Jain *et al. Molecular Cancer* (2024) 23:211
https://doi.org/10.1186/s12943-024-02133-5

Molecular Cancer

CORRESPONDENCE Open Access

# Charting the metabolic biogeography of the colorectum in cancer: challenging the right sided versus left sided classification

Abhishek Jain[1], Montana T. Morris[2], Domenica Berardi[1], Trisha Arora[3], Xavier Domingo-Almenara[3], Philip B. Paty[4], Nicholas J. W. Rattray[5], Daniel Kerekes[2], Lingeng Lu[6], Sajid A. Khan[2*] and Caroline H. Johnson[1*]

## Abstract

**Objective** Colorectal cancer (CRC) is conventionally classified as right sided, left sided, and rectal cancer. Clinicopathological, molecular features and risk factors do not change abruptly along the colorectum, and variations exist even within the refined subsites, which may contribute to inconsistencies in the identification of clinically relevant CRC biomarkers. We generated a CRC metabolome map to describe the association between metabolites, diagnostic and survival heterogeneity in cancers of different subsites of the colorectum.

**Design** Utilizing 372 patient-matched tumor and normal mucosa tissues, liquid chromatography-mass spectrometry was applied to examine metabolomic profiles along seven subsites of the colorectum: cecum ($n=63$), ascending colon ($n=44$), transverse colon ($n=32$), descending colon ($n=28$), sigmoid colon ($n=75$), rectosigmoid colon ($n=38$), and rectum ($n=92$).

**Results** 39 and 70 significantly altered metabolites (including bile acids, lysophosphatidylcholines and lysophosphatidylethanolamines) among tumors and normal mucosa, respectively, showed inter-subsite metabolic heterogeneity between CRC subsites. Gradual changes in metabolite abundances with significantly linear trends from cecum to rectum were observed: 23 tumor-specific metabolites, 30 normal mucosa-specific metabolites, and 15 metabolites in both tumor and normal mucosa, had concentration gradients across the colorectum, and is disease status dependent. The metabolites that showed a linear trend included bile acids, amino acids, lysophosphatidylcholines, and lysophosphatidylethanolamines. Comparison of tumors to patient-matched normal mucosa revealed metabolite changes exclusive to each subsite, thereby further highlighting differences in cancer metabolism across the 7 subsites of the colorectum. Furthermore, metabolites associated with survival were different and unique to each subsite. Finally, an interactive and publicly accessible CRC metabolome database was designed to enable access and utilization of this rich data resource (https://colorectal-cancer-metabolome.com/yale-university).

**Conclusions** Gradual changes exist in metabolite abundances from the cecum to the rectum. The association between patient survival and distinct metabolites with anatomic subsite of the colorectum, reveals differences

*Correspondence:
Sajid A. Khan
sajid.khan@yale.edu
Caroline H. Johnson
caroline.johnson@yale.edu

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

between cancers across the colorectum. These inter-subsite metabolic heterogeneities enrich the current understanding and substantiate previous studies that have challenged the conventional classification of right-sided, left-sided, and rectal cancers, by identifying specific metabolites that offer new biological insights into CRC subsite heterogeneity. The database designed in this study will enable researchers to delve into granular information on the CRC metabolome, which until now has not been available.

**Keywords** Colorectal cancer, Metabolomics, Biomarkers, Metabolome database, Subsite heterogeneity

## Introduction

Colorectal cancer (CRC) is a highly heterogeneous disease with different types of risk factors, prognostic indicators, chemotherapeutic susceptibility, and survival outcomes. This heterogeneity leads to different etiologically and clinically distinct subgroups [1]. Evidence suggests that embryological, histological, genetic, molecular, and immunological characteristics of CRC differ by anatomic subsite defined as right-sided CRC (RCC), left-sided CRC (LCC), and rectal cancer [2], which supports the classification of CRC as three distinct tumor entities [2, 3]. Despite these findings, which illuminated the heterogeneous nature of CRC, the current simplistic RCC versus LCC classifications are insufficient to show the greater anatomic regional variations that exist even within subsites categorized by RCC, LCC, and rectal locations [4]. Clinicopathological, molecular features, and risk factors, change gradually rather than abruptly between the refined subsites of CRC from the cecum to rectum [5]. In a previous study, Wang et al. [6] showed that demographic, clinical, anthropometric, lifestyle, and nutritional risk variables differ for cancers along the colorectum in a continuum manner. Similarly, the key molecular hallmarks of CRC, including BRAF mutations, CpG island methylator phenotype (CIMP), and microsatellite instability (MSI), have been found to rise steadily in frequency in nine anatomic subsites from the rectum to the ascending colon [7]. Furthermore, accumulating evidence points to the differences in gut microbiome composition between different anatomic sublocations in the colorectum [8, 9]. These findings call into question the simple division of CRC into primary subsites and emphasize the necessity for further research into refined CRC subsites. If not addressed, subsite-associated variability may contribute to the failure of mechanism-targeted therapies for CRC as well as poor accuracy of potential CRC biomarkers [10].

In a previous study, we unraveled the biological differences between CRC liver metastasis originating from left versus right sided CRCs [11]. This led to a novel hypothesis regarding the mechanism of EGFRi resistance in RCC [11]. Therefore, to further investigate the diagnostic and survival differences among the refined subsites of the colorectum, we conducted the first study aimed at characterizing the metabolome along the length of the colorectum in CRC patients. We identified unique subsite specific metabolite markers (normal versus tumor), inter-subsite metabolic heterogeneity (tumor versus tumor), and unique subsite specific survival markers. Furthermore, the changes in metabolite abundances from the cecum to the rectum were gradual rather than abrupt. Our results highlighted metabolic differences between cancers of different subsites and thus enriched the current understanding of CRC anatomic subsite heterogeneity, aligning with and substantiating the previously proposed continuum, which challenged the conventional classification of CRC into RCC, LCC and rectal cancer. In addition to validating earlier studies on the CRC continuum, our study identifies metabolite markers that provide a new biological perspective on CRC subsite heterogeneity. By completing the gaps left by previous studies, these newly discovered metabolites offer a deeper understanding of the metabolic variations along the colorectum. To facilitate interrogation of the extensive CRC metabolome data generated in this study, we designed an interactive, publicly accessible online platform (https://colorectal-cancer-metabolome.com/yale-university) that will allow user-driven exploration and hypothesis generation. An overview of the study is presented in Fig. 1A.

