JYX
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Rajamäki, Kristiina; Taira, Aurora; Katainen, Riku; Välimäki, Niko; Kuosmanen, Anna; Plaketti, Roosa-Maria; Seppälä, Toni T.; Ahtiainen, Maarit; Wirta, Erkki-Ville; Vartiainen, Emilia; Sulo, Päivi; Ravantti, Janne; Lehtipuro, Suvi; Granberg, Kirsi J.; Nykter, Matti; Tanskanen, Tomas; Ristimäki, Ari; Koskensalo, Selja; Renkonen-Sinisalo, Laura; Lepistö, Anna; Böhm, Jan; Taipale, Jussi; Mecklin,

Title: Genetic and Epigenetic Characteristics of Inflammatory Bowel Disease–Associated Colorectal Cancer

Year: 2021

Version: Published version

Copyright: © 2021 the AGA Institute

Rights: CC BY-NC-ND 4.0

Rights url: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the original version:

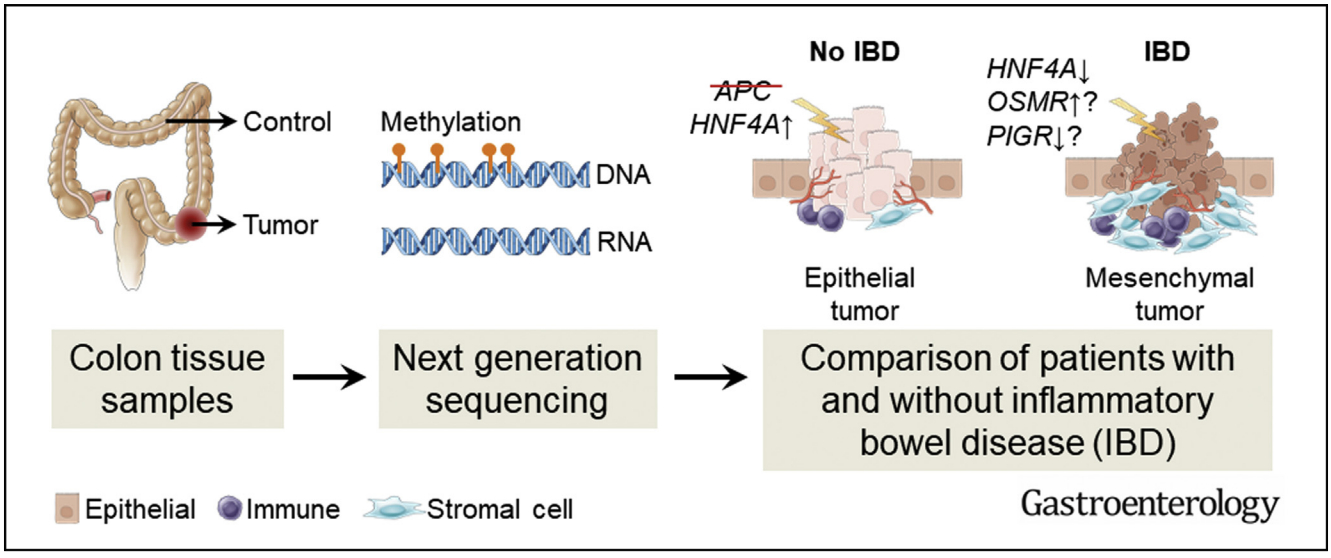
Rajamäki, K., Taira, A., Katainen, R., Välimäki, N., Kuosmanen, A., Plaketti, R.-M., Seppälä, T. T., Ahtiainen, M., Wirta, E.-V., Vartiainen, E., Sulo, P., Ravantti, J., Lehtipuro, S., Granberg, K. J., Nykter, M., Tanskanen, T., Ristimäki, A., Koskensalo, S., Renkonen-Sinisalo, L., . . . Aaltonen, L. A. (2021). Genetic and Epigenetic Characteristics of Inflammatory Bowel Disease–Associated Colorectal Cancer. Gastroenterology, 161(2), 592-607. https://doi.org/10.1053/j.gastro.2021.04.042

Gastroenterology 2021;161:592–607

Genetic and Epigenetic Characteristics of Inflammatory Bowel Disease–Associated Colorectal Cancer

Kristiina Rajamäki,[1,2,*] **Aurora Taira**,[1,2,*] Riku Katainen,[1,2] Niko Välimäki,[1,2] Anna Kuosmanen,[1,2] Roosa-Maria Plaketti,[1,2] Toni T. Seppälä,[2,3,4] Maarit Ahtiainen,[5] Erkki-Ville Wirta,[6] Emilia Vartiainen,[1,2] Päivi Sulo,[1,2] Janne Ravantti,[1,2] Suvi Lehtipuro,[7,8] Kirsi J. Granberg,[7,8] Matti Nykter,[7,8] Tomas Tanskanen,[9] Ari Ristimäki,[2,10] Selja Koskensalo,[11] Laura Renkonen-Sinisalo,[11] Anna Lepistö,[11] Jan Böhm,[5] Jussi Taipale,[2,12,13] Jukka-Pekka Mecklin,[14,15] Mervi Aavikko,[1,2,16] Kimmo Palin,[1,2] and Lauri A. Aaltonen[1,2]

[1]Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; [2]Applied Tumor Genomics Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland; [3]Department of Surgery, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; [4]Department of Surgical Oncology, Johns Hopkins University, Baltimore, Maryland; [5]Department of Pathology, Central Finland Health Care District, Jyväskylä, Finland; [6]Department of Gastroenterology and Alimentary Tract Surgery, Tampere University Hospital, Tampere, Finland; [7]Prostate Cancer Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [8]Tays Cancer Center, Tampere University Hospital, Tampere, Finland; [9]Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland; [10]Department of Pathology, HUSLAB, HUS Diagnostic Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [11]Department of Gastrointestinal Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland; [12]Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; [13]Department of Biochemistry, University of Cambridge, Cambridge, UK; [14]Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland; [15]Department of Education and Research, Central Finland Central Hospital, Jyväskylä, Finland; and [16]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland



See Covering the Cover synopsis on page 380.

BACKGROUND & AIMS: Inflammatory bowel disease (IBD) is a chronic, relapsing inflammatory disorder associated with an elevated risk of colorectal cancer (CRC). IBD-associated CRC (IBD-CRC) may represent a distinct pathway of tumorigenesis compared to sporadic CRC (sCRC). Our aim was to comprehensively characterize IBD-associated tumorigenesis integrating multiple high-throughput approaches, and to compare the results with in-house data sets from sCRCs. **METHODS:** Whole-genome sequencing, single nucleotide polymorphism arrays, RNA sequencing, genome-wide methylation analysis, and immunohistochemistry were performed using fresh-frozen and formalin-fixed tissue samples of tumor and corresponding normal tissues from 31 patients with IBD-CRC. **RESULTS:** Transcriptome-based tumor subtyping revealed the complete absence of canonical epithelial tumor subtype associated with WNT signaling in IBD-CRCs, dominated instead by mesenchymal stroma-rich subtype. Negative WNT regulators *AXIN2* and *RNF43* were strongly down-regulated in IBD-CRCs and chromosomal gains at *HNF4A*, a negative regulator of WNT-induced epithelial–mesenchymal transition (EMT), were less frequent compared to sCRCs. Enrichment of hypomethylation at HNF4α binding sites was detected solely in sCRC genomes.

PIGR and *OSMR* involved in mucosal immunity were dysregulated via epigenetic modifications in IBD-CRCs. Genome-wide analysis showed significant enrichment of noncoding mutations to 5′untranslated region of *TP53* in IBD-CRCs. As reported previously, somatic mutations in *APC* and *KRAS* were less frequent in IBD-CRCs compared to sCRCs. **CONCLUSIONS:** Distinct mechanisms of WNT pathway dysregulation skew IBD-CRCs toward mesenchymal tumor subtype, which may affect prognosis and treatment options. Increased OSMR signaling may favor the establishment of mesenchymal tumors in patients with IBD.

Keywords: Colorectal Cancer; Inflammatory Bowel Disease; Epithelial–Mesenchymal Transition; DNA Methylation; Consensus Molecular Subtype.

Inflammatory bowel disease (IBD), comprising ulcerative colitis (UC) and Crohn's disease (CD), involves a complex interplay of genetic predisposition and environmental factors that alter host–microbiota interactions causing dysregulation of gut immune responses.[1] The growing prevalence and diminishing age at onset of IBD amplify the risk of comorbidities and the associated economic burden.[2] Patients with IBD have an elevated risk of colorectal cancer (CRC)[3] attributed to chronic inflammation, yet the detailed mechanisms remain elusive.[4]

Accumulating evidence suggests that IBD-associated CRC (IBD-CRC) may emerge through a distinct pathway of tumorigenesis compared to sporadic CRC (sCRC). Patients with IBD are younger at CRC diagnosis and the tumors develop at inflamed areas of the colon with characteristic clinicopathologic features.[4,5] IBD-CRCs show lower frequency of somatic *APC* and *KRAS* mutations, whereas *TP53* mutations occur earlier in tumorigenesis compared to sCRCs.[4,6–10] Despite reduced *APC* mutations, nuclear accumulation of β-catenin is prevalent in IBD-CRCs,[11] suggesting an alternative mechanism of WNT pathway activation. Additional suggested driver genes have varied across studies,[6–10] while the frequency of hypermutated[9] and microsatellite unstable tumors,[12] level of somatic copy number alterations,[9,10] and distribution of somatic mutational signatures[6,9,10] appear similar to sCRCs. Studies on DNA methylation patterns characterizing IBD-CRC have focused on a limited number of genes[13–15] or microarray data,[16] warranting further genome-wide analyses.

Transcriptome-based classification of CRCs has emerged as a powerful tool to describe tumor transcriptional, genetic, epigenetic, and microenvironment characteristics.[17,18] A large-scale international effort resulted in amalgamation of 4 consensus molecular subtypes (CMSs).[17] CMS distribution remains unknown in IBD-CRC; in sCRC, epithelial WNT-associated CMS2 is the most common and mesenchymal CMS4 associates with poor prognosis.[17]

Here, we integrate multiple high-throughput sequencing approaches to comprehensively characterize IBD-CRC and to identify differences compared to sCRC. Our most striking finding was the complete absence of CMS2 tumors among IBD-CRCs that were instead skewed toward CMS4.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Inflammatory bowel disease increases the risk of colorectal cancer and the tumors developing in patients may be genetically and epigenetically distinct compared to sporadic tumors.

NEW FINDINGS

Transcriptomic analyses of colorectal cancer specimens from patients with inflammatory bowel disease revealed the absence of canonical epithelial tumor subtype and predominance of mesenchymal tumors associated with oncostatin M receptor overexpression.

LIMITATIONS

The intensive surveillance of colorectal cancer in patients with inflammatory bowel disease makes these tumors rare, limiting the sample size of the study.

IMPACT

The results suggest that colon inflammation may favor the development of mesenchymal tumor subtype, which has previously been associated with poor survival in large colorectal cancer cohorts.

