Ikuta *et al. Molecular Cancer* (2024) 23:249
https://doi.org/10.1186/s12943-024-02159-9

Molecular Cancer

RESEARCH Open Access

Variability in non-tumor areas of colorectal cancer patients as revealed by endoscopic intestinal step biopsies

Shoko Ikuta[1,2], Yutaka Saito[3], So Takata[1,2], Yoichiro Nakatani[1], Izumi Nagatomo[2,4], Satoshi Shiba[5], Yoshito Takeda[2], Yasushi Totoki[1], Sayaka Mizutani[6], Hironori Sunakawa[7], Hiroaki Ikematsu[7], Hiroyuki Takamaru[3], Atsushi Kumanogoh[2,8,9,10,11,12] and Shinichi Yachida[1,8*]

Abstract

A comprehensive endoscopic small and large intestinal untargeted step biopsy procedure was conducted to compare gene expression between the normal intestinal mucosa of healthy individuals and that of patients with colorectal tumors. From 78 participants (healthy individuals [*n* = 17], patients with colorectal conventional adenomas [*n* = 6], patients with Tis–T1 colorectal cancer [*n* = 41], patients with T2–4 colorectal cancer [*n* = 14]), biopsies of normal mucosa of the terminal ileum, right-sided colon (cecum and ascending colon), and left-sided colorectum (descending colon, sigmoid colon, and rectum) were obtained using a lower gastrointestinal endoscope. RNA was extracted from all samples, and total transcriptome sequencing was performed. Transcriptome data from 388 samples was analyzed. DNA was also extracted from tumor biopsy tissues and analyzed for whole-exome sequencing. In healthy individuals, gene expression differed significantly among the terminal ileum, right-sided colon, and left-sided colorectum, presumably linked to embryological factors. There were differences in gene expression in the normal mucosa in colorectal cancer patients, compared to healthy controls. Patients with tumors, especially T2–4 colorectal cancer, showed considerable variation in gene expression in non-tumor tissues, even in the terminal ileum distant from the tumor site. Based on endoscopic biopsies, the results imply cancer-predisposing conditions in seemingly normal tissues. The present study points to the importance of small intestine and cancer-predisposing conditions in the colon of colorectal cancer patients, with possible implications for developing novel immunotherapy and other therapeutic modalities.

Keywords Colorectal cancer, Adenoma, Transcriptome, Field cancerization, Right-sided colon, Left-sided colorectum, Terminal ileum, Immune cell, CIBERSORTx

*Correspondence:
Shinichi Yachida
syachida@cgi.med.osaka-u.ac.jp

Full list of author information is available at the end of the article

Introduction

Colorectal cancer (CRC) affects approximately 1.9 million people annually [1]. A gradual accumulation of mutations in the normal mucosa through to adenoma, early-stage CRC, and advanced CRC is thought to be etiologically involved in a process known as the adenoma-carcinoma sequence [2]. Cancer-causing changes may occur even before any adenomas arise, and identification of examples would have obvious implications for early detection and potentially prevention of cancer development.

Field cancerization was first proposed for oral squamous cell carcinoma in 1953 [3]. This concept of certain cancer-prone areas or fields of epithelium might also be pertinent for the CRC case [4, 5]. CRC patients undertaking total lesion resection are known to have a persistent elevated risk of newly developing advanced adenomas [6] or a second CRC [7, 8]. While knowledge as to epigenetic alteration [9] and somatic mutations is accumulating [10], gene expression characteristics mostly remain to be determined, particularly when taking into account anatomic differences in the colorectum.

It is widely recognized that the characteristics of CRC differ depending on the anatomic locations. CRC has a higher incidence in the left-sided colorectum compared to the right-sided colon, with the adenoma-carcinoma pathway being the predominant pathway. Mutations in the DNA mismatch repair pathway play a significant role in right-sided colon cancer [11]. The occurrence of primary carcinoma of the ileum is exceptionally uncommon. Those differences indicate that normal intestine mucosa has different carcinogenic factors depending on the anatomic sites.

Therefore, we first assessed whether gene expression in the intestinal mucosa of healthy individuals differed depending on the anatomical site. Next, we compared the normal intestinal mucosa of healthy individuals and CRC patients for each anatomic site. We found that gene expression in non-tumorous tissue in CRC patients differed from that of healthy individuals, especially in patients with advanced CRC. Interestingly, gene expression in patients with T2–4 CRC proved significantly altered in the terminal ileum, anatomically distant from tumors.

Materials and methods

Study subjects and sample collection

This study was conducted on individuals who underwent colonoscopy at the National Cancer Center Hospital (NCCH), Tokyo, Japan, and the National Cancer Center Hospital East (NCCHE), Chiba, Japan. The study's inclusion criteria were as follows: (i) those individuals who are diagnosed with colorectal diseases such as colorectal tumors (including CRC and colorectal polyps) and inflammatory bowel disease, have a history of such ailments, or are suspected of having such diseases and having an upcoming colonoscopy, (ii) individuals 20 years or older, (iii) individuals not having any blood coagulation disorders, including hemophilia, idiopathic thrombocytopenic purpura, liver cirrhosis, or the use of antithrombotic medication, and (iv) written consent to partake in the study, regardless of their racial background. The subjects were provided with a low residue diet the day before the colonoscopy procedure [12]. Sodium-calcium-ascorbic acid combination powder (Moviprep, EA Pharma, Tokyo, Japan) was used for bowel preparation [13]. Biopsies of the normal mucosa of the terminal ileum, right-sided colon (cecum and ascending colon), and left-sided colorectum (descending colon, sigmoid colon and rectum) were performed on each participant without causing physical discomfort. Biopsies were also taken from CRCs and adenomas, along with normal mucosa 1–2 cm away. Where possible, two biopsies were taken from each site, and immediately frozen on dry ice. Subsequently, the biopsy tissue was cryopreserved at -80 °C. Normal mucosa biopsies from the transverse colon were obtained in only a few cases, when tumors were present in the transverse colon. Biopsies were not obtained from all target sites in patients where passage through the intestinal tract was difficult due to the presence of the tumor or when there was considered to be high risk for bleeding.

The CRC patients included in the study had no prior treatment history at the time of biopsy. The recording of CRC stages was based on the TNM Classification of Malignant Tumors by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC). Out of the 738 participants collected, a careful selection was made to include 17 healthy controls, 6 with conventional colorectal adenomas, 41 with Tis–T1 CRC, and 14 with T2–4 CRC. As this is an observational study, the number of participants was not determined from a predetermined sample size calculation or power analysis. Instead, cases with comprehensive clinical information were selected. The analysis excluded patients with autoimmune diseases and those who were currently receiving or had previously received immunosuppressive or immunomodulatory therapies. Given the broad age range of participants, and the potential for age to act as a confounding variable, we applied regression analysis to adjust for its influence. This selection aimed to investigate genetic abnormalities associated with the transition from normal to tumorigenesis, excluding cases with limited specimens or insufficient clinical data (Supplementary Fig. 1A). The study did not include individuals with hereditary colorectal tumors such as familial adenomatous polyposis and Lynch syndrome.