## Methods

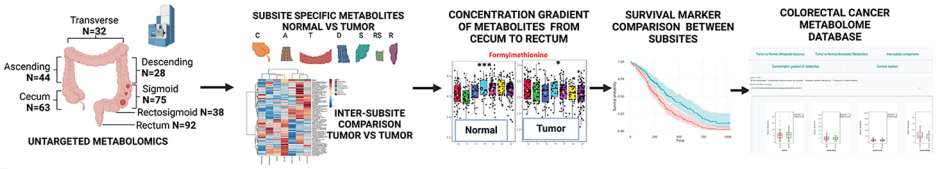
A detailed overview of Materials and Methods can be found in the Supplementary information 1 (Supplementary Materials and Methods).

## Results

### Patient characteristics

In this study, our analysis consists of tissues collected from 372 CRC patients. Primary tumor tissues with an equal number of patient-matched normal mucosa tissues were obtained from seven different regions of the colorectum: 63 cecum, 44 ascending, 32 transverse, 28 descending, 75 sigmoid, 38 rectosigmoid, and 92 rectum. We implemented a rigorous procedure to ensure that the normal mucosa tissues were indeed free of malignant cells and belonged to the same anatomical subsite as the matched tumor tissue (Supplementary information 1). We did not observe any statistically significant differences in age ($p=0.997$), sex ($p=0.073$), and chemotherapy status ($p=0.237$) between cases from different anatomic subsites (Table S4). We also compared cancer stages between all subsites and found the overall ANOVA

**A** WORKFLOW




**B** UNIQUE SUBSITE SPECIFIC METABOLITES (NORMAL VS TUMOR)

| | Cecum | Ascending | Transverse | Descending | Sigmoid | Rectosigmoid | Rectum |
|---|---|---|---|---|---|---|---|
| **Alanyl-Leucine** | ***3.50E-03(.68)*** | 7.90E-01 | 2.90E-01 | 4.00E-01 | 6.90E-01 | 6.90E-01 | 2.10E-01 |
| **Alanine** | ***3.60E-02(.44)*** | 9.70E-02 | 8.80E-01 | 7.30E-01 | 2.30E-01 | 4.30E-01 | 1.10E-01 |
| **Valyl-Methionine** | ***3.00E-02(.70)*** | 4.50E-01 | 2.00E-01 | 2.10E-01 | 9.10E-02 | 7.70E-01 | 1.90E-01 |
| **Maltotriose** | 2.50E-01 | ***3.50E-02(.47)*** | 8.00E-02 | 4.30E-01 | 3.30E-01 | 7.30E-01 | 6.00E-01 |
| **Succinate** | 1.20E-01 | ***3.90E-02(.45)*** | 8.40E-01 | 2.20E-01 | 6.60E-02 | 6.80E-01 | 1.00E-01 |
| **Valerylcarnitine** | 2.20E-01 | ***1.00E-02(.57)*** | 7.40E-01 | 9.00E-01 | 2.00E-01 | 1.56E-01 | 7.70E-02 |
| **LysoPE (P18:0)** | 9.10E-01 | ***2.00E-02(-.63)*** | 3.50E-01 | 1.43E-01 | 9.70E-02 | 4.50E-01 | 1.45E-01 |
| **Decanoylcarnitine** | 6.90E-01 | 4.30E-01 | ***3.80E-02(.83)*** | 1.30E-01 | 3.40E-01 | 2.70E-01 | 9.20E-01 |
| **Hexanoylcarnitine** | 2.60E-01 | 8.40E-01 | 4.00E-01 | ***2.00E-02(1.08)*** | 2.70E-01 | 9.50E-01 | 9.00E-02 |
| **Octanoylcarnitine** | 2.80E-01 | 6.00E-01 | 8.80E-01 | ***4.70E-02(.99)*** | 7.90E-02 | 2.30E-01 | 1.15E-01 |
| **Epsilon-(gamma-Glutamyl)lysine** | 3.00E-01 | 8.90E-01 | 7.10E-02 | 2.50E-02(-.86) | 1.60E-01 | 8.40E-01 | 1.40E-01 |
| **Estradiol-17B-glucuronide** | 1.80E-01 | 1.90E-01 | 4.80E-01 | 2.00E-01 | ***4.20E-02(.47)*** | 2.20E-01 | 9.20E-01 |
| **Glycocholic acid** | 7.10E-01 | 3.60E-01 | 6.10E-01 | 4.10E-01 | ***6.70E-03(-.50)*** | 6.50E-01 | 2.30E-01 |
| **Glycodeoxycholic acid** | 4.70E-01 | 6.60E-01 | 7.00E-01 | 3.20E-01 | ***2.80E-02(-.32)*** | 6.30E-01 | 6.30E-01 |
| **Vitamin A** | 4.40E-01 | 2.00E-01 | 1.50E-01 | 4.00E-01 | ***2.20E-02(-.39)*** | 7.20E-01 | 4.50E-01 |
| **Arachidonic acid** | 2.30E-01 | 2.50E-01 | 8.90E-02 | 4.70E-01 | ***2.80E-02(-.6)*** | 2.60E-01 | 2.20E-01 |
| **LysoPE(18:0)** | 5.90E-01 | 5.00E-01 | 1.00E+00 | 6.20E-01 | 2.10E-01 | ***2.70E-02(.42)*** | 1.50E-01 |
| **Cysteinylglycine** | 9.40E-02 | 1.00E-01 | 2.40E-01 | 1.90E-01 | 8.50E-02 | 2.40E-01 | ***9.00E-04(.6)*** |
| **Diethyl succinate** | 5.40E-02 | 1.60E-01 | 4.90E-01 | 7.30E-01 | 1.15E-01 | 6.50E-01 | ***1.60E-02(-.39)*** |
| **Cytidine** | 6.60E-01 | 4.20E-01 | 5.20E-01 | 6.20E-01 | 2.70E-01 | 1.12E-01 | ***1.60E-02(.39)*** |
| **Cytidine-5'-monophosphate** | 9.10E-01 | 6.40E-01 | 8.90E-01 | 8.20E-01 | 5.60E-01 | 6.00E-01 | ***2.30E-02(.29)*** |
| **Ethyl-4-aminobenzoate** | 5.40E-02 | 2.70E-01 | 1.20E-01 | 9.70E-01 | 1.00E-01 | 4.80E-01 | ***2.50E-02(-.27)*** |
| **Glucose-6-phosphate** | 7.59E-01 | 8.15E-01 | 1.58E-01 | 4.62E-01 | 1.09E-01 | 1.51E-01 | ***4.19E-03(.34)*** |
| **Glucosamine-6-phosphate** | 2.28E-01 | 9.60E-01 | 2.15E-01 | 1.05E-01 | 3.77E-01 | 3.31E-01 | ***3.70E-02(.22)*** |
| **Mevalonate** | 2.91E-01 | 4.64E-01 | 9.42E-01 | 6.15E-01 | 5.66E-01 | 2.39E-01 | ***4.35E-02(.21)*** |
| **Phenylacetylglutamine** | 2.19E-01 | 7.73E-01 | 5.16E-01 | 4.40E-01 | 1.04E-01 | 4.68E-01 | ***4.35E-03(-.59)*** |
| **Alanyl-glutamine** | 5.17E-01 | 4.64E-01 | 8.15E-01 | 5.04E-01 | 1.14E-01 | 1.26E-01 | ***8.99E-04(-.66)*** |
| **Palmitoleic acid** | 6.17E-01 | 7.44E-02 | 7.08E-01 | 4.18E-01 | 1.35E-01 | 7.17E-01 | ***1.09E-04(-.53)*** |
| **Valyl-Serine** | 2.02E-01 | 1.98E-01 | 3.20E-01 | 9.85E-01 | 5.20E-02 | 7.34E-01 | ***1.88E-04(-.49)*** |

