BBA - Molecular Basis of Disease 1870 (2024) 167311

Contents lists available at ScienceDirect

BBA - Molecular Basis of Disease

journal homepage: www.elsevier.com/locate/bbadis

# Differences in the molecular organisation of tumours along the colon are linked to interactions within the tumour ecosystem

Tiziano Dallavilla [a], Serena Galiè [a], Gaia Sambruni [a], Simona Borin [b], Nicola Fazio [c], Uberto Fumagalli-Romario [b], Teresa Manzo [a], Luigi Nezi [a], Martin H. Schaefer [a,*]

[a] Department of Experimental Oncology, European Institute of Oncology-IRCCS, Milano, Italy
[b] Digestive Surgery, European Institute of Oncology-IRCCS, Milano, Italy
[c] Division of Gastrointestinal Medical Oncology and Neuroendocrine Tumors, European Institute of Oncology-IRCCS, Milano, Italy

A B S T R A C T

Tumours exhibit significant heterogeneity in their molecular profiles across patients, largely influenced by the tissue of origin, where certain driver gene mutations are predominantly associated with specific cancer types. Here, we unveil an additional layer of complexity: some cancer types display anatomic location-specific mutation profiles akin to tissue-specificity.

To better understand this phenomenon, we concentrate on colon cancer. While prior studies have noted changes of the frequency of molecular alterations along the colon, the underlying reasons and whether those changes occur rather gradual or are distinct between the left and right colon, remain unclear. Developing and leveraging stringent statistical models on molecular data from 522 colorectal tumours from The Cancer Genome Atlas, we reveal disparities in molecular properties between the left and right colon affecting many genes. Interestingly, alterations in genes responsive to environmental cues and properties of the tumour ecosystem, including metabolites which we quantify in a cohort of 27 colorectal cancer patients, exhibit continuous trends along the colon.

Employing network methodologies, we uncover close interactions between metabolites and genes, including drivers of colon cancer, showing continuous abundance or alteration profiles. This underscores how anatomic biases in the composition and interactions within the tumour ecosystem help explaining gradients of carcinogenesis along the colon.

## 1. Introduction

Tumours show a high degree of tissue-specificity, where distinct mutations manifest selectively in particular cancer types while remaining absent in others [1]. Both cell internal as well as environmental factors have been proposed to contribute to the tissue-specificity of cancer molecular alteration profiles [1]. Notably, colon cancer exhibits an anatomical location-specificity where tumours from different colon segments are molecularly different [2], yet the underlying determinants of these distinctions remain elusive. Furthermore, a comprehensive exploration into whether similar anatomic location-based differences exist across various cancer types has not been systematically undertaken.

Epidemiological and molecular studies have suggested that left colon cancers (i.e. distal) are physiologically, pathologically, and genetically different from right colon cancers (i.e. proximal) [2–5]. These studies led to the development of the two-colon paradigm, which proposes classifying CRC according to their location relative to the splenic flexure. While the contribution of this division to a better understanding and treatment of CRC development is widely recognised, recent studies are questioning if this model best captures the regional variation of tumour properties. Indeed, some critical molecular features of CRC change gradually along the colon rather than undergoing a sudden change at the splenic flexure [6–9]. There is no definitive answer yet on whether such a continuous model would describe variations of CRC properties along the colon better than the two-colon paradigm. However, this observation has potential clinical implications since treatment decisions might benefit from considering the tumour's position with higher granularity.

The colon houses the most diverse microbial community of the human body [9–11], and single microbes, as well as microbiome composition, have been linked to carcinogenesis. Closely associated with the microbiome is the intestinal metabolome, and certain metabolites of the colon have been associated with specific colon cancer subtypes [12]. This raises the question of whether there is a link between local composition of microbiome as well as metabolome and differences between tumours along the proximal-to-distal axis along the colon. However, spatially resolved microbiome and metabolomics data is largely missing. We address this by performing spatially resolved

* Corresponding author.
*E-mail address:* martin.schaefer@ieo.it (M.H. Schaefer).

https://doi.org/10.1016/j.bbadis.2024.167311
Received 8 March 2024; Received in revised form 22 May 2024; Accepted 14 June 2024
Available online 22 June 2024

metabolomics in a cohort of 27 CRC patients and we implemented our recent computational approach to reconstruct the microbiome from tumour sequencing data [13] to infer spatial trends of the microbiome along the colon.

In this study, we first show that several cancer types including CRC show substantial differences in driver gene mutation frequencies at different anatomic locations. In order to better understand this phenomenon, we systematically compare molecular, clinical and tissue-environmental properties of CRC and ask how well the two-colon or continuum colon models explain molecular variation along the colon's proximal-distal axis. We developed a workflow to test the goodness of the two models in describing the spatial dependency of a large number of molecular and environmental features (Fig. 1A). We then investigated the interplay between the profiled molecular and environmental features of CRC.