Materials and Methods

Samples

The study adhered to the Declaration of Helsinki and was approved by the local ethics committee (details are in Supplementary Material). CRC patient samples were collected in 1994–2017 at 9 regional hospitals in Finland.[19–21] Thirty-one IBD-CRC cases were identified from our collection comprising approximately 2,500 CRC patients (UC: n = 27, CD: n = 2, unclassified IBD: n = 2 sharing features of UC and CD) (Table 1). The availability of sample material for downstream analyses is summarized in Supplementary Figure 1.

Whole-Genome Sequencing

DNA was isolated from fresh-frozen tumor, normal colon, or blood of 29 patients with IBD-CRC. Libraries were prepared using TruSeq Nano DNA HT or TruSeq PCR-Free Kit (Illumina), followed by paired-end sequencing using Illumina platform (HiSeqXTen/HiSeq2000). sCRCs were sequenced as described previously.[22]

***Authors share co-first authorship.**

Abbreviations used in this paper: **AI, allelic imbalance; CD, Crohn's disease; CMS, consensus molecular subtype; CRC, colorectal cancer; DE, differentially expressed; DML, differentially methylated loci; EMT, epithelial-mesenchymal transition; FDR, false discovery rate; IBD, inflammatory bowel disease; IBD-CRC, inflammatory bowel disease–associated colorectal cancer; MSI, microsatellite instability; MSS, microsatellite stable; SBS, single base substitution; sCRC, sporadic CRC; SV, structural variant; TGF-β, transforming growth factor β; TSS, transcription start site; UC, ulcerative colitis; UTR, untranslated region; WGS, whole-genome sequencing.**

Most current article

© 2021 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

0016-5085
https://doi.org/10.1053/j.gastro.2021.04.042

Table 1. Clinical Characteristics of the Patients With Inflammatory Bowel Disease and Colorectal Cancer

Sample	Sex	IBD diagnosis	Age at IBD diagnosis, y[a]	Age at CRC diagnosis, y	Tumor location	Tumor histology	TNM stage	Grade	MSI status
c174.1T	F	UC	38	54	Cecum	AC	I	2	MSS
c175.1T	M	UC	33	48	Rectum	AC	II	3	MSS
c269.1T	F	UC	41	41	Cecum	ACPM	II	2	MSS
c3.1T	M	IBD-U	27	48	Transverse colon	AC	II	1	MSS
c424.1T1	M	UC	61	72	Rectosigmoid junction	AC	III	3	MSS
c461.1T	M	UC	82	82	Ascending colon	ACPM	III	3	MSS
c492.1T	M	UC	57	57	Rectum	AC	III	2-3	MSS
c589.1T	F	UC	43	50	Transverse colon	AC	III	1	MSS
c596.1T	M	UC	25	51	Rectum	AC	II	2	MSS
c696.1T	F	UC	19	42	Cecum	AC	III	3	MSS
c745.1T	M	CD	25	55	Rectum	AC	II	2	MSS
c989.1T	F	UC	26	62	Cecum	AC	III	3	MSS
s1111.1T	M	UC	36	64	Rectum	AC	II	1	MSS
s1138.1T	M	UC	41	61	Rectum	AC	II	2	MSS
s1170.1T	M	UC	65	65	Sigmoid colon	AC	III	NA	MSS
s1179.1T	M	UC	15	50	Transverse colon	ACPM	III	2	MSS
s205.1T	M	UC	22	60	Descending colon	AC	II	2	MSS
s576.1T	M	UC	22	41	Ascending colon	AC	III	3	NA
s617.1T	M	UC	20	33	Rectum	ACM	IV	NA	MSS
s649.1T	M	UC	20	33	Cecum	ACM	III	NA	MSS
s660.1T	F	IBD-U	56	66	Ascending colon	AC	II	2	MSI
s683.1T	M	UC	8	46	Cecum	ACM	II	NA	MSS
s700.1T	M	UC	41	64	Transverse colon	AC	I	1	MSS
s703.1T	F	UC	11	22	Descending colon	ACPM	III	2	MSS
s750.1T	F	UC	13	30	Cecum	ACPM	III	3	MSS
s751.1T	M	UC	61	61	Sigmoid colon	AC	III	3	MSS
s763.1T	M	CD	18	44	Rectum	AC	I	3	MSS
s814.1T	F	UC	42	72	Rectum	ACPM	I	3	MSS
s842.1T	F	UC	16	53	Rectum	ACPM	III	3	MSS
s85.1T	M	UC	64	64	Cecum	AC	I	NA	MSS
s982.1T1[b,c]	M	UC	34	34	Cecum	ACPM	III	2	MSI
s982.1T2[b,c]	M	UC	34	34	Splenic flexure	ACPM	III	2	MSI

AC, adenocarcinoma; ACM, mucinous adenocarcinoma; ACPM, partially mucinous adenocarcinoma; IBD-U, unclassified IBD; NA, not available.
[a]In cases where IBD was diagnosed together with CRC based on pathologic findings from the surgical resection, there was often a history of several years of undiagnosed gastrointestinal symptoms mentioned in the clinical data.
[b]Two tumors sampled from the same individual.
[c]The individual was also diagnosed with hereditary nonpolyposis CRC.

Somatic Variant Calling and Analysis

Sequence alignment to GRCh38 reference genome, other data preprocessing steps, and somatic variant calling were performed with GenomeAnalysisToolkit GATK4 best practices workflow (version 4.0.4.0.) for all tumor/normal pairs. Gene annotation (Ensembl genes release 89) for somatic single

nucleotide variants and small insertions/deletions was performed with BasePlayer.[23]

Mutational Signatures

Somatic mutational signatures were extracted from whole-genome sequencing (WGS) data of 237 colorectal tumor/normal pairs, including the 27 paired IBD-CRCs. Briefly, single base substitutions (SBSs) were classified based on their flanking sequence context (±1 bp) into 96 possible mutation types. These mutational spectra were analyzed with standard non-negative matrix factorization, as described previously.[22,24]

OncodriveFML

OncodriveFML (version 2.2.0)[25] was used to analyze the coding sequence, 3′untranslated region (UTR), and 5′UTR for signals of positive selection using the somatic mutations from 27 microsatellite stable (MSS) IBD-CRCs.

Allelic Imbalance

Single nucleotide polymorphism array data were analyzed previously from 1699 colorectal tumor/normal pairs,[26] including 23 IBD-CRCs; allelic imbalance (AI) regions of somatic allelic loss and gain were processed with the same pipeline.[26] Variant calls (HaplotypeCaller) from WGS data were used to calculate AI for 6 additional IBD-CRCs lacking single nucleotide polymorphism array data. The presence of AI in each tumor was evaluated at loci found heterozygous in the corresponding normal sample, as described previously.[26] Analysis was limited to MSS IBD-CRCs (n = 27) and MSS sCRCs (n = 1360; microsatellite instability [MSI] statuses from Palin et al[26]).

RNA Sequencing

Trizol-extracted RNA from 64 CRCs, including 17 IBD-CRCs, underwent HiSeq LncRNA-Seq library preparation and paired-end sequencing using Illumina HiSeqXTen. Raw sequences were mapped onto the human transcriptome (ensembl release 79) using Salmon (version 0.12.0).[27] Gene-level quantification was done with DESeq2 (version 1.18.1),[28] followed by limma (version 3.34.9)[29] correction of sequencing batch effects.

Differential Expression Analysis

Differential gene expression was analyzed using Partek Genomics Suite 6.6 (Partek Inc) Gene Expression workflow, followed by pathway analysis using PANTHER (version 15.0).[30]

Consensus Molecular Subtypes

Random forest classifier of CMSClassifier R package[17] was used to call CMS for each RNA-sequenced tumor, expressed here as the nearest CMS (RF.1) predicted by the classifier.

Deconvolution

The proportions of tumor-infiltrating immune cells were estimated from RNA sequencing data using CIBERSORT.[31,32] Reference was created as described previously[33] by combining bulk RNA sequencing data from isolated blood immune cells (accession GSE60424) and representative median expression profiles from CRCs and normal colon samples from an independent data set.[34] See Supplementary Table 16.

Cell type–specific gene expression profiles were inferred from RNA sequencing data using PRISM,[35] combining single-cell RNA sequencing data of healthy colon[36] and CRCs[37] as reference.

Immune Cell Score

Whole-section slides from 265 formalin-fixed, paraffin-embedded CRCs, including 26 IBD-CRCs, were stained with anti-CD3 (LN10, 1:200; Novocastra) and anti-CD8 (SP16, 1:400; Thermo Scientific) antibodies. Positively stained cells were analyzed using QuPath,[38] as described previously.[39] The immune cell score was formulated as described previously,[40] following the original method by Galon et al.[41]

Nanopore Long-Read Sequencing

Libraries were prepared for 20 IBD-CRCs, 36 sCRCs, and 12 normal colon samples from IBD-CRC patients following Genomic DNA by Ligation (SQK-LSK109) protocol (Oxford Nanopore Technologies). Sequencing and base-calling on PromethION platform employed Live base-calling with MinKNOW. Reads were aligned against the reference genome GRCh38 using minimap2 (version 2.16; preset: map-ont).[42] Structural variants (SVs) were identified using Sniffles (version 1.0.11)[43] and merged together from all tumors and normals with SURVIVOR (version 1.0.6)[44] to filter out SVs found in any normal sample. Genome-wide methylation patterns were obtained using Nanopolish.[45] Differentially methylated loci (DMLs) at autosomal regions were identified and analyzed with R, version 3.5.1 using R packages DSS (version 2.28.0,)[46] bsseq (version 1.16.1), annotatr (version 1.12.1),[47] and Locus Overlap Analysis (version 1.16.0).[48] Chromatin immunoprecipitation sequencing and DNase-sequencing data and 15 chromatin states provided by Roadmap Epigenomics[49] project (Supplementary Table 12) and transcription factor binding sites from CRC cell lines[50] served as region set databases in Locus Overlap Analysis.

Data Availability

Genome-wide somatic single nucleotide variant and insertion/deletion calls (GRCh38) are deposited in the EGA database under accession code EGAS00001004710.

Results

Somatic Point Mutations in Known Colorectal Cancer Driver Genes and Immunity-Related Genes Characterize Inflammatory Bowel Disease–Associated Colorectal Cancers

The study comprised 31 cases of IBD-CRC (Table 1); 29 underwent WGS. A total of 1,104,175 somatic single nucleotide variants and insertion/deletions were identified (Supplementary Figure 2). Two outliers with high somatic variant counts were explained by MSI, a distinct pathway of

CRC tumorigenesis driven by mismatch repair deficiency. We focused on the more typical MSS IBD-CRCs.

The 27 MSS IBD-CRCs harbored a median of 18,194 somatic variants per tumor (3387–73,003), which was highly similar to 259 MSS sCRCs previously whole-genome sequenced in-house (median, 17,319; range, 253–84,646) (Supplementary Figure 2); genome-wide mutation densities showed similar distributions. Coding sequences displayed 4702 variants, with 2816 genes affected by nonsynonymous variants. We ranked the 100 genes mutated in 3 or more tumors by mutation density (Figure 1*A*). Top 20 genes featured several known CRC driver genes, including *TP53* and *KRAS* having the highest mutation densities. Analysis of mutual exclusivity and co-occurrence of mutations in these genes revealed no significant results (Supplementary Table 1). As expected, *KRAS* and *APC* mutations were few in IBD-CRCs compared to sCRCs (22% vs 49% and 22% vs 73%, respectively, Fisher exact test, $P = .0087$ and $P = 2.5 \times 10^{-7}$); *TP53* was frequently mutated in both groups (63% vs 62%) (Figure 1*A*, Supplementary Table 2). A comparison of all variants in COSMIC Cancer Gene Census[51] genes revealed paucity of genes mutated uniquely in IBD-CRC (Supplementary Table 2).