**C** UNIQUE METABOLITES IN RIGHT VERSUS LEFT CRC SUBSITES (NORMAL VS TUMOR)

| | Cecum | Ascending | Transverse | Descending | Sigmoid | Rectosigmoid | Rectum |
|---|---|---|---|---|---|---|---|
| **Oleic acid** | ***1.80E-02(-.45)*** | ***4.30E-02(-.5)*** | 9.90E-01 | 4.50E-01 | 2.80E-01 | 8.40E-01 | 1.50E-01 |
| **Acetylmethionine** | ***9.00E-03(.53)*** | ***2.40E-02(.55)*** | 7.50E-01 | 4.70E-01 | 5.30E-02 | 8.30E-01 | 5.10E-01 |
| **Formylmethionine** | ***2.20E-02(.54)*** | ***8.00E-04(.81)*** | 4.50E-01 | 2.40E-01 | 1.12E-01 | 9.40E-01 | 6.50E-01 |
| **Isovalerylglycine** | ***1.90E-03(.70)*** | ***5.40E-03 (.63)*** | 8.20E-01 | 8.90E-01 | 5.40E-01 | 6.90E-01 | 9.60E-01 |
| | | | | | | | |
| **NAD** | 6.20E-02 | 3.40E-01 | 1.40E-01 | ***1.70E-02(.55)*** | **3.20E-03(.32)** | **1.30E-02(.44)** | **1.40E-03(.36)** |
| **Tryptophanamide** | 8.50E-01 | 6.50E-02 | 1.80E-01 | ***4.20E-02(-.99)*** | **9.00E-05(-.92)** | **1.30E-03(-1.11)** | **4.00E-05(-1.02)** |
| **LysoPE(22:5)** | 4.40E-01 | 7.70E-01 | 7.50E-02 | 9.20E-02 | **1.00E-04(.63)** | **9.00E-05(1.08)** | **1.10E-04(.85)** |
| **Pantothenate** | 6.70E-02 | 1.00E-01 | 1.12E-01 | 2.80E-01 | **1.10E-03(.38)** | **1.60E-02(.43)** | **2.00E-04(.41)** |
| **Sn-glycero-3-phosphocholine** | 5.30E-01 | 2.40E-01 | 2.20E-01 | 6.80E-01 | **1.00E-02(.67)** | **2.50E-02(.78)** | **7.30E-03(.54)** |
| **LysoPC (22:6)** | 3.90E-01 | 3.10E-01 | 3.90E-01 | 1.00E-01 | **2.60E-02(.58)** | **1.80E-03(1.02)** | **3.00E-05(.82)** |
| **LysoPC (20:4)** | 3.10E-01 | 2.40E-01 | 8.90E-01 | 5.60E-02 | **2.90E-02(.44)** | **2.20E-03(.72)** | **4.00E-06(.75)** |
| **Thymidine** | 2.30E-01 | 3.00E-01 | 3.50E-01 | 7.80E-01 | **1.00E-05(-1.46)** | **1.29E-02(-1.23)** | **2.00E-04(-1.25)** |
| **Bilirubin** | 2.60E-01 | 9.10E-01 | 5.50E-01 | 5.00E-01 | **3.00E-05(-.87)** | **2.50E-02(-.45)** | **2.00E-03(-.66)** |

**Fig. 1** **A** Workflow to investigate diagnostic and survival heterogeneity among cancers of different subsites of the colorectum. C-Cecum, A-Ascending, T-Transverse, D-Descending, S-Sigmoid, RS-Rectosigmoid, R-Rectum. **B** Unique metabolites altered in tumor tissue compared to normal mucosa tissues in each specific subsite of the colorectum. C Metabolites uniquely altered in the left (descending, sigmoid, rectosigmoid and rectum) and right (cecum, ascending) subsites of the colorectum. FDR-corrected p-values with $\log_2$ fold changes in brackets are shown for statistically significant metabolites. The significant changes (FDR-corrected $p < 0.05$) are highlighted in blue. Only FDR-corrected p-values and not fold changes are shown for non-significant metabolites

p-value to be significant ($p=0.006$). However, after performing Bonferroni post hoc analysis we observed only one significant difference between sigmoid and rectum subsites ($p=0.004$) concerning cancer stage (Table S4 and S5). Notably, there was no difference between the remaining subsite comparisons in terms of stages.

### Metabolites commonly altered for cancers across colorectum subsites but their association with each subsite is different

Untargeted metabolome profiles were compared between all patient-matched tumors and normal mucosa from the seven subsites, which revealed 26 identified metabolites that were commonly altered in all the subsites (Table S6). We observed significantly reduced levels of hydroxyindoleacetic acid (HIAA), uric acid, linolenic acid, and dopamine sulfate, in tumors across all subsites.