Results show that each model successfully describes a subset of features undergoing alterations at the molecular and environmental levels. However, while tumours exhibit continuous behaviours in both their molecular and environmental properties, sigmoid behaviours are only observed in the molecular properties of tumours and are less frequent than continuous behaviours. We identified close interactions between continuous genes and CRC-associated metabolites suggesting environmentally-induced variation in selection acting on those genes (Fig. 1B).

## 2. Materials and methods

### 2.1. Data

#### 2.1.1. Molecular tumour data

Transcriptome and clinical data of tumour samples were obtained from The Cancer Genome Atlas Program (TCGA) (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). We collected 522 cancer samples from the colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) projects. HTSeq [14] FPKM counts were downloaded from the GDC data portal and converted to TPM. Transcriptome and clinical data of healthy colon samples were retrieved from the GTEx portal (https://gtexportal.org/). We collected 268 samples from sigmoid and transverse colon as TPM counts files.

We retrieved promoter methylation levels for 23,436 promoters of 14,856 genes [15]. We excluded from the analysis promoters with a methylation level outside the $\pm 1\sigma$ range.

Four different mutation annotation outputs from four different callers were retrieved from TCGA as VCF files. We considered the union of the different caller outputs. The following mutation types were discarded: Intragenic region, Intron, RNA, Silent, Splice Region, Splice Site, and Translation Start Site. We also excluded from the analysis genes with an average mutation frequency per section <5 %.

Mutational signature counts were extracted from TCGA VCF files using mutSigExtractor (https://github.com/UMCUGenetics/mutSigExtractor, v1.28).

The CIMP, MSI and CIN status of TCGA samples were obtained [16]. We retrieved CMS data [17]. TCGA samples with no classification were discarded from this analysis.

#### 2.1.2. Immune cells

CIBERSORTx [18] was used to estimate immune infiltrates from transcriptomic data. The analysis was done with the following settings:

- Job type: Impute Cell Fractions
- Batch correction: Enabled
- Batch correction mode: B-mode
- Disable quantile normalisation: True
- Run mode: Absolute
- Permutations: 1000.

We used a signature matrix [18] containing barcode genes for 22 immune cell types to estimate immune cell infiltrates.

#### 2.1.3. Gut microbiota

Microbiota data were estimated from the unmapped reads from the BAM files obtained from TCGA samples. We adapted our previously published method to extract bacterial reads from human sequencing data [13], following a controversy regarding the misclassification of human reads as bacterial [19,20]. Unmapped reads from the BAM files were aligned against the most current human T2T-CHM13 reference genome [21] using STAR, version 2.7.10b [22]. Those which were still not mapped served as input for PathSeq [23]. In addition, we eliminated ambiguous sequences that aligned to both the human genome and bacterial genomes. Subsequently, we quantified the presence of bacteria using the "unambiguous scores" method, as detailed in [13]. We detected a total of 7,903,173 bacterial reads and classified 8793 bacterial species. We used the data repository for human gut microbiota (GMrepo) to exclude non-human species or those observed in <100 human samples.

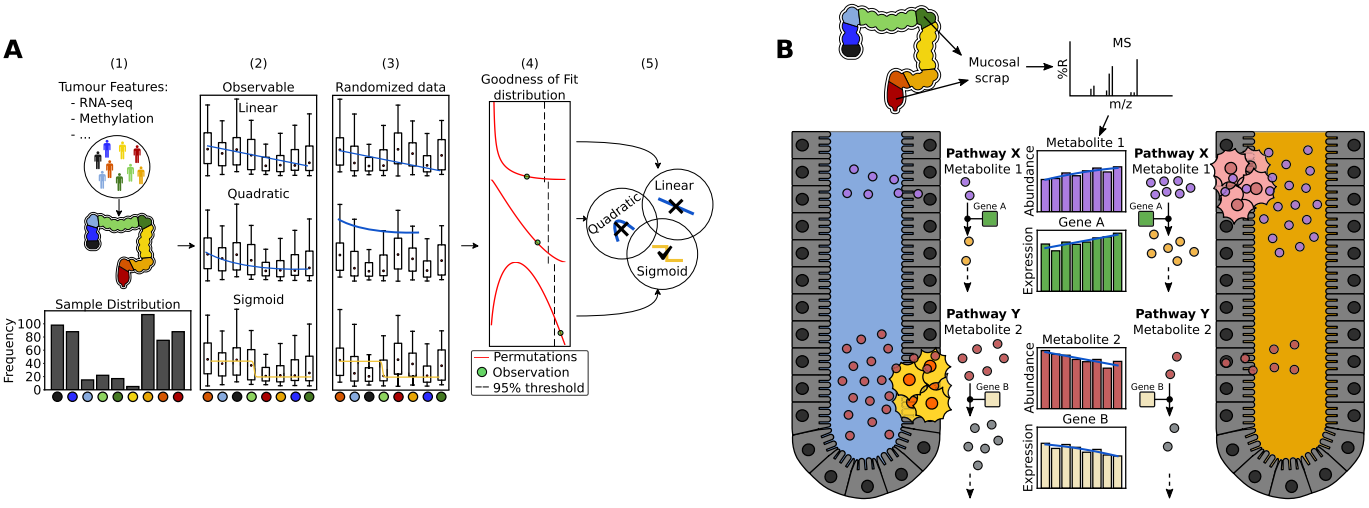


**Fig. 1.** Overview of the study methods and hypothesis. Graphical schemes of (A) the workflow used to investigate the spatial dependency of tumour features along the colon to identify subsets of continuous or sigmoid tumour features and (B) the hypothetical link between environmental tumour features and alterations in tumour molecular features.