Immunity-related *CARD8* and *PIGR* ranked high by mutation density (Figure 1*A*). All 3 *CARD8* variants were missense, and 2 of the 4 *PIGR* variants present in 3 of 27 IBD-CRCs (11% vs 1% in MSS sCRCs) were truncating (Supplementary Table 3).

Noncoding TP53 *Mutations Are Enriched in Inflammatory Bowel Disease–Associated Colorectal Cancers*

To discover candidate driver genes, we applied OncodriveFML[25] to all somatic point mutations from the 27 MSS IBD-CRCs, totaling 622,366 variants. Coding regions of 2 genes showed significant evidence of positive selection after false discovery rate (FDR) correction ($Q < 0.1$), *TP53* ($P < 1.1 \times 10^{-6}$, $Q = .00061$) mutated in 16 tumors and *GBA2* ($P = .00014$, $Q = .043$) mutated in 2 tumors (Supplementary Figure 2, Supplementary Table 4). We further analyzed 3′UTR and 5′UTR of all protein-coding genes. Overlapping 5′UTR of *TP53* and *WRAP53* genes was the only region showing significant evidence of positive selection ($P = .00036$, $Q = .022$) in IBD-CRCs, while remaining nonsignificant in sCRCs ($P = .015$, $Q = 1.00$) (Supplementary Figure 2, Supplementary Table 4). This region harbored 4 mutations in 3/27 IBD-CRCs and 6 mutations in 5 of 239 sCRCs, resulting in low *TP53* expression (Figure 1*B* and *C*).

Uncoupling of Age-Related Mutational Signature From Age in Inflammatory Bowel Disease–Associated Colorectal Cancers

Detection of genome-wide mutational signatures[22,24] in IBD-CRCs revealed 5 somatic SBS signatures that resembled the signatures SBS1 (age-related spontaneous deamination of 5-methylcytosine), SBS8 (unknown), SBS17 (CTCF/cohesin binding sites), and SBS15 and SBS20 (defective DNA mismatch repair) in human cancers described previously[22,52] (Figure 2*A*). Compared to MSS sCRCs, MSS IBD-CRCs showed a decreased exposure to SBS1 ($P = .0037$) (Figure 2*B*) and a significantly lower rate of SBS1 mutations during life before CRC diagnosis (ordinary least squares, $P = .023$). Neither the mean difference nor difference in SBS1 mutation rate could be explained in Bayesian analysis by the younger age at onset in IBD-CRC patients (mean ± SD, 54 ± 11 years vs 69 ± 11 years in IBD-CRC vs sCRC) (Figure 2*C*). SBS17 exposure was elevated ($P = .0082$) in IBD-CRCs (Figure 2*B*); tumors dominated by SBS17 showed no obvious correlations to clinical characteristics or known driver mutations.

Analysis of Chromosomal Rearrangements Reveals Tumor Type–Specific Allelic Imbalance

Chromosomal stability of the MSS IBD-CRCs was inspected using both nanopore WGS data (n = 19) and single nucleotide polymorphism array data (n = 27). Numbers of somatic nanopore-detected SVs did not differ between IBD-CRCs and sCRCs (Mann-Whitney U test $P > .76$) (Figure 3*A*–*C*). Coding region SVs in IBD-CRC revealed 5 genes with a recurring somatic breakpoint: *CCSER1* (n = 2 tumors), *FHIT* (n = 2), *IMMP2L* (n = 3), *MACROD2* (n = 2), and *PIBF1* (n = 2) (Supplementary Table 5), all known fragile site genes.[53,54]

We previously characterized AI in 1699 CRCs[26] and now compared MSS IBD-CRCs (n = 27) to MSS sCRCs (n = 1360). No overall difference between IBD-CRCs (mean, 9.2×10^8 bp of AI per tumor) and sCRCs (mean, 1.1×10^9 bp) was observed (Mann-Whitney U test $P = .29$) (Figure 3*D*). The only outstanding difference genome-wide (at FDR < 10%) was found at 5p13.1-p12, where IBD-CRC was enriched for gains (10 of 27 tumors (37%); Figure 3*E*; Supplementary Table 6). Of the genes in this region, *OSMR* and *LIFR* cytokine receptors sharing the ligand oncostatin M[55] were significantly up-regulated in IBD-CRCs compared to sCRCs (Supplementary Table 7). Closer inspection of 38 previously characterized AI target genes in CRC[26] revealed 4 genes with differential AI (FDR < 10%): *CDKN2B* and *SMARCA2* enriched for allelic losses, *FOXA1* for allelic gains, and *HNF4A* having fewer gains in IBD-CRC compared to sCRC (Supplementary Table 8).

Differential Gene Expression Highlights Changes Related to Stromal and Immune Cells in Inflammatory Bowel Disease–Associated Colorectal Cancers

Bulk RNA sequencing revealed 870 significantly differentially expressed (DE) genes between MSS IBD-CRCs and MSS sCRCs (FDR < 10%; Supplementary Table 7, Supplementary Figure 3). In IBD-CRCs, pathway analysis showed a strong overenrichment (overexpression) of gene sets related to complement activation and extracellular matrix organization (Figure 4*A*, Supplementary Table 9). Very few gene sets were significantly underenriched,

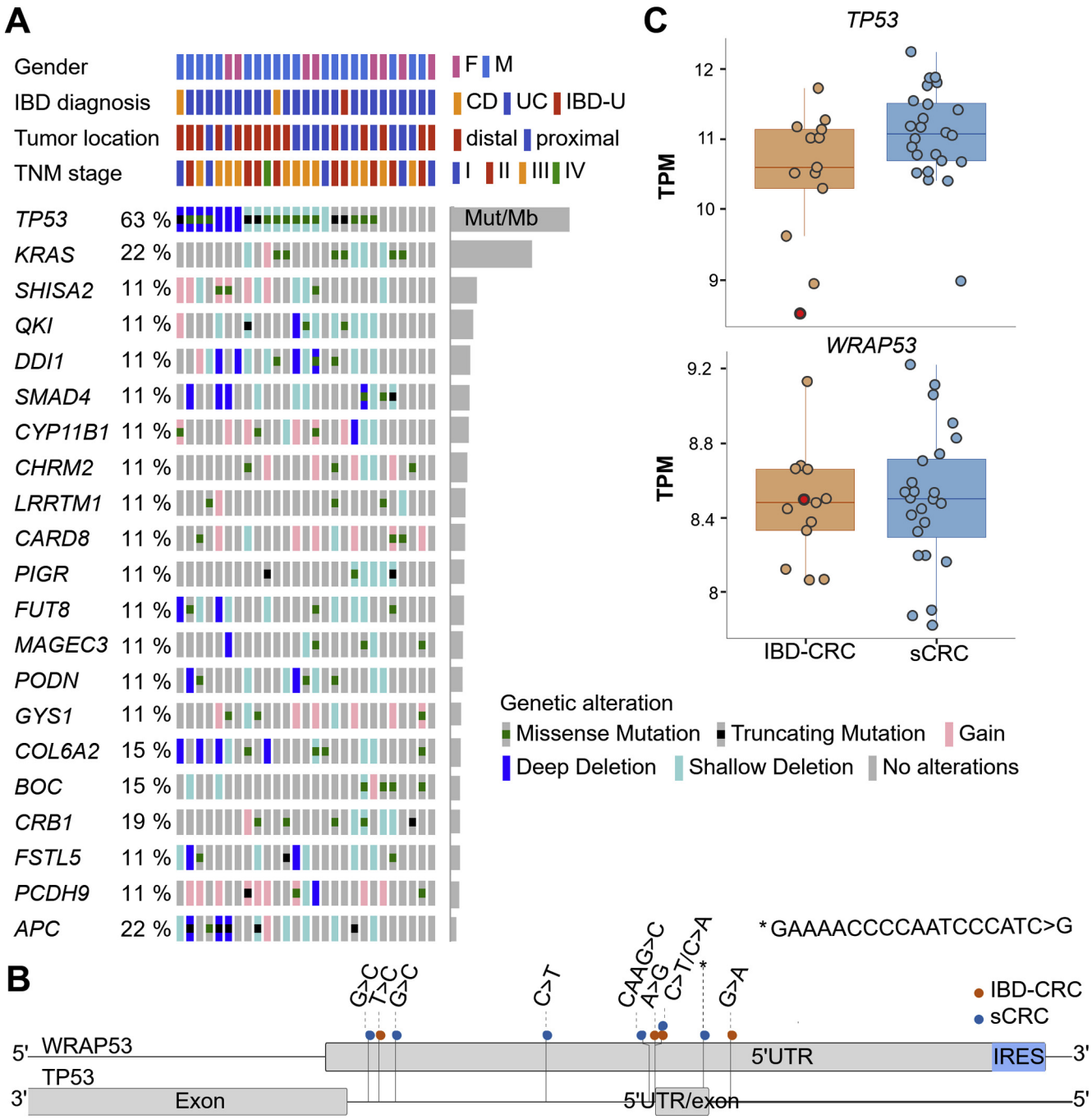


Figure 1. Somatic point mutations. (*A*) OncoPrint (https://www.cbioportal.org/oncoprinter) showing somatic point mutations (*filled squares*) and copy number alterations (*filled bars*) in genes mutated in 3 or more MSS IBD-CRCs and ranked by mutation density (mutations/Mb; top 20 genes and *APC* ranking no. 41 presented with percentages of mutated tumors). F, female; IBD-U, unclassified inflammatory bowel disease; M, male; TNM, tumor-nodes-metastases. (*B*) Variants observed in the overlapping 5′UTR of *TP53* and *WRAP53*. IRES, internal ribosome entry site. (*C*) Gene expression in 64 RNA-sequenced CRCs as transcripts per million (TPM). *Red dot* signifies an IBD-CRC carrying *TP53/WRAP53* 5′UTR variant (no coding *TP53* variants).

including only 1 of 44 significant results for Reactome database, TCF-dependent signaling in response to WNT (R-HSA-201681). Genes in this set included *RNF43* and *AXIN2*, the most significantly down-regulated genes in IBD-CRCs. The most significantly up-regulated gene, *OSMR*, appeared in a large overenriched Immune System gene set (R-HSA-168256) comprising 146 DE genes related to myeloid/lymphoid immune cells, complement, and, interestingly, EMT (*TWIST1*, *ZEB1*, *VIM*) (Supplementary Table 9).