The association between these metabolites and CRC differed depending on the subsite location. A negative association with tumors strengthened from cecum to rectum for HIAA (OR ranges from 0.42 to 0.11), uric acid (OR ranges from 0.48 to 0.17), and linolenic acid (OR ranges from 0.58 to 0.37). The highest risk of CRC with a reduced level of HIAA was observed in the cecum. The risk was generally higher in the transverse and ascending subsites (which are part of the proximal colon). The risk was reduced in descending, rectosigmoid, and rectum tissues, but not in sigmoid subsites. The magnitude of risk was lowest in the rectosigmoid, and highest in the cecum, with a 3.6-fold difference in the OR. A reduced level of uric acid was associated with a higher risk in the cecum, ascending colon, and descending colon, as compared to other subsites; the risk was reduced in the sigmoid, rectosigmoid, and rectum. The magnitude of risk was lowest in the rectum and highest in the cecum, with a 2.88-fold difference in the OR. The reduced level of linoleic acid in tumors was associated with a higher risk in the regions from the cecum to descending and then the risk was reduced from the sigmoid to the rectum. Dopamine sulfate showed heterogeneity in the association with CRC (OR ranges from 0.54 to 0.34) in different regions of the colorectum but there was no pattern observed.

A significantly increased level of 15 metabolites was observed in tumors from each of the seven subsites when compared to their matched normal mucosa. The majority of these metabolites had an OR that differed between subsites, suggesting the heterogeneous association of these metabolites among the seven subsites of colorectum. The positive association of metabolites with CRC strengthened from cecum to rectum for kynurenine (OR ranges from 2.06 to 6.60) and reduced nicotinamide adenine dinucleotide (NADH) (OR ranges from 2.06 to 3.72). The risk was higher from the sigmoid to rectum subsites as compared to other subsites. The magnitude of risk was lowest in the cecum and highest in the rectosigmoid, with a 3.2-fold difference in the OR in the case of kynurenine. The magnitude of risk was very similar within proximal subsites and distal subsites, but the ORs were generally substantially higher in distal subsites for NADH. The magnitude of risk for hydroxyproline was lowest in the rectum and highest in the sigmoid, with a 2.6-fold difference in the OR. The ORs were very similar within proximal subsites for hydroxyproline. N-acetyl-L-cysteine was highly heterogeneous with no specific pattern across the subsites. The magnitude of risk was lowest in descending and highest in ascending subsites, with a 3-fold difference in the OR. We also performed a linear regression between CRC subsites and the OR for each metabolite. It was observed that the strength of association between CRC risk and metabolites was significantly linear ($P_{linear}<0.05$) and gradually changed from cecum to rectum for NADH, kynurenine, and HIAA.

The ORs for uridine 5'-diphosphate (UDP) and guanosine diphosphate (GDP)-glucose did not differ across the subsites which shows that these metabolites have low heterogeneity for OR along the colorectum. For all the remaining metabolites, the risk association with CRC was generally heterogeneous with no specific patterns from cecum to rectum.

#### Distinct metabolite changes in tumors compared to matched normal mucosa in each anatomic subsite

Subsite-specific metabolite markers associated with CRC were discerned through a comparative analysis of tumors from each of the colorectal subsites with their respective matched normal mucosa. The unique metabolite changes in each subsite are presented in Fig. 1B and Figures S3-S4. Notable metabolite changes were identified for each CRC subsite, and metabolite identification was confirmed using analytical standards. Specifically, in the cecum, two dipeptides; alanyl-leucine, valyl-methionine, and an amino acid alanine, exhibited significant upregulation. In ascending tumors, a marked increase in valerylcarnitine and a decreased level of LysoPE(P-18:0) were observed. Transverse tumors displayed a specific elevation in decanoylcarnitine. Notably, in the descending colon, two carnitine metabolites, hexanoylcarnitine and octanoylcarnitine were elevated. Epsilon-(gamma-glutamyl)lysine exhibited a decrease in tumor compared to normal tissue in the descending colon. Interestingly, in sigmoid tumors, two bile acids, glycocholic acid and glycodeoxycholic acid were downregulated. Rectosigmoid tumors showed a significant upregulation of LysoPE(18:0). In rectal tumors, increased levels of six metabolites; cysteinylglycine, cytidine, cytidine-5'-monophosphate, glucose-6-phosphate, glucosamine-6-phosphate, and mevalonate were noted, alongside decreased levels of six metabolites; ethyl-4-aminobenzoate, phenylacetylglutamine, alanyl glutamine, palmitoleic acid, valyl-serine, and diethyl succinate. Furthermore, DDA and $MS^E$ data were processed to predict structures of additional unique metabolites (designated level 3 identifications) in each subsite. Our analysis enabled the identification of additional distinct metabolites in each subsite; 8 in cecum, 3 in ascending, 2 in descending, 9 in sigmoid, 2 in rectosigmoid and 19 in rectum (Figure S4). This enhanced metabolome coverage provides a more comprehensive insight into the unique tumor biology of each subsite. Overall, findings shown in this section underscore subsite-specific metabolic alterations in CRC, revealing potential markers that contribute to a nuanced understanding of the molecular landscape of colorectal tumors.

### Comparative metabolite changes in tumors compared to normal mucosa and specificity to right versus left CRC subsites

Next, we identified the metabolites that were uniquely altered in the tumors of proximal subsites (cecum and ascending) as compared to their respective normal mucosa, which were not similarly altered in distal subsites (i.e. descending, sigmoid, rectosigmoid and rectum), and vice versa. The unique metabolite changes in the proximal subsites or distal subsites are presented in Fig. 1C, S3, and S5. Formylmethionine, N-acetylmethionine and isovalerylglycine were significantly upregulated and oleic acid was significantly downregulated in the cecum and ascending tumors when compared to their matched normal mucosa, but there was no change in any of the subsites in the distal regions. Next, we identified the metabolites specifically altered in the tumors of distal subsites when compared to their matched normal mucosa. Nicotinamide adenine dinucleotide (NAD) and tryptophanamide were significantly upregulated in the tumors of all the distal subsites including descending, sigmoid, rectosigmoid, and rectum. Five metabolites, LysoPC (20:4), LysoPC (22:6), LysoPE (22:5), sn-glycero-3-phosphocholine, and pantothenic acid, were specifically upregulated and two metabolites (thymidine and bilirubin) were specifically downregulated in the tumors of the sigmoid, rectosigmoid, and rectum. Furthermore, the level 3 identification of additional metabolites revealed some additional distinct metabolites in proximal or distal subsites; three in cecum and ascending, one in descending, sigmoid, rectosigmoid and rectum, three in sigmoid, rectosigmoid and rectum (Figure S5).