RNA extraction and sequencing

Total RNA was extracted using either an RNeasy Mini kit (QIAGEN, Hilden, Germany) with an RNase-Free DNase Set (QIAGEN) or an AllPrep DNA/RNA Mini kit (QIAGEN). RNA was quantified using NanoDrop (Thermo Fisher Scientific, Waltham, MA), and RNA integrity was assessed with a 4200 TapeStation (Agilent Technologies, Santa Clara, CA) and a 2100 BioAnalyser (Agilent Technologies) before preparing RNA-seq libraries. Samples with a concentration 20 ng/μL or greater and an RNA integrity number (RIN) of 5.5 or higher as measured by BioAnalyser underwent RNA-seq library preparation. Finally, a total of 397 samples underwent RNA-seq library preparation using TruSeq Stranded mRNA (Illumina, San Diego, CA) and sequencing on a 100 bp paired-end Illumina Novaseq 6000 platform.

RNA-seq data analysis

RNA-seq data were acquired in FASTQ format. Trim Galore (version 0.6.6), a wrapper tool around Cutadapt and FastQC, was employed to remove low-quality reads and trim adapters. The peak of the per base quality score was established at Q30, while the cut-off for per sequence GC content was below 30%. Kallisto (version 0.46.2) quantification was performed for the trimmed files. A Kallisto index was built with reference to transcriptome GRCh38 with a k-mer length of 31. The count data were subsequently analyzed with R (version 4.0.5), using the R package tximport (version 1.18.0) to convert transcript levels to gene-level expression.

Outlier samples identified through principal component analysis (PCA) or read count calculation were excluded. Samples of hyperplastic polyps occurring simultaneously with colorectal adenomas or in patients with CRC were excluded from the analysis. Finally, a total of 388 samples were analyzed. Genes that did not encode a protein or had a total transcript per million (TPM) of less than 10 for all samples were excluded, leaving 17,038 genes.

Differentially expressed gene (DEG) analysis was performed using edgeR (version 3.32.1) and limma (version 3.46.0) to exclude the effects of multiple sampling from individuals. Gene ontology (GO) analysis was performed with clusterProfiler (version 3.18.1) and org.Hs.eg.db (version 3.12.0). The DEG thresholds were established based on the analysis objectives. A stringent criterion was established, requiring a log2 fold change (log2FC) absolute value > 1.5 and FDR (False Discovery Rate) < 0.001 when comparing intestinal sites. To identify slight distinctions in healthy mucosa of CRC patients, the thresholds were set at log2FC absolute value > 0.8 and FDR < 0.01. Regression analysis was performed to examine the correlation with age for each DEG, and genes correlated with age were excluded from the DEG. In addition, regression analysis was employed to investigate the relationship between age and sex when comparing the normal mucosa of CRC patients and healthy individuals. This was necessary due to the limited sample size, which precluded the elimination of any potential impact arising from sex differences.

Immune cell abundance was analyzed with the CIBERSORTx algorithm on the web portal (https://cibersortx.stanford.edu/). We input the TPM normalized gene expression data and calculated immune cell abundance using LM22, which was the default signature matrix to calculate the abundance of 22 types of immune cells. All sample results yielded a *P*-value, computed using the default CIBERSORTx program, below 0.05.

DNA extraction and whole-exome sequencing

DNA was extracted from frozen biopsy samples using a QIAamp DNA Mini kit (QIAGEN) or an AllPrep DNA/RNA Mini kit (QIAGEN) and quantified with NanoDrop (Thermo Fisher Scientific). A whole-exome sequence library was prepared with 200 ng gDNA using Illumina DNA prep with Enrichment (Illumina) and the Twist Comprehensive Exome Panel (Twist Bioscience, South San Francisco, CA). Libraries were sequenced using a 150 bp paired-end Illumina Novaseq 6000 platform. Data were acquired in FASTQ format. We aligned whole-exome, paired-end reads (150 bp) to the human genome (GRCh37) using BWA (version 0.7.15, https://github.com/lh3/bwa), and generated BAM files by computing MarkDuplicates, BaseRecalibrator, and ApplyBQSR with GATK (version 4.1.9.0, https://github.com/broadinstitute/gatk). We used Manta (version 1.6.0, https://github.com/Illumina/manta) and Strelka2 (version 2.9, https://github.com/Illumina/strelka) for somatic mutation calls and Ensembl Variant Effect Predictor (release 102, https://github.com/Ensembl/ensembl-vep) for functional annotation. Driver genes were selected from the significantly mutated genes presented by Myer et al. [14].

Results

Endoscopic intestinal step biopsy procedure

A total of 78 cases who underwent lower gastrointestinal endoscopy were enrolled for this study: 17 healthy controls, 6 with conventional colorectal adenomas, 41 with Tis–T1 CRC, and 14 with T2–4 CRC. According to the AJCC and UICC TNM classifications, which are identical to the T classification, CRC patients were categorized into Tis–T1 and T2–4 stage groups. Patient characteristics are shown in Supplementary Tables S1-S3.

Biopsies were taken from various sections of the intestinal mucosa in the participants, such as the terminal ileum, right-sided colon, left-sided colorectum, CRC, adenoma, and adjacent normal mucosa 1–2 cm away from CRC or adenoma (Fig. 1A). We termed this