We estimated cell type–specific gene expression profiles for epithelial, stromal, and immune cells using PRISM deconvolution tool35 (Supplementary Figure 4). PRISM-estimated cell-type proportions for each tumor consensus molecular subtype (Supplementary Figure 5) matched those

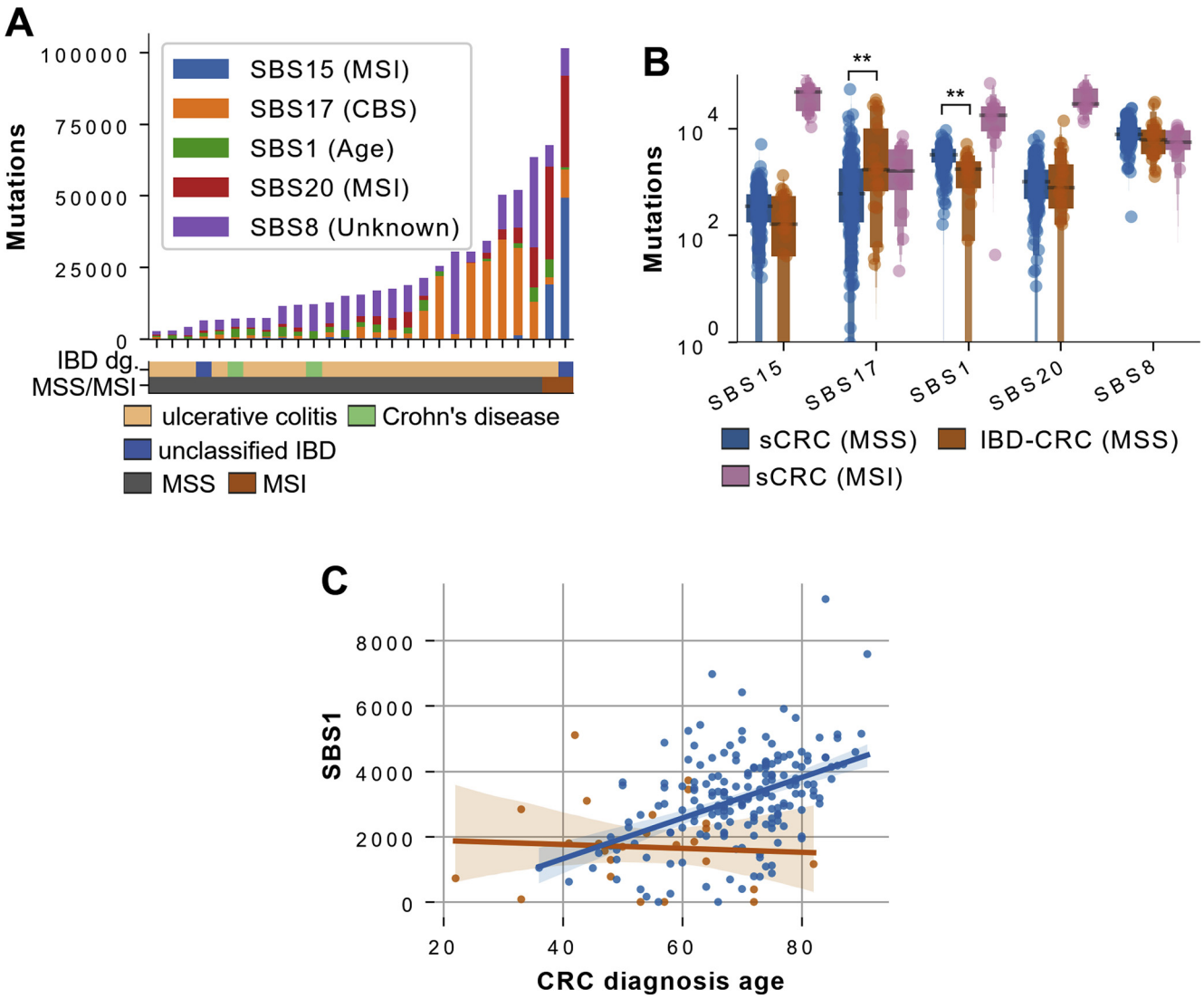


Figure 2. Somatic mutational signatures. (*A*) Contributions of SBS signatures to somatic mutations in IBD-CRCs. CBS, CTCF/cohesin binding site. (*B*) SBS signatures in IBD-CRCs (MSS n = 25) and sCRCs (MSS n = 194, MSI n = 12). Robust linear model regression was used to compare the MSS groups (**$P < .01$). (*C*) Analysis of age-dependence of SBS1. After Bayesian analysis with and without intercept, depending on the IBD status, the slope for IBD-CRCs for SBS1 exposure is significantly lower than for sCRCs. On average, the MSS IBD-CRCs had 7.4–61 fewer SBS1-associated mutations per year (95% credible interval).

reported previously using a different algorithm.[17] Differential expression analysis between MSS IBD-CRCs and MSS sCRCs yielded 376, 626, and 66 DE genes for epithelial, stromal, and immune cells, respectively (Supplementary Table 7). Pathway analysis of these separate gene lists allowed connecting each pathway change to a particular cell type (Supplementary Table 9). Epithelial DE genes, predominantly down-regulated in IBD-CRCs, were enriched for pathways related to epithelial cell differentiation and development. Stromal DE genes, predominantly up-regulated, were enriched for extracellular matrix organization, integrin interactions, vasculature development, and insulin-like growth factor metabolism. Immune DE genes, predominantly up-regulated, were enriched for classical complement activation and other antibody-mediated immune responses.

Common Epithelial Colorectal Cancer Subtype Associated With WNT Signaling Is Absent in Inflammatory Bowel Disease–Associated Colorectal Cancers

We defined the nearest CMS for each of the 64 RNA-sequenced CRCs, as described previously.[17] As reported, MSI sCRCs were highly enriched for CMS1 (Supplementary Table 10). Based on PRISM deconvolution, immune cell proportion was highest in CMS1, stromal in CMS4, and epithelial in CMS2/CMS3 tumors (Supplementary Figure 5), as reported previously.[17] Comparison of MSS tumors in IBD-CRCs revealed a striking paucity of the canonical epithelial CMS2 subtype associated with WNT and MYC signaling (0% of IBD-CRCs vs 39% of sCRCs; $P = .0048$, $P_{adj} = .019$) (Figure 4*B*, Supplementary Table 10). IBD-CRCs were

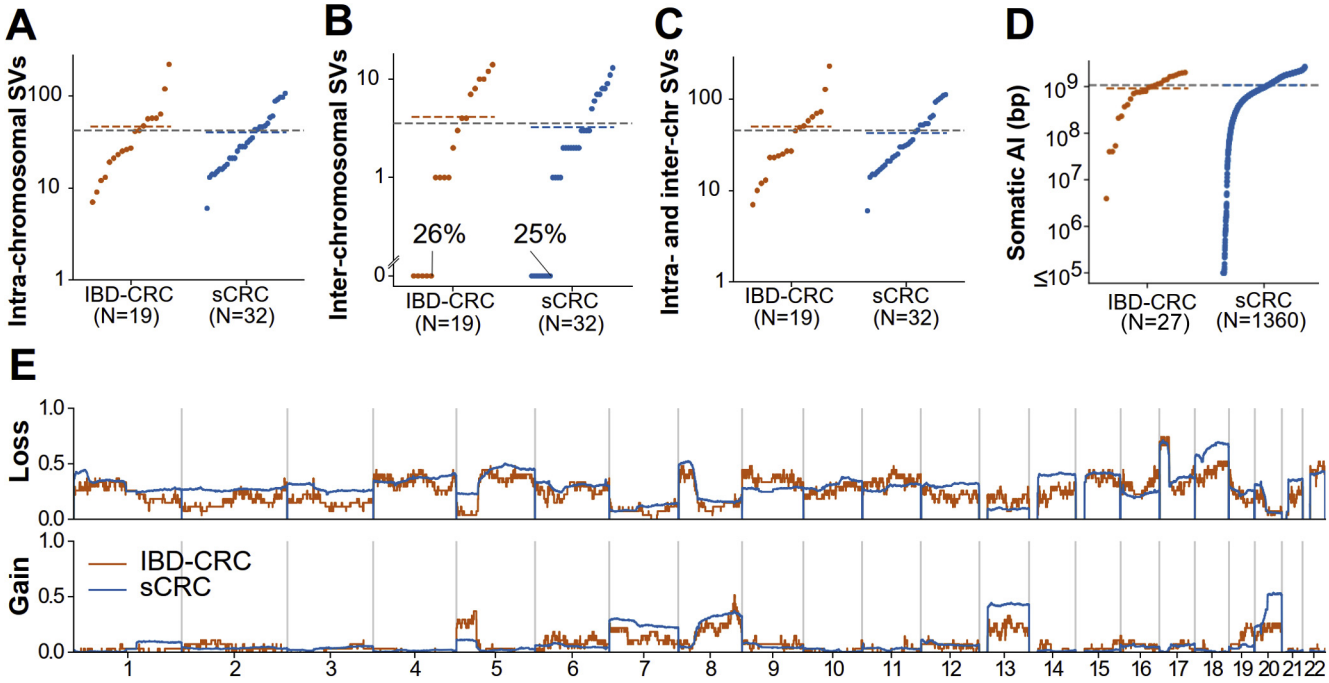


Figure 3. Overview of somatic structural aberrations. Numbers of somatic (*A*) intra- and (*B*) inter-chromosomal, and (*C*) total SVs in 19 nanopore-sequenced MSS IBD-CRCs and 32 MSS sCRCs. The percentages in (*B*) refer to the proportion of tumors lacking inter-chromosomal SVs. (*D*, *E*) Somatic AI in 27 MSS IBD-CRCs and 1360 MSS sCRCs across the autosomes. (*D*) Total amount of base pairs affected by somatic AI. *Dashed lines* denote the mean per group (*orange/blue*) and overall mean (*gray*). Y-axis is logarithmic and truncated to a minimum 100 kbp. (*E*) Observed proportion of tumors with allelic loss (or gain) at each genomic position.

dominated by mesenchymal CMS4 (57% vs 21%; $P = .019$, $P_{adj} = .057$), with concomitant up-regulation of transcription factors mediating EMT (*TWIST1, TWIST2, SNAI2, ZEB1, ZEB2*) (Supplementary Table 7).

Mesenchymal Tumors Show a Distinct Pattern of Immune Cell Infiltration

We interrogated the immune cell contexture of RNA-sequenced CRCs using CIBERSORT deconvolution.[31,32] MSI and CMS1 tumors were found in clusters with high estimated proportions of $CD8^+$ cytotoxic T cells (Supplementary Figure 5), as reported previously.[17] MSS tumors showed 3 clusters dominated by $CD4^+$ T, $CD8^+$ T, and B cells, respectively (Figure 4*C*). IBD-CRCs were divided between the B cell and the $CD4^+$ T cell clusters, forming in the latter a distinct CMS4-enriched subcluster distinguished by high proportions of monocytes.

We further analyzed immune cell score, a prognostic measure of tumor T cell infiltration reflecting numbers of total ($CD3^+$) and cytotoxic ($CD8^+$) T cells.[41] Rates of high immune cell score (3–4) were similar between 24 MSS IBD-CRCs and 196 MSS sCRCs (54% vs 43%; Fisher exact test, $P = .39$) (Supplementary Table 11). There were no significant differences in the 4 individual stainings (Figure 4*D*) or in the ratios of $CD8^+$ to $CD3^+$ T cells (Supplementary Figure 5) between these groups (Mann-Whitney U test, Holm-Bonferroni correction).

