.74	20	4742334 Gene body
BPIL3	$1.1e^{-8}$	-0.35	0.48	0.83	$3.4e^{-18}$	-0.44	0.30	0.74	20	1822339 Gene body
HBBP1	$1.0e^{-8}$	-0.30	0.59	0.90	$1.7e^{-32}$	-0.44	0.40	0.84	11	4161236 Gene body
SAMSN1	$6.5e^{-12}$	-0.36	0.38	0.73	$7.8e^{-15}$	-0.48	0.25	0.73	21	13951664 Gene body
ASCC2	$2.7e^{-9}$	-0.31	0.37	0.68	$8.4e^{-34}$	-0.35	0.31	0.66	22	147442445 Gene body
IL22RA2	$5.8e^{-8}$	-0.30	0.40	0.71	$1.3e^{-25}$	-0.35	0.32	0.67	6	23507945 Gene body
CAPSL	$2.8e^{-9}$	-0.30	0.54	0.84	$1.9e^{-29}$	-0.39	0.45	0.84	5	1084270 3´UTR
BMPR1B	$4.9e^{-8}$	-0.32	0.46	0.90	NA	NA	NA	NA	4	18836661 5´UTR

Negative Δβ-values reflect an hypomethylated status while positive Δβ an hypermethylated on RC, rectal cancer; Δβ, delta beta is the value of the differential methylation; Chr, Chromosome number; CpG ID (Location), the coordinate of the CpG location according to the human genome build 37; CGI, CpG island; CGI shore, regions 2000 bp away from the CpG island; Body, intragenic CpG sites; NA, data missing. The adjusted *p*-value was considered significant when < 0.05.