#### 2.1.4. Colon metabolome

We assembled a cohort of 27 CRC patients at the European Institute of Oncology (IEO) hospital and quantified the abundances of 346 metabolites by an untargeted metabolomics approach in colon tissue samples. Metabolomics analysis was performed on the mucosal scrap of tumour samples by UPLC-MS/MS. A normalisation step was included to take into account the initial weight of the colon tissue received. Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. After removing proteins, the resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI. For the downstream statistical analysis, only metabolites with non-zero measurements in seven or more sections were included. A filtering data step with 30 % cut-off for missing values was considered.

In addition to our own metabolomics experiments, we collected 105 metabolites found differentially abundant in normal/cancer or left/right colon samples in previous studies [24–26].

#### 2.1.5. Comparative Toxicogenomics Database

Chemical-gene interaction data were retrieved from the Comparative Toxicogenomics Database (CTD) at: http://ctdbase.org/downloads/#cg. We filtered the dataset to include only interactions observed in '*homo sapiens*' and removed chemicals interacting with >10,000 genes.

### 2.2. Statistical analysis

#### 2.2.1. Analysis of feature frequency

In this work we analysed how the frequencies of features like, SNP, CNV, CMS, and mutational signatures behave along the proximal-to-distal axis of the colon. In all the datasets considered the frequencies in each colon section are calculated as the number of samples with a particular feature property (i.e. a CMS state or the presence of a CNV) divided by the total number of samples in a section.

#### 2.2.2. Analysis of mutation frequencies in TCGA projects

For the analysis of mutation frequencies, we considered TCGA projects with at least 200 samples. For each of these projects, we discarded those sections (which were retrieved from clinical data from the field 'site of resection or biopsy') with <50 samples. We discarded from the analysis all the mutations classified as silent. PCA analysis was performed on the mutation frequency of 18 genes that were mutated in all the sections of all the tissues considered in at least one sample (CD209, CDKN2A, COL4A6, DNMT3A, EGFR, FLG, KEL, MECOM, NF1, NOTCH1, PIK3CA, PIK3R1, PLCG2, POLR2B, PRDM1, PTEN, RASA1, TP53).

For each tissue, we calculated the average Euclidean distance between sections by using as coordinates the mutation frequency of all the genes.

The distribution of differences in mutation frequency between sections was calculated by computing the difference in frequency between all the possible combinations of sections within a tissue. The distribution of differences in mutation frequencies between tissues was calculated by computing the difference in frequency between all the possible combinations of tissues.

#### 2.2.3. Design of the profiling workflow

TCGA clinical data of COAD and READ primary tumour samples determines the number of sections of our colon model. The field 'site of resection or biopsy' reported 12 categories. We discarded 3 categories: 'Colon NOS', 'Connective, subcutaneous and other soft tissues of abdomen', 'Unknown primary site' and grouped the samples into nine colon segments (Fig. 1A(1)). For the analysis of TCGA samples, we combined the splenic flexure with the descending colon section due to the low number of samples in the former (five). Similarly, in our analysis of IEO data, we excluded the Transverse colon section from the model as no samples belonged to that section.

For continuous datasets, we calculated the median per section of each feature. Instead, for binary datasets, we computed the percentage of samples with a certain property (e.g. mutation present) out of the total samples per section. Features with zeroes or missing measurements in >30 % of the samples were discarded. The resulting dataset contained the observed median value of each feature for each colon section (Fig. 1A(2)).

In the next step, we created random datasets for bootstrapping to test the ability of each model to fit observed data better than random ones. To generate random data sets, we performed random permutations of the colon sections from the observed dataset. We created 100 sets for each dataset analysed (Fig. 1A(3)).

The observed and random datasets were fitted with three models: linear, sigmoid and polynomial (from quadratic to quartic depending on the data type analysed) (Fig. 1A(2–3)). We used the coefficient of determination (R2) to measure the GoF of each feature.

To identify those models in which the R2 score distribution of the observed data was significantly greater than random data, we tested the observed and random distribution scores for each model using the one-sided Mann–Whitney *U* test (testing the null hypothesis that two related paired distribution, observed and random data, came from the same distribution). We discarded models with a $p$-value $> 0.05$ (Supplementary Fig. 3).