Similar Genome-Wide Methylation Patterns in Inflammatory Bowel Disease–Associated Colorectal Cancers and Sporadic Colorectal Cancer

Methylation analyses were carried out using whole-genome nanopore sequencing data. Based on methylation values at CpG islands, IBD-CRCs and a pool of nondysplastic normal colon samples from patients with IBD ("IBD-normals") mainly clustered separately from sCRCs (Supplementary Figure 6). All subsequent methylation analyses focused on samples from patients with MSS tumors. IBD-CRCs showed, on average, higher genome-wide methylation compared to sCRCs (Figure 5*A*). Neither age nor cancer type was significantly associated with the average methylation level (joint model $P = .078$). Differentially methylated loci (DMLs) were studied in autosomes comparing IBD-CRCs to IBD-normals and sCRCs to IBD-normals, resulting in 553,390 DMLs (4.4% hypermethylated) and 2,413,663 DMLs (3.7% hypermethylated), respectively. Five of the 10 IBD-normals were matched with the studied IBD-CRCs, likely decreasing the IBD-CRC DML count, potentially affected also by shared IBD-derived methylation changes.

In both tumor groups, genomic annotation of the DMLs (Figure 5*B–D*, Supplementary Figure 6) revealed the majority of hypomethylated loci at non-CpG island ("CpG inter") areas and modest enrichment only on quiescent/low chromatin areas, out of 15 chromatin states studied.

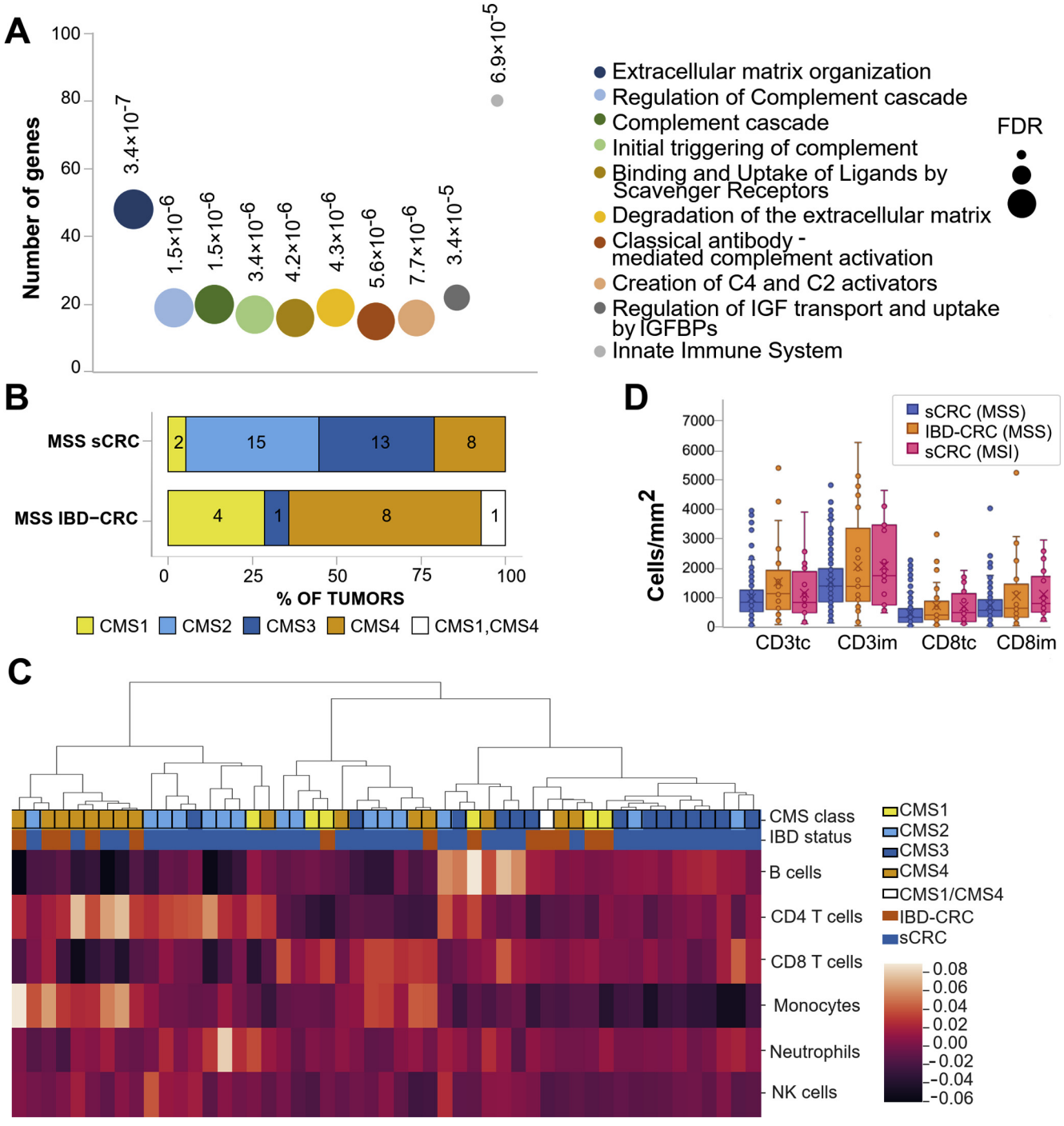


Figure 4. Patterns of gene expression and immune cell infiltration. (*A–C*) RNA sequencing data from 14 MSS IBD-CRCs and 38 MSS sCRCs was compared. (*A*) Top 10 pathways significantly enriched among DE genes (n = 870) in PANTHER Statistical Enrichment Test[30] against the Reactome database. (*B*) CMS17 presented as the nearest CMS having the highest posterior probability (equal for CMS1 and CMS4 in one tumor). (*C*) Clustering of tumors based on the estimated proportions of immune cells from CIBERSORT deconvolution. NK, natural killer. (*D*) $CD3^+$ and $CD8^+$ T cell numbers determined by immunohistochemistry at the tumor core (tc) and invasive margin (im) in 25 MSS IBD-CRCs, 217 MSS sCRCs, and 20 MSI sCRCs.

Conversely, most of the hypermethylated loci were located on CpG islands and intronic areas (Figure 5*B*), showing enrichment on several chromatin states related to bivalent chromatin (Figure 5*C*). From chromatin states with average odds ratio >20, the enrichment of hypermethylation relative to IBD-normals was lower in IBD-CRC compared to sCRC on bivalent/poised transcription start sites (TSS) (Mann-Whitney U test, $P_{adj} < 1.6 \times 10^{-6}$) and flanking bivalent TSS/enhancers ($P_{adj} = .00011$). In both groups, the enrichment of hypermethylated loci was strongest on H3K27me3 marked chromatin (Figure 5*D*). We further inspected whether the DE genes were commonly annotated

with bivalent TSS, which was indeed the case more often than expected by chance (Supplementary Figure 7).

HNF4A *Binding Sites Are Hypomethylated Solely in Sporadic Colorectal Cancer*

We studied the enrichment of DMLs in chromatin immunoprecipitation sequencing peaks of 382 transcription factors and DNA binding proteins obtained from LoVo CRC cell line.[50] Significant enrichment was detected at multiple binding sites in both tumor groups compared to IBD-normals (Supplementary Table 13). However, the significance often originated from a small total number of binding sites harboring DMLs, diminishing the biologic relevance, especially in results unique to IBD-CRCs. In sCRCs, the most significant enrichment of hypomethylation relative to IBD-normals was found at *HNF4A* binding sites, with DMLs detected at >3000 separate binding sites. This enrichment was absent in IBD-CRCs, in agreement with fewer *HNF4A* gains. As reported previously,[17] the frequency of *HNF4A* gains and *HNF4A* expression, together with *APC* mutations, were highest in CMS2 tumors (Supplementary Figure 8). Interestingly, binding sites of *TCF7L2*, key transcription factor mediating WNT/β-catenin signaling, showed hypomethylation only in sCRC when compared to IBD-normals, and binding sites of *MYC*, a critical WNT/β-catenin target gene, were enriched for hypermethylation in IBD-CRCs when compared to sCRCs (Supplementary Table 13).

PIGR *and* OSMR *Involved in Mucosal Immunity Show Dysregulated Promoter Methylation in Inflammatory Bowel Disease–Associated Colorectal Cancers*

We next identified genes containing DMLs at promoter region (1 kbp upstream of TSS) in IBD-CRCs, but not sCRCs, compared to IBD-normals. Thirteen and 7 protein coding genes contained >1 hyper- or hypomethylated loci, respectively, at their promoter. Visualization of methylation patterns around these transcripts, together with their expression data (Supplementary Figure 9) revealed distinctly high *PIGR* promoter methylation in IBD-CRCs, with significantly reduced epithelial expression (Figure 6*A* and *B*; Supplementary Table 7; Supplementary Figure 10). An IBD-CRC harboring 2 somatic *PIGR* mutations displayed lower methylation values than IBD-CRCs with wild-type *PIGR* (n = 18) (Figure 6*A*).

