

Original Article

# IBD-associated Colon Cancers Differ in DNA Methylation and Gene Expression Profiles Compared With Sporadic Colon Cancers

Joel Pekow,<sup>a</sup> Kyle Hernandez,<sup>b,c</sup> Katherine Meckel,<sup>a</sup> Zifeng Deng,<sup>a</sup> Haider I. Haider,<sup>a</sup> Abdurahman Khalil,<sup>a</sup> Chunling Zhang,<sup>b</sup> Nitya Talisila,<sup>a</sup> Shivi Siva,<sup>a</sup> Farzana Jasmine,<sup>d</sup> Yan Chun Li,<sup>a</sup> David T. Rubin,<sup>a</sup> Neil Hyman,<sup>e</sup> Marc Bissonnette,<sup>a</sup> Christopher Weber,<sup>f</sup> Muhammad G. Kibriya<sup>d</sup>

<sup>a</sup>University of Chicago, Section of Gastroenterology, Hepatology, and Nutrition <sup>b</sup>University of Chicago, Center for Research Informatics <sup>c</sup>University of Chicago, Department of Pediatrics <sup>d</sup>University of Chicago, Department of Public Health Sciences <sup>e</sup>University of Chicago, Department of Surgery <sup>f</sup>University of Chicago, Department of Pathology

Corresponding author: Joel Pekow, MD, 900 East 57th St, MB #9, University of Chicago Medicine, Section of Gastroenterology, Hepatology, and Nutrition, Chicago, IL 60637, USA. Tel: 773-702-2774; Fax: 773-702-2281; email: [jpekow@medicine.bsd.uchicago.edu](mailto:jpekow@medicine.bsd.uchicago.edu)

## Abstract

**Background and Aims:** As ulcerative colitis [UC]-associated colorectal cancer [CRC] and sporadic CRC differ in presentation and molecular features, we sought to evaluate differences in the impact of DNA methylation on gene expression.

**Methods:** DNA methylation was assessed in 11 UC-CRCs and adjacent tissue and 11 sporadic CRCs and adjacent tissue, using Illumina arrays. RNA sequencing was performed on 10 UC-CRCs and adjacent tissue and eight sporadic CRCs and adjacent tissues. Differences in DNA methylation and transcript expression, as well as their correlation in the same tissues, were assessed. Immunohistochemistry was performed for three proteins, ANPEP, FAM92A1, and STK31, all of which exhibited an inverse correlation between DNA methylation and transcript expression in UC.

**Results:** Thirty three loci demonstrated differences in DNA methylation between UC-CRC and adjacent tissue. In contrast, there were 4204 differentially methylated loci between sporadic colon cancer and adjacent tissue. Eight hundred eighty six genes as well as 10 long non-coding RNAs [lncRNA] were differentially expressed between UC-CRC and adjacent tissues. Although there were no differentially methylated loci between UC and sporadic CRC, 997 genes and 38 lncRNAs were differentially expressed between UC-CRC and sporadic CRC. In UC, 18 genes demonstrated a negative correlation between DNA methylation and transcript expression. Evaluation of protein expression related to three genes, ANPEP, FAM92A1, and STK31, confirmed down-regulation of ANPEP and up-regulation of STK31 in UC-CRC.

**Conclusions:** Regulation of transcript expression by DNA methylation involves genes key to colon carcinogenesis and may account for differences in presentation and outcomes between inflammatory bowel disease and sporadic colon cancer.

**Key Words:** Ulcerative colitis; inflammatory bowel disease; colon cancer; DNA methylation; gene expression

## 1. Introduction

Chronic colonic mucosal inflammation can lead to colon cancer development in patients with inflammatory bowel disease [IBD].<sup>1,2</sup> Although population-based studies reporting the incidence of cancer in IBD differ in their risk assessments, likely related to heterogeneity in patient populations, surveillance strategies, and management of dysplasia, there is a consistently reported significantly increased risk of colon cancer development in patients with IBD-associated colitis compared with individuals without IBD.<sup>3-5</sup> Improving our understanding of mechanisms of IBD-associated cancer are, therefore, critical in these high-risk patients in order to prevent the onset of neoplasia.

In accordance with the increase in risk, several biological differences exist between IBD-associated and sporadic colon cancer. Unlike sporadic colon cancer, which is typically isolated and arises in non-dysplastic mucosa, ulcerative colitis [UC]-associated neoplasia is often multifocal. This finding reflects a broader 'field effect' of involved mucosa at risk in patients with IBD. Tumour development and progression in sporadic colorectal cancer result from accumulation of genetic alterations.<sup>6,7</sup> Similar genetic changes influence carcinogenesis in IBD, although several of these events, including p53 mutations, microsatellite instability, loss of heterozygosity, and methylation of CpG islands, can occur before the development of dysplasia.<sup>8,9</sup> Furthermore, aneuploidy, chromosomal alterations, and p53 loss of heterozygosity for example, can occur in normal-appearing mucosa in IBD patients who harbour cancer elsewhere.<sup>10,11</sup> In previous studies, we also demonstrated that gene expression changes were present in non-dysplastic mucosa in the rectum from UC patients harbouring a proximal neoplastic lesion. These changes can, therefore, serve as surrogate markers of more advanced dysplastic changes in other areas of the colon.<sup>12</sup>

DNA methylation plays a critical role in the control of gene expression that regulates normal cellular processes. In cancer, DNA is globally hypomethylated, although promoter hypermethylation of tumour suppressor genes also occurs.<sup>13,14</sup> Previous studies examining DNA methylation in patients with IBD offer several lines of evidence to support the hypothesis that these epigenetic modifications contribute to inflammation and IBD-associated tumorigenesis. First, increased methylation patterns are seen in patients with chronic colitis.<sup>15,16</sup> Several of these genes are also hypermethylated in IBD-associated neoplasia.<sup>8,15,17,18</sup> Second, animal studies demonstrate that a methylation field defect at several CpG islands occurs early in the course of inflammation before neoplasia develops.<sup>19</sup> The contribution of these epigenetic changes to tumorigenesis remains unknown; however, as recent studies have demonstrated, IBD-associated cancers have less frequent methylation changes than sporadic colon cancers, and a functional analysis of hyper- and hypo-methylated genomic regions in chronic colitis has not been performed.<sup>20,21</sup> In the current study, we sought to identify differences in DNA methylation and gene expression between sporadic colon cancers and IBD-associated cancers, in order to identify epigenetically regulated genes specific to inflammation-induced colon cancer.

## 2. Methods

### 2.1. Patients

Subjects were enrolled at the time of surgery, and consented under University of Chicago Institutional Review Board numbers 10-209A, 15573A, and 12758A. Subjects were included if they had either sporadic colon cancer or ulcerative colitis-associated colon cancer.

A diagnosis of ulcerative colitis was confirmed by the treating gastroenterologist and histological review by a gastrointestinal pathologist. All UC-associated colon cancers used in the analysis were confirmed histologically to arise in an area of the colon involved by IBD. Samples were excluded from the analysis if a patient had a known hereditary colon cancer syndrome. Tissue was obtained immediately following colectomy by mucosal stripping, and placed in RNAlater by the Human Tissue Research Core facility at the University of Chicago.