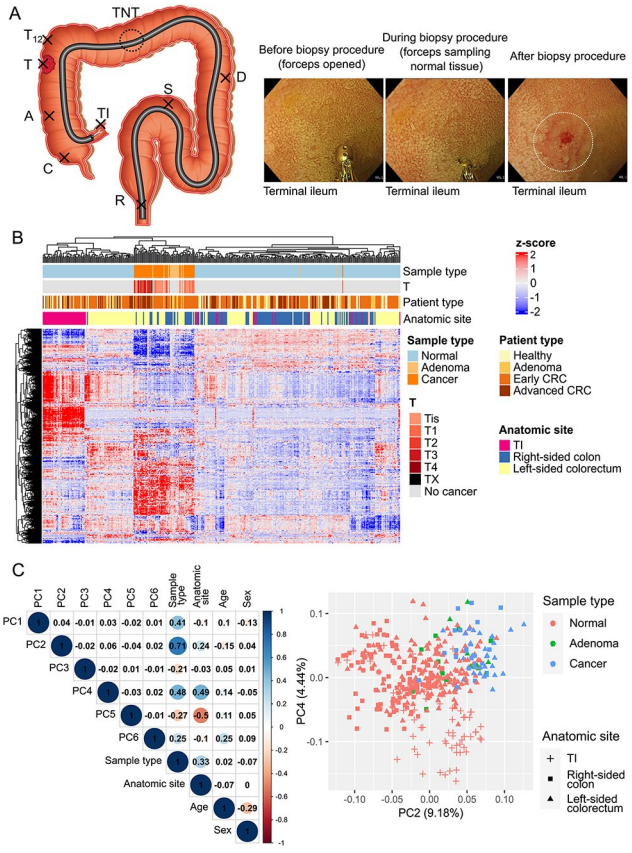

Fig. 1 Overview of the endoscopic intestinal step biopsy procedure. (**A**) Endoscopic biopsies of normal appearing mucosa of the terminal ileum and colorectum were performed for all individuals. Tumor samples and 1–2 cm of surrounding normal mucosa (T_{12}) were also obtained where present. Biopsies from the transverse colon were only taken in a few cases because of the higher perforation risk than in other locations. The endoscopic images on the right show a biopsy procedure of the terminal ileum. The white dotted circle indicates the site after the biopsy. A, ascending colon; C, cecum; D, descending colon; R, rectum; S, sigmoid colon; T, tumor; TI, terminal ileum; TNT, transverse colon; T_{12}, normal mucosa 1–2 cm away from the tumor. (**B**) Heatmap of all 388 samples using unsupervised clustering (top 2,500 genes). Colorectal cancer (CRC) and conventional adenoma cases belonged to the same cluster. (**C**) Principal component analysis (PCA) of 388 samples. PC2 strongly correlated with sample type, and PC4 with sample type and anatomic site (left panel). In PCA plots using PC2 and PC4, the terminal ileum samples are generally plotted separately from colorectal tumor samples. Colors (pink, green, blue) indicate "Sample type", and shapes (+, ■, ▲) "Anatomic site"

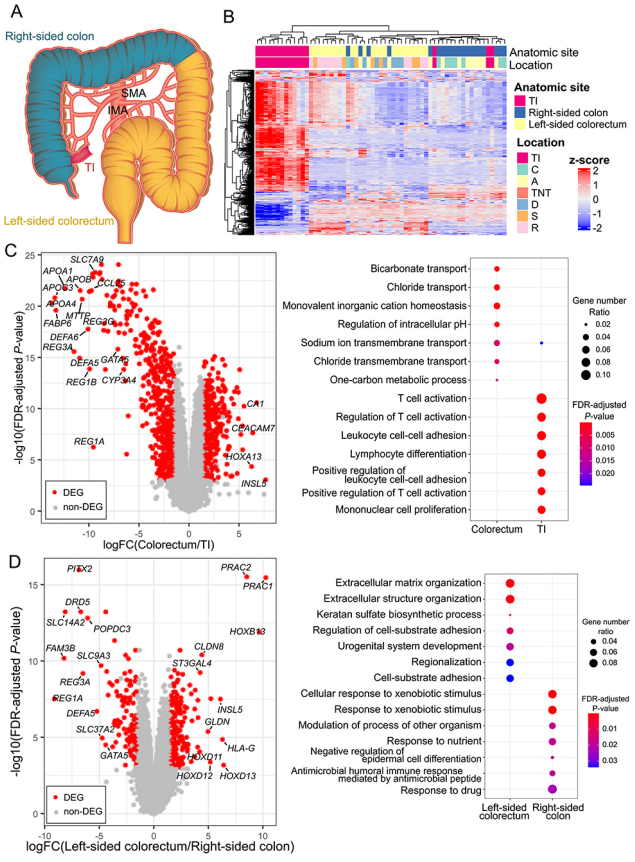

Fig. 2 Anatomic site-specific differences in the normal mucosa of healthy individuals. (**A**) Schema showing the anatomic sites of the terminal ileum (TI), right-sided colon, and left-sided colorectum. The superior mesenteric artery (SMA) dominates the terminal ileum and right-sided colon, embryologically derived from the midgut. The inferior mesenteric artery (IMA) dominates the left-sided colorectum, embryologically derived from the hindgut. (**B**) Heatmap for 61 normal mucosa samples from 17 healthy individuals using unsupervised clustering (top 500 genes). Note: separate clustering of terminal ileum (TI), right-sided colon, and left-sided colorectum. (**C**, **D**) Discrepancies in gene expression between the terminal ileum and the colorectum (**C**) and between the right-sided colon and left-sided colorectum (**D**) in healthy individuals. Volcano plots (left panels) show differences in gene expression in normal mucosa of different anatomic sites, with gene ontology (GO) enrichment analysis results in the right panels. An absolute value of log2 fold change (|log2FC|) > 1.5 and a false discovery rate (FDR) < 0.001 were used as cut-offs for expression variation genes

procedure of collecting untargeted step biopsies of the intestinal mucosa from multiple locations in addition to biopsies from the targeted lesion(s) in the same individual "Endoscopic intestinal step biopsy."

A heatmap of all samples using the top 2,500 genes with high variability showed that CRC and conventional adenoma samples belonged to the same cluster (Fig. 1B). PCA revealed that PC2 and PC4 correlated strongly with sample type (CRC, conventional adenoma, or normal mucosa) and anatomic site (terminal ileum, right-sided colon, or left-sided colorectum) (Fig. 1C **left**). In contrast, age or sex did not exhibit any correlation with any axis. The terminal ileum tissues were distinct from both colonic tissue (tumor and non-tumor areas) when visualized using PC2 and PC4 (Fig. 1C **right**).

Gene expression in the intestinal mucosa of healthy individuals varies with the anatomic site

DEG analysis was performed to examine anatomic site-specific differences in gene expression in the intestine. Each anatomic site is shown in Fig. 2A. Unsupervised clustering of the top 500 variable genes showed that the terminal ileum, right-sided colon (cecum, ascending

colon, transverse colon), and left-sided colorectum (descending colon, sigmoid colon, rectum) belonged to different clusters (Fig. 2B), indicating differences in embryological origin. The data obtained from PCA analysis did not indicate any discernible distinctions between the two hospitals (Supplementary Figure S1B). Age showed no correlation with the primary axis in the PCA analysis (Supplementary Figure S1C). However, considering the influence of age on transcriptome analysis data, we excluded DEGs that, based on regression analysis, appeared to have a potential association with age.

In the comparison between the colorectum and the terminal ileum, there were 877 DEGs up-regulated in the terminal ileum and 338 in the colorectum (Fig. 2C). In the terminal ileum, there was a significant enrichment of GO terms related to immune cell proliferation and activity. The expression of *CCL25*, a chemokine produced by the small intestinal epithelium and involved in lymphocyte homing, significantly differed between the terminal ileum and colorectum. Although not detected as a GO term, genes related to lipid transport, such as *APOA*, *FABP6*, and *MTTP*, were more highly expressed in the terminal ileum. In the colorectum, GO terms related to substance transport were enriched (Fig. 2C).