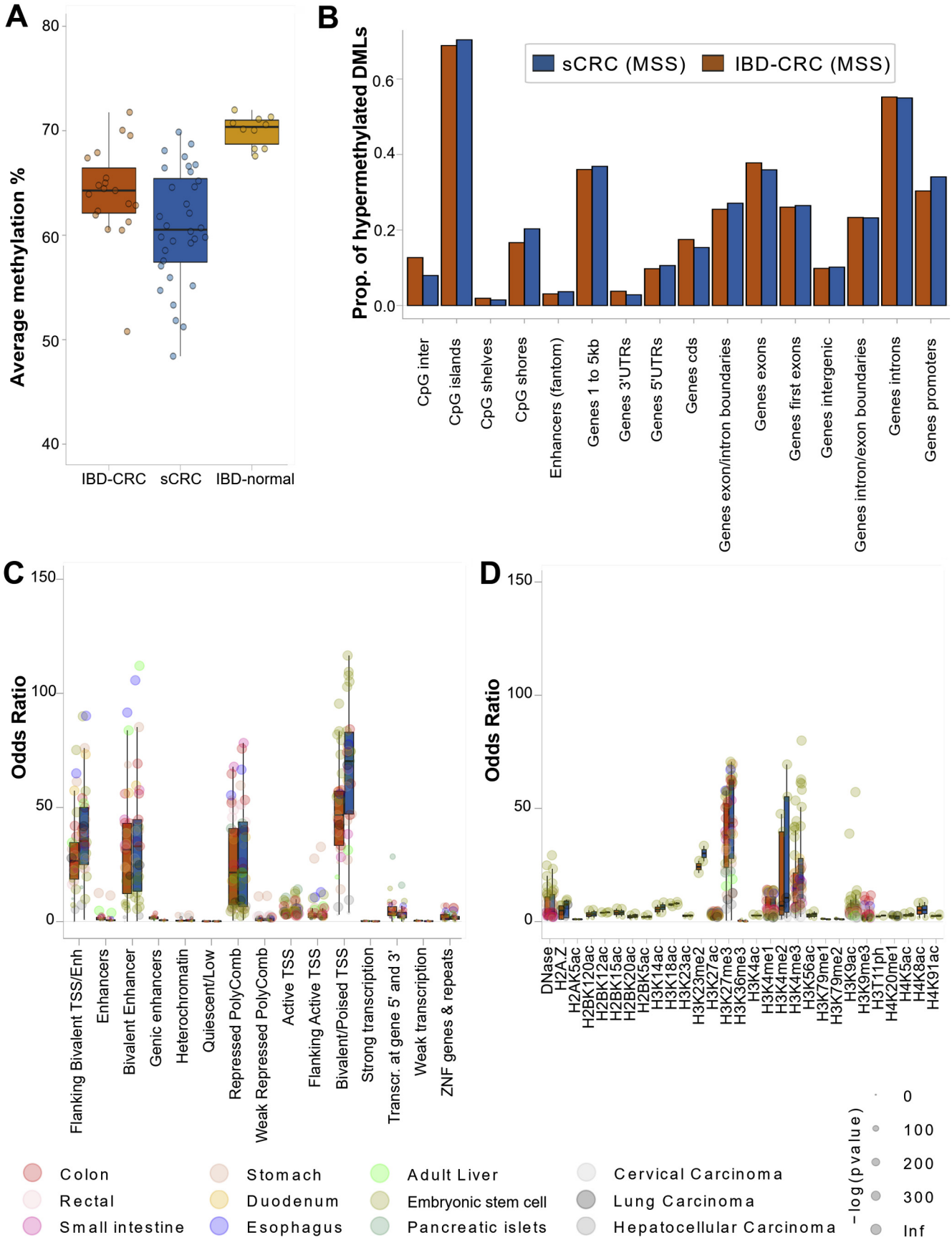
DML calling and annotation comparing the 2 tumor groups (Supplementary Figures 11 and 12) revealed multiple hypomethylated DMLs at *OSMR* promoter region in IBD-CRCs compared to sCRCs (Figure 6*C*, Supplementary Figures 10 and 12). Stromal *OSMR* expression was strongly elevated in IBD-CRCs and particularly high in CMS4 tumors (Figure 6*D*), showing a strong correlation with expression of EMT-related genes especially in IBD-CRCs (Supplementary Figure 13). Inspection of protein coding genes having DMLs at gene bodies (exons, introns, and 5′/3′UTRs) revealed enrichment of WNT pathway genes for gene body hypermethylation in IBD-CRC compared to sCRC (Supplementary Table 14), including *AXIN2* and *RNF43* strongly down-regulated in IBD-CRCs (Supplementary Table 15, Supplementary Table 7).

Discussion

Increased risk of CRC in IBD patients has been attributed to long-standing chronic inflammation, yet the underlying mechanisms remain poorly understood.[3,4] Here, we integrated data from multiple experimental approaches to gain a deeper understanding of IBD-CRC. While many tumor characteristics, including somatic AI, SVs, and genome-wide methylation patterns, showed an overall similarity to sCRCs, several distinct features were identified, as summarized in Supplementary Table 17.

Constitutive activation of WNT/β-catenin signaling via loss-of-function mutations in *APC* is a hallmark of sCRC.[4] The pathway regulates cell fate, proliferation, polarity, and stemness via β-catenin–dependent and –independent mechanisms.[56] The most common transcriptomic CRC subtype, CMS2, displays up-regulation of WNT and MYC target genes and an epithelial differentiation signature.[17] Among our MSS sCRCs, 39% were CMS2. In striking contrast, CMS2 was completely absent among IBD-CRCs, dominated instead by mesenchymal CMS4 previously associated with EMT, matrix remodeling, transforming growth factor β (TGF-β) signaling, complement activation, and depletion of WNT/MYC-related expression signatures.[17] Differential expression analysis reflected this CMS2 vs CMS4 predominance, revealing in IBD-CRCs enrichment of gene sets involved in extracellular matrix organization and complement activation, up-regulation of key transcription factors mediating EMT, and down-regulation of genes related to epithelial differentiation. Moreover, binding sites of TCF7L2 and MYC, mediators of canonical WNT signaling, were hypomethylated in sCRCs relative to IBD-CRCs. Thus, an alternative mode of WNT dysregulation may govern tumorigenesis in IBD-CRCs that still show accumulation of nuclear/cytoplasmic β-catenin.[11]

EMT involves loss of tumor cell polarity and cell–cell adhesion to gain more migratory and invasive properties.[57] WNT signaling is a prominent regulator of EMT in CRC[57] and could thus contribute to the observed mesenchymal skewing of IBD-CRCs. One mechanism could involve the differential activity of HNF4α transcription factor, essential for embryonic development of colonic epithelium.[58] We showed that HNF4α binding sites were highly significantly hypomethylated in sCRCs but not in IBD-CRCs, implying increased binding in sCRC tumorigenesis. Chromosomal gains at *HNF4A* locus were more common in sCRCs, enriching in CMS2 tumors. In hepatocellular carcinoma, HNF4α maintains epithelial tumor phenotype by facilitating corepression of EMT-related WNT/β-catenin target genes,[59] putting a break on EMT. *HNF4A* gains could enforce this mechanism in sCRCs with constitutive WNT/β-catenin activation, promoting the establishment of epithelial CMS2 tumors. Conversely, in a mouse model of colitis, inflammation-induced epigenetic change resulted in reduced HNF4α occupancy.[60] Thus, a state of low HNF4α


A
Average methylation %
80
70
60
50
40
IBD-CRC
sCRC
IBD-normal
B
Prop. of hypermethylated DMLs
0.6
0.4
0.2
0.0
sCRC (MSS)
IBD-CRC (MSS)
CpG inter
CpG islands
CpG shelves
CpG shores
Enhancers (fantom)
Genes 1 to 5kb
Genes 3'UTRs
Genes 5'UTRs
Genes cds
Genes exon/intron boundaries
Genes exons
Genes first exons
Genes intergenic
Genes intron/exon boundaries
Genes introns
Genes promoters
C
Odds Ratio
150
100
50
0
Flanking Bivalent TSS/Enh
Enhancers
Bivalent Enhancer
Genic enhancers
Heterochromatin
Quiescent/Low
Repressed PolyComb
Weak Repressed PolyComb
Active TSS
Flanking Active TSS
Bivalent/Poised TSS
Strong transcription
Transcr. at gene 5' and 3'
Weak transcription
ZNF genes & repeats
D
Odds Ratio
150
100
50
0
DNase
H2A.Z
H2AK5ac
H2BK120ac
H2BK12ac
H2BK15ac
H2BK20ac
H2BK5ac
H3K14ac
H3K18ac
H3K23ac
H3K23me2
H3K27ac
H3K27me3
H3K36me3
H3K4ac
H3K4me1
H3K4me2
H3K4me3
H3K56ac
H3K79me1
H3K79me2
H3K9ac
H3K9me3
H3T11ph
H4K20me1
H4K5ac
H4K8ac
H4K91ac
Colon
Rectal
Small intestine
Stomach
Duodenum
Esophagus
Adult Liver
Embryonic stem cell
Pancreatic islets
Cervical Carcinoma
Lung Carcinoma
Hepatocellular Carcinoma
−log(pvalue)
0
100
200
300
Inf

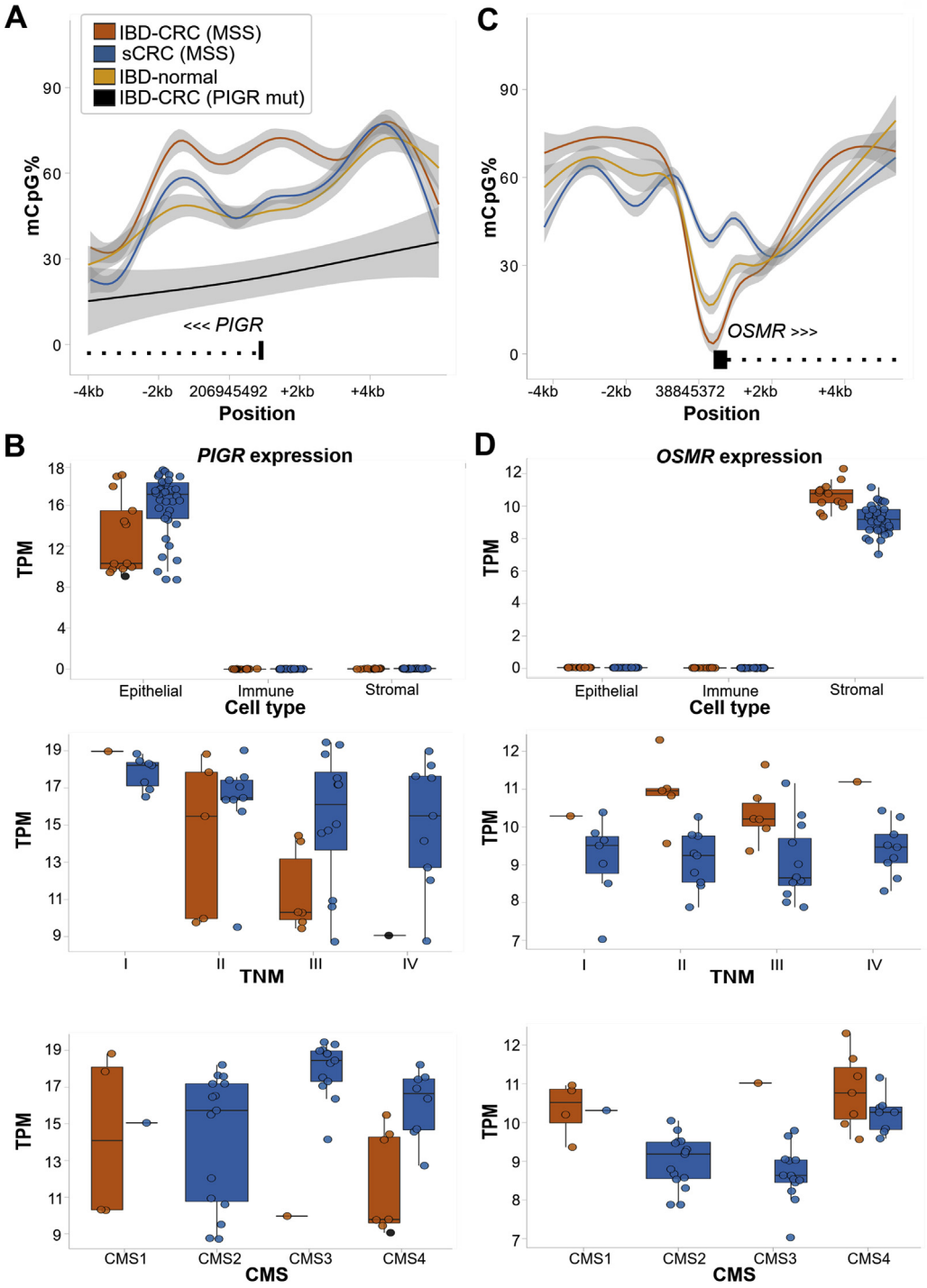


Figure 6. Patterns of PIGR and OSMR methylation and expression. (*A*) Smoothing of the methylation measurements around *PIGR* from 18 *PIGR* wild-type vs 1 *PIGR*-mutated IBD-CRC, 32 *PIGR* wild-type sCRCs, and 10 IBD-normals. (*B*) Deconvoluted *PIGR* expression data; data from epithelial cells are plotted against tumor-node-metastases (TNM) stage and CMS. (*C*) Smoothing of the *OSMR* methylation data. (*D*) Deconvoluted *OSMR* expression data; data from stromal cells are plotted against TNM stage and CMS.

Figure 5. Genome-wide annotation of methylation patterns. (*A*) Average genome-wide methylation levels from 19 MSS IBD-CRCs, 32 MSS sCRCs, and 10 nondysplastic normal colon tissue samples from IBD patients. (*B*) Proportion of hypermethylated DMLs in tumors compared to IBD normals at different genomic regions. (*C*, *D*) Enrichment analysis of hypermethylated loci using Roadmap Epigenomics chromatin annotations from tissue samples of the normal human gastrointestinal tract and samples relevant to tumorigenesis. Different *colors* depict different tissue origins of the cells (Supplementary Table 12).

occupancy may pre-exist in patients with IBD. Once tumorigenesis is set forth by *TP53* mutations and aberrant WNT signaling, it may favor the emergence of mesenchymal CMS4 tumors through EMT from the earliest stages of tumorigenesis.

The paucity of *APC* and *KRAS* mutations in IBD-CRC has been identified consistently.[4,6–10] Novel driver mutations explaining the relative independence of IBD-CRC from these key sCRC driver events have not emerged. We revealed evidence for positive selection of somatic mutations at noncoding 5′UTR of *TP53* in IBD-CRCs, resulting in low *TP53* expression. A recent pan-cancer analysis described similar variants in this region as a novel rare driver event reducing *TP53* expression,[61] further highlighting the importance of *TP53* loss in IBD-CRC.[4] We could not identify a distinct subgroup of MSS IBD-CRCs with elevated genome-wide mutation densities, as suggested earlier by panel sequencing.[62]

Two negative feedback regulators induced by WNT signaling, *AXIN2* and *RNF43*, were strongly down-regulated in IBD-CRCs compared to sCRCs. *AXIN2* participates with APC in the β-catenin destruction complex,[63] while *RNF43* inhibits upstream WNT signaling by inducing the degradation of frizzled receptors.[64] Both genes are commonly mutated in MSI sCRCs showing mutual exclusivity with APC mutations[63,65] and down-regulated in *APC* wild-type compared to *APC*-mutated MSS sCRCs.[66,67] Thus, loss of WNT regulation by *AXIN2* and *RNF43* could trigger aberrant WNT signaling in *APC* wild-type CRCs enriched among MSI-CRCs and IBD-CRCs. Moreover, disruption of *AXIN2* and *RNF43* in serrated adenomas and carcinomas[66,68,69] may link them with nonconventional types of precursor lesions common in IBD-CRC.[4] Intriguingly, model organoids mimicking sessile serrated adenomas, but not conventional tubular adenomas, are responsive to TGF-β–induced development toward CMS4.[70]

Tumor immune cell infiltrate in CMS4 IBD-CRCs was dominated by $CD4^+$ T cells and monocytes. In agreement, CMS4 has been linked with an "immune inflamed" immunosuppressive microenvironment enriched in $CD4^+$ regulatory T cells and monocyte-derived macrophages.[71] The remaining IBD-CRCs were dominated by B-cell infiltration. While the role of B lymphocytes in cancer remains poorly understood, this could result from an attempt to compensate for reduced *PIGR* expression and disrupted epithelial transport of B cell–derived IgA.

IBD-CRCs tended to show higher genome-wide methylation compared to sCRCs, but decreased exposure to mutational signature linked with age-driven deamination of 5'methylcytosines; the latter was found also when compared with tumors from patients with Lynch syndrome.[62] Neither were explained by the younger age at diagnosis or by differential expression of DNMT or TET enzymes regulating DNA methylation, suggesting inflammation-driven methylation changes.

Aberrant methylation patterns were detected in 2 genes related to mucosal immunity solely in IBD-CRCs. *PIGR* mediates IgA and IgM transport across intestinal epithelium[72] and accumulates somatic loss-of-function mutations in IBD-affected colon.[73–75] *PIGR* showed promoter hypermethylation in IBD-CRCs compared to IBD-affected non-dysplastic colon, with down-regulated expression. This may imply further selection for *PIGR* loss during tumorigenesis, which could promote epithelial barrier dysfunction and sustained inflammation.[72] Alternatively, the progressive loss of *PIGR* with increasing tumor stage could reflect gradual loss of epithelial properties in tumor cells. *OSMR*, encoding a receptor for cytokine oncostatin M (*OSM*), showed promoter hypomethylation, frequent chromosomal gain, and strong overexpression in IBD-CRCs. Deconvolution analysis pointed toward *OSMR* overexpression originating from stromal cells. In agreement, OSMR-positive inflammatory stromal cells expand in IBD-affected colon.[76] Importantly, elevated intestinal *OSMR* and *OSM* expression is associated with a subgroup of IBD patients showing poor response to TNFα blockers.[76] As OSM stimulation triggers EMT and mesenchymal characteristics in breast cancer cells,[77] increased OSM-OSMR signaling in IBD could favor the establishment of mesenchymal CRC subtype, although this remains to be studied further.

Our study represents one of the largest next-generation sequencing datasets on both IBD-CRC and sCRC patients published so far. Multiple data layers allowed deep characterization of these tumors. Even larger studies would be needed to control for more detailed clinical parameters, such as UC/CD differences, changes in anti-inflammatory medication for IBD over time, and different patient population backgrounds. Further limitations include lack of non-IBD normal colon in methylation analyses, and potential batch effects. The rich layers of data created, focus on results showing connections between different data sets, and rigorous batch-to-batch quality control mitigate many of these effects. For methylation analyses, IBD-CRCs were compared with both IBD normal colon and sCRCs to better capture differences distinguishing IBD-associated tumorigenesis.

Taken together, our results highlight skewing of IBD-associated tumorigenesis toward increased importance of the dynamic process of EMT, driven by distinct mechanisms of WNT pathway dysregulation compared to sCRC. The mesenchymal tumor subtype enriched in IBD-CRCs has been associated with drug resistance[18] and worse relapse-free and overall survival in large CRC patient cohorts,[17] which may affect prognosis and treatment options in IBD-CRC. *OSM-OSMR* signaling could contribute to establishment of this mesenchymal subtype in IBD patients, potentially linking a TNFα blocker–resistant subgroup of patients with increased CRC risk. Cost-effective and clinically feasible alternatives for CMS subtyping, including immunohistochemistry, are reaching high accuracy.[18] First treatment combinations targeted toward CMS4-like TGF-β–activated CRC have entered clinical trials,[18] paving the way for overcoming the tumor-microenvironment crosstalk that drives the mesenchymal phenotype. Further studies are needed to determine whether similar approaches are beneficial in treatment of IBD-CRC.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://doi.org/10.1053/j.gastro.2021.04.042

References

1. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature 2007;448:427–434.
2. Windsor JW, Kaplan GG. Evolving epidemiology of IBD. Curr Gastroenterol Rep 2019;21:40.
3. **Annese V, Beaugerie L, Egan L**, et al. European evidence-based consensus: inflammatory bowel disease and malignancies. J Crohns Colitis 2015;9:945–965.
4. Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. Gastroenterology 2011;140:1807–1816.
5. Reynolds IS, O'Toole A, Deasy J, et al. A meta-analysis of the clinicopathological characteristics and survival outcomes of inflammatory bowel disease associated colorectal cancer. Int J Colorectal Dis 2017;32:443–451.
6. **Robles AI, Traverso G, Zhang M**, et al. Whole-exome sequencing analyses of inflammatory bowel disease-associated colorectal cancers. Gastroenterology 2016;150:931–943.
7. 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.
8. Fujita M, Matsubara N, Matsuda I, et al. Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. Oncotarget 2018;9:969–981.
9. **Din S, Wong K**, Mueller MF, et al. Mutational analysis identifies therapeutic biomarkers in inflammatory bowel disease-associated colorectal cancers. Clin Cancer Res 2018;24:5133–5142.
10. **Baker A-M, Cross W, Curtius K**, et al. Evolutionary history of human colitis-associated colorectal cancer. Gut 2019;68:985–995.
11. Claessen MMH, Schipper MEI, Oldenburg B, et al. WNT-pathway activation in IBD-associated colorectal carcinogenesis: potential biomarkers for colonic surveillance. Cell Oncol 2010;32:303–310.
12. **Schulmann K, Mori Y**, Croog V, et al. Molecular phenotype of inflammatory bowel disease-associated neoplasms with microsatellite instability. Gastroenterology 2005;129:74–85.
13. Konishi K, Shen L, Wang S, et al. Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias. Gastroenterology 2007;132:1254–1260.
14. Sanchez JA, Dejulius KL, Bronner M, et al. Relative role of methylator and tumor suppressor pathways in ulcerative colitis-associated colon cancer. Inflamm Bowel Dis 2011;17:1966–1970.
15. **Kisiel JB, Klepp P**, Allawi HT, et al. Analysis of DNA methylation at specific loci in stool samples detects colorectal cancer and high-grade dysplasia in patients with inflammatory bowel disease. Clin Gastroenterol Hepatol 2019;17:914–921.e5.
16. 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;18:641–648.
17. **Guinney J, Dienstmann R, Wang X**, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–1356.
18. Dienstmann R, Vermeulen L, Guinney J, et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat Rev Cancer 2017;17:79–92.
19. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998;338:1481–1487.
20. Salovaara R, Loukola A, Kristo P, et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. J Clin Oncol 2000;18:2193–2200.
21. Tanskanen T, Gylfe AE, Katainen R, et al. Exome sequencing in diagnostic evaluation of colorectal cancer predisposition in young patients. Scand J Gastroenterol 2013;48:672–678.
22. **Katainen R, Dave K, Pitkänen E**, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet 2015;47:818–821.
23. Katainen R, Donner I, Cajuso T, et al. Discovery of potential causative mutations in human coding and non-coding genome with the interactive software BasePlayer. Nat Protoc 2018;13:2580–2600.
24. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3:246–259.
25. Mularoni L, Sabarinathan R, Deu-Pons J, et al. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol 2016;17:128.
26. **Palin K, Pitkänen E, Turunen M**, et al. Contribution of allelic imbalance to colorectal cancer. Nat Commun 2018;9:3664.
27. Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417–419.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
29. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
30. Mi H, Muruganujan A, Huang X, et al. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc 2019;14:703–721.
31. **Newman AM, Liu CL**, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–457.

32. Chen B, Khodadoust MS, Liu CL, et al. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 2018;1711:243–259.
33. Luoto S, Hermelo I, Vuorinen EM, et al. Computational characterization of suppressive immune microenvironments in glioblastoma. Cancer Res 2018;78:5574–5585.
34. Ongen H, Andersen CL, Bramsen JB, et al. Putative cis-regulatory drivers in colorectal cancer. Nature 2014;512:87–90.
35. Häkkinen A, Zhang K, Alkodsi A, et al. PRISM: recovering cell type specific expression profiles from composite RNA-seq data [published online ahead of print March 15, 2021]. Bioinformatics https://doi.org/10.1093/bioinformatics/btab178.
36. **Smillie CS, Biton M, Ordovas-Montanes J**, et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. Cell 2019;178:714–730.e22.
37. **Li H, Courtois ET**, Sengupta D, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 2017;49:708–718.
38. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.
39. **Ahtiainen M, Wirta E-V**, Kuopio T, et al. Combined prognostic value of CD274 (PD-L1)/PDCDl (PD-1) expression and immune cell infiltration in colorectal cancer as per mismatch repair status. Mod Pathol 2019;32:866–883.
40. Wirta E-V, Seppälä T, Friman M, et al. Immunoscore in mismatch repair-proficient and -deficient colon cancer. Hip Int 2017;3:203–213.
41. Galon J, Mlecnik B, Bindea G, et al. Towards the introduction of the "Immunoscore" in the classification of malignant tumours. J Pathol 2014;232:199–209.
42. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–3100.
43. **Sedlazeck FJ, Rescheneder P**, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15:461–468.
44. Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun 2017;8:14061.
45. Simpson JT, Workman RE, Zuzarte PC, et al. Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 2017;14:407–410.
46. Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics 2016;32:1446–1453.
47. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics 2017;33:2381–2383.
48. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics 2016;32:587–589.
49. Roadmap Epigenomics Consortium, **Kundaje A, Meuleman W**, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–330.
50. **Yan J, Enge M**, Whitington T, et al. Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. Cell 2013;154:801–813.
51. Sondka Z, Bamford S, Cole CG, et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer 2018;18:696–705.
52. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415–421.
53. **Kumar R, Nagpal G**, Kumar V, et al. HumCFS: a database of fragile sites in human chromosomes. BMC Genomics 2019;19:985.
54. **Rajaram M, Zhang J, Wang T**, et al. Two distinct categories of focal deletions in cancer genomes. PLoS One 2013;8:e66264.
55. West NR, Owens BMJ, Hegazy AN. The oncostatin M-stromal cell axis in health and disease. Scand J Immunol 2018;88:e12694.
56. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene 2017;36:1461–1473.
57. Vincan E, Barker N. The upstream components of the Wnt signalling pathway in the dynamic EMT and MET associated with colorectal cancer progression. Clin Exp Metastasis 2008;25:657–663
58. Garrison WD, Battle MA, Yang C, et al. Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. Gastroenterology 2006;130:1207–1220.
59. **Yang M, Li S-N, Anjum KM**, et al. A double-negative feedback loop between Wnt-β-catenin signaling and HNF4α regulates epithelial-mesenchymal transition in hepatocellular carcinoma. J Cell Sci 2013;126:5692–5703.
60. Chahar S, Gandhi V, Yu S, et al. Chromatin profiling reveals regulatory network shifts and a protective role for hepatocyte nuclear factor 4 during colitis. Mol Cell Biol 2014;34:3291–3304.
61. **Rheinbay E, Nielsen MM, Abascal F**, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature 2020;578:102–111.
62. Mäki-Nevala S, Ukwattage S, Olkinuora A, et al. Somatic mutation profiles as molecular classifiers of ulcerative colitis-associated colorectal cancer. Int J Cancer 2021;148:2997–3007.
63. Liu W, Dong X, Mai M, et al. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. Nat Genet 2000;26:146–147.
64. Koo B-K, Spit M, Jordens I, et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. Nature 2012;488:665–669.
65. **Giannakis M, Hodis E**, Jasmine Mu X, et al. RNF43 is frequently mutated in colorectal and endometrial cancers. Nat Genet 2014;46:1264–1266.
66. **Jorissen RN, Christie M**, Mouradov D, et al. Wild-type APC predicts poor prognosis in microsatellite-stable proximal colon cancer. Br J Cancer 2015;113:979–988.

67. **Grasso CS, Giannakis M, Wells DK**, et al. Genetic mechanisms of immune evasion in colorectal cancer. Cancer Discov 2018;8:730–749.
68. **Muto Y, Maeda T**, Suzuki K, et al. DNA methylation alterations of AXIN2 in serrated adenomas and colon carcinomas with microsatellite instability. BMC Cancer 2014;14:466.
69. Bond CE, McKeone DM, Kalimutho M, et al. RNF43 and ZNRF3 are commonly altered in serrated pathway colorectal tumorigenesis. Oncotarget 2016;7:70589–70600.
70. Fessler E, Drost J, Hooff SR, et al. TGFβ signaling directs serrated adenomas to the mesenchymal colorectal cancer subtype. EMBO Mol Med 2016;8:745–760.
71. Picard E, Verschoor CP, Ma GW, et al. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. Front Immunol 2020;11:369.
72. Kaetzel CS. The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. Immunol Rev 2005;206:83–99.
73. **Nanki K, Fujii M**, Shimokawa M, et al. Somatic inflammatory gene mutations in human ulcerative colitis epithelium. Nature 2020;577:254–259.
74. Kakiuchi N, Yoshida K, Uchino M, et al. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. Nature 2020;577:260–265.
75. Olafsson S, McIntyre RE, Coorens T, et al. Somatic evolution in non-neoplastic IBD-affected colon. Cell 2020;182:672–684.e11.
76. **West NR**, Oxford IBD Cohort Investigators, **Hegazy AN**, et al. Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor–neutralizing therapy in patients with inflammatory bowel disease. Nat Med 2017;23:579–589.
77. West NR, Murray JI, Watson PH. Oncostatin-M promotes phenotypic changes associated with mesenchymal and stem cell-like differentiation in breast cancer. Oncogene 2014;33:1485–1494.

Author names in bold designate shared co-first authorship.

Received January 18, 2021. Accepted April 16, 2021.

Correspondence
Address correspondence to: Kristiina Rajamäki, PhD, Department of Medical and Clinical Genetics and Applied Tumor Genomics Research Program, Research Programs Unit, PO Box 63 (Haartmaninkatu 8) FI-00014 University of Helsinki, Helsinki, Finland. e-mail: ekristiina.rajamaki@helsinki.fi; or Lauri A. Aaltonen, MD, PhD, Department of Medical and Clinical Genetics and Applied Tumor Genomics Research Program, Research Programs Unit, PO Box 63 (Haartmaninkatu 8) FI-00014 University of Helsinki, Helsinki, Finland. e-mail: lauri.aaltonen@helsinki.fi.

Acknowledgments
The authors thank Marjo Rajalaakso, Sini Marttinen, Sirpa Soisalo, Inga-Lill Åberg, Iina Vuoristo, Alison London, Justyna Kolakowska, and Heikki Metsola for excellent technical support, and Iikka Järvinen for helping in the immune cell scoring. The authors acknowledge the computational resources provided by the ELIXIR node, hosted at the CSC–IT Center for Science, Finland. Kristiina Rajamäki and Aurora Taira contributed equally to this work.

CRediT Authorship Contributions

Aurora Taira, MSc (Conceptualization: Lead; Data curation: Equal; Formal analysis: Lead; Investigation: Lead; Methodology: Equal; Software: Equal; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead).

Kristiina Rajamäki, PhD (Conceptualization: Lead; Data curation: Equal; Formal analysis: Lead; Investigation: Lead; Project administration: Lead; Visualization: Equal; Writing – original draft: Lead; Writing – review & editing: Lead).

Riku Katainen, PhD (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Equal; Investigation: Supporting; Methodology: Equal; Software: Equal; Visualization: Supporting; Writing – original draft: Equal; Writing – review & editing: Supporting).

Niko Välimäki, PhD (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Software: Equal; Visualization: Supporting; Writing – original draft: Equal; Writing – review & editing: Supporting).

Anna Kuosmanen, PhD (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Software: Equal; Visualization: Equal; Writing – original draft: Equal; Writing – review & editing: Supporting).

Roosa-Maria Plaketti, BSc (Data curation: Supporting; Formal analysis: Equal; Investigation: Supporting; Methodology: Supporting; Software: Equal; Visualization: Supporting; Writing – original draft: Equal; Writing – review & editing: Supporting).

Toni T. Seppälä, MD, PhD (Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Equal).

Maarit Ahtiainen, PhD (Investigation: Supporting; Methodology: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting).

Erkki-Ville Wirta, MD, PhD (Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting).

Emilia Vartiainen, BSc (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Software: Supporting; Visualization: Supporting; Writing – review & editing: Supporting).

Päivi Sulo, MSc (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Methodology: Supporting; Software: Supporting; Visualization: Supporting; Writing – review & editing: Supporting).

Janne Ravantti, PhD (Methodology: Supporting; Resources: Equal; Software: Supporting; Writing – review & editing: Supporting).

Suvi Lehtipuro, PhD (Methodology: Supporting; Software: Supporting; Writing – review & editing: Supporting).

Kirsi J. Granberg, PhD (Methodology: Supporting; Writing – review & editing: Equal).

Matti Nykter, PhD (Methodology: Supporting; Supervision: Supporting; Writing – review & editing: Supporting).

Tomas Tanskanen, MD, PhD (Data curation: Supporting; Investigation: Supporting; Writing – review & editing: Supporting).

Ari Ristimäki, MD, PhD (Data curation: Equal; Resources: Equal; Writing – review & editing: Supporting).

Selja Koskensalo, MD, PhD (Resources: Equal; Writing – review & editing: Supporting).

Laura Renkonen-Sinisalo, MD, PhD (Resources: Equal; Writing – review & editing: Supporting).

Anna Lepistö, MD, PhD (Resources: Equal; Writing – review & editing: Supporting).

Jan Böhm, MD, PhD (Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting).

Jussi Taipale, PhD (Conceptualization: Supporting; Resources: Supporting; Writing – review & editing: Supporting).

Jukka-Pekka Mecklin, MD, PhD (Funding acquisition: Equal; Supervision: Supporting; Writing – review & editing: Supporting).

Mervi Aavikko, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Supporting; Investigation: Equal; Project administration: Equal; Supervision: Lead; Writing – review & editing: Equal).

Kimmo Palin, PhD (Conceptualization: Supporting; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Software: Equal; Supervision: Lead; Visualization: Supporting; Writing – review & editing: Supporting).

Lauri A. Aaltonen, MD, PhD (Conceptualization: Equal; Funding acquisition: Lead; Project administration: Equal; Supervision: Lead; Writing – review & editing: Equal).

Conflicts of interest
The authors disclose no conflicts.

Funding
This study was supported by The Finnish Center of Excellence in Tumor Genetics and other Academy of Finland grants 312041, 335823, 250345, 319083, 320149, and 320185. Cancer Foundation Finland (Lauri A. Aaltonen, Jukka-Pekka Mecklin, Toni T. Seppälä), iCAN Digital Precision Cancer Medicine Flagship (Lauri A. Aaltonen, Kimmo Palin), Sigrid Jusélius Foundation (Lauri A. Aaltonen, Ari Ristimäki, Toni T. Seppälä), Doctoral Programme in Biomedicine, University of Helsinki (Aurora Taira), Jane and Aatos Erkko Foundation (Jukka-Pekka Mecklin), UEF state research funding (Jukka-Pekka Mecklin), Emil Aaltonen Foundation (Toni T. Seppälä), Finnish Medical Foundation (Toni T. Seppälä), Finnish Cancer Organizations (Ari Ristimäki), Finska Läkaresällskapet (Ari Ristimäki), Helsinki University Central Hospital Research Funds (Ari Ristimäki), and Instrumentarium Science Foundation (Toni T. Seppälä).