Then for each significant model, we determined which features obtained a better score than random data. For each feature, we fitted the R2 scores of the 100 random sets with a normal and assigned a p-value calculated as the complement of the cumulative distribution function of the fitted distribution at x equal to the R2 score of the observed feature. We corrected the $p$-values by False discovery rate (FDR) and, for each model, features with a q-value $< 0.2$ were classified as potentially following that type of behaviour (Fig. 1A(4–5)). In this way, a single feature could be assigned to more than one model. In such cases, we assign a feature to a profile only if one of the models has a q-value $> 10$ % with respect to all the other model scores. Otherwise, we discard that feature from the analysis.

Due to the zero-inflated nature of the microbiome and mutational signature data, a slightly different approach was used when analysing those datasets. Since these matrices are sparse, instead of calculating the median abundance or counts per section (which would typically give a zero value), we calculated the mean per section.

For the signature dataset, before calculating the median counts per section, we removed samples with zeroes in >70 % of the signatures. For the microbiome dataset once we calculated the mean abundance per section we removed from the analysis bacterial species with zero mean in four or more colon segments.

#### 2.2.4. Design of the network search workflow

The workflow takes as input a set of genes and metabolites of interest. All the 2083 continuous genes found during analysis of gene expression spatial behaviour, along with a list of metabolites composed by CRC associated metabolites taken from literature and metabolites that we found experimentally to follow a continuous trend (see below) were queried in Reactome to retrieve the pathways containing at least one gene/metabolite pair from the input sets (Fig. 5A(12)).

For every pathway found, we used NetworkX [27] to compute the shortest path between the gene and metabolite of each pair. We considered a pair to be close when the distance between the gene and metabolite nodes was ≤3. We counted the number of close pairs contained in each pathway and defined that as the observed test statistics (TS) (Fig. 5A(3)).

To determine if the number of close reactions found in a pathway was significantly higher than the average number of close interactions

between the input set of genes and random metabolites, we compared the observed TS of a pathway with the distribution of random TS generated by bootstrapping. For each pathway, we generated 500 random groups of metabolites, with the group size matching the number of input metabolites found in each pathway and for each group, we calculated our TS (Fig. 5A(4)). The TS values from bootstrapping were used to estimate a distribution of random TS (Fig. 5A(5)). We set a confidence interval at 5 % of the random TS distribution area. If the observed network test statistic was above 95 % of the area of the random distribution, we concluded that the number of interactions between the genes and metabolites of interest was higher than what was expected by randomly picking metabolites inside the pathway (Fig. 5A(6)). We assigned a *p*-value to each pathway depending on the position of its TS with respect to the area of the random distribution. We then corrected the *p*-values of all the selected pathways by false discovery rate (FDR). Pathways with a q-value $\leq 0.2$ were identified as pathways with groups of gene/metabolite of interest closely interacting (Fig. 5A(6)).

#### 2.2.5. Filtering transcriptomic data with GTEX dataset

The aim of this filtering step was to identify those genes with a trend in cancer samples (TCGA), but no trend in normal samples (GTEX).

We used GTEX metadata to determine whether samples from GTEX belonged to the left or right colon. Samples labelled as 'Colon - Sigmoid' were classified as left, while those labelled as 'Colon - Transverse' were classified as right. Next, we calculated the median distribution for each side, which allowed us to calculate the right/left expression ratio for each gene. We excluded from the analysis genes with a GTEX right/left ratio outside $\pm 1\sigma$ from the median of TCGA expression distribution. We then repeated this procedure for TCGA expression data, excluding genes in between $\pm 0.5\sigma$ from the median of TCGA expression distribution.

#### 2.2.6. Quantification of colibactin reads in COAD samples

Resulting BAM files from PathSeq analysis described in the work of [13] were used to estimate colibactin reads in COAD samples. Using Bowtie 2 the output of PathSeq for each COAD sample was aligned on the FASTA sequence of *Escherichia coli* colibactin polyketide biosynthesis gene cluster (GenBank: AM229678), a cluster of genes responsible for the synthesis of colibactin. Each read from the PathSeq output mapping to this region was counted as a colibactin read.

We then computed the percentage of colibactin-positive samples in each colon section. A sample was considered positive for colibactin if it contained 3 or more colibactin reads.

#### 2.2.7. Test enrichment in metabolites interacting with continuous genes

We tested CTD metabolites for enrichment in interaction with continuous genes for each chemical using Chi2. We then used significant chemicals to test enrichment in compounds associated with colorectal neoplasm (MESH ID: D015179). From the list of chemicals associated with colorectal neoplasm, we removed chemicals where the association with the disease has been observed in <10 studies.

#### 2.2.8. Gene ontology analysis

GO analyses were performed using ConsensusPathDB [28]. For gene set analysis queries were filtered by biological process terms from gene ontology levels 2–5 and a q-value cutoff of 0.2. Results are available in Supplementary Table 2.