The comparison between the left-sided colorectum and the right-sided colon revealed an up-regulation of 210 DEGs in the left-sided colorectum and 122 DEGs in the right-sided colon (Fig. 2D). In the left-sided colorectum, GO terms associated with extracellular matrix organization and cell adhesion were highly enriched. The left-sided colorectum exhibited high expression levels of *PRAC1* and *PRAC2*, which are specifically expressed in the human prostate and distal colorectum [15]. Homeobox genes, specifically *HOXB13*, *HOXD11*, *HOXD12*, and *HOXD13* were upregulated in the left-sided colon. In the right-sided colon, GO terms related to response to xenobiotic stimulus and antimicrobial peptide were enriched (Fig. 2D). The REG (regenerating gene) family with various physiological activities in the intestinal tract, including antibacterial, anti-inflammatory, and antiapoptotic potential [16], exhibited significantly higher expression in the right-sided colon. Lists of DEGs are shown in Supplementary Tables S4 and S5.

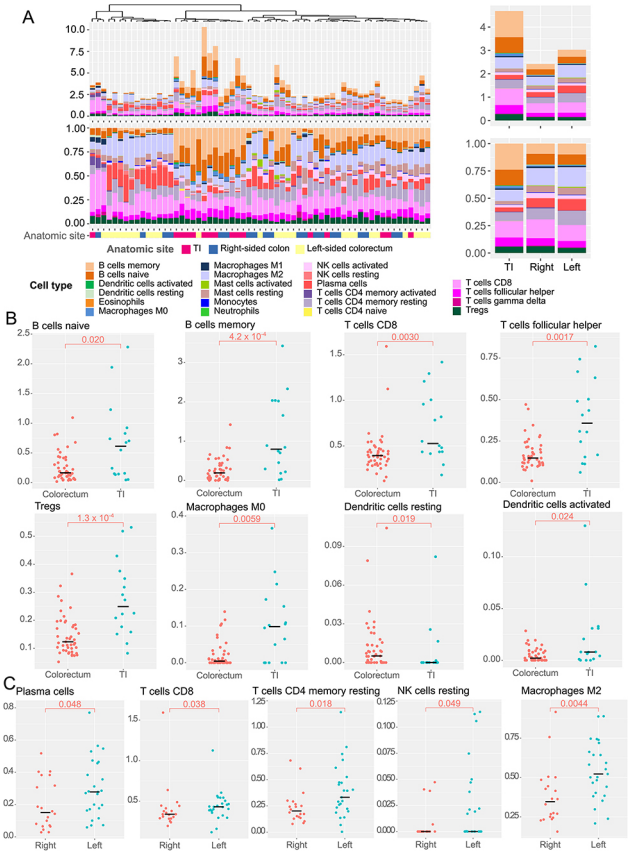

Fig. 3 Differences in immune cell profiles of healthy individuals by anatomic site. (**A**) Bar plot (left panel) showing the results of unsupervised clustering of immune profiles of intestinal mucosa in healthy individuals calculated with CIBERSORTx. Bar plot (right panel) showing the mean values for the terminal ileum (TI), right-sided colon (Right), and left-sided colorectum (Left). (**B**, **C**) Depending on the anatomic site, immune cells were found in differing amounts (Mann-Whitney *U* test). The Y-axis indicates absolute abundance. TI, terminal ileum; Right, right-sided colon; Left, left-sided colorectum

Immune profiles of intestinal mucosa of healthy individuals differ by anatomic site

To investigate how immune cells differ by anatomic site, we compared the immune profiles of the terminal ileum, right-sided colon, and left-sided colorectum of the normal mucosa of healthy individuals using CIBERSORTx. Unsupervised clustering showed that the terminal ileum samples were predominantly characterized by high B cell immunity (Fig. 3A). Among the immune cells, lymphocytes, followed by macrophages, accounted for a large proportion. M2 macrophages were found to be the prevailing type among macrophages in the terminal ileum, right-sided colon, and left-sided colorectum (Fig. 3A). In contrast, γδT cells, eosinophils, and neutrophils were observed in only a few samples.

The terminal ileum exhibited significantly more immune cells than the colorectum ($P=0.0039$, Mann-Whitney *U* test) (Fig. 3A, Supplementary Figure S2A), with more B cells than the colorectum and more $CD8^+$ T cells, follicular helper T cells, regulatory T cells, and M0 macrophages. There were fewer resting dendritic cells and more active ones (Fig. 3B). The abundance of B cells, active dendritic cells, and follicular helper T cells in the terminal ileum presumably reflects abundant Peyer's patches and active antigen presentation. Data on the abundance of cell types are provided in Supplementary Figure S2B.

A site-specific comparison showed that the left-sided colorectum displayed a significantly elevated presence of immune cells compared to the right-sided colon ($P=0.022$, Mann-Whitney U test) (Fig. 3C, Supplementary Figure S3A). The left-sided colorectum exhibited a higher presence of plasma cells, $CD8^+$ T cells, resting memory $CD4^+$ T cells, resting NK cells, and macrophage M2. The data regarding the abundance of cell types is presented in Supplementary Figure S3B. We also examined the disparity in gene expression between rectal and colonic tissue in healthy individuals. The rectum exhibited a higher count of immune cells, denoted as Absolute score, compared to the colon (Supplementary Figure S4A). In the rectum, B cell memory, T cell CD4 memory resting, macrophage M2, and resting mast cells were more prevalent, whereas activated T cell CD4 memory was more prevalent in the colon (Supplementary Figure S4B).