### Characterization of the metabolic landscape across the different subsites of colorectum

To systematically chart metabolome composition along the colorectum, we compared metabolite abundances of 223 annotated metabolites in normal mucosa and tumor tissue separately, between seven different subsites of the colorectum. Spatially resolved analysis of the colorectum revealed the heterogeneous metabolic landscape of normal mucosa and tumor tissue of refined subsites. Kruskal-Wallis test was applied to detect the metabolic differences between seven subsites, and the metabolites that were significantly different in at least one pairwise comparison was used to perform hierarchical clustering analysis (Tables S7-S8). The results revealed that metabolic heterogeneity exists among tumors and normal mucosa of the seven CRC subsites. There was a statistically significant difference in the mean abundances of 70 metabolites in at least one pairwise comparison between the normal mucosa of all the colorectum subsites (Figure S6 and Table S7). The major differentiating metabolites included 13 lipid metabolites (seven LysoPCs and six LysoPEs). Other metabolites that varied significantly among subsites were two bile acids (glycocholic acid and glycoursodeoxycholic acid), N-acetylneuraminic acid, kynurenine, succinic acid, lactic acid, and S-adenosylhomocysteine (SAH). Similarly, we identified statistically significant differences in the mean abundances of 39 metabolites in at least one pairwise comparison between the tumor tissues of all the colorectum subsites (Fig. 2A and Table S8). Normal and tumor colorectum tissue had 21 common metabolites (including four LysoPCs, four LysoPEs, and two bile acids (glycocholic acid, glycoursodeoxycholic acid)) that were significantly different between subsites; yet the distribution of these metabolites varied depending on the tissue types (tumor and normal) (Fig. 2A, S6 and Tables S7-S8). For example, LysoPC(18:1) was significantly different between cecum and descending in both the tumor and normal tissues but only in normal tissues between ascending and rectosigmoid.

Next, the subsites were grouped into three major regions for further analysis: (1) right sided CRC (RCC); cecum, ascending, transverse, and hepatic flexure, (2) left-sided CRC (LCC); splenic flexure, descending, sigmoid, rectosigmoid, and (3) rectum. The metabolites were compared between these three major regions in the normal mucosa and tumor tissues independently. In the tumor tissues, we identified 62 metabolites that were significantly different (FDR < 0.05) between LCC versus RCC, 13 metabolites significantly different between LCC versus rectum, and 54 metabolites significantly different between RCC versus rectum (Fig. 2B). Similarly, in the normal mucosa, 83 metabolites were significantly different (FDR < 0.05) between LCC versus RCC, 33 metabolites were significantly different between LCC versus rectum, and 88 metabolites were significantly different between RCC versus rectum (Figure S7). The metabolites that were significantly different in at least one pairwise comparison was used to perform hierarchical clustering analysis, and the distribution of these metabolites across three regions (RCC, LCC and rectum) is shown in Fig. 2B and S7. We discovered several intriguing metabolic differences between RCC and LCC tumors. Specifically, 11 phospholipid metabolites (five LysoPCs and six LysoPEs) and three bile acids (glycocholic acid, glycodeoxycholic acid, and glycoursodeoxycholic acid) were significantly upregulated in RCC tumors compared to LCC tumors. Comparing tumors from RCC and LCC locations, a purine metabolism pathway was markedly activated, with seven metabolites downregulated (NAD, NADH, NADP+, FAD, ADP, cyclic AMP, and inosine) and one metabolite elevated (uric acid) in RCC tumors. Additionally, we observed downregulation of six metabolites (glucose-6-phosphate, glucosamine-6-phosphate, GDP-glucose, fucose-1-phosphate, UDP, and N-acetylneuraminic acid) involved in the amino acid sugar and

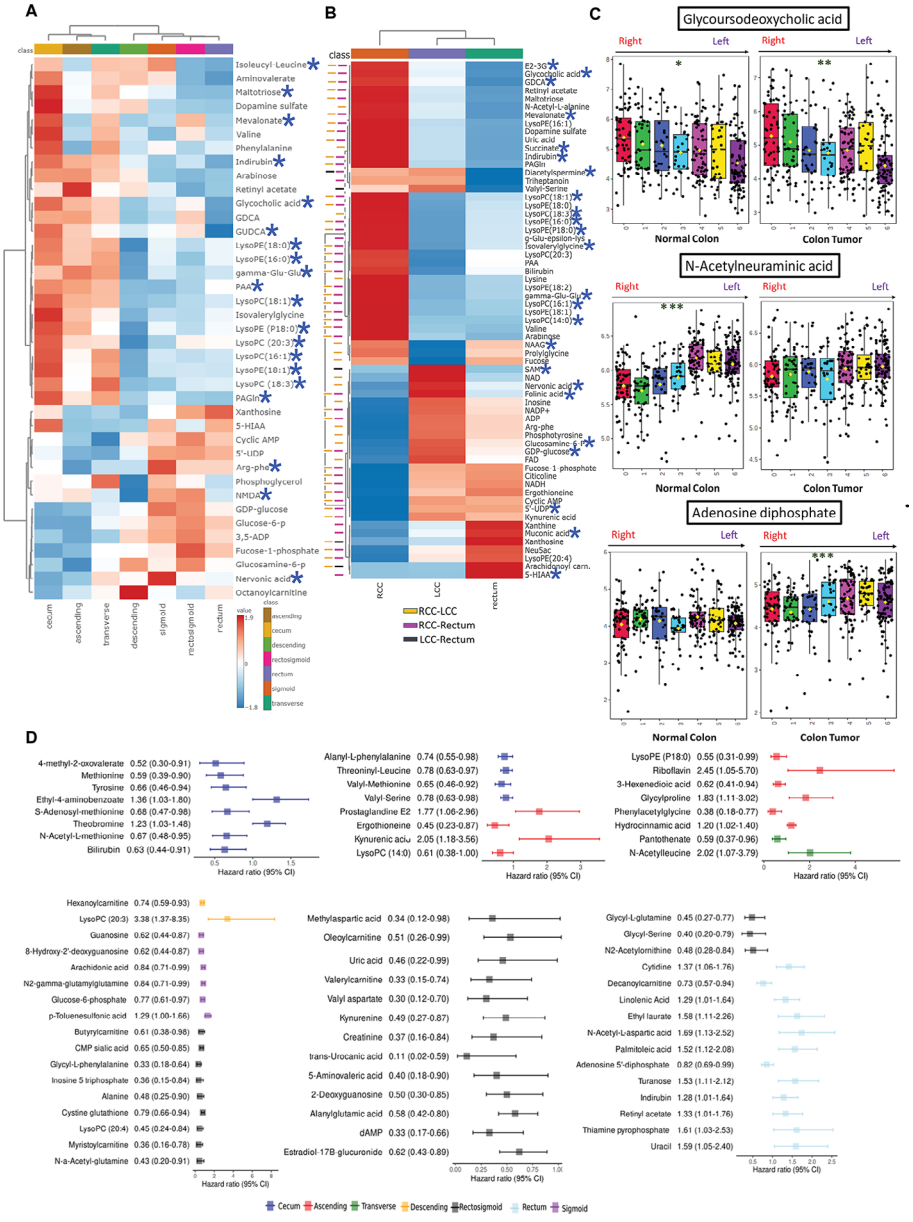



**Fig. 2** **A** Hierarchical clustering analysis of metabolite abundances from colorectum tissue samples. Abundances for metabolites significantly differentiating in at least one pairwise comparison between colorectum subsites are shown across seven colorectum subsites in tumors (39 metabolites). Abundances were averaged from each subsite. Metabolites highlighted with blue asterisk (19 metabolites) were commonly altered in both normal and tumor tissues. **B** Hierarchical clustering analysis of metabolite abundances from colorectum tissue samples. Abundances for metabolites significantly differentiating in at least one pairwise comparison between three major regions are shown across the left side colon, right side colon, and rectum in tumors. Abundances were averaged from each subsite. Metabolites highlighted with blue asterisk (25 metabolites) were commonly altered in both normal and tumor tissues. **C** Distribution of metabolite abundances showing concentration gradient exist from cecum to rectum in both tumors and normal mucosa. Linear regression analysis was performed between metabolite abundances and seven subsites. Three representative metabolites showing significantly linear trend ($P_{linear}$<0.05) from cecum to rectum. The concentration gradient is similar in both normal colon and colon tumor (glycoursodeoxycholic acid), concentration gradient is different between normal colon and colon tumor (N-acetylneuraminic acid and adenosine diphosphate). Each subsite is presented in a unique color. Red-Cecum, green-Ascending, blue-Transverse, cyan- Descending, light violet-Sigmoid, yellow-Rectosigmoid and dark violet-Rectum. **D** Subsite specific unique metabolites markers significantly ($p < 0.05$) associated with CRC prognosis. Hazard ratio (HR) for each metabolite ($log_2$-transformed abundance) was calculated by Cox proportional hazard regression analysis between $log_2$ abundances of individual metabolites and 5-year overall survival in each subsite, adjusting for age, sex, chemotherapy and stage. Error bars represent 95% confidence intervals (CIs). The unique survival metabolites for each subsite are color coded differently. Cecum-blue, Ascending-pink, Transverse-green, Descending-yellow, Sigmoid-purple, Rectosigmoid-grey, Rectum-ice blue. GDCA - Glycodeoxycholic acid, GUDCA -Glycoursodeoxycholic acid, 5-HIAA - 5-Hydroxyindole-3-acetic acid, N-acetylneuraminic acid, PAA - Phenylacetylglycine, PAGln - Phenylacetylglutamine, NMDA - N-Methyl-D-aspartic acid, E2-3G - Estradiol-3-glucuronide, gamma-Glu-Glu - Gamma-glutamylglutamate, g-Glu-epsilon-lys - Epsilon-(gamma-Glutamyl)lysine, SAM - S-Adenosyl-methionine, NAAG - N-Acetylaspartylglutamic acid

nucleotide sugar metabolism pathway, in RCC tumors compared to LCC tumors.

### Gradual changes in metabolite abundances from the cecum to the rectum challenge the simple classification of CRC into right-sided and left-sided cancer

We analyzed the distribution of the annotated metabolites (those identified at level 1 and level 2) across all seven subsites of the colorectum in both normal and tumor tissues. The analysis showed that many of the metabolites did not change abruptly and that these markers gradually increased or decreased in abundance from the cecum to the rectum along the colorectum, challenging the conventional classification of CRC into right-sided and left-sided cancers. By linearity test, 50 metabolites were significantly linear from the cecum to rectum in normal mucosa tissues with 23 metabolites showing an increasing gradient and 27 metabolites showing a decreasing gradient. We found that 25 metabolites were significantly linear from the cecum to the rectum in tumor tissues with 10 metabolites showing an increasing gradient and 15 metabolites showing a decreasing gradient. There were 14 metabolites that showed a significant linear trend in both normal and tumor tissue (Figure S8 and Table S9). The distribution of three representative metabolites showing significantly linear trends ($P_{linear}<0.05$) from cecum to rectum in each case are shown in Fig. 2C. For instance, glycoursodeoxycholic acid displayed a decreasing concentration gradient from cecum to rectum, with similar trends followed in both colon tumor and normal tissues (Fig. 2C). Glycoursodeoxycholic acid had $P_{linear}<0.05$ for both normal mucosa and tumor tissue. Figure 2C shows that adenosine diphosphate (ADP) exhibited a gradual increase ($P_{linear}< 0.05$) from cecum to rectum in tumor tissue but remains relatively stable across subsites in normal tissues. N-acetylneuraminic acid showed exactly the opposite trend, gradually increasing ($P_{linear}<0.05$) from cecum to rectum in normal tissue but remains relatively similar across subsites in tumor tissues (Fig. 2C). These results also validated our previous study which showed significantly higher levels of glycocholic acid and glycoursodeoxycholic acid in RCC tumor as compared to LCC tumors [12]. Overall, our findings indicate that metabolite concentration gradient exists across the colorectum with these gradients contingent upon the disease status.