### 2.2. Tissue processing

Tissue was homogenised using the Bullet Blender<sup>R</sup> [Next Advance, Averill Park, NY] and extraction performed using the AllPrep DNA/RNA/miRNA kit [Qiagen, Hilden, Germany]. DNA and RNA concentrations were assessed using the NanoDrop ND-1000 [NanoDrop Technologies, Wilmington, DE] and RNA quality with the Agilent 2100 bioanalyser [Agilent Technologies, Santa Clara, CA].

### 2.3. DNA methylation analysis

A total of 44 DNA samples [11 sporadic colon cancers and adjacent tissue, 11 UC-associated colon cancers and adjacent tissue] were used for the methylation assay. Bisulphite modification was performed on 1 µg of DNA from each sample, using the EZ DNA Methylation kit [Zymo Research, Irvine CA]. Bisulphite-converted DNA was profiled by by Illumina Infinium HumanMethylation450 to interrogate 485 000 methylation sites per sample at single-nucleotide resolution. The raw iDAT format data were processed using the SWAN normalisation algorithm in the Partek Genomics suite.<sup>22</sup> Probes with single nucleotide polymorphisms [SNPs], with cross-hybridisation and not on autosomes, were filtered out of analysis. Batch effects of slides were removed by the ComBat program. The processed  $\beta$  values (average intensity for methylated/[average intensity for methylated and unmethylated +  $\alpha$ ]) for 267 996 methylation loci were used for further statistical analysis. Paired or unpaired t tests and multiple testing corrections were used to detect differentially methylated loci [DML] between IBD-associated cancer or sporadic colon cancer, and adjacent non-dysplastic tissue, as well between cancer groups [IBD cancer vs sporadic colon cancer] and normal groups [non-dysplastic normal tissue adjacent to a sporadic cancer vs non-dysplastic UC tissue adjacent to a UC-associated cancer].

'Multiple test corrections' is a statistical method to control false discoveries when multiple hypotheses are tested simultaneously in large-scale analysis. The q-value measures hypothesis testing significance with respect to the positive false-discovery rate [pFDR] in multiple hypothesis testing.<sup>23</sup> For each pairwise comparison, DML were detected with false-discovery rate [FDR; q]  $\leq 0.05$ , and mean difference between groups  $\geq 0.2$ . Lists of unique related genes were generated for these loci, and their relationship was visualised by a Venn diagram. Hierarchical clustering visualisation of the DML showed that samples from the same group clustered together.

### 2.4. RNA sequencing

RNA was prepared for sequencing using the Ribo-zero rRNA Removal kit [Illumina] with 1 µg of total RNA on a total of 36 samples [eight sporadic colon cancers and adjacent tissue, and 10 UC-associated cancers and adjacent tissue]. The rRNA-depleted samples were then converted to cDNA with reverse transcriptase and random primers using the TruSeq RNA Library Preparation Kit [Illumina]. Each library was then sequenced according to manufacturer's specifications using six samples per lane in the

Illumina HiSeq platform, producing a total of 881 090 887 75-bp, paired-end reads. Base conversion was performed using Illumina's OLB v1.9.

Sequence quality was assessed using FastQC v0.11.2 [Babraham Bioinformatics: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>] to explore quality distribution and other sequence characteristics. Then, reads were aligned to the human genome [Gencode GRCh37.p13 v19] using the splice-aware aligner STAR v2.4.0k [optional parameters: `-outSAMtype BAM SortedByCoordinate -outSAMstrandField intronMotif -outSAMattributes NH HI NM AS nM -outFilterIntronMotifs RemoveNoncanonical`].<sup>24</sup> Aligned reads were then post-processed using Sambamba v0.5.1,<sup>25</sup> and various quality metrics were collected using Picard [Broad Institute: <http://broadinstitute.github.io/picard/>]. Alignments were quantified using FeatureCounts v1.4.6 [optional parameters: `-p -t exon -Q 1 -g gene id -s 2`].<sup>26</sup> The counts were filtered to include only expressed protein coding genes. Differences were assessed using various linear models in DESeq2, with an adjusted *p*-value cutoff of 0.05.<sup>27</sup>

## 2.5. Within-group model

To control for genotype effects, a linear model encompassing both the tissue [CRC vs adjacent] and group variables [IBD vs sporadic] was employed. Thus, the IBD and sporadic groups were analysed separately using the following model:  $\text{expression} = \text{intercept} + \text{patient} + \text{tissue}$ . After fitting the model, all further analyses focused on the [tumour] – [adjacent non-dysplastic tissue] specific contrast. Basic principal component analysis [PCA] plots revealed that one IBD patient was an outlier, and this was dropped from all analyses.

In order to evaluate for the impact of inflammation on differential expression in UC-associated colon cancers, separate models were fitted for active and inactive IBD samples with contrasts between the tumour and adjacent tissues [ $\text{expression} = \text{patient} + \text{tissue}$ ]. The significantly differentially expressed gene sets were then compared between the active and inactive samples.

## 2.6. Global difference model

To detect genes where the difference between the tumour and normal samples depended on whether the patient was from the IBD or the sporadic group, the following intercept-free model was fitted:  $\text{expression} = \text{IBD}_{\text{adjacent}} + \text{IBD}_{\text{tumour}} + \text{sporadic}_{\text{adjacent}} + \text{sporadic}_{\text{tumour}}$ . Here, the dependent variables represent binary vectors containing the membership [or non-membership] of each sample in the various group-tissue combinations. After fitting the model, further analyses focused on the following contrast.

## 2.7. Within-tissue models

To compare differences in expression between tumour or normal samples across patient groups, the tumour and normal tissues were analysed separately using the following model:  $\text{expression} = \text{intercept} + \text{group}$ . After fitting the model, all further analyses focused on the [IBD] – [sporadic] specific contrast.

## 2.8. Enrichment analysis

After detecting differentially expressed genes [DEGs], the log<sub>2</sub> fold change for all genes was used to detect up- and down-regulated pathways and ontologies using the Generally Applicable Gene-set Enrichment for pathway analysis [GAGE] and Pathway R/bioconductor packages.<sup>28,29</sup>

## 2.9. DNA methylation-transcript expression integration analysis

Two datasets were created using the within-group model comparing IBD cancer and adjacent tissue: 1] normalised expression and methylation data of all significant DEGs regardless of methylation significance [SET1]; and 2] normalised expression and methylation data of all significant methylation probes/genes regardless of expression significance [SET2]. Normalised expression values were correlated [Pearson's] with the normalised methylation values of corresponding probes across all overlapping samples. Significant negative correlations [adjusted *p*-value <0.05] were plotted to reveal any potential clustering of normal and tumour samples for candidate genes.