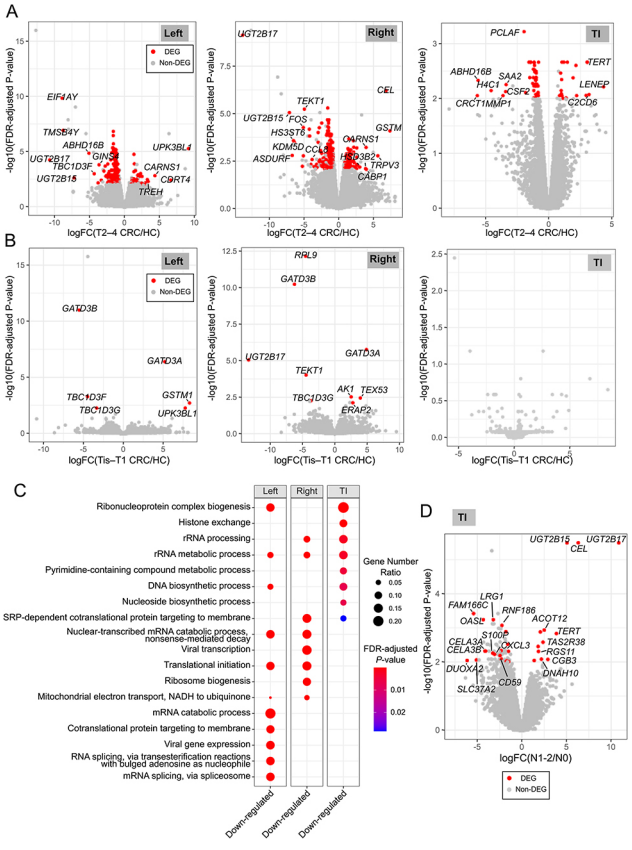

Fig. 4 Differences in gene expression between normal mucosa of CRC patients and healthy individuals. (**A**, **B**) Volcano plots show gene expression differences between normal mucosa of CRC patients and healthy controls (HC). The comparison for T2–4 CRC patients is provided in **A**, and for Tis–T1 CRC patients (Tis and T1), in **B**. Left-sided colorectum (Left), right-sided colon (Right), and terminal ileum (TI) are shown in order from left to right. After conducting a multiple regression analysis that accounted for age and sex, the differential expression analysis omitted genes suspected of introducing confounding variables. These genes are visually represented as gray dots in the figure. (**C**) Results of gene ontology (GO) enrichment analysis of expression variation genes in normal mucosa of T2–4 CRC patients and normal mucosa of healthy controls. FDR, false discovery rate; Left, left-sided colorectum; logFC, log2 fold change; Right, right-sided colon; TI, terminal ileum. (**D**) Volcano plot showing gene expression differences between two groups of cases: those with lymph node metastasis ($n=11$) and those without ($n=11$). The thresholds were set at log2FC absolute value > 0.8 and FDR < 0.01. Regression analysis was performed to examine the correlation with age, sex, and T stage for each DEG, and genes correlated with them were excluded from the DEG

Normal-appearing intestinal mucosa of CRC patients exhibits different gene expression from healthy individuals

A DEG analysis was conducted in order to examine potential differences in gene expression between the normal intestinal mucosa of CRC patients and healthy individuals. Due to the significant variation in gene expression across different anatomic sites in the intestinal tract, a differential expression analysis was performed for each site. The DEG analysis did not factor in the distance from the CRC, as PCA did not establish a distinction between normal colonic mucosa close to or distant from the CRC (Supplementary FigureS5). In order to take account of the age and sex disparities among healthy individuals, Tis–T1 CRC patients, and T2–4 CRC patients, regression analysis was conducted to eliminate any DEGs that could be attributed to age or sex. When comparing T2–4 CRC patients with healthy individuals, more DEGs were observed across all anatomical sites, in contrast to the comparison between Tis–T1 CRC patients and healthy individuals. In the left-sided colorectum, right-sided colon, and terminal ileum, 259, 316, and 71 DEGs were detected in T2–4 CRC patients, respectively (Fig. 4A and B). Notably, many DEGs were also identified in the terminal ileum of T2–4 CRC patients, even when the terminal ileum was distant from the CRC anatomically. A detailed compilation of the DEGs is presented in Supplementary Tables S6-S10.

GO analysis was performed on the DEGs of T2–4 CRC patients' normal mucosa compared to healthy individuals' normal mucosa (Fig. 4C). The down-regulation of transcription and translation-related GO terms, which indicate the involvement of ribosome-related genes (e.g., *RPS18*, *RPS17*, *RPL24*), was observed in patients with T2–4 CRC across all anatomic sites.

Furthermore, we analyzed the disparities in gene expression in the terminal ileum between two groups of cases: those with lymph node metastasis ($n=11$) and those without ($n=11$), among T1 to T4 cases (total of 22 cases) (Fig. 4D). In cases with lymph node metastasis, the terminal ileum exhibited high expression levels of genes encoding glycosyltransferases involved in the metabolism and excretion of both endogenous and exogenous toxic compounds (e.g., *UGT2B17*, *UGT2B15*). On the contrary, the expression of immune-related genes, specifically *CXCL3* and *CD59*, was decreased in these patients.

Differences in mutations between Tis–T1 and T2–4 CRCs on whole-exome sequencing (WES) analysis

Tumor samples from 47 CRC patients were subjected to WES, with 23 patients classified as having Tis–T1 CRCs, 24 with T2–4 CRCs, and 7 with concomitant conventional adenomas. Note: three Tis–T1 CRC patients and ten T2–4 CRC patients were not included in this RNA-seq analysis.

Figure 5A illustrates an oncoplot of somatic mutations of known driver genes in CRCs. Somatic mutations of *TP53* were detected in T2–4 CRCs more frequently than in Tis–T1 CRCs ($P=0.0010$, Fisher's exact test). Interestingly, there was a mutual exclusivity observed between somatic mutations in *TP53* and *KRAS* in all CRCs ($P=0.020$, Fisher's exact test), as well as in CRCs at the Tis–T1 stage ($P=0.021$, Fisher's exact test). This is likely since the study specifically concentrated on early-stage CRC. The typical progression of abnormalities in CRC involves the sequential accumulation of *APC*, *KRAS*, and *TP53* mutation [14, 17]. The presence of *TP53* mutations in colorectal tumors with *APC* abnormalities may obviate the need for *KRAS* mutations in early-stage CRC [18]. *BRAF* mutations were observed in only two cases [19]. In addition, three cases were found to have a hyper-mutated phenotype.