## 3. Results

### 3.1. Many cancer types show high anatomic location-specificity in their driver gene mutation frequencies

We wondered if the mutation profiles of tumours from different anatomic locations of the same tissue-of-origin tended to be more similar to each other as compared to those of tumours of different cancer types. We observed that for some cancer types, segments corresponding to one cancer type clustered in a space that was spanned by cancer driver mutation frequencies. However, some cancer types showed strong differences between tumours from different anatomic locations within the same tissue-of-origin comparable to differences between cancer types (Fig. 2A). Colon cancer showed the highest average distance between tumours from different anatomic locations (Fig. 2B).

We then considered each frequently mutated cancer driver gene [29] one by one and asked if the distribution of mutation frequency differences was higher between different cancer types than between different anatomic positions of the same cancer type. Intriguingly, we found that of 132 cancer genes, 127 showed no significant difference ($q > 0.2$; Wilcoxon test; Supplementary Table 1). This includes tumour suppressor genes such as PBRM1 or BRCA2 and oncogenes like BRAF (Fig. 2C).

### 3.2. Molecular properties of tumours show spatial trends along the proximal-to-distal axis of the colon

We compared the continuum and the two-colon cancer hypotheses leveraging all major available molecular data types from The Cancer Genome Atlas (TCGA) by developing stringent statistical models testing how well the occurrence of molecular alterations could be explained by one model or the other (Fig. 3A, Supplementary Table 1). We retrieved gene expression data for 60,483 protein-coding and non-coding genes in 9 colon sections from 522 TCGA samples.

As we were primarily interested in cancer-specific patterns and not genes that might show differences in expression along the healthy colon too, we used data from The Genotype-Tissue Expression (GTEx) [30] quantifying gene expression of normal samples and excluded genes that exhibited a similar trend in both cancer and normal samples from all subsequent analyses relating to transcriptome data.

We then used non-linear least squares to fit each expression profile. We used linear and polynomial models to represent a continuous colon, while a sigmoid model was used to represent the two-colon paradigm (Fig. 1A). Analysis of expression profiles identified 846 continuous and 259 sigmoid genes (Supplementary Table 1). Of the 259 sigmoid genes, 160 had an inflexion point between the transverse and descending colon, as expected from the two-colon paradigm (Supplementary Fig. 1A, Supplementary Table 1). Interestingly some genes with an established role in colon carcinogenesis, like *APC* or *RPN2*, follow a continuous expression profile along the proximal to distal axis of the colon (Fig. 3B).

We then performed gene ontology (GO) enrichment analysis [28] on sigmoid genes with an inflexion point between the transverse and descending colon. Only a few GO terms were significantly enriched and those were associated with cell cycle (Supplementary Table 2). We repeated the GO analysis on genes with continuous profiles, discovering that many terms were related to various types of metabolism (Supplementary Table 2).

We wondered which of the two models would best describe the distribution of other molecular properties of the tumour. We observed that promoter methylation levels were significantly higher in the right colon ($p = 1.36e\text{-}8$; Wilcoxon rank sum test, Supplementary Fig. 1D), in agreement with previous observations showing that the CpG island methylator phenotype (CIMP) is mostly present in proximal tumours [31,32]. Fitting methylation data revealed 7976 promoters showing a continuous profile, while 1435 followed a sigmoid one (Supplementary Table 1). As predicted by the two-colon model, most of the promoters with a sigmoid profile had inflexion points between the transverse and descending colon (Supplementary Fig. 1B). We asked how many promoters following a sigmoid or continuous trend would be under positive selection for methylation change during tumour evolution and, hence, are likely drivers of colon carcinogenesis [15]. We found a significant overlap between those genes under selection for methylation changes and those where the methylation change along the colon is well described by one of our models ($p = 0{,}0002$; Fisher's exact test).