## 2.10. Immunohistochemistry

Immunohistochemistry for Alanyl membrane aminopeptidase [ANPEP] and Serine/Threonine Kinase 31 [STK31] was performed on nine UC-colon cancers, 10 non-dysplastic tissue samples adjacent to a UC-colon cancer, and 10 normal controls, and for Family with Sequence Similarity 92 Member A1 [FAM92A1] on five UC-colon cancers, nine non-dysplastic tissues adjacent to a UC cancer, and nine normal controls. In addition, protein expression for each was assessed on a tissue array containing 27 UC-associated colon cancers. Immunostaining was performed as previously described, following antigen retrieval with a sodium citrate buffer using primary antibodies for ANPEP [1:100; Lifespan Biosciences, Seattle, WA], FAM92A [1:200; One World Lab, San Diego, CA], and STK31 [1:50; One World Lab].<sup>12</sup> These proteins were selected as their respective genes demonstrated a significant negative correlation between DNA methylation and transcript expression in IBD patients with significant differential expression between IBD cancer and adjacent tissue. Tight junction expression of ANPEP was visually scored as positive or negative, and results between groups were compared using a Fisher's exact test. For FAM92A, staining was visualised using the Aperio Scanscope [Leica Biosystems, Wetzlar, Germany] and captured using Aperio ImageScope software [Leica Biosystems]. The mean intensity of 20 nuclei per slide using ImageScope algorithms software was calculated and compared using a Student's *t* test. Staining of STK31 was captured using Image-Pro Plus [Media Cybernetics, Rockville, MD] and processed by ImageScope software. Cytoplasmic expression of STK31 was visually scored on five different sections per immunostained tissue by three blinded investigators. A scoring scheme of 0 to 2, based on the colour intensity gradient analysed using ImageScope algorithms, was applied as 0: negative; 1: weak positive; and 2: strong positive. The mean score for each patient was calculated. Differences between groups were compared using the Student's *t* test.

## 3. Results

### 3.1. Patient population

DNA methylation was analysed in 11 patients with IBD-associated colon cancer and 11 patients with sporadic colon cancer. Patients were matched by tumour location and age within 10 years. Ten of the 11 patients with IBD cancer and four of the 11 patients with sporadic colon cancer who were included in the methylation analysis also had concordant transcript expression examined by RNA sequencing. Among the 10 UC-associated colon cancer patients with RNA-seq data, one subject was removed from the analysis after initial quality control identified their RNA expression as an outlier in an initial principle component analysis, leaving a total of nine IBD

patients included in the final analysis. Because of superior RNA quality compared with seven of the sporadic colon cancer tissues analysed for DNA methylation, four additional patients with sporadic colon cancer were included in the RNA sequencing analysis who were not investigated by initial DNA methylation array [eight total sporadic patients included in the final analysis]. Patient, tumour, and disease-specific characteristics are included in [Table 1](#).

### 3.2. DNA methylation

Comparing UC-associated cancer with adjacent, non-dysplastic tissue, there were 33 loci with significant differences in methylation representing 19 unique genes [ $q \leq 0.2$ , mean differential [dif]  $\geq 0.2$ ]. The location of the differentially methylated CpG sites included an island for seven genes, a north shelf for one, a north shore for four, a south shelf for five, and a south shore for three. Fourteen of the genes had differential methylation in the body of the gene, and three were located from 200 nt to 1500 nt from the transcription start site [TSS1500].

There were 4204 significantly differentially methylated loci involving 1485 unique genes in sporadic colon cancers compared with adjacent non-dysplastic tissue [ $q \leq 0.05$ , mean dif  $\geq 0.2$ ]; 514 CpG sites were located in an island, 210 on a north shelf, 324 on a north shore, 195 on a south shelf, and 204 on a south shore. A total of 448 methylated sites were in a gene body, 30 in the 3' untranslated region [UTR], 135 in the 5'UTR, 164 within TSS1500, and 80 within 200 nt of the transcription start site. Of the differentially methylated loci identified in sporadic CRC compared with adjacent tissue, 13 loci representing 10 unique genes were also differentially expressed between IBD cancer and adjacent tissue, and 506 loci

were differentially expressed between IBD cancer and normal tissue [non-dysplastic tissue adjacent to sporadic colon cancer] [[Figure 1](#)]. Similar unique genes with differentially methylated loci between IBD cancer and adjacent tissue, as well as between sporadic colon cancer and adjacent tissue, included: GALNT9, HS3S2, TBCD, FBXL7, WDR67, BAI1, ZNF469, BAI1, PTPRN2, and LOC100133991. A list of differentially methylated loci between UC-CRC and adjacent non-dysplastic UC tissue, as well as the location of methylation, is listed in [Supplementary Table 1](#), available as [Supplementary data at ECCO-JCC online](#) and a similar list between sporadic CRC and adjacent normal tissue is provided in [Supplementary Table 2](#), available as [Supplementary data at ECCO-JCC online](#).

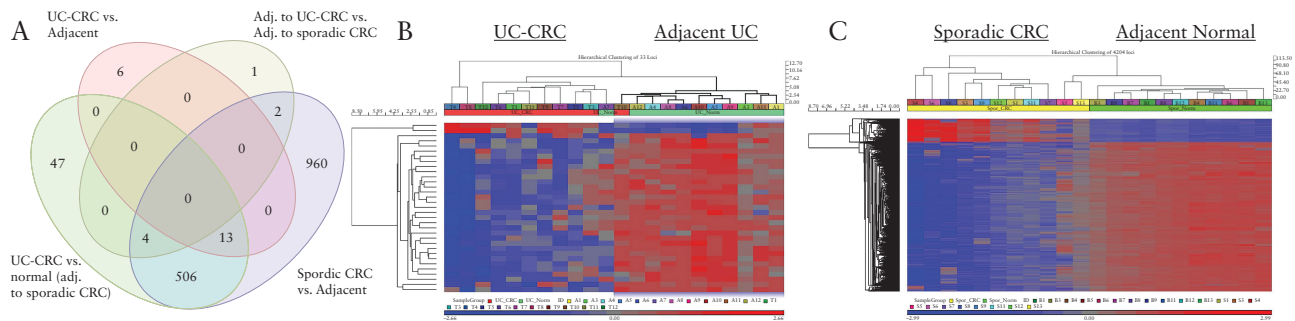
### 3.3. RNA sequencing

In all, 886 genes, as well as 10 long noncoding RNAs [lncRNA], were differentially expressed between IBD cancers and adjacent tissues [[Supplementary Table 3](#), available as [Supplementary data at ECCO-JCC online](#)]; and 4258 genes, as well as 95 lncRNA, were differentially expressed between sporadic colon cancers and adjacent tissue [[Supplementary Table 4](#), available as [Supplementary data at ECCO-JCC online](#), adjusted  $p$ -value of 0.05]. A total of 997 genes and 38 lncRNAs were differentially expressed between IBD cancers and sporadic cancers [[Supplementary Table 5](#), available as [Supplementary data at ECCO-JCC online](#)], and 3423 genes, including 68 lncRNA, were differentially expressed between non-dysplastic tissue adjacent to IBD cancers and non-dysplastic tissue adjacent to a sporadic colon cancer [[Supplementary Table 6](#), available as [Supplementary data at ECCO-JCC online](#)]. In the global difference model, a total of 16 genes and five lncRNAs were differentially expressed [[Supplementary Table 7](#),

**Table 1.** Clinical characteristics of subjects with colon cancer included in the analysis.

	DNA methylation analysis		RNAseq analysis	
	IBD-CRC [ $n = 11$ ]	Sporadic CRC [ $n = 11$ ]	IBD-CRC [ $n = 9$ ]	Sporadic CRC [ $n = 8$ ]
Age [mean]	47.2	65.5	49.2	57.25
Sex [male %]	100%	45%	100%	25%
Location				
Left colon	7	7	6	5
Right colon	4	4	3	3
Race				
Caucasian	8	5	7	5
AA	0	3	0	2
Unknown	3	3	2	1
Ethnicity				
Hispanic	2	0	1	0
Tumour grade				
Low	4	6	4	6
High	4	5	5	2
Mucinous	4	1	4	0
Signet ring	0	1	0	0
Tumour multiplicity				
Adenocarcinoma	3	0	3	0
Dysplasia	5	N/A	4	N/A
PSC	1	0	1	0
Inflammation				
Quiescent	4	N/A	4	
Mild	1		1	
Moderate	4		3	
Severe	2		1	
Pancolitis	100%	N/A	100%	N/A

IBD, inflammatory bowel disease; CRC, colorectal cancer; AA, African American; PSC, primary sclerosing cholangitis; N/A, not available.



**Figure 1.** Differentially methylated loci by tissue phenotype. A) Venn diagram demonstrating the number of overlapping methylated loci between each comparison. B) Heatmap representing unsupervised hierarchical clustering between UC-associated colon cancer and adjacent non-dysplastic UC tissue. C) Heatmap representing unsupervised hierarchical clustering between sporadic CRC and adjacent normal tissue. UC, ulcerative colitis; CRC, colorectal cancer.

available as [Supplementary data at ECCO-JCC online](#)]. The GAGE pathways significantly enriched in IBD-cancers compared with sporadic colon cancers [IBD-cancer compared to adjacent vs. sporadic colon cancer compared to adjacent] are shown in [Figure 2D](#); and in [Supplementary Table 8](#), available as [Supplementary data at ECCO-JCC online](#). A list of GAGE pathways enriched in IBD and sporadic cancers compared with adjacent tissue is provided in [Supplementary Table 9](#), available as [Supplementary data at ECCO-JCC online](#).

In order to investigate the impact of inflammation on differentially expressed genes, we compared inflamed UC cancers [ $n = 5$ ] with uninfamed cancers [ $n = 4$ ]. For this comparison, differentially expressed genes were identified in IBD cancers with and without active inflammation, using the following comparisons: ( $IBD_{\text{tumour[active]}}$  vs  $IBD_{\text{adjacent[active]}}$ ) and ( $IBD_{\text{tumour[quiescent]}}$  vs  $IBD_{\text{adjacent[quiescent]}}$ ). In total, there were 77 genes that were commonly differentially expressed between active UC cancers and adjacent tissues as well as between quiescent UC cancers and adjacent tissues. There were 244 genes that were differentially expressed only between quiescent UC cancers and adjacent tissue, and 181 genes that were unique to active IBD cancers compared with adjacent tissue [[Figure 2E](#); and [Supplementary Table 10](#), available as [Supplementary data at ECCO-JCC online](#)]. The GAGE pathways significantly enriched in active disease alone included: protein degradation and absorption [ $q = 0.0009$ ], retinol metabolism [ $q = 0.007$ ], chemical carcinogenesis [ $q = 0.01$ ], fat digestion and absorption [ $q = 0.04$ ], and drug metabolism-cytochrome p450 [ $q = 0.04$ ]. Pathways that were common to both quiescent and active IBD cancers included proximal tubule bicarbonate reclamation [ $q = 0.0005$ ], bile secretion [ $q = 0.02$ ], and PPAR signalling pathway [ $q = 0.02$ ].

### 3.4. Methylation-expression correlation

In IBD, a total of 796 genes were compared with values from 10 662 methylation array probes in SET1 [normalised expression and methylation data of all significant DEGs regardless of methylation significance]. Of these, 32 probes representing 18 unique genes had significant negative correlations [ACADS, ANPEP, BRX1, CCND1, CDH11, ENPP3, FAM160A1, FAM92A1, GRHL2, HM13, KIAA1549, NR3C2, NSMCE2, PUF60, RGMB, SATB2, SLC26A2, STK31] [[Figure 3](#); [Supplementary Table 11](#), available as [Supplementary data at ECCO-JCC online](#)]. In SET2 [normalised expression and methylation data of all significant methylation probes/genes regardless of expression significance] only a single probe was significant, CSPP1.

### 3.5. Immunohistochemistry

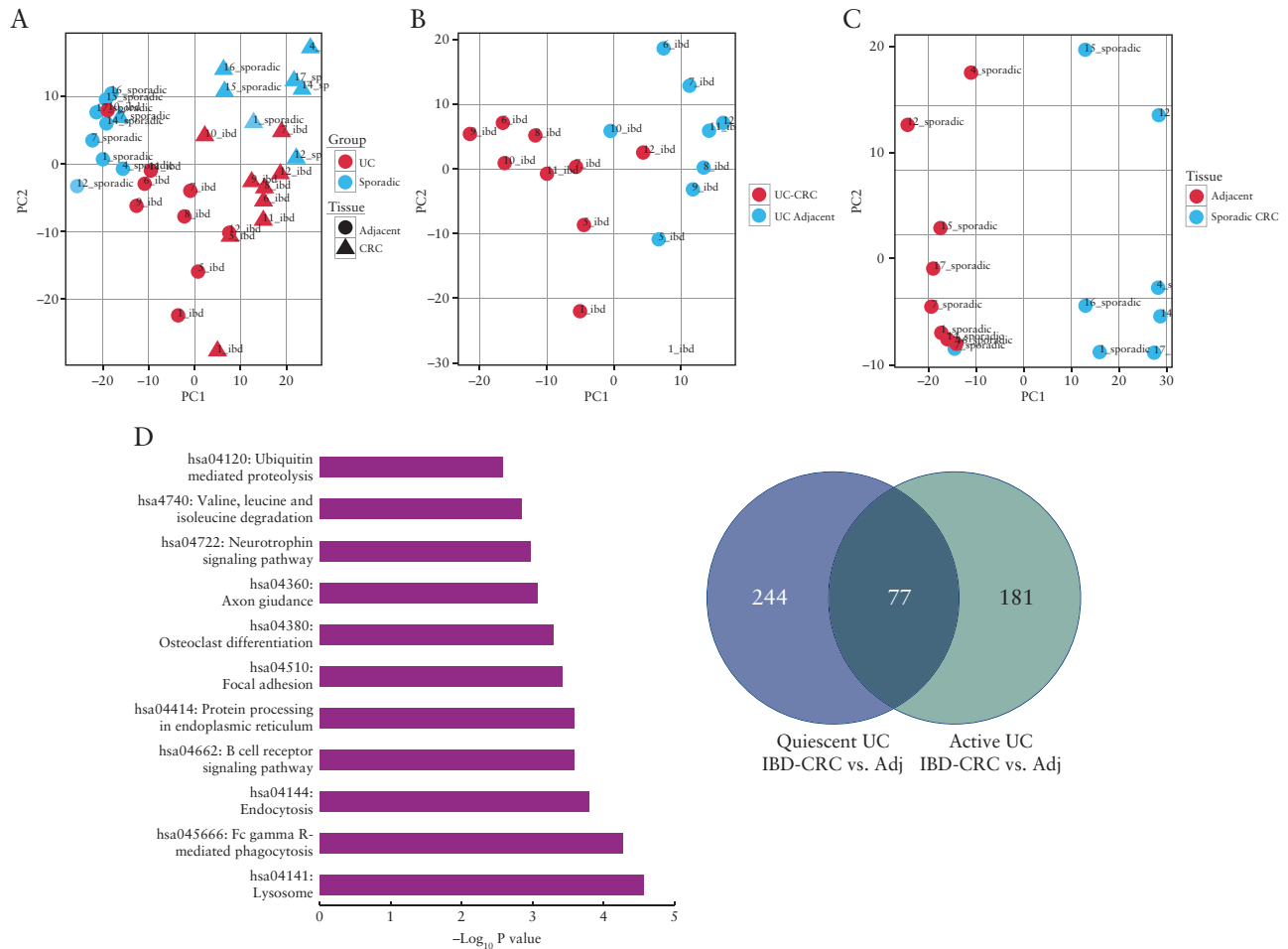
Expression and localisation of three proteins [ANPEP, FAM92A, and STK31], whose gene products had an inverse correlation with

methylation levels in IBD-associated colon cancer and adjacent tissue with significant differential transcript expression between cancer and adjacent tissue, were investigated by immunostaining. These proteins were evaluated since commercial antibodies were available that showed staining specific for epithelial tissue. Significantly fewer UC cancers had brush border and tight junction expression of ANPEP than did tissues adjacent to UC-associated colon cancer [ $p = 0.001$ ] and normal controls [ $p < 0.0001$ ] [[Figure 4A](#)]. As opposed to RNA expression levels, FAM92A1 nuclear expression was unchanged between IBD cancers compared with adjacent tissues [ $p = 0.67$ ], and decreased compared with normal controls [ $p = 0.02$ ] [[Figure 4B](#)]. In contrast, cytoplasmic expression of STK31 was significantly increased in UC-associated colon cancers compared with the epithelium of adjacent non-dysplastic tissue [ $p = 0.02$ ] and trended towards an increase in UC-associated colon cancers compared with normal controls [ $p = 0.07$ ] [[Figure 4C](#)].

Expression of all three proteins was also assessed in a tissue array containing 27 IBD-associated colon cancers. ANPEP was negative in all the cancers except for four well-differentiated carcinomas. FAM92A1 expression was absent in nine colon cancers, weakly positive in 17 cancers, and strongly positive in one cancer. Expression of STK31 was negative in two cancers, weakly positive in 11 cancers, and strongly positive in 14 cancers.

## 4. Discussion

DNA methylation is known to play an important role in the development of sporadic colon cancer where previous studies have demonstrated a large number of differentially methylated loci.<sup>30</sup> A subset of sporadic colon cancers exhibit a CpG island methylator phenotype [CIMP] in which methylation of CpG islands mediates inactivation of tumour suppressor genes.<sup>31</sup> Tumours that display CIMP are often located in the proximal colon, possess BRAF mutations, and arise from sessile serrated adenomas.<sup>32</sup> In contrast, other colon cancers are known to display DNA hypomethylation. DNA hypomethylation can occur early in the transition from adenoma to neoplastic tissue, and leads to chromosomal instability.<sup>33</sup> These differences highlight the influence of epigenetic modifications on different phenotypes of sporadic colorectal cancer, although less is known regarding the impact of DNA methylation on the development of cancer in patients with inflammatory bowel disease. As epigenetic modifications can be influenced by environmental factors, identifying site-specific DNA methylation could provide insights into carcinogenic mechanisms and may ultimately uncover environmental influences that contribute to cancer development unique to patients with inflammatory bowel disease.

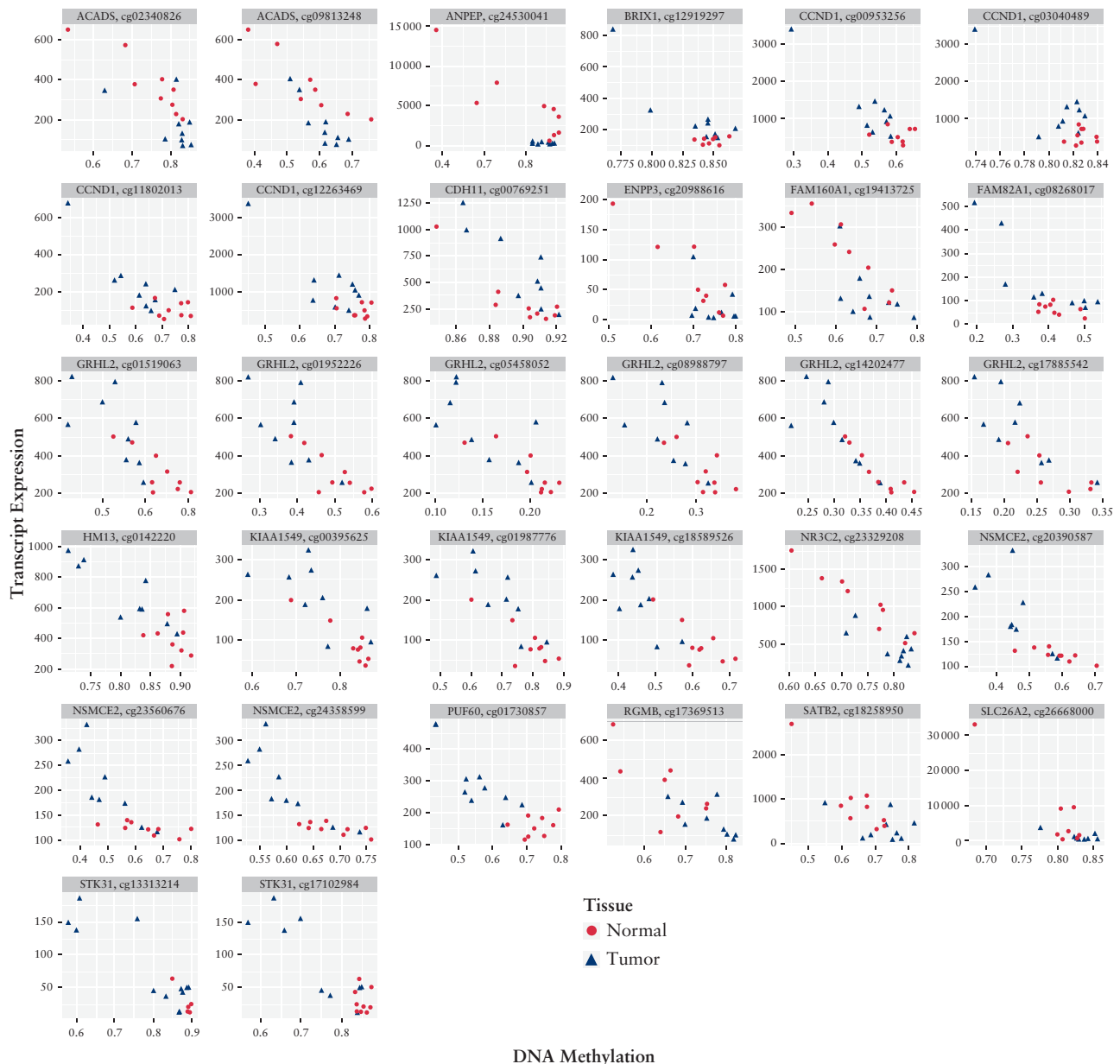


**Figure 2.** Genome-wide transcriptomic analysis distinguishes UC from sporadic colon cancer. Principle component analyses using RNA sequencing data comparing: A) UC colon cancer, non-dysplastic tissue adjacent to a UC cancer, sporadic colon cancer, and normal tissue adjacent to a sporadic colon cancer; B) UC-associated colon cancer and adjacent non-dysplastic UC tissue; and C) sporadic colon cancer and adjacent normal tissue. D) Significantly enriched GAGE pathways in IBD cancers compared with sporadic colon cancers. E) Venn diagram demonstrating the number of significantly differentially expressed genes in UC-associated colon cancers compared with adjacent tissues with active inflammation and UC-associated colon cancers compared with no inflammation [quiescent] in the adjacent tissue. UC, ulcerative colitis; IBD, inflammatory bowel disease.

The results of this study support the hypothesis that there are key epigenetic differences between sporadic and IBD-associated colon cancer. It is known that IBD-associated cancer has several distinct clinical features compared with non-IBD cancers, including occurring at a younger age, more likely being multifocal, arising in the proximal colon, appearing in a field of dysplasia, often exhibiting mucinous or signet-ring histology, and having an association with poorer outcomes.<sup>34–36</sup> Reflecting these clinical dissimilarities, recent work has demonstrated that several genetic mutations occur at different frequencies in IBD compared with sporadic colon cancer.<sup>37,38</sup> Similarly, we have identified differences in global DNA methylation and transcript expression between IBD and sporadic colon cancer. Together, these findings suggest that there are mechanisms unique to inflammatory bowel disease which drive development of colorectal neoplasia and likely account for differences in clinical presentation and molecular appearance when compared with non-IBD colon cancers.

As far as we know, there are no previously published studies or publicly available data examining genome-wide RNA expression in IBD-associated cancers compared with UC without neoplasia or compared with normal controls. One previous published study

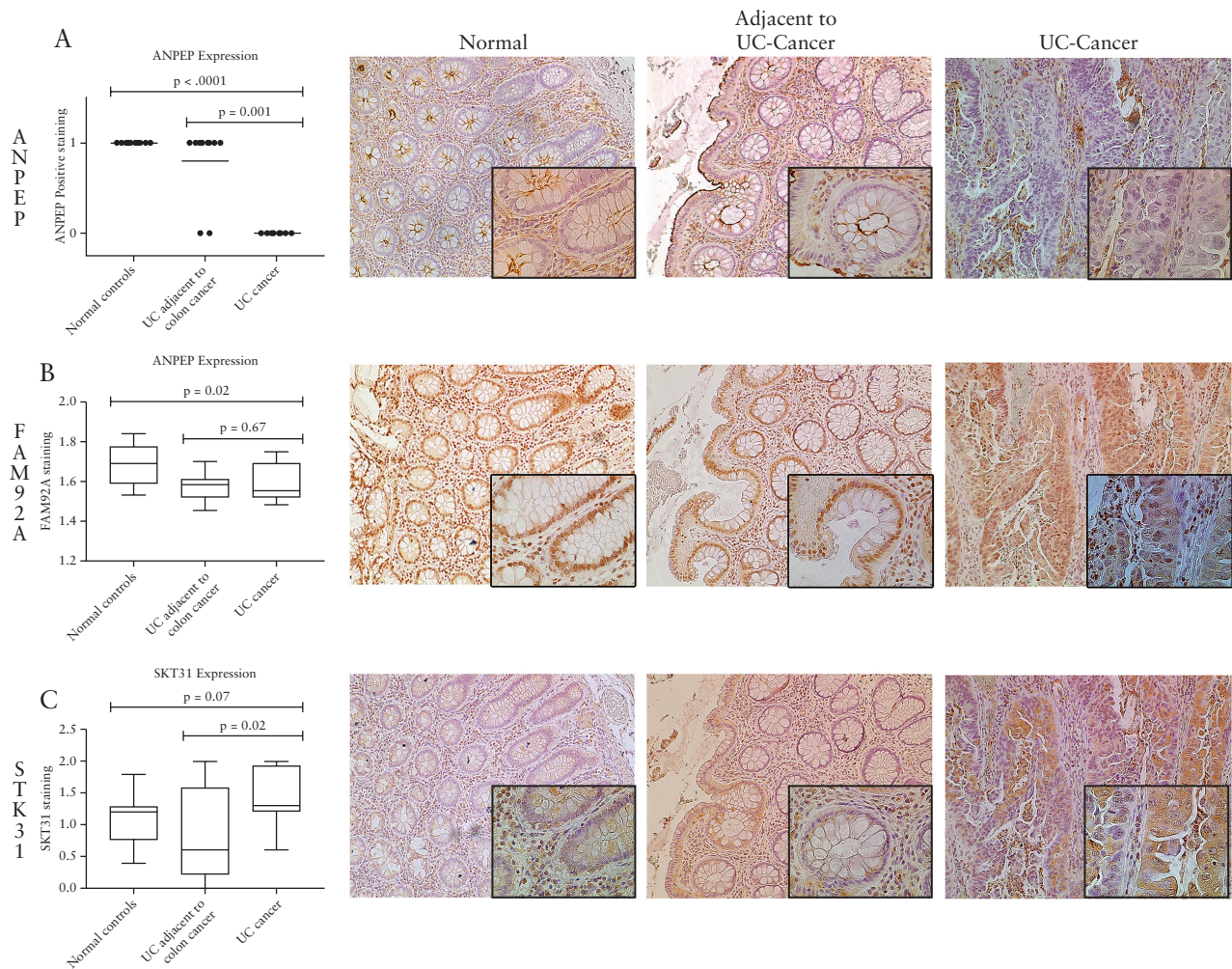
examined cDNA microarrays in sporadic colon cancer compared with IBD-associated neoplasia, although the study included only five IBD cancers, used array technology, and did not examine expression in UC cancers compared with UC without neoplasia, an important comparison to distinguish pathways that are unique to inflammation and carcinogenesis.<sup>39</sup> By analysing concomitant DNA methylations and RNA transcript levels in the same tissues, we were able to identify genes dysregulated in colon cancer which are likely to be epigenetically regulated. Although previous studies have examined DNA methylation in IBD-associated colon cancer,<sup>21,40–43</sup> as far as we know this is the first study to perform RNA sequencing on IBD-associated colon cancers, compare differences with sporadic colon cancer, and assess the correlation with DNA methylation in the same tissue. Several genes that have putative roles in the development of colon cancer were both significantly dysregulated in IBD-associated colon cancer compared with adjacent tissue, and had expression that was inversely correlated with DNA methylation. These include genes which encode for proteins that are involved in DNA repair [NSMCE2], cell cycle regulation [CCND1, STK31], and transcription factor binding [ANPEP, SATB, NR3C2], or target extracellular matrix function [GRHL2, ENPP3].



**Figure 3.** Genes with an inverse relationship between DNA methylation and RNA transcript expression in patients with ulcerative colitis. Correlation plots demonstrating transcript expression on the y-axis and normalised DNA methylation values for 32 loci encompassing 18 genes on the x-axis which had a significant difference in transcript expression between UC cancer and adjacent tissue [adjusted  $p$ -value  $<0.05$ ] as well as a statistically significant negative correlation between methylation and expression [Pearson's  $p$ -value  $<0.05$ ]. Blue triangles represent tissue from UC-associated colon cancers and red circles represent tissue from adjacent non-dysplastic mucosa. UC, ulcerative colitis.

In this study, we evaluated tissue expression and localisation of three proteins [ANPEP, FAM92A1, and STK31] whose respective genes demonstrated an inverse correlation between methylation and transcript expression in the colonic mucosa of IBD patients with colon cancer. ANPEP encodes for a membrane-bound zinc-dependent metalloprotease, aminopeptidase N, that is involved in angiogenesis, proliferation, and as differentiation and was significantly decreased in IBD-associated colon cancer compared with adjacent tissues.<sup>44,45</sup> Although it is up-regulated in several malignancies, previous studies have demonstrated that gene and protein expression of ANPEP is decreased overall in sporadic colon cancer, and increased expression is associated with improved survival.<sup>45,46</sup> FAM92A1, which was significantly increased in IBD-associated colon cancer compared

with adjacent tissue, belongs to a family of proteins with conserved DUF1208 domains.<sup>47</sup> Little is known about the role of FAM92A1 in colon cancer, although previous studies demonstrate that it increases cell migration and growth in other cancers, including renal cell carcinoma and cervical cancer.<sup>48,49</sup> The fact that FAM92A1 demonstrated increased transcript expression, although it was not significantly up-regulated in IBD cancers by immunohistochemistry [IHC], may potentially be related to several factors including: post translational modifications, limitations in quantification related to different tissue preservation techniques, and the semi-quantitative scoring employed to assess protein expression by IHC. STK31 is a novel cancer testis antigen that was significantly increased UC-associated colon cancer compared with adjacent tissue. Previous data support that



**Figure 4.** Expression of ANPEP, FAM92A, and STK31 in UC-associated cancer. Immunohistochemistry was performed with antibodies targeting: A] ANPEP, B] FAM92A, and C] STK31 in normal controls, UC-associated colon cancer, and UC tissue adjacent to colon cancer. Representative images for UC-associated cancer and adjacent non-dysplastic tissue are captured at 10x magnification with insets at 40x magnification. UC, ulcerative colitis.

STK31 is increased in colorectal cancer, and that increases in its expression may be associated with metastatic disease.<sup>50</sup> In vitro studies have demonstrated a role in cell cycle regulation where ectopic expression of STK31 increases colorectal tumour cell invasion and migration and maintains an undifferentiated state.<sup>51,52</sup> Taken together, these findings demonstrate the potential impact of DNA methylation on functional properties of cancer growth through regulation of key proteins involved in cellular proliferation, migration, and invasion.

A major conclusion from this analysis is that compared with sporadic colon cancer, there were far fewer genes that were differentially methylated or demonstrated significant changes in expression in IBD-associated colon cancer when matched to adjacent non-dysplastic tissue. Ten of 19 genes that were differentially methylated in UC-associated colon cancer, however, were also found to be differentially methylated in sporadic colon cancer. The finding that a limited number of loci displayed differential methylation between IBD-cancers and adjacent tissue likely reflects a broader 'field effect' observed in inflammatory bowel disease, where molecular changes observed in colon cancers also occur more widely in non-dysplastic colon mucosa. We previously showed that similar findings occurred in adjacent non-dysplastic mucosa in IBD patients, in studies examining cDNA microarrays as well as miRNA expression. Others have

similarly demonstrated changes in p53 expression, loss of heterozygosity, and aneuploidy occurring in adjacent non-dysplastic mucosa adjacent or remote from a neoplastic lesion.<sup>10-12,53-55</sup>

This study was limited by the relatively small number of cancers analysed. Given the sample size, we were not able to control for other variables including age, sex, and medications that might affect DNA methylation. In addition, the results of the comparison of IBD cancers with and without active inflammation were likely limited by sample size and need to be validated in a larger analysis. Because of poor RNA quality in a subset of the sporadic colon cancers for which we assessed DNA methylation, we were further limited in the number of samples with which to compare concomitant DNA methylation and transcript expression in sporadic colon cancers. Given the inherent challenges in identifying and banking IBD-associated colon cancers from a single centre, we plan to conduct a similar analysis in a larger multicentre collaboration, to validate the findings of this work and address these limitations.

In conclusion, we have identified several epigenetically regulated genes that are unique to IBD-associated colon cancer. Although these genes have putative roles in the development of inflammation-associated neoplasia, the functional role of these genes does need to be evaluated further in future studies. As patients with IBD are at risk for the development of neoplasia, these findings are likely



to be important in identifying environmental triggers of neoplasia, biomarkers to improve colon cancer surveillance, and targets for chemopreventive efforts in this high-risk population.

## Funding

American Cancer Society, Scholz Family Foundation, Shapiro Family Foundation.

## Conflicts of Interest

No relevant conflicts of interest.

## Author Contributions

Conceptualisation of the study: JP, YCL, DTR, NH, MB, MK. Collection of tissue/extraction of DNA/RNA: KM. DNA methylation arrays: FJ. DNA methylation analysis: CZ, MK. RNA sequencing analysis: KH. Immunohistochemistry: ZD, HH, AK, NT, SS, CW. Analysis of data: JP, CW, MB, KH, CZ, MK, DTR, NH. Preparation of manuscript: JP. Critical review of manuscript: all authors.

## Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

## References

- Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001;48:526–35.
- Rutter MD, Saunders BP, Wilkinson KH, et al. Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis. *Gastroenterology* 2006;130:1030–8.
- Söderlund S, Brandt L, Lapidus A, et al. Decreasing time-trends of colorectal cancer in a large cohort of patients with inflammatory bowel disease. *Gastroenterology* 2009;136:1561–7; quiz 1818–9.
- Lutgens MW, van Oijen MG, van der Heijden GJ, Vleggaar FP, Siersema PD, Oldenburg B. Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. *Inflamm Bowel Dis* 2013;19:789–99.
- Lakatos L, Mester G, Erdelyi Z, et al. Risk factors for ulcerative colitis-associated colorectal cancer in a Hungarian cohort of patients with ulcerative colitis: results of a population-based study. *Inflamm Bowel Dis* 2006;12:205–11.
- Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* 1997;275:1787–90.
- Oshima M, Oshima H, Kitagawa K, Kobayashi M, Itakura C, Taketo M. Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. *Proc Natl Acad Sci U S A* 1995;92:4482–6.
- Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res* 2001;61:3573–7.
- Xie J, Itzkowitz SH. Cancer in inflammatory bowel disease. *World J Gastroenterol* 2008;14:378–89.
- Brentnall TA, Crispin DA, Rabinovitch PS, et al. Mutations in the p53 gene: an early marker of neoplastic progression in ulcerative colitis. *Gastroenterology* 1994;107:369–78.
- Rubin CE, Haggitt RC, Burmer GC, et al. DNA aneuploidy in colonic biopsies predicts future development of dysplasia in ulcerative colitis. *Gastroenterology* 1992;103:1611–20.
- Pekow J, Dougherty U, Huang Y, et al. Gene signature distinguishes patients with chronic ulcerative colitis harboring remote neoplastic lesions. *Inflamm Bowel Dis* 2013;19:461–70.
- Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 2011;11:726–34.
- Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008;9:465–76.
- Hsieh CJ, Klump B, Holzmann K, Borchard F, Gregor M, Porschen R. Hypermethylation of the p16INK4a promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. *Cancer Res* 1998;58:3942–5.
- Wang FY, Arisawa T, Tahara T, et al. Aberrant DNA methylation in ulcerative colitis without neoplasia. *HepatoGastroenterology* 2008;55:62–5.
- Moriyama T, Matsumoto T, Nakamura S, et al. Hypermethylation of p14 [ARF] may be predictive of colitic cancer in patients with ulcerative colitis. *Dis Colon Rectum* 2007;50:1384–92.
- Tominaga K, Fujii S, Mukawa K, et al. Prediction of colorectal neoplasia by quantitative methylation analysis of estrogen receptor gene in nonneoplastic epithelium from patients with ulcerative colitis. *Clin Cancer Res* 2005;11:8880–5.
- Katsurano M, Niwa T, Yasui Y, et al. Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction. *Oncogene* 2012;31:342–51.
- Konishi K, Shen L, Wang S, Meltzer SJ, Harpaz N, Issa JP. Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias. *Gastroenterology* 2007;132:1254–60.
- Olaru AV, Cheng Y, Agarwal R, et al. Unique patterns of CpG island methylation in inflammatory bowel disease-associated colorectal cancers. *Inflamm Bowel Dis* 2012;18:641–8.
- Maksimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol* 2012;13:R44.
- Storey JD. The positive false discovery rate: a bayesian interpretation and the q-value. *Ann Stat* 2003;31:2013–35.
- Widmann J, Stombaugh J, McDonald D, et al. RNASTAR: an RNA Structural Alignment Repository that provides insight into the evolution of natural and artificial RNAs. *RNA* 2012;18:1319–27.
- Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 2015;31:2032–4.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* 2009;10:161.
- Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 2013;29:1830–1.
- Kibriya MG, Raza M, Jasmine F, et al. A genome-wide DNA methylation study in colorectal carcinoma. *BMC Med Genomics* 2011;4:50.
- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 1999;96:8681–6.
- Kambara T, Simms LA, Whitehall VL, et al. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* 2004;53:1137–44.
- Matsuzaki K, Deng G, Tanaka H, Kakar S, Miura S, Kim YS. The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. *Clin Cancer Res* 2005;11:8564–9.
- Jewel Samadder N, Valentine JF, Guthery S, et al. Colorectal cancer in inflammatory bowel diseases: a population-based study in Utah. *Dig Dis Sci* 2017;62:2126–32.
- Leowardi C, Schneider ML, Hinz U, et al. Prognosis of ulcerative colitis-associated colorectal carcinoma compared with sporadic colorectal carcinoma: a matched pair analysis. *Ann Surg Oncol* 2016;23:870–6.
- Higashi D, Futami K, Ishibashi Y, et al. Clinical course of colorectal cancer in patients with ulcerative colitis. *Anticancer Res* 2011;31:2499–504.
- Yaeger R, Shah MA, Miller VA, et al. Genomic alterations observed in colitis-associated cancers are distinct from those found in sporadic colorectal

- cancers and vary by type of inflammatory bowel disease. *Gastroenterology* 2016;151:278–87.e6.
38. Robles AI, Traverso G, Zhang M, *et al.* Whole-exome sequencing analyses of inflammatory bowel disease-associated colorectal cancers. *Gastroenterology* 2016;150:931–43.
  39. Selaru FM, Xu Y, Yin J, *et al.* Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions. *Gastroenterology* 2002;122:606–13.
  40. Sanchez JA, Dejulus KL, Bronner M, Church JM, Kalady MF. Relative role of methylator and tumour suppressor pathways in ulcerative colitis-associated colon cancer. *Inflamm Bowel Dis* 2011;17:1966–70.
  41. Azuara D, Rodriguez-Moranta F, de Oca J, *et al.* Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients. *Inflamm Bowel Dis* 2013;19:165–73.
  42. Svrcek M, Buhard O, Colas C, *et al.* Methylation tolerance due to an O6-methylguanine DNA methyltransferase [MGMT] field defect in the colonic mucosa: an initiating step in the development of mismatch repair-deficient colorectal cancers. *Gut* 2010;59:1516–26.
  43. Toiyama Y, Okugawa Y, Tanaka K, *et al.* A panel of methylated micro-RNA biomarkers for identifying high-risk patients with ulcerative colitis-associated colorectal cancer. *Gastroenterology* 2017;153:1634–46.e8.
  44. Menrad A, Speicher D, Wacker J, Herlyn M. Biochemical and functional characterization of aminopeptidase N expressed by human melanoma cells. *Cancer Res* 1993;53:1450–5.
  45. Hashida H, Takabayashi A, Kanai M, *et al.* Aminopeptidase N is involved in cell motility and angiogenesis: its clinical significance in human colon cancer. *Gastroenterology* 2002;122:376–86.
  46. Sanz B, Perez I, Beitia M, *et al.* Aminopeptidase N activity predicts 5-year survival in colorectal cancer patients. *J Invest Med* 2015;63:740–6.
  47. Ruan XZ, Yang HS, Yao SH, *et al.* Isolation and characterization of a novel *Xenopus* gene [xVAP019] encoding a DUF1208 domain containing protein. *Mol Reprod Dev* 2007;74:1505–13.
  48. Liang S, Gong F, Zhao X, *et al.* Prokaryotic expression, purification of a new tumour-relative protein FAM92A1-289 and its characterization in renal cell carcinoma. *Cancer Lett* 2009;276:81–7.
  49. Gui H, Guo XR, Fang J, *et al.* The tumour-promoting effects of FAM92A1-289 in cervical carcinoma cells. *Anticancer Res* 2016;36:5197–204.
  50. Zhong L, Liu J, Hu Y, *et al.* STK31 as novel biomarker of metastatic potential and tumourigenicity of colorectal cancer. *Oncotarget* 2017;8:24354–61.
  51. Kuo PL, Huang YL, Hsieh CC, Lee JC, Lin BW, Hung LY. STK31 is a cell-cycle regulated protein that contributes to the tumourigenicity of epithelial cancer cells. *PLoS One* 2014;9:e93303.
  52. Fok KL, Chung CM, Yi SQ, *et al.* STK31 maintains the undifferentiated state of colon cancer cells. *Carcinogenesis* 2012;33:2044–53.
  53. Pekow J, Hutchison AL, Meckel K, *et al.* miR-4728-3p functions as a tumour suppressor in ulcerative colitis-associated colorectal neoplasia through regulation of focal adhesion signaling. *Inflamm Bowel Dis* 2017;23:1328–37.
  54. Chen R, Rabinovitch PS, Crispin DA, *et al.* DNA fingerprinting abnormalities can distinguish ulcerative colitis patients with dysplasia and cancer from those who are dysplasia/cancer-free. *Am J Pathol* 2003;162:665–72.
  55. Rabinovitch PS, Dziadon S, Brentnall TA, *et al.* Pancolonial chromosomal instability precedes dysplasia and cancer in ulcerative colitis. *Cancer Res* 1999;59:5148–53.