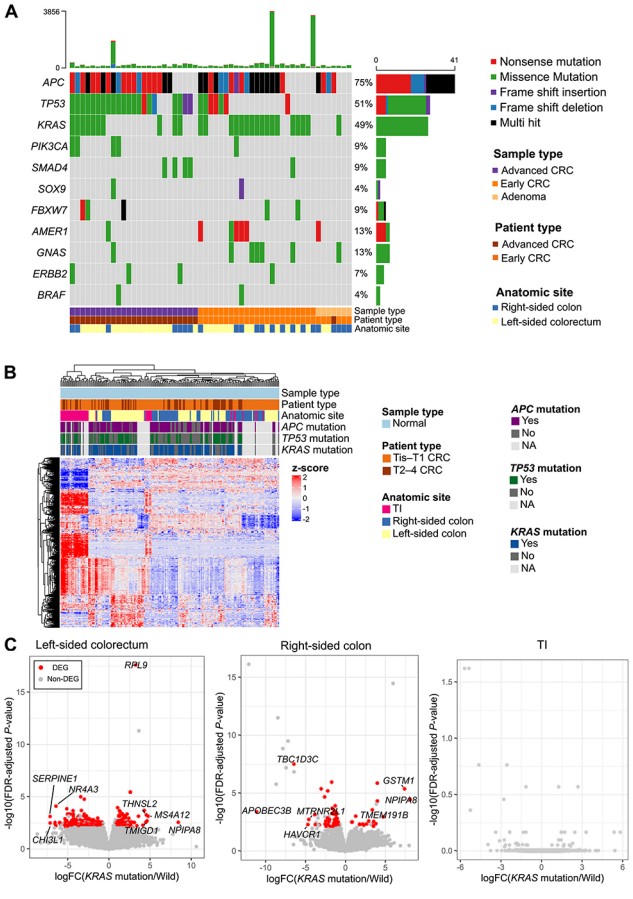

Fig. 5 Mutations revealed by whole-exome analysis and their association with the transcriptome. (**A**) Oncoplot of mutations revealed by whole-exome sequencing analysis of Tis–T1 and T2–4 CRC and adenomas. Note: the inclusion of somatic mutations in known driver genes. The upper panel shows the number of somatic mutations. The most advanced lesion is classified as Patient type. (**B**) Heatmap for normal mucosa samples of 55 CRC patients using unsupervised clustering (top 1,000 genes). No obvious associations were identified between major CRC mutation (i.e., *APC*, *TP53*, and *KRAS*) and gene expression. (**C**) Volcano plots show differences in gene expression between normal mucosa of CRC patients with *KRAS* mutation and that of CRC patients with *KRAS* wild-type. The analysis of the left-sided colorectum is shown in the left panel, the right-sided colon in the center, and terminal ileum on the right. After conducting a multiple regression analysis that accounted for age and sex, the differential expression analysis omitted genes suspected of introducing confounding variables. These genes are visually represented as gray dots in the figure. logFC, log2 fold change; NA, not available; TI, terminal ileum

Examination of the relationship between tumor mutations and gene expression

An assessment was performed to examine the relationship between tumor mutations and gene expression in normal mucosa. The heatmap generated using normal mucosa samples from CRC patients, focusing on the top 1,000 variable genes, revealed no discernible correlation between major three CRC mutations (i.e., *APC*, *TP53*, and *KRAS*) and gene expression (Fig. 5B).

Subsequently, DEG analysis was conducted in order to evaluate potential disparities between normal mucosa of CRC patients with significant mutations and that of CRC patients without mutations. Due to the high prevalence of *APC* mutations in CRC patients and the need to consider genetic abnormalities such as LOH (loss of heterozygosity) in the case of the *TP53* gene, DEG analysis was restricted to the presence or absence of *KRAS* gene mutations. DEG analysis was conducted for each anatomic site, comparing the normal appearing mucosa of CRC

patients with *KRAS* mutations and the *KRAS* wild-type. The study revealed the presence of 404 DEGs in the left-sided colorectum, 71 DEGs in the right-sided colon, and no DEGs in the terminal ileum (Fig. 5C). The Consensus Molecular Subtypes (CMS) framework is being developed as a tool for classifying CRC [20]. CMS1, characterized by hypermutation, microsatellite instability, and pronounced immune activation, is frequently associated with *BRAF* mutations, whereas CMS3, distinguished by epithelial features and marked metabolic dysregulation, often harbors *KRAS* mutations. The findings of this study indicate that the gene expression profiles of normal tissues vary depending on the presence or absence of *KRAS* mutations, suggesting that distinct predisposition states might exist prior to tumor development, as classified by the CMS framework.

Discussion

Analysis of endoscopic intestinal step biopsies in the present study revealed dissimilarities of gene expression between the right colon and left colorectum among healthy individuals. This is likely attributed to the embryological derivation of the former from the midgut and the latter from the hindgut. We also compared gene expression between the intestinal tissues of healthy individuals and those of patients with colorectal tumors. There were significant differences in gene expression between healthy individuals and colorectal tumor patients even in their normal tissues. Notably, patients with CRC exhibited dissimilarities in gene expression at the terminal ileum, separate from the lesion, compared to healthy individuals.

Transcriptome analysis of the intestinal mucosa of healthy individuals showed gene expression to vary significantly depending on anatomic sites. The small intestine has an absorptive epithelium covered with villi and contains many enzymes and transporters necessary for digesting dietary components [21]. The ileum absorbs bile salts and vitamin B12 but contributes little to nutrient absorption and has short villi. The large intestine is responsible for water absorption and excretion of undigested food. In this study, genes related to lipid transport were highly expressed in the terminal ileum, and genes involved in the transport of small water-soluble molecules were highly expressed in the large intestine. Notably, the expression of genes related to xenobiotic stimulus and antimicrobial peptide was high in the right-sided colon, indicating that foreign substances not digested and absorbed in the small intestine may undergo metabolism there [22]. Microarray analysis has also observed variations in gene expression between the left-sided colorectum and the right-sided colon [23, 24]. Differences in anatomical location can be attributed to the developmental derivation of the right-sided colon from the midgut and the left-sided colorectum from the hindgut. Studies have indicated more pronounced anatomical variations in the adult colon's transcriptome than in the fetal colon [24]. The differences in anatomy may result from changes in the postnatal intestinal environment, such as the microbiome and food antigens.

The gene expression in the normal intestinal mucosa of patients with CRC was distinct from that of healthy individuals, particularly in patients with T2–4 CRCs. The normal mucosa of the colorectum in patients with T2–4 CRC exhibited down-regulated expression of genes associated with ribosomal protein. There is a close relationship between ribosome biogenesis and cell proliferation, and it is recognized that up-regulated ribosomal protein is a risk factor for CRC development [25]. Previous works indicates that the process of adenoma and CRC development from normal colon tissue results in an enhancement of ribosomal biogenesis [26]. Additional research is necessary to clarify whether these variations in gene expression are due to modifications in field cancerization, which would suggest a predisposition to CRC in the intestinal environment, or if they are a reaction to CRC. Nevertheless, there were only minimal changes observed in Tis–T1 CRC patients. T2–4 CRC patients and Tis–T1 CRC patients might possess distinct predispositions to CRC. Additionally, the terminal ileum exhibits a down-regulation in rRNA processing and DNA biosynthetic process. This phenomenon is intriguing due to the potential suppression of abundant inflammatory cells at the terminal ileum. Despite previous disregard for the significance of biopsies of the terminal ileum [27], our findings indicate a correlation with the development of CRC, necessitating additional investigation.

A distinctive feature of this study is its utilization of endoscopic multi-sampling of the normal intestinal mucosa, encompassing the terminal mucosa of the ileum. The study included samples from diverse anatomic locations in order to determine gene expression not only in patients with colorectal tumors but also in healthy individuals. In this study, all specimens were collected endoscopically. Previous research has predominantly examined frozen surgical specimens. Nevertheless, surgically frozen specimens have usually been extracted from the body several hours after the ligation of major blood vessels. The effect of surgically induced warm ischemia on CRC tissues in a mouse model was investigated by Atkin et al. [28], who found that mRNA expression changes began within 20 min and increased considerably after 4 h. Conversely, endoscopic biopsy specimens are extracted from the living subject and promptly cryopreserved, facilitating more accurate analysis.

There are several limitations to this study. First, the entire cohort was from two hospitals in the Tokyo metropolitan area, which may include biases in patient

backgrounds such as diet, lifestyle, and age. Since the two hospitals are national centers specializing in cancer treatment, they may have a younger patient population than general area hospitals. In addition, because the sample includes patients who visited the endoscopy division, it includes a large number of patients who were treated for early colorectal lesions by colonoscopy. Second, due to the restricted number of cases, this study could have been susceptible to intra- and interindividual confounding variables that might have impacted gene expression. In order to minimize bias, intra-individual effects were carefully accounted for, specifically specimen collection site, age, and sex. However, the potential influence of other confounding factors cannot be disregarded. Third, this study is a case-control study in which biopsies were taken from one patient at only one time point. It did not look at changes in the intestinal mucosa over time, such as before and after cancer development in the same patient. Therefore, it is not possible to completely rule out that the changes identified in the CRC patients are not confounded by individual differences. To solve this problem, endoscopic intestinal step biopsy of the same patient over time, especially before and after tumor onset, is needed. It remains unclear if the predisposing conditions in this study are directly associated with the onset of CRC. To address this issue, it will be necessary to conduct additional analysis, focusing on the interplay between longitudinal changes in gene expression within the colorectal mucosa and the accumulation of gene mutations and epigenomic abnormalities. Finally, molecular biological validation of gene expression changes in CRC patients may also be needed.

In conclusion, this study provides novel insights into physiological variation with the anatomical site and possible contributions of gene expression changes in normal-appearing mucosa in CRC patients. Since patients with CRC, especially advanced cases, show significant changes even in the anatomically distant terminal ileum, future research and development should focus on the small intestine. Control of the terminal ileum, the center of immunity, may lead to the prevention of colorectal tumorigenesis and the development of stratification and new treatment approaches, such as immunotherapy.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12943-024-02159-9.

Supplementary Material 1: figure S1. Overview of the research conducted. (**A**) Chart illustrating the individuals involved in the study. The study included a total of 78 cases, consisting of 17 healthy controls, 6 individuals with conventional colorectal adenomas, 41 patients with Tis-T1 colorectal cancer (CRC), and 14 patients with T2-4 CRC. (**B**) Principal component analysis (PCA) included 61 samples from 17 healthy individuals. Colors (pink, green) indicate "Facility," and shapes (+, ■, ▲) "Anatomic site." (**C**) Correlation of clinical information with the principal components (PCs) of the transcriptome in 17 healthy individuals. CRC, colorectal cancer; L, left-sided colorectum; NCCH, National Cancer Center Hospital; NCCHE, National Cancer Center Hospital East; R, right-sided colon; TI, terminal ileum; WES, whole-exome sequencing.

Supplementary Material 2: figure S2. Disparity in immune cell profiles between the terminal ileum and colorectum in healthy individuals. (**A**) The CIBERSORTx absolute scores exhibit a notable increase in the terminal ileum (TI) compared to the colorectum. (**B**) The absolute abundance of each immune cell type in healthy individuals. Black lines show the median. Right, right-sided colon; Left, left-sided colorectum.

Supplementary Material 3: figure S3. Disparity in immune cell profiles between the right-sided colon and left-sided colorectum in healthy individuals. (**A**) The CIBERSORTx absolute scores exhibit a notable increase in the left-sided colorectum compared to the right-sided colon. (**B**) The absolute abundance of each immune cell type in healthy individuals. Black lines show the median. Right, right-sided colon; Left, left-sided colorectum.

Supplementary Material 4: figure S4. Disparities in the immune cell compositions between the rectum and colon in healthy individuals. (**A**) There is a significant increase ($P=0.015$) in immune cells in the rectum compared to the colon, as indicated by Absolute scores obtained from CIBERSORTx analysis. (**B**) Absolute abundance of each immune cell type in healthy individuals. The median is represented by the black line. Left, colon; Right, rectum.

Supplementary Material 5: figure S5. Principal component analysis (PCA) of normal colorectal mucosa samples. (**A**) Correlation plot. No principal components (PCs) correlated with distance from the tumor. (**B**) PCA plot. Colorectal mucosa close and far from tumors did not demonstrate different clusters. Colors (pink, green, blue) indicate "Distance from tumor", and shapes (■, ▲) "Anatomic site".

Supplementary Material 6

Supplementary Material 7

Acknowledgements
We thank all patients and their families who participated in this study, Dr. M. Horie and Dr. D. Maeda (Kanazawa University, Kanazawa, Japan) for expert advice, and Ms. N. Sezawa, Ms. R. Ohira, Ms. H. Sekiguchi, and Ms. R. Itsuki (National Cancer Center Hospital, Tokyo, Japan) for expert technical assistance.

Author contributions
S.I. and S.Y. wrote the main manuscript text and prepared figures. Y.S., S.S., H.S., and H.I. designed the work. S.T., Y.N., Y.Totoki, S.M., and. H.Y. contributed to the analysis. I.N., Y.Takeda, and A.K. contributed to interpreting data. S.Y. contributed to the conception. All authors reviewed the manuscript.

Funding
This work was supported by grants from the National Cancer Center Research and Development Fund (2020-A-4 to Y.S. and S.Y. and 2023-A-6 to Y.S. and S.Y.); Practical Research for Innovative Cancer Control from the Japan Agency for Medical Research and Development (AMED) (JP22ck0106546 to Y.S., and S.Y.; JP23ck0106799 to Y.S., and S.Y.); Project for Cancer Research and Therapeutic Evolution (P-CREATE) from AMED (JP21cm0106477 to S.S., and S.Y.); Project for Promotion of Cancer Research and Therapeutic Evolution (P-PROMOTE) from AMED (JP23ama221404 to Y.S., H.S., and S.Y.; JP24ama221430 to Y.S., H.S., and S.Y.); United States-Japan Cooperative Medical Science Program from AMED (JP23jk0210009 to S.Y.); JSPS KAKENHI (20H03662 to S.Y.; 23H02892 to S.Y.); JST-AIP Acceleration Research (JPMKCR19U3 to S.Y.); Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University (to S.Y.); Joint Research Project of the Institute Medical Science, the University of Tokyo (to S.Y.); the Takeda Science Foundation (to S.Y.); the Yasuda Medical Foundation (to S.Y.); the Mitsubishi Foundation (to S.Y.); the Princess Takamatsu Cancer Research Fund (to S.Y.); Yakult Bio-Science Foundation (to S.Y.).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
This study was approved by the Research Ethics Committees of the National Cancer Center, Osaka University and Tokyo Institute of Technology as meeting the ethical guidelines for medical and health research involving human individuals (National Cancer Center, 2018-009; Osaka University, 20218: Tokyo Institute of Technology, 2014018).

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita 565-0871, Osaka, Japan
[2]Department of Respiratory Medicine and Clinical Immunology, Graduate School of Medicine, Osaka University, Suita 565-0871, Osaka, Japan
[3]Endoscopy Division, National Cancer Center Hospital, Chuo-ku, Tokyo 104-0045, Japan
[4]Health and Counseling Center, Osaka University, Suita 565-0871, Osaka, Japan
[5]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan
[6]School of Life Science and Technology, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan
[7]Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa 277-8577, Chiba, Japan
[8]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, Suita 565-0871, Osaka, Japan
[9]Department of Immunopathology, Immunology Frontier Research Center (IFReC), World Premier International Research Center Initiative (WPI), Osaka University, Suita 565- 0871, Osaka, Japan
[10]Center for Infectious Diseases for Education and Research (CiDER), Osaka University, Suita 565-0871, Osaka, Japan
[11]Japan Agency for Medical Research and Development – Core Research for Evolutional Science and Technology (AMED–CREST), Chiyoda-ku 100-0004, Tokyo, Japan
[12]Center for Advanced Modalities and DDS (CAMaD), Osaka University, Suita 565- 0871, Osaka, Japan

Received: 8 September 2024 / Accepted: 15 October 2024
Published online: 07 November 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67.
3. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953;6:963–8.
4. Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics. 2014;103:211–21.
5. Park SK, Song CS, Yang HJ, Jung YS, Choi KY, Koo DH, Kim KE, Jeong KU, Kim HO, Kim H, et al. Field cancerization in sporadic Colon cancer. Gut Liver. 2016;10:773–80.
6. Sekiguchi M, Matsuda T, Saito Y. Surveillance after endoscopic and surgical resection of colorectal cancer. Best Pract Res Clin Gastroenterol. 2016;30:959–70.
7. Shureiqi I, Cooksley CD, Morris J, Soliman AS, Levin B, Lippman SM. Effect of age on risk of second primary colorectal cancer. J Natl Cancer Inst. 2001;93:1264–6.
8. Levi F, Randimbison L, Blanc-Moya R, Maspoli-Conconi M, Rosato V, Bosetti C, La Vecchia C. High constant incidence of second primary colorectal cancer. Int J Cancer. 2013;132:1679–82.
9. Grady WM. Epigenetic events in the colorectum and in colon cancer. Biochem Soc Trans. 2005;33:684–8.
10. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, Georgakopoulos N, Torrente F, Noorani A, Goddard M, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature. 2019;574:532–7.
11. Baran B, Mert Ozupek N, Yerli Tetik N, Acar E, Bekcioglu O, Baskin Y. Difference between left-sided and right-sided colorectal Cancer: a focused review of literature. Gastroenterol Res. 2018;11:264–73.
12. Nishimoto Y, Mizutani S, Nakajima T, Hosoda F, Watanabe H, Saito Y, Shibata T, Yachida S, Yamada T. High stability of faecal microbiome composition in guanidine thiocyanate solution at room temperature and robustness during colonoscopy. Gut. 2016;65:1574–5.
13. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H, Masuda K, Nishimoto Y, Kubo M, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat Med. 2019;25:968–76.
14. Myer PA, Lee JK, Madison RW, Pradhan K, Newberg JY, Isasi CR, Klempner SJ, Frampton GM, Ross JS, Venstrom JM, et al. The Genomics of Colorectal Cancer in populations with African and European ancestry. Cancer Discov. 2022;12:1282–93.
15. Liu XF, Olsson P, Wolfgang CD, Bera TK, Duray P, Lee B, Pastan I. PRAC: a novel small nuclear protein that is specifically expressed in human prostate and colon. Prostate. 2001;47:125–31.
16. Sun C, Wang X, Hui Y, Fukui H, Wang B, Miwa H. The potential role of REG Family Proteins in inflammatory and inflammation-Associated diseases of the gastrointestinal tract. Int J Mol Sci 2021, 22.
17. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA. Jr., Kinzler KW: Cancer genome landscapes. Science. 2013;339:1546–58.
18. Matsumoto K, Urabe Y, Oka S, Inagaki K, Tanaka H, Yuge R, Hayashi R, Kitadai Y, Arihiro K, Shimamoto F, et al. Genomic Landscape of early-stage colorectal neoplasia developing from the Ulcerative Colitis Mucosa in the Japanese Population. Inflamm Bowel Dis. 2021;27:686–96.
19. Yachida S, Mudali S, Martin SA, Montgomery EA, Iacobuzio-Donahue CA. Beta-catenin nuclear labeling is a common feature of sessile serrated adenomas and correlates with early neoplastic progression after BRAF activation. Am J Surg Pathol. 2009;33:1823–32.
20. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21:1350–6.
21. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. Nat Rev Immunol. 2014;14:667–85.
22. Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. Annu Rev Pharmacol Toxicol. 2003;43:149–73.
23. Diez-Obrero V, Dampier CH, Moratalla-Navarro F, Devall M, Plummer SJ, Diez-Villanueva A, Peters U, Bien S, Huyghe JR, Kundaje A, et al. Genetic effects on Transcriptome profiles in Colon epithelium provide functional insights for genetic risk loci. Cell Mol Gastroenterol Hepatol. 2021;12:181–97.
24. Glebov OK, Rodriguez LM, Nakahara K, Jenkins J, Cliatt J, Humbyrd CJ, DeNobile J, Soballe P, Simon R, Wright G, et al. Distinguishing right from left colon by the pattern of gene expression. Cancer Epidemiol Biomarkers Prev. 2003;12:755–62.
25. Derenzini M, Montanaro L, Trere D. Ribosome biogenesis and cancer. Acta Histochem. 2017;119:190–7.
26. Guo H, Zeng W, Feng L, Yu X, Li P, Zhang K, Zhou Z, Cheng S. Integrated transcriptomic analysis of distance-related field cancerization in rectal cancer patients. Oncotarget. 2017;8:61107–17.
27. McHugh JB, Appelman HD, McKenna BJ. The diagnostic value of endoscopic terminal ileum biopsies. Am J Gastroenterol. 2007;102:1084–9.
28. Atkin G, Daley FM, Bourne S, Glynne-Jones R, Northover J, Wilson GD. The effect of surgically induced ischaemia on gene expression in a colorectal cancer xenograft model. Br J Cancer. 2006;94:121–7.

Publisher's note