### Metabolites associated with survival were different and unique to each subsite

Cox proportional hazard regression analysis was performed to identify the association between $\log_2$ abundances of individual metabolites and 5-year overall survival in each subsite, adjusting for age, sex, chemotherapy, and stage. A two-sided $p<0.05$ was considered statistically significant. Notably, we observed inter-subsite differences in survival markers and distinct metabolites were associated with survival in each subsite (Fig. 2D). We identified unique metabolite markers significantly ($p<0.05$) associated with 5-year overall survival, with 13 metabolites in the cecum, 10 metabolites in the ascending, 2 metabolites in the transverse, 2 metabolites in the descending, 6 metabolites in the sigmoid, 28 metabolites in the rectosigmoid and 14 metabolites in the rectum that showed exclusive associations (Fig. 2D). Next, we grouped the subsites and identified the association between metabolites and 5-year overall survival in RCC, LCC, and the rectum regions. We identified unique metabolite markers significantly ($p<0.05$) associated with 5-year overall survival in each of these regions, with 6 metabolites in RCC, 15 metabolites in LCC, and 19 metabolites in the rectum region (Figure S9).

### Development of an interactive colorectal cancer metabolome platform

To facilitate further exploration of the extensive data generated in this study, we designed an interactive, publicly accessible online platform to enable researchers and healthcare professionals to probe the CRC metabolic landscape across seven subsites (https://colorectal-cancer-metabolome.com/yale-university). This user-friendly interface permits the investigation of metabolite distributions, diagnostic and survival heterogeneity, and unique metabolite markers of each subsite, with highlights for data visualization and statistical analysis. Furthermore, the interactive nature of the database allows for user-driven exploration and hypothesis generation, empowering deeper insights into CRC heterogeneous biology and illuminating future research directions in the underlying biological mechanisms driving colorectal cancer progression and response to therapy.

## Discussion

Our comprehensive CRC metabolome map provides a novel and profound insight into CRC heterogeneity. Prior studies have reported that clinicopathological, molecular features, and risk factors do not change abruptly along the colorectum. Our study examining subsite specific metabolic changes offers a new dimension into CRC biology and may enable the development of more precise biomarkers. Moreover, our results emphasize the metabolic interactions that may be responsible for the subtle but noteworthy biological differences among subsites, which may influence clinical outcomes. This study paves the way for more precise and tailored therapeutic approaches that are designed to address the distinct metabolic environments of individual tumor microenvironments. Intertumoral heterogeneity, subsite-specific alterations in tumors relative to matched normal mucosa,

and unique subsite-specific survival markers collectively challenge the conventional classification of CRC into RCC, LCC, and rectal cancers, supporting and complementing previous proposals of the CRC continuum [5, 6, 13, 14]. Our study shows that right-sided vs. left-sided CRC classifications are too broad for understanding CRC heterogeneous biology and accurate identification of clinically relevant biomarkers.

Gradual changes in metabolites along the length of the colorectum reinforce its continuum nature and undermine the notion of abrupt variations between anatomic boundaries such as the splenic flexure, sigmoid, and rectosigmoid. Metabolic heterogeneity was further validated with 39 and 70 significantly different metabolites between CRC subsites in tumor and normal mucosa tissues, respectively. Metabolic heterogeneity across the colorectum could be related to subsite-associated differences in tumor development and the survival of patients. For instance, LysoPCs and LysoPEs showed significant heterogeneity between colorectum subsites and were significantly upregulated only in tumors from sigmoid, rectosigmoid and rectum subsites when compared to their matched normal mucosa, but they remain non-significant in other subsites. Similarly, LysoPC (20:3) was significantly associated with poor survival in descending tumors only. The identification of metabolites significantly altered in the specific subsites of the colorectum can help gain insights into how metabolic pathways differ and how their regulation changes along the length of the colorectum.

Knowledge of the differences in metabolite levels in normal tissues across the colorectum is also crucial, as the basal metabolism of the normal tissue of origin influences the metabolism of cancer cells. The biology of the normal colon tissue, which also contributes to the tumor microenvironment, is temporally and geographically heterogeneous across the colorectum due to differences in blood flow, nutrition availability, oxygen levels, and microbial and cellular composition [15]. Certain metabolites abundant in normal tissue may affect cancer development, e.g. short chain fatty acids produced by the microbiome are anti-inflammatory and anti-proliferative [16]. The gut microbial changes along the length of the colorectum from the cecum to the rectum have been shown previously [8, 9] which can influence the creation and utilization of different metabolites and can also explain the linear trends of metabolites in our study. Knowing these mechanisms can aid targeted therapies that capitalize on metabolic heterogeneity between CRC subsites, potentially improving treatment efficacy while safeguarding normal tissue function.

### Inter-subsite heterogeneity in survival metabolite markers across colorectum

Finally, we explored the clinical relevance of the discovered metabolites by examining their association with survival of the CRC patients. Previous studies have reported that the survival rate varies between CRC subsites, but the reason for such differences remained largely unknown [14]. Inter-subsite differences in survival markers in our study not only validate the previous studies but also add substantial value by revealing distinct tumor biology that could govern these differences in each subsite. Our study is the first to identify metabolite markers associated with survival in each specific subsite which, in turn, provides a robust foundation for future studies to gain a deeper understanding of biological mechanisms governing subsite specific differences in CRC.

## Conclusion

This is the first comprehensive CRC metabolome map describing diagnostic and survival heterogeneity in cancers of different subsites of the colorectum. To facilitate further interrogation of the extensive data generated in this study, we designed an interactive, publicly accessible online platform (https://colorectal-cancer-metabolome.com/yale-university) that will allow user-driven exploration and hypothesis generation, empowering deeper insights into CRC heterogeneous biology. The right-sided vs. left-sided CRC classifications are too broad for understanding CRC biology and accurate identification of clinically relevant biomarkers. Metabolite profiles vary for tumors in different regions of the colorectum. Our study underscores the need to investigate subsite specific differences in CRC more in detail. This can ultimately be critical for designing better treatments tailored to each patient according to their tumor type.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12943-024-02133-5.

Supplementary Material 1

Supplementary Material 2

**Acknowledgements**

We would like to thank Mr. Prateek Jain and Mr. Akshat Jain from Divrit Consultancy Pvt Ltd, India for their assistance in coding and maintining the CRC Metabolome Database.

**Author contributions**

Abhishek Jain: Conceptualization, Methodology, Investigation, Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Data Curation, MS/MS Fragmentation, Metabolomics Data Analysis and Statistical Analysis , Funding acquisition, Construction of Large Scale Metabolomics Pipeline and Database, Designing of the colorectal-cancer-metabolome Database. Montana T. Morris: Data Curation, Visualization, Writing – review & editing Domenica Berardi: MS/MS fragmentation, Visualization, Writing – review & editing. Trisha Arora: Statistical analysis, Visualization, Writing – review & editing. Xavier Domingo Almenara:

Statistical analysis, Visualization, Writing – review & editing Philip B. Paty: Writing – review & editing, Resources, Project administration, Investigation, Data curation, Conceptualization. Nicholas J W Rattray: Data curation, Visualization, Writing – review & editing. Daniel Kerekes: Statistical analysis, Visualization, Writing – review & editing. Lingeng Lu: Statistical analysis, Visualization, Writing – review & editing. Caroline H. Johnson: Writing – review & editing, Supervision, Resources, Project administration, Conceptualization, Methodology, Investigation, Funding acquisition, Data curation. Sajid A. Khan: Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

**Funding**
This work was supported by the National Institutes of Health [1RM1GM141649-01] and American Cancer Society [RSG-20-065-01-TBE] both awarded to CJ. Additionally, this publication was made possible by CTSA Grant Number UL1 TR001863 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH), awarded to AJ. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. XD was supported by "La Caixa" Foundation (ID 100010434) via the Junior Leader Fellowship LCF/BQ/PR21/11840001. XD and TA were supported by the European Commission's Horizon 2020 via the Innovative Training Network Marie Skłodowska-Curie COL_RES project under the grant agreement No 956279. NJW was supported with funding from the UK Royal Society - IF\R1\231034 and Medical Research Council - MR/Y010736/1 grants.

**Data availability**
We chose to make our data available in the format of this database, other data requests, along with protocols and codes can be made by email.

## Declarations

**Ethics approval and consent to participate**
The Yale University Institutional Review Board (IRB) determined that the study conducted in this publication was not considered to be Human Subjects Research and did not require an IRB review (IRB/HSC# 1612018746).

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Department of Environmental Health Sciences, Yale School of Public Health, 60 College Street, New Haven, CT 06510, USA
[2]Department of Surgery/Surgical Oncology, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA
[3]Omics Sciences Unit, EURECAT - Technology Centre of Catalonia, Avda. Universitat 1, Reus 43204, Catalonia, Spain
[4]Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10065, USA
[5]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK
[6]Department of Chronic Disease Epidemiology, Yale School of Public Health, 60 College Street, New Haven CT06510, USA

Received: 6 September 2024 / Accepted: 20 September 2024
Published online: 28 September 2024

**References**

1. Chan DKH, Buczacki SJA. Tumour heterogeneity and evolutionary dynamics in colorectal cancer. Oncogenesis. 2021;10(7):53.
2. 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(4):264–73.
3. Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. Ann Intern Med. 1990;113(10):779–88.
4. Huyghe JR, Harrison TA, Bien SA, Hampel H, Figueiredo JC, Schmit SL, Conti DV, Chen S, Qu C, Lin Y, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. Gut. 2021;70(7):1325–34.
5. Loree JM, Pereira AAL, Lam M, Willauer AN, Raghav K, Dasari A, Morris VK, Advani S, Menter DG, Eng C, et al. Classifying Colorectal Cancer by Tumor Location rather than sidedness highlights a Continuum in Mutation profiles and Consensus Molecular subtypes. Clin Cancer Res. 2018;24(5):1062–72.
6. Wang L, Lo CH, He X, Hang D, Wang M, Wu K, Chan AT, Ogino S, Giovannucci EL, Song M. Risk factor profiles differ for cancers of different regions of the Colorectum. Gastroenterology. 2020;159(1):241–e256213.
7. Yamauchi M, Morikawa T, Kuchiba A, Imamura Y, Qian ZR, Nishihara R, Liao X, Waldron L, Hoshida Y, Huttenhower C. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. Gut. 2012;61(6):847–54.
8. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD. Bacterial biogeography of the human digestive tract. Sci Rep. 2011;1(1):170.
9. Mima K, Cao Y, Chan AT, Qian ZR, Nowak JA, Masugi Y, Shi Y, Song M, Da Silva A, Gu M. Fusobacterium Nucleatumin Colorectal Carcinoma tissue according to Tumor Location. Clin Translational Gastroenterol. 2016;7(11):e200.
10. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in colorectal cancer: a challenge for personalized medicine? Int J Mol Sci. 2018;19(12):3733.
11. Morris MT, Jain A, Sun B, Kurbatov V, Muca E, Zeng Z, Jin Y, Roper J, Lu J, Paty PB. Multi-omic analysis reveals metabolic pathways that characterize right-sided colon cancer liver metastasis. Cancer Lett. 2023;574:216384.
12. Cai Y, Shen X, Lu L, Yan H, Huang H, Gaule P, Muca E, Theriot CM, Rattray Z, Rattray NJ. Bile acid distributions, sex-specificity, and prognosis in colorectal cancer. Biology Sex Differences. 2022;13(1):61.
13. Yamauchi M, Lochhead P, Morikawa T, Huttenhower C, Chan AT, Giovannucci E, Fuchs C, Ogino S. Colorectal cancer: a tale of two sides or a continuum? In. Volume 61. BMJ Publishing Group; 2012. pp. 794–7.
14. Jess P, Hansen IO, Gamborg M, Jess T, Danish Colorectal Cancer G. A nationwide Danish cohort study challenging the categorisation into right-sided and left-sided colon cancer. BMJ Open. 2013;3(5):e002608.
15. Sullivan MR, Danai LV, Lewis CA, Chan SH, Gui DY, Kunchok T, Dennstedt EA, Vander Heiden MG, Muir A. Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. Elife. 2019;8:e44235.
16. Kim CH. Control of lymphocyte functions by gut microbiota-derived short-chain fatty acids. Cell Mol Immunol. 2021;18(5):1161–71.

## Publisher's note