We then retrieved mutation data from TCGA, and for each gene, we

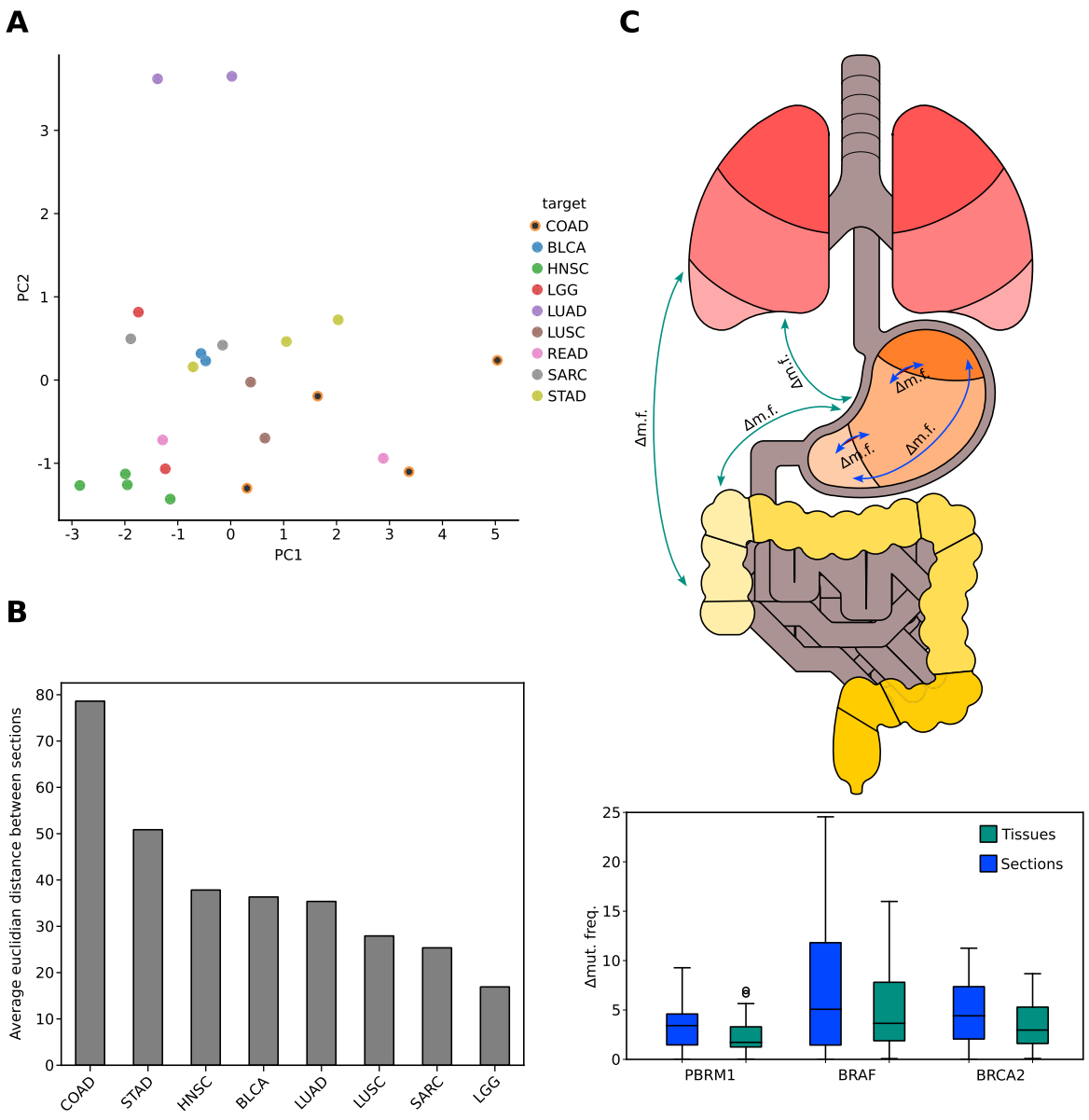


**Fig. 2.** Intra-tissue variations in gene mutation frequencies can be comparable to or even higher than inter-tissue variations. (A) PCA visualisation of mutation frequencies in all tissue sections for 18 genes showing mutations in each section of each TCGA project considered. (B) Average Euclidean distance between sections of each tissue considered. Distances were calculated using the mutation frequency of each gene in a section as coordinates. (C) Comparison between the distribution of PBRM1, BRAF and BRCA2 frequency difference within tissue sections and the frequency difference between the nine different tissues analysed. The distributions are not significantly different ($q > 0.05$; Wilcoxon test).

calculated the mutation frequency per colon section. Results revealed that genes had higher mutation frequencies in the right sections of the colon ($p < 0{,}0001$; Wilcoxon test, Supplementary Fig. 1E), in agreement with previous observations showing that hypermutated CRCs are more common in the proximal colon [33]. Fitting of the 1211 genes with an average mutation frequency per section ≥5 % revealed that 202 showed a sigmoid mutation frequency profile, while 306 were continuous (Supplementary Table 1). The distribution of inflexion points of sigmoidal profiles resembled what was found for transcriptome and methylation data (Supplementary Fig. 1C). Notably, *APC* mutation frequency followed a continuous trend (Fig. 3C); which is remarkable as *APC* is one of the most frequently mutated tumour suppressor genes in CRC [34]. Moreover, our profile analysis indicates that *APC* expression and mutation frequency show opposite gradual trends (Fig. 3B–C). A continuous mutation frequency was observed for the oncogene BRAF (Fig. 3C).

CRC can be subdivided into four consensus molecular subtypes (CMS) based on transcriptome data [2,17]. We analysed the relative frequency of CMSs in each colon section (Supplementary Table 1). CMS2 followed a linear trend increasing from the cecum to the rectum (Supplementary Fig. 1F).

CIMP, CIN and MSI status have been previously reported to differ between tumours from the left vs. those from the right colon [2]. We asked if their frequencies could be better described by continuous or sigmoid models. Fitting results showed that CIN and MSI molecular classifications followed a sigmoid behaviour, CIN with a higher frequency in distal sections and the changing point at the splenic flexure, the opposite was observed for MSI. CIMP followed a continuous trend, with a higher frequency in the cecum gradually decreasing towards the rectum (Supplementary Fig. 1G).

