Journal of Crohn's and Colitis, 2025, **19(3)**, jjaf026
https://doi.org/10.1093/ecco-jcc/jjaf026
Advance access publication 13 February 2025
Original Article

A transcriptomic score to classify the inflammation-dysplasia-cancer sequence lesions in inflammatory bowel disease

Anneline Cremer[*,1,2], Nicolas Rosewick[2], Maxfield Kelsey[3], Eric Trépo[1,2], Frédérick Libert[4], Martine De Vos[5], Filip Baert[6], Tom Moreels[7], Edouard Louis[8], Jean-François Rahier[9], Pieter Demetter[2], John M. Sedivy[3], Séverine Vermeire[10], Denis Franchimont[1,2]

[1]Department of Gastroenterology, HUB Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium
[2]Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium
[3]Center on the Biology of Aging, and the Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI, United States
[4]Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIBHM), Université Libre de Bruxelles, Brussels, Belgium
[5]Department of Gastroenterology, University Hospital Ghent, Ghent, Belgium
[6]Department of Gastroenterology, AZ Delta, Roeselare, Belgium
[7]Department of Gastroenterology, University Hospital Antwerp, Edegem, Belgium
[8]Department of Gastroenterology, University Hospital Liège, Liège, Belgium
[9]Department of Gastroenterology, CHU UCL Namur site Mont-Godinne, Université Catholique de Louvain, Yvoir, Belgium
[10]Department of Gastroenterology, University Hospital Leuven, Leuven, Belgium

[*]Corresponding author: Anneline Cremer, Department of Gastroenterology and Hepatology, Erasme University Hospital, Route de Lennik 808, 1070 Brussels, Belgium (anneline.cremer@hubruxelles.be).

Abstract

Background and aims: Inflammatory bowel disease (IBD) is associated with a higher risk of developing colorectal cancer, according to the inflammation-dysplasia-cancer (IDC) sequence from inflammation to colitis-associated colorectal cancer (CAC). The objective of this study was to identify and generate a transcriptomic signature and score, related to the IDC sequence, that could ultimately classify dysplasia and cancer in IBD.

Methods: Demographics, clinical parameters, histological characteristics, and RNA-sequencing data were evaluated on 134 formalin-fixed paraffin-embedded lesions from 2 independent cohorts of IBD patients with low- or high-grade dysplasia (LGD, HGD) and/or CAC. An ordinal logistic regression screened for significant IDC sequence-associated genes that were computed in a transcriptomic signature score.

Results: Principal component analysis and unsupervised clustering on 1% of the most variable genes showed a good clustering between the 4 lesion groups (Normal Mucosa, Inflamed Mucosa, LGD/HGD, and CAC). A gene signature was identified on 27 genes that correlated with the lesion groups in the exploratory cohort. The most weighted gene in this transcriptomic signature was the long non-coding regulatory RNA KCNQ1OT1, a gatekeeper against genomic instability and transposon activation. Based on the expression of these 27 genes, we built and validated a transcriptomic signature score to classify dysplasia and CAC. The overall accuracy of the transcriptomic signature score was 85.71% in the exploratory cohort and 90.91% in the validation cohort.

Conclusion: We identified a tissue-based transcriptomic score to classify IDC lesions in IBD patients and uncovered some of the pivotal genes in carcinogenesis related to inflammation in IBD.

Key words: RNA expression; carcinogenesis; inflammatory bowel disease; dysplasia; colorectal cancer.

1. Introduction

Patients with inflammatory bowel disease (IBD) (Crohn's disease [CD] and ulcerative colitis) have an increased risk of developing dysplasia and colorectal cancer (CRC), namely colitis-associated colorectal cancer (CAC) compared to the general population.[1] Carcinogenesis related to IBD follows the inflammation-dysplasia-cancer (IDC) sequence from inflammation to low-grade (LGD), high-grade dysplasia (HGD), and cancer. Chronic mucosal inflammation is the obvious trigger as disease extent, duration, and activity are the most prominent clinical risk factors for CAC.[2]

The management of patients with dysplasia, more specifically with LGD lesions, remains a challenge for most clinicians as some patients/LGD lesions are associated with a higher risk of progression to CAC. Lesion (flat or polypoid, multifocal lesions) or patient (sclerosing cholangitis) characteristics are the only markers to guide therapeutic decisions toward endoscopic resection or surgery.[1] A signature of the IDC sequence could potentially help stratify lesions at risk for developing CAC and adapt follow-up and treatment decisions

accordingly. This requires in the first place the description of molecular features of these mucosal lesions before further prognostic validation in prospective cohorts.

This IDC progression results from the transformation and progressive accumulation of chromosomal abnormalities, somatic mutations, and epigenetic modifications that is distinct from the adenoma-carcinoma sequence in the general population.[3] The same driver genes of the canonical pathways of carcinogenesis are present and may contribute to the development of CAC,[4] but their occurrence has been reported repeatedly to differ in timing and frequency compared to sporadic CRC.[5–7] Most notably, *TP53* mutations are typically early events during tumor progression in CAC, whereas *TP53* mutations rarely occur in the adenomatous precursors of sporadic CRC. On the contrary, *APC* and *KRAS* mutations are reported to be less prevalent and occur later in CAC than in sporadic CRC.[8–10] Mutations may not be the first event predisposing to cancer development in IBD. The IDC sequence seems to result more from a transcriptional rewiring, through non genetic events, such as epigenetic changes, early during tumorigenesis.[11] Indeed, careful analysis of the clonality of genomic changes of the surrounding mucosa of associated early lesions and/or CAC indicates substantial genetic heterogeneity and the absence of a clear genetic field effect for cancer risk.[12] Today, it is difficult to reconcile all these reported findings of the IDC sequence in a comprehensive model.

Early and late transcriptional changes during the IDC sequence can highlight some of the transcriptional rewiring pathways that are dysregulated and participate to the tumorigenesis related to inflammation. The best illustration is the transcriptome-based classification of CRC with the large-scale international effort resulting in an amalgamation of 4 consensus molecular subtypes (CMS).[13] To the same extent, some transcriptomic changes during the IDC sequence must help the classification of dysplasia and CAC with some having a mechanistic input in the transformation of the colonocytes while others only reflecting the consequences of dedifferentiated colonocytes and their environment.

The objectives of this study were to detect in the IDC sequence genes which expression is correlated with lesion groups to identify/generate a transcriptomic gene signature and score that best classify IDC lesions in a large multicenter retrospective cohort of IBD patients.

2. Materials and methods

This is a large national cohort study conducted across 7 Belgian tertiary centers within the Belgian Inflammatory Bowel Disease Research and Development Group. We retrospectively evaluated all patients with histologically confirmed IBD, diagnosed with at least 1 episode of dysplasia (LGD or HGD) and/or CAC between January 1, 1990, and December 31, 2016. Ethics Committee reference number approval: P2013/331 (February 25, 2014). We have built one of the world's largest cohorts of dysplasia and CAC in IBD patients with 1183 lesions in 541 patients between 1990 and 2016. For characterization of Dysplastic/CAC lesions, see Cremer et al.[14] Among those, 150 formalin-fixed paraffin-embedded (FFPE) samples underwent RNA-sequencing and 134 were analyzed (flowchart in Figure S1). Clinical characteristics and demographic details of the patients were collected by electronic chart review. Two cohorts of patients were analyzed. An exploratory cohort was used to build the transcriptomic score, which was validated on the samples of our validation cohort. Normal Mucosa and Inflamed Mucosa are defined in Supplementary Materials. Methodology about the histological classification of the lesions, CMSs analysis, and Gene set enrichment analysis is described in Supplementary Materials.

2.1. RNA-seq preprocessing

RNA extraction, quality assessment of RNA, library preparation, and RNA-sequencing methodology are described in Supplementary Materials. Sequence quality was assessed using FastQC to explore quality distribution and other sequence characteristics. Raw reads were aligned on human hg38 genome using the STAR aligner (v2.7.5a).[15] Read counts per gene (ENSEMBL v100) were extracted using featureCounts.[16] Gene expression analysis was performed using Differential Expression Analysis using Sequence Counts version 2 (DESeq2) R package.[17] Samples with high levels of rRNA read count (>25% of total read count) were further discarded from analysis. Only genes expressed (>10reads) in a minimum of 10 samples were used for further analysis (36 193 out of 60 668). DESeq2's normalization using Variance Stabilizing Transformation (vst) was used for clustering, visualization, and gene signature construction. Unsupervised hierarchical clustering was performed using Euclidean distance and complete linkage on DESeq2's vst read counts. Analysis of retrotransposon expression is described in Supplementary Materials.

2.2. Gene signature and transcriptomic score

In order to detect genes which expression is correlated with lesion groups (Normal Mucosa→Inflamed Mucosa→LGD/HGD→CAC) in the IDC sequence, we performed an ordinal logistic regression (olm) for each 36 193 genes (expression(vst)~group). The construction of the gene signature and the transcriptomic score is described in Figure 2 and in Supplementary Materials.

2.3. The Cancer Genome Atlas

The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) data were retrieved using TCGAbiolinks R package.[18] Data processing is described in Supplementary Materials.

3. Results

3.1. Study population

A total of 95 neoplastic lesions inside diseased areas (56LGD, 7HGD, and 32CAC), 25 histologically inflamed mucosa biopsies taken during follow-up colonoscopy of IBD patients, and 14 histologically normal mucosa biopsies taken during routine screening colonoscopy of healthy individuals were included in the study. The exploratory cohort included 73 lesions (45LGD, 3HGD, and 25CAC) from the same number of IBD patients, while the validation cohort included 22 lesions (11LGD, 4HGD, and 7CAC) from 16 IBD patients (Table S1). Demographics and clinical parameters of the study population are summarized in Table 1. Histological and transcriptional classification of dysplasia and CAC are described in Supplementary Data (Table S2).

Table 1. Demographics and clinical parameters of the study population (most advanced grade) (*n* = number of patients).

Variables	LGD/HGD (*n* = 50)	CAC (*n* = 27)	Normal mucosa(*n* = 14)	Inflamed mucosa(*n* = 25)
Type of IBD, *n* (%)				
CD	16 (32%)	14 (52%)	NA	6 (24%)
UC	33 (66%)	12 (44%)		18 (72%)
Unclassified IBD	1 (2%)	1 (4%)		1 (4%)
Male, *n* (%)	35 (70%)	14 (52%)	4 (29%)	11 (44%)
Age (yr) at IBD diagnosis, median (IQR)	50 (36-65) (*n* = 47)	33 (25-49) (*n* = 27)	NA	24 (22-39) (*n* = 25)
Follow-up duration (yr) after IBD diagnosis, median (IQR)	15 (8-22) (*n* = 47)	28 (16-33) (*n* = 27)	NA	NA
Deceased, *n* (%)	3 (6%)	10 (37%)	NA	NA
Smoking status, *n* (%)				
Active	2 (4%)	4 (15%)		
Stopped	16 (32%)	5 (18.5%)		
No	17 (34%)	13 (48%)		
Unknown	15 (30%)	5 (18.5%)	14 (100%)	25 (100%)
Age (yr) at diagnosis of the index lesion, median (IQR)	57 (48-71) (*n* = 50)	56 (46-63) (*n* = 27)	49 (34-58)	27 (23-46) (*n* = 25)
Duration (yr) of IBD at time of index lesion diagnosis, median (IQR)	8 (1-14) (*n* = 47)	18 (9-26) (*n* = 27)	NA	1 (0-5)
Follow-up duration (mo) after diagnosis of the index lesion, median (IQR)	5 (2-8) (*n* = 50)	4 (2-9) (*n* = 27)	NA	NA
Family history of CRC, *n* (%)				
Yes	3 (6%)	1 (4%)	0	0
No or unknown	47 (94%)	26 (96%)	14	25
Associated PSC, *n* (%)	4 (8%)	3 (11%)	0	0

*Patients were classified according to the most advanced lesion during colonoscopy or at surgery performed at follow-up.
Abbreviations: CAC, colitis-associated colorectal cancer; CD, Crohn disease; CRC, colorectal cancer; HGD, high-grade dysplasia; IBD, inflammatory bowel disease; IQR, interquartile range; LGD, low-grade dysplasia; PSC, primary sclerosing cholangitis; UC, ulcerative colitis.

3.2. Global gene expression profiling according to stages of neoplasia

After alignment on the human genome, an average of 22 768 770 (min:625 082–max:63 056 801) and 14 879 093 (min:553 740–max:40 332 689) RNA-seq reads aligned on ENSEMBL genes for the exploratory and validation cohorts, respectively, and were further used in gene expression analysis. On the entire cohort, 36 193 genes were identified in the sequence (Normal Mucosa-Inflamed Mucosa-LGD/HGD-CAC). Principal component analysis (PCA) (Figure 1A) and unsupervised clustering of the top 1% (362 genes) of the most variable genes from 36 193 genes and 134 samples (Figure 1B, Table S3) showed an overall good clustering between the 4 lesion groups (Normal Mucosa-Inflamed Mucosa-LGD/HGD-CAC). PCA and unsupervised clustering of CAC samples highlighted a good segregation especially between CMS4 and CMS 1, 2, and 3 (Figure S2). Gene set enrichment analyses are described in Supplementary Data (Figure S3 and S4).

3.3. A transcriptomic signature score specific to the IDC sequence

Using an ordinal logistic regression (Figure 2A), a gene signature of 27 genes that correlated with lesion groups was identified based on our exploratory lesions (Figure 3). A gene signature weight for the 27 genes was computed using a backward elimination strategy (Table S4). Among these 27 genes, 23 were downregulated, and 4 were upregulated. Heatmap showed that 4 groups were clustered together and very well separated from each other when considering only the 27 genes of the gene signature (Figure 4). Based on the expression of the 27 genes, we generated a transcriptomic score to classify dysplasia (LGD/HGD) and CAC in IBD patients. Coefficients for all genes are available in Supplementary Data (Table S5). The overall correct classification rate based on the olm probabilities of the transcriptomic score was 85.71% in our exploratory cohort. We then validated this transcriptomic score in an independent validation cohort with an overall correct classification rate of 90.91% (Table 2, Figure S5). The sensitivity and specificity of the score are available in Supplementary Data.

3.4. Replication of the transcriptomic signature score in the TCGA-COAD cohort

The transcriptomic signature score was applied to the publicly available TCGA-COAD cohort with an overall correct classification of 16.00% (Table S6, Figure S6). Interestingly TCGA normal samples were perfectly classified as normal samples, while TCGA-COAD samples were misclassified highlighting transcriptomic differences between CAC and sporadic CRC.

3.5. KCNQ1OT1 and retrotransposable element expression in the IDC sequence

The gene signature weight for the 27 genes using a backward elimination strategy (Table S4) showed that the most

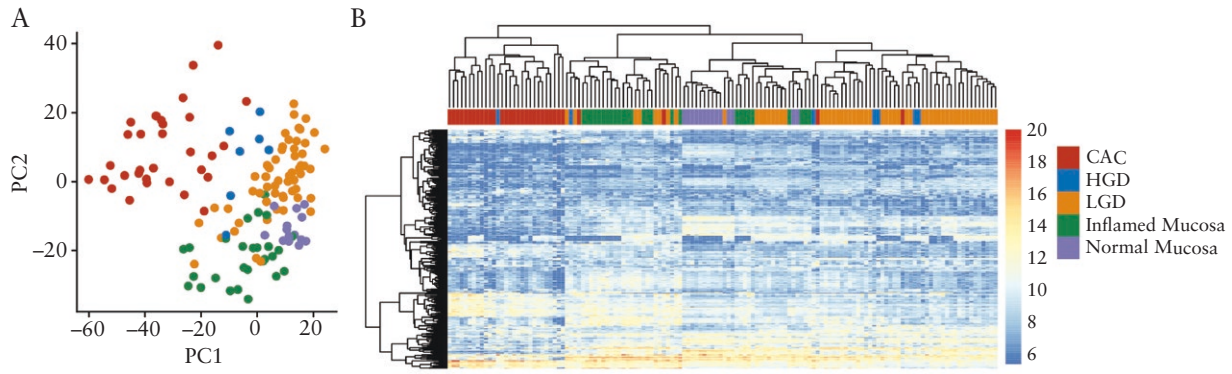

Figure 1. Unsupervised transcriptomic profile of the IDC sequence. (**A**) PCA bi-plot of all samples colored by lesion groups and (**B**) heatmap of the top 1% of most variable genes (n = 362 genes) within all samples reflects lesion groups of the IDC sequence. IDC, inflammation-dysplasia-cancer; PCA, principal component analysis.

weighted gene in this transcriptomic signature is KCNQ1OT1 (KCNQ1 opposite strand/antisense transcript 1 of CDKN1C [cyclin-dependent kinase inhibitor 1C]), an un-spliced long non-coding regulatory RNA (lncRNA). KCNQ1OT1 is a paternally expressed lncRNA that is involved, in *cis*, in the transcriptional silencing of 8 to 10 nearby protein-coding genes and, in *trans*, in heterochromatin reorganization and repression of transposon activation and retrotransposition.[19,20] KCNQ1OT1 guards against genomic instability and senescence through sequence-specific DNA methylation and hence repression of retrotransposable elements (RTEs), particularly evolutionarily young LINE1 (L1HS) and Alu subfamilies (AluY).[19] Indeed, RTEs promote genomic instability, insertional mutagenesis, and therefore tumorigenesis, as has been reported in a wide variety of cancers.[21,22] In order to evaluate the potential relationship of KCNQ1OT1 with the expression of RTEs in the IDC sequence, we first looked at the expression of RTEs according to the lesion group. Interestingly, there was a significant increase of L1HS, AluY, and Human Endogenous Retrovirus type K (HERVK) subfamily expression in LGD/HGD and CAC groups compared to the control group (Figure S7A). HERVK expression was significantly increased as early as in the inflammation group compared to the control group. To demonstrate a direct association between KCNQ1OT1 and these RTEs, we then looked at the expression of KCNQ1OT1 and RTEs in the different lesion groups (Figure S7B) and in each of the samples (Figure S7C) and found a direct correlation in the expression of KCNQ1OT1 and RTEs across the lesion groups of IDC sequence.

Since KCNQ1OT1 is a negative regulator of RTEs, a repressor being positively correlated with its target is counterintuitive. One possible explanation is that the concomitant increase of RTEs and KCNQ1OT1 results from a positive regulatory feedback loop of RTEs on KCNQ1OT1 expression. Inflammation has been associated with RTE de-repression[23]; indeed, as found here with HERVK expression. This would suggest that a primary increase of RTEs during inflammation and later in tumorigenesis might secondarily induce KCNQ1OT1 expression. Inversely, it is well known that LINE1 expression can trigger a type I interferon (IFN-I) response and promote inflammation.[24] The L1 retrotransposition cycle produces cytosolic cDNA, which the cell takes as evidence of an invading viral pathogen via sensing by Cyclic GMP-AMP Synthase (cGAS).[24] The cGAS-Stimulator of Interferon Genes axis then orchestrates an IFN-I response through the activation of Interferon Regulatory Factor 3 (IRF3) and Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-kB).[25] We thus interrogated the Encyclopedia of DNA Elements Consortium (National Human Genome Research Institute) for transcription factor binding to the KCNQ1OT1 promoter (Table S7). Examining transcription factor Chromatin Immunoprecipitation Sequencing data, we found that both IRF3 and NF-kB bind the KCNQ1OT1 promoter in the majority of cell lines assayed (Figure S8).[26] Thus, our data, together with publicly available TF ChIP results, suggest that RTEs may promote KCNQ1OT1 expression through its IFN-responsive promoter, resulting in the concomitant increase of RTEs and KCNQ1OT1 along the IDC progression sequence.

4. Discussion

In this study, we evaluated the transcriptome of 134 different lesions of the IDC sequence in IBD. We identified a gene signature of 27 genes of which the expression was correlated with the IDC sequence in an exploratory cohort, from normal and inflamed mucosa to dysplastic lesions and CAC. This transcriptomic signature helped generate a transcriptomic score that showed a good prediction rate in reclassifying the lesions in a small validation cohort. It is important to point out that this score is not intended to predict future progression from inflammation to dysplasia and CAC. It is a class prediction score that helps the correct classification of the lesion. Remarkably, this IDC signature score failed to reclassify the sporadic CRC lesions from the publicly available TCGA cohort suggesting that the IDC gene score is specific or only representative of the inflammation-induced tumorigenesis. Yet, some of these 27 genes have already been reported in separate studies as relevant prognostic marker for disease progression and/or overall survival in sporadic CRC and in cancer overall.[19,20,27–31]

This 27-gene signature may reveal some of the relevant players in the IDC sequence. All these genes are related to colonocytes or infiltrated immune cells in the dysplasia or adenocarcinoma specimens. Obviously, it is uncertain whether they reflect causes or consequences of the tumorigenesis process. Among the 27 genes in the signature, several have notable biological relevance. For example, CD163L1 is likely associated with macrophage activity,[32] CD177 may relate to neutrophil infiltration,[33] and PYY could be linked

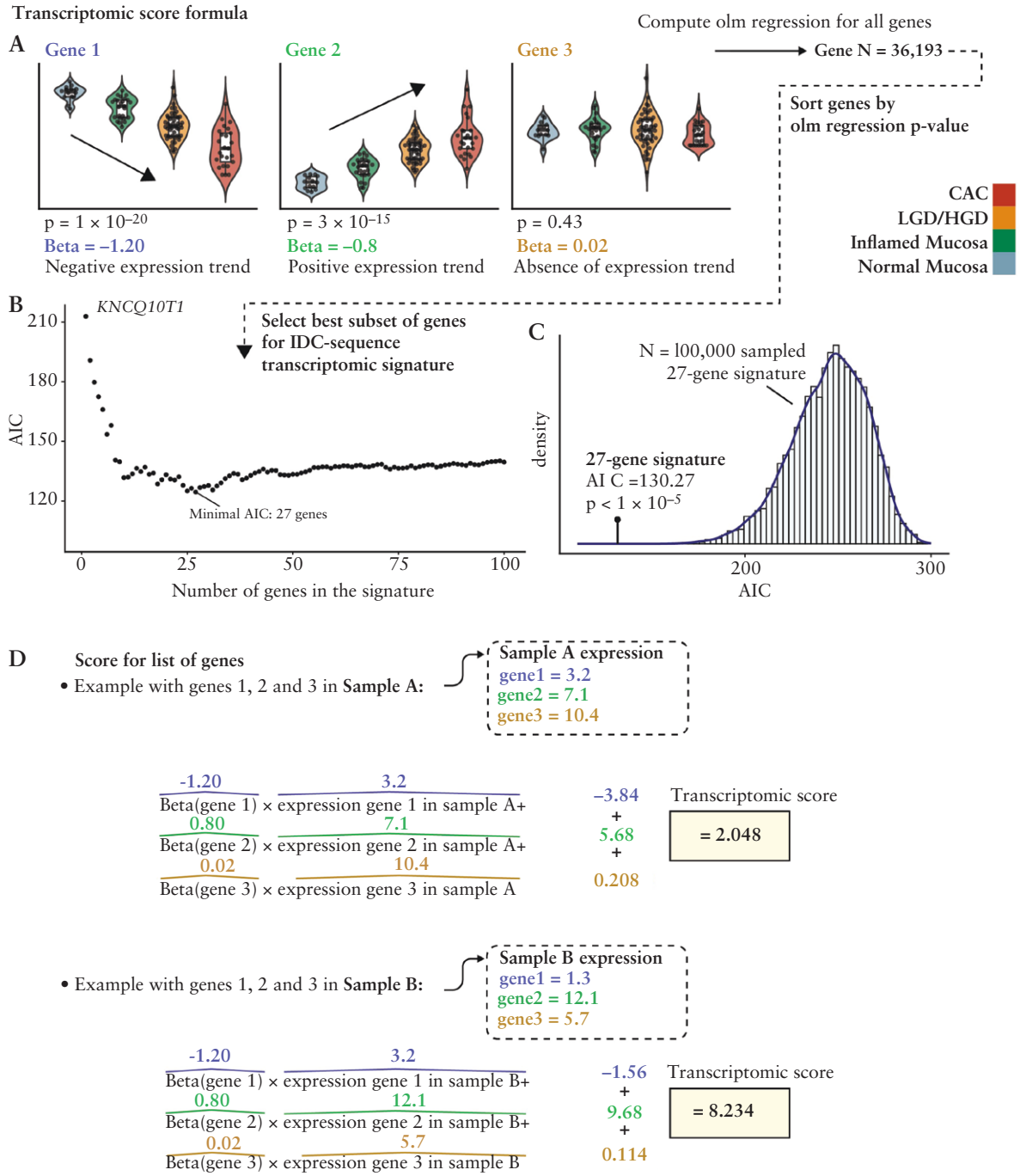

Figure 2. Schematic description of the transcriptomic score formula. (**A**) For each gene, a *P*-value and a beta regression are computed using an ordinal logistic regression. The beta represents the expression trend of the associated gene relative to IDC sequence (ie, if positive the gene follows a positive trend from Normal Mucosa toward CAC eg Gene2; in opposite, if negative, the gene follows a negative trend from Normal Mucosa toward CAC eg Gene1). Non-significant *P*-value (ie, $P > .05$) represents an absence of expression trend (eg Gene3). (**B**) Selection of the best gene signature based on the Akaike information criterion (AIC). Each point (x,y) represents a gene list of length x with an AIC of y. Gene lists were built based on genes ordered by olm P-value computed on the IBD-IDC sequence (Normal Mucosa→Inflamed Mucosa→LGD/HGD→CAC). The transcriptional signature score was computed for 100 incremental gene lists (from $n = 1$gene to $n = 100$ genes in the signature). An AIC for each gene list assessing the quality of the model (ie, the transcriptional signature score reflects the IBD-IDC sequence) was computed. The gene list with the lowest AIC (first 27 genes ordered by olm *P*-value) was retained as IBD-IDC gene signature. (**C**) To assess the selected 27-gene signature robustness, we computed an empirical *P*-value for the $N = 27$genes in the IBD-disease signature by counting the number of size-matched sampled gene list ($n = 100\,000$) with a lower AIC than the 27-gene IBD-disease signature. (**D**) In order to build a transcriptomic score using a list of genes for a sample of interest, we multiply each gene and its associated beta with the gene expression value in the sample of interest. Hereby 2 examples for Sample A and B using Gene 1, 2, and 3. CAC, colitis-associated colorectal cancer; IBD, inflammatory bowel disease; IDC, inflammation-dysplasia-cancer; HGD, high-grade dysplasia; LGD, low-grade dysplasia.

to **short-chain fatty acids** and the glucagon-like peptide-1 axis[34] The most weighted gene in this transcriptomic signature is KCNQ1OT1, an un-spliced lncRNA, that is involved in transposon activation and heterochromatin reorganization.[19] KCNQ1OT1 was initially identified in the imprinting control region 2 on human chromosome 11p15,[35] which is

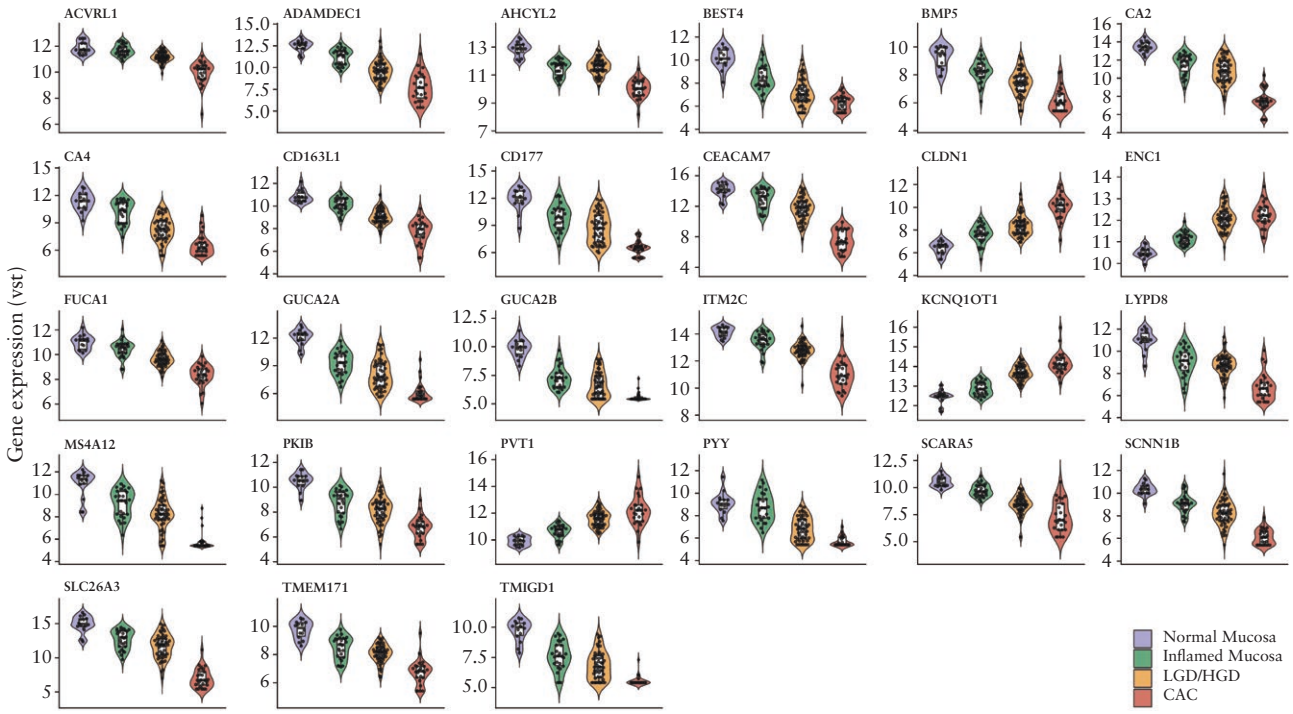

Figure 3. Expression of the 27 genes included in the IBD-IDC gene expression signature in the exploratory cohort. The expression scale is in DESeq2's vst read count. IBD, inflammatory bowel disease; IDC, inflammation-dysplasia-cancer.

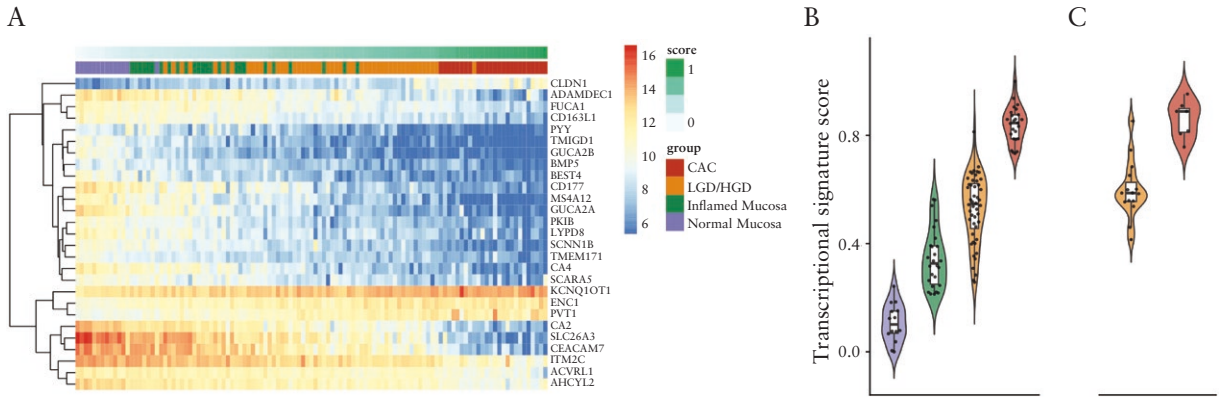

Figure 4. Transcriptional signature score based on the 27-gene signature. (**A**) Heatmap of the 27 genes of the gene signature in the exploratory cohort. The top bar shows the signature score and the lesion groups of the IDB-IDC sequence. (**B**) Transcriptional signature score of the exploratory cohort stratified by IBD-IDC sequence lesion groups. (**C**) Transcriptional signature score of the validation cohort stratified by IBD-IDC sequence lesion groups. IBD, inflammatory bowel disease; IDC, inflammation-dysplasia-cancer.

the most epi-mutated region of the congenital imprinting disorder, Beckwith-Wiedemann Syndrome, associated with an increased tumor risk.[36,37] KCNQ1OT1 seems clinically deleterious as it represents a poor prognostic marker for disease progression in these cancers.[20] A recent study highlights the specific role of KCNQ1OT1 in guarding against genomic instability and senescence through sequence-specific DNA methylation and hence repression of RTEs, particularly evolutionarily young LINE1 (L1HS) and Alu subfamilies (AluY).[19] Indeed, RTEs promote genomic instability, insertional mutagenesis, and therefore tumorigenesis, as has been reported in a wide variety of cancers.[21,22] Widespread LINE1 insertions have been documented as an early event in gastrointestinal cancer evolution.[21,38] In this study, we showed an increase in the expression of RTEs along the IDC sequence and a direct correlation between KCNQ1OT1 and RTE expression across lesions of the IDC sequence. We speculate that KCNQ1OT1 may initially be able to restrain RTE de-repression and expression early on during chronic inflammation, but as insults to epigenomic stability accrue throughout the IDC sequence, despite the increased expression of KCNQ1OT1, RTEs can ultimately surpass this defense mechanism and promote inflammation-induced tumorigenesis.[12,39] Thus, the predominant weight assigned to KCNQ1OT1 in this transcriptomic signature may in fact make it a surrogate marker of global retrotransposon activation, suggesting in turn a pathogenic role for RTEs in the IDC sequence.

In parallel, the transcriptomic signature identified genes that code for structural and/or functional proteins in various cell systems. Interestingly, most transcripts, except 3 of them,

Table 2. Classification rates of the transcriptomic signature score in the exploratory and validation cohorts.

Exploratory cohort		Lesion groups classification based on ordinal logistic probabilities				% correct classification
		Normal mucosa	Inflamed mucosa	LGD/HGD	CAC	
Real group	Normal mucosa	13	1	0	0	**92.86%**
	Inflamed mucosa	0	18	7	0	**72.00%**
	LGD/HGD	0	7	40	1	**83.33%**
	CAC	0	0	0	25	**100.00%**
Total correct classification						**85.71%**
Validation cohort		Lesion groups classification based on ordinal logistic probabilities				
		Normal mucosa	Inflamed mucosa	LGD/HGD	CAC	
Real group	LGD/HGD	0	0	13	2	**86.67%**
	CAC	0	0	0	7	**100%**
Total correct classification						**90.91%**

Abbreviations: CAC, colitis-associated colorectal cancer; HGD, high-grade dysplasia; LGD, low-grade dysplasia.

in the signature are inversely associated with CAC, suggesting that the downregulation of key pathways is a hallmark of tumorigenesis. Claudin-1, a tight junction-specific protein, is one of the 3 upregulated genes in the signature. Claudin-1 overexpression has already been reported in active IBD and in IBD-associated dysplasia and CAC[30] and was correlated with inflammation. Several of the downregulated genes have already been reported so far in IBD and IBD-related dysplasia and cancer such as the carbonic anhydrases CA2 and CA4 genes.[27] Interestingly, CA2 and CA4 may reflect functional interrelationships that are relevant to the IDC sequence. The observed downregulation of CA2 in our study contrasts with findings in sporadic CRC, where CA2 overexpression has been linked to poor prognosis.[40] This discrepancy may reflect differences between colitis-associated cancer (CAC) and sporadic CRC or variations in gene vs protein expression. Humoral immune response, immunoglobulin complex, and antigen binding were the most enriched pathways in the top 1% of mostly regulated genes. ADAMDEC1 is an anti-inflammatory secreted peptidase belonging to the disintegrin metalloproteinase family that plays a role in the crosstalk between dendritic cells and germinal center T-helper cells and is almost exclusively expressed in the gastrointestinal tract.[31,41] TMIGD1 is part of the immunoglobulin domain-containing cell adhesion molecules involved in several processes, including cell differentiation and apoptosis. TMIGD1 expression is lower in active CD, progressively lost in sporadic CRC, and associated with poor CRC overall survival.[29] SCNN1B, a sodium transporter and known target of amiloride included in the signature, is notable for its therapeutic implications. While amiloride has been reported to inhibit colon cancer cell growth in vitro,[42] there are conflicting data regarding diuretic use and colon cancer mortality.[43] These findings highlight the need for further research into its potential role in CAC progression and its utility as a therapeutic target. BMP5 gene modulates epithelial-mesenchymal transition and its loss of expression correlates with recurrence and poor prognosis in human CRC.[28]

One limitation of our study is the use of FFPE-based sequencing, which can introduce biases; however, rigorous quality control and consistent processing across samples ensured robust differential expression, and future validation in fresh or Trizol-preserved specimens will confirm its reproducibility. Another limitation is the possibility of overfitting when deriving a 27-gene signature from the analyzed sample size. However, the robust performance of the signature in the validation cohort supports its generalizability. It is important to note that we observed significant differences in gene expression between normal control tissue and inflamed mucosa, which underscores the relevance of including inflamed mucosa in our analysis. This choice was made to capture the full spectrum of transcriptional changes in the IDC sequence, reflecting the natural history of CAC.[44] While inflammation introduces variability in gene expression, driven by immune cell infiltration and widespread transcriptional disturbances, this variability represents a critical early step in the IDC sequence.[45,46] Excluding inflammation could lead to a different set of genes with more linear expression patterns but would fail to represent the early events initiated by inflammation that drive tumorigenesis in IBD.

In conclusion, we have identified a transcriptomic 27-gene signature associated with the IDC sequence and generated promising hypotheses about pivotal genes and underlying mechanisms involved in its progression, providing a basis for future functional validation studies. The most discriminant top gene in this signature, the non-coding RNA KCNQ1OT1, is consistent with the activation of RTEs that may be involved in the progression of the IDC sequence. The transcriptomic score generated here helps the classification of the lesions of the IDC sequence but should be further evaluated for its prognostic value in prospective independent cohorts.

Acknowledgments

This study was promoted by the Belgian Inflammatory Bowel Disease Research and Development (BIRD) group.

Author contributions

Anneline Cremer: conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article. Denis Franchimont: conception and design of the study, analysis and interpretation of data, drafting the article. Nicolas Rosewick, Maxfield Kelsey: analysis and interpretation of data, revising the article critically for important intellectual content. Other authors: acquisition of data, revising the article critically for important intellectual content. All authors: final approval of the version to be submitted.

Funding

This study was funded through research grants from Erasme Foundation (A.C.), Research Foundation against Cancer-Belgium (A.C.), Televie (A.C.), and National Institutes of Health grant (grants P01 AG051449 and R01 AG016694) (J.M.S.).

Conflicts of interest

A.C. obtained a doctoral research grant from the Fonds Erasme for Medical Research. D.F. and S.V. are senior scientists of the FNRS/FWO. J.M.S. is a cofounder and SAB chair of Transposon Therapeutics, holds equity in PrimeFour and Atropos Therapeutics, and consults for Atropos Therapeutics and Longaevus Technologies. All other authors have no conflict of interest to disclose relevant to the manuscript.

Data availability

Data and analytic methods underlying this article will be made available to researchers upon reasonable request to the corresponding author.

Writing assistance

The authors did not have any writing assistance in producing this manuscript.

Profiling accession numbers

Data will be available via the NCBI's Gene Expression Omnibus after approval of the manuscript.

Supplementary material

Supplementary material is available at *ECCO-JCC* online.

References

1. Beaugerie L, Itzkowitz SH. Cancers complicating inflammatory bowel disease. *N Engl J Med*. 2015;**372**:1441–1452. doi:10.1056/nejmra1403718
2. Rutter M, Saunders B, Wilkinson K, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology*. 2004;**126**:451–459. doi:10.1053/j.gastro.2003.11.010
3. Shah SC, Itzkowitz SH. Colorectal cancer in inflammatory bowel disease: mechanisms and management. *Gastroenterology*. 2022;**162**:715–730.e3. doi:10.1053/j.gastro.2021.10.035
4. Itzkowitz SHI, Cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *AJP Gastrointest Liver Physiol*. 2004;**287**:G7–17. doi:10.1152/ajpgi.00079.2004
5. Rajamäki K, Taira A, Katainen R, et al. Genetic and epigenetic characteristics of inflammatory bowel disease–associated colorectal cancer. *Gastroenterology*. 2021;**161**:592–607. doi:10.1053/j.gastro.2021.04.042
6. Baker AM, Cross W, Curtius K, et al. Evolutionary history of human colitis-associated colorectal cancer. *Gut*. 2019;**68**:985–995. doi:10.1136/gutjnl-2018-316191
7. Chatila WK, Walch H, Hechtman JF, et al. Integrated clinical and genomic analysis identifies driver events and molecular evolution of colitis-associated cancers. *Nat Commun*. 2023;**14**:110. doi:10.1038/s41467-022-35592-9
8. 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**:278–287.e6. doi:10.1053/j.gastro.2016.04.001
9. Du L, Kim JJ, Shen J, Chen B, Dai N. KRAS and TP53 mutations in inflammatory bowel disease associated colorectal cancer: a meta-analysis. *Oncotarget*. 2017;**8**:22175–22186. doi:10.18632/oncotarget.14549
10. Chatila WK, Walch HS, Benhamida J, et al. Genomic alterations in colitis-associated cancers in comparison to those found in sporadic colorectal cancer and present in precancerous dysplasia. *J Clin Oncol*. 2020;**38**:191–191. doi:10.1200/jco.2020.38.4_suppl.191
11. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and intercellular rewiring of the human colon during ulcerative colitis. *Cell*. 2019;**178**:714–730.e22. doi:10.1016/j.cell.2019.06.029
12. Chatila WK, Walch H, Hechtman JF, et al. Integrated clinical and genomic analysis identifies driver events and molecular evolution of colitis-associated cancers. *Nat Commun*. 2023;**14**:1–13. doi:10.1038/s41467-022-35592-9
13. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;**21**:1350–1356. doi:10.1038/nm.3967
14. Cremer A, Demetter P, De Vos M, et al.; Belgian Inflammatory Bowel Disease Research and Development (BIRD) Group. Risk of development of more-advanced lesions in patients with inflammatory bowel diseases and Dysplasia. *Clin Gastroenterol Hepatol*. 2020;**18**:1528–1536.e5. doi:10.1016/j.cgh.2019.05.062
15. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;**29**:15–21. doi:10.1093/bioinformatics/bts635
16. Liao Y, Smyth GK, Shi WF. An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;**30**:923–930. doi:10.1093/bioinformatics/btt656
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;**15**:550. doi:10.1186/s13059-014-0550-8
18. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;**44**:e71. doi:10.1093/nar/gkv1507
19. Zhang X, Jiang Q, Li J, et al. KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA–DNA triplexes and HP1 binding. *Nat Cell Biol*. 2022;**24**:1617–1629. doi:10.1038/s41556-022-01008-5
20. Xia F, Wang Y, Xue M, et al. LncRNA KCNQ1OT1: molecular mechanisms and pathogenic roles in human diseases. *Genes Dis*. 2022;**9**:1556–1565. doi:10.1016/j.gendis.2021.07.003
21. Ewing AD, Gacita A, Wood LD, et al. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res*. 2015;**25**:1536–1545. doi:10.1101/gr.196238.115
22. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, et al.; PCAWG Structural Variation Working Group. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet*. 2020;**52**:306–319. doi:10.1038/s41588-019-0562-0
23. Russ E, Mikhalkevich N, Iordanskiy S. Expression of human endogenous retrovirus Group K (HERV-K) HML-2 correlates with immune activation of macrophages and type I interferon response. *Microbiol Spectr*. 2023;**11**:e0443822. doi:10.1128/spectrum.04438-22
24. De Cecco M, Ito T, Petrashen AP, et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*. 2019;**566**:73–78. doi:10.1038/s41586-018-0784-9
25. Motwani M, Pesiridis S, Fitzgerald KA. DNA sensing by the cGAS–STING pathway in health and disease. *Nat Rev Genet*. 2019;**20**:657–674. doi:10.1038/s41576-019-0151-1
26. Luo Y, Hitz BC, Gabdank I, et al. New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res*. 2020;**48**:D882–D889. doi:10.1093/nar/gkz1062
27. Zhang J, Tsoi H, Li X, et al. Carbonic anhydrase IV inhibits colon cancer development by inhibiting the Wnt signalling pathway

through targeting the WTAP-WT1-TBL1 axis. *Gut*. 2016;**65**:1482–1493. doi:10.1136/gutjnl-2014-308614
28. Chen E, Yang F, He H, et al. Alteration of tumor suppressor BMP5 in sporadic colorectal cancer: a genomic and transcriptomic profiling based study. *Mol Cancer*. 2018;**17**:176. doi:10.1186/s12943-018-0925-7
29. Zabana Y, Lorén V, Domènech E, et al. Transcriptomic identification of TMIGD1 and its relationship with the ileal epithelial cell differentiation in Crohn's disease. *Am J Physiol Gastrointest Liver Physiol*. 2020;**319**:G109–G120. doi:10.1152/ajpgi.00027.2020
30. Mees ST, Mennigen R, Spieker T, et al. Expression of tight and adherens junction proteins in ulcerative colitis associated colorectal carcinoma: upregulation of claudin-1, claudin-3, claudin-4, and β-catenin. *Int J Colorectal Dis*. 2009;**24**:361–368. doi:10.1007/s00384-009-0653-y
31. Jiang L, Wang P, Su M, Yang L, Wang Q. Identification of mRNA signature for predicting prognosis risk of rectal adenocarcinoma. *Front Genet*. 2022;**13**:880945. doi:10.3389/fgene.2022.880945
32. Skytthe MK, Graversen JH, Moestrup SK. Targeting of cd163+ macrophages in inflammatory and malignant diseases. *Int J Mol Sci*. 2020;**21**:5497–5431. doi:10.3390/ijms21155497
33. Zheng C, Li J, Chen H, et al. Dual role of CD177 + neutrophils in inflammatory bowel disease: a review. *J Transl Med*. 2024;**22**:1–13. doi:10.1186/s12967-024-05539-3
34. Tolhurst G, Heffron H, Lam YS, et al. Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. *Diabetes*. 2012;**61**:364–371. doi:10.2337/db11-1019
35. Du M, Zhou W, Beatty LG, Weksberg R, Sadowski PD. The KCNQ1OT1 promoter, a key regulator of genomic imprinting in human chromosome 11p15.5. *Genomics*. 2004;**84**:288–300. doi:10.1016/j.ygeno.2004.03.008
36. Eggermann T, Elbracht M, Schröder C, et al. Congenital imprinting disorders: a novel mechanism linking seemingly unrelated disorders. *J Pediatr*. 2013;**163**:1202–1207. doi:10.1016/j.jpeds.2013.05.017
37. Gorgoulis V, Adams PD, Alimonti A, et al. Cellular senescence: defining a path forward. *Cell*. 2019;**179**:813–827. doi:10.1016/j.cell.2019.10.005
38. Solyom S, Ewing AD, Rahrmann EP, et al. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res*. 2012;**22**:2328–2338. doi:10.1101/gr.145235.112
39. McKerrow W, Wang X, Mendez-Dorantes C, et al. LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. *Proc Natl Acad Sci U S A*. 2022;**119**:e2115999119. doi:10.1073/pnas.2115999119
40. Mboge MY, Mahon BP, McKenna R, Frost SC. Carbonic anhydrases: role in pH control and cancer. *Metabolites*. 2018;**8**:19. doi:10.3390/metabo8010019
41. O'Shea NR, Chew TS, Dunne J, et al. Critical role of the disintegrin metalloprotease ADAM-like DECysin-1 [ADAMDEC1] for intestinal immunity and inflammation. *J Crohn's Colitis*. 2016;**10**:1417–1427. doi:10.1093/ecco-jcc/jjw111
42. Koo JY, Parekh D, Townsend CM, et al. Amiloride inhibits the growth of human colon cancer cells in vitro. *Surg Oncol*. 1992;**1**:385–389. doi:10.1016/0960-7404(92)90040-R
43. Tenenbaum A, Grossman E, Fisman EZ, et al. Long-term diuretic therapy in patients with coronary disease: increased colon cancer-related mortality over a 5-year follow-up. *J Hum Hypertens*. 2001;**15**:373–379. doi:10.1038/sj.jhh.1001192
44. Porter RJ, Arends MJ, Churchhouse AMD, Din S. Inflammatory bowel disease-associated colorectal cancer: translational risks from mechanisms to medicines. *J Crohns Colitis*. 2021;**15**:2131–2141. doi:10.1093/ecco-jcc/jjab102
45. Nardone OM, Zammarchi I, Santacroce G, Ghosh S, Iacucci M. Inflammation-driven colorectal cancer associated with colitis: from pathogenesis to changing therapy. *Cancers (Basel)*. 2023;**15**:2389. doi:10.3390/cancers15082389
46. Grivennikov SI. Inflammation and colorectal cancer: colitis-associated neoplasia. *Semin Immunopathol*. 2013;**35**:229–244. doi:10.1007/s00281-012-0352-6