Chromosomal amplification or loss is a common CRC feature primarily affecting left-sided cancers [2]. We calculated each chromosome

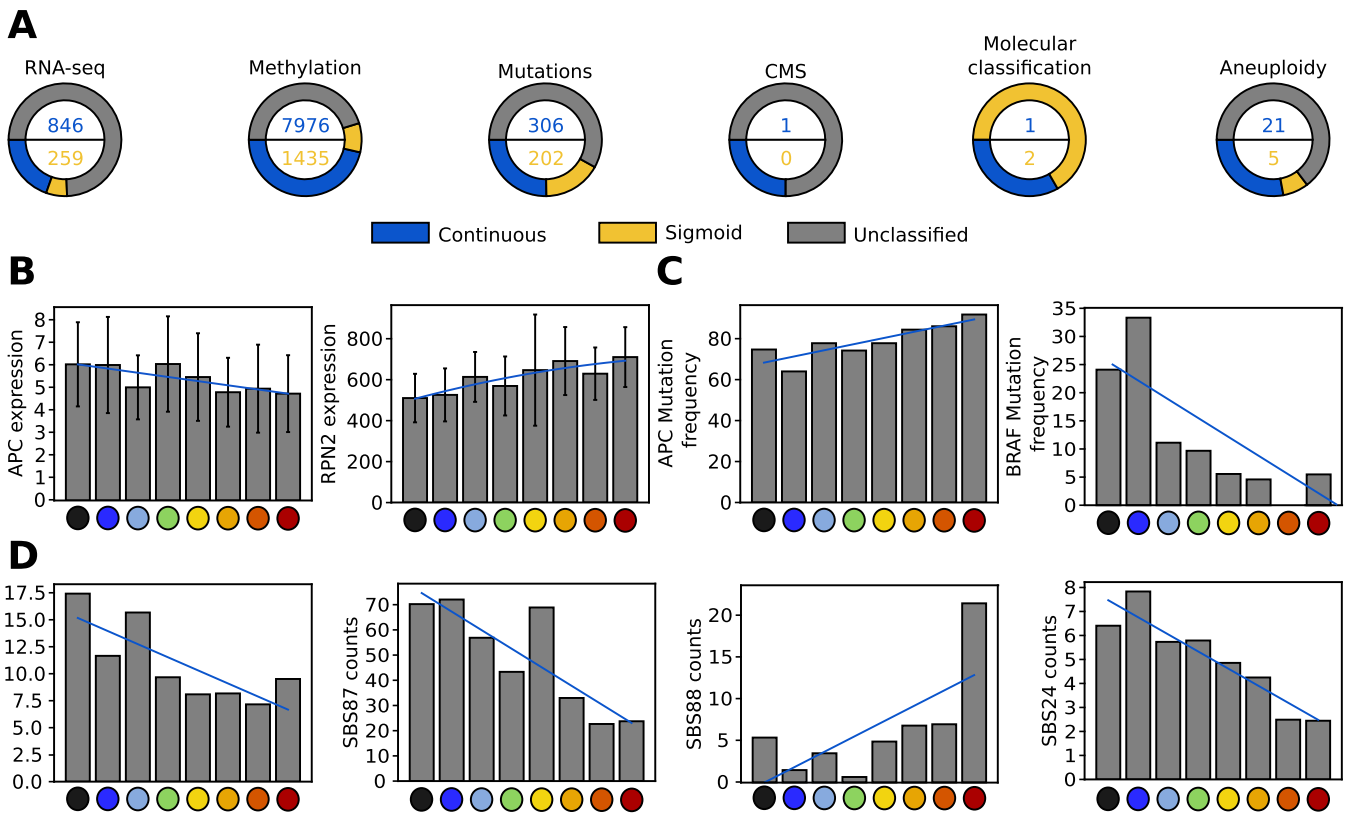


**Fig. 3.** Different molecular tumour features show continuous and sigmoid trends. (A) Proportion of sigmoid and continuous features identified for each molecular tumour property analysed. (B) Expression profiles of APC and RPN2 along the proximal to distal axis of the colon. Expression is measured in transcripts per million. (C) Mutation frequency of APC and BRAF. The frequency is calculated as the fraction of samples with a gene mutation over the total samples in a section. (D) Mutational signatures profile of continuous signatures SBS29, SBS87, SBS88, SBS24. Data are reported as mean counts per section.

arm's amplification and deletion frequency in each colon section. Analysis of deletions revealed 11 continuous and 5 sigmoid deletions. We identified eight amplifications following a continuous trend (Supplementary Table 1).

Finally, we tested trends for tumour mutational signatures, which are mutation patterns linked to distinctive mutagenesis processes, sometimes characteristic of specific environmental factors. We analysed the abundance profile of 60 mutational signatures. We found four signatures following a continuous behaviour (Fig. 3D, Supplementary Table 1) and four following a sigmoid behaviour (Supplementary Fig. 1H, Supplementary Table 1), showing the inflexion point in between the transverse and descending colon. Interestingly the four continuous signatures are found to be associated with environmental factors, namely chemotherapy treatment (SBS87), tobacco chewing (SBS29), aflatoxin (SBS24) and colibactin exposure (SBS88). Sigmoid signatures are found to be associated with molecular processes, like defective DNA mismatch repair (SBS6, SBS26), and clock-like signatures associated with ageing (SBS1).

### 3.3. Environmental factors follow a continuous trend along the colon

Our results show that the distribution of molecular features of tumours along the colon can be often well described by either sigmoid or continuous models. Intriguingly, genes associated with metabolism or interactions with the local environment tend to be associated with continuous trends. Previous studies revealed that the colon's chemical and physical properties, like pH, oxygen concentration and mucus thickness, follow continuous trends [10]. We tested if the genes with a continuous expression profile were enriched in interactions with environmental chemicals, as defined by the Comparative Toxicogenomics Database (CTD) [35], 37 chemicals had more interactions than expected by chance with continuous genes ($q < 0.2$; Chi-squared test; Supplementary Table 3). Of those, seven metabolites were involved in colorectal neoplasms, which is a higher number than expected by chance ($p$-value = 2.55e-11; Chi-squared test; annotation by CTD).

These results led us to wonder if the continuous distributions of environmental factors of the tumour ecosystem could play a role in shaping the observed continuous trends.

Aiming to quantify and reconstruct the spatial organisation of the tumour ecosystem along the colon by experimental and computational means first we charted changes in the tumoural tissue metabolome. Since publicly accessible metabolomic data with matching positional information within the colon are sparse, we performed global untargeted metabolomic analysis on 27 surgically resected tumours from a new cohort of patients ($n = 27$) with CRC collected at the IEO Hospital (Supplementary Fig. 2A).

Results from our spatial dependency analysis revealed that 10 out of 346 measured metabolites followed a continuous profile while no metabolite showed a sigmoid profile (Fig. 4A, B). Among the continuous metabolites identified, fatty acids (eicosenoate, erucate, oleate/vaccenate, stereate) are noteworthy since different studies revealed an increasingly important role of these compounds and their metabolism in tumorigenesis and cancer progression [36,37]. Other metabolites like polyamines (spermine), acylcarnitines (butyrylcarnitine, propionylcarnitine), and sulphate metabolites (S-adenosylhomocysteine) also showed continuous behaviour.

The microbiome has been demonstrated to have an impact on cancer progression [38] and is closely linked to the colon metabolome, in this regard, we have recently demonstrated that quantification of the intratumour microbiota can be inferred by extracting microbial reads from RNA-seq data and used this approach to reconstruct the colon cancer microbiota from TCGA COAD and READ samples [13,39]. We used this

data to profile the abundance of the microbes along the proximal-to-distal colon axis. Strikingly we found 17 continuous human microbes and none with sigmoid behaviour (Fig 4A, Supplementary Table 1). Interestingly, among the genera identified, *Porphyromonas gingivalis* and *Ruminococcus gnavus* have been already associated to CRC [40], while a CRC protective role has been proposed for *Bifidobacterium adolescentis* and the butyrate-producing bacteria *Roseburia hominis*, *Coprococcus comes* and *Blautia hydrogenotrophica* [41–45].

To further investigate the interplay of gradients observed in microbiome and metabolome data, we correlated average abundances per segment between the two, resulting in 102 significant correlations ($q < 0.2$; Pearson correlation test; Supplementary Table 3), mainly involving S-adenosine homocystein (SAH), acetylcarnitines (butyrylcarnitine), long chain fatty acids (erucate, stearate) and citrate. We tested if the number of observed correlations was larger than expected by chance by randomly permutating the colon sections for metabolites and microbe measurements, revealing that the original set showed a significantly higher number of correlations ($p = 0.0099$ and $p < 0.0001$; randomisation test; Supplementary Fig. 2B).

Long-chain fatty acids have been extensively associated with CRC progression and an immune suppressive tumour immune microenvironment [46,47]. The increased fatty acid metabolism in the tumour microenvironment is linked to the carnitine system levels, as a consequence of the increased mitochondrial protein acetylation and subsequent Randle effect [48]. Intriguingly, we observe a significant positive correlation of *Porphyromonas gingivalis*, a well known pathogen in CRC progression with SAH which acts as a metabolic fuel for this asaccharolytic bacteria through the activation of its methyl-cycle metabolism. Intriguingly, this metabolic pathway is strictly linked to the *LuxS* gene signalling, a fundamental transcription regulator in *Porphyromonas gingivalis* involved in bacterial growth, biofilm formation, stress resistance and virulence promotion [49]. At the same time, positive correlations have been found between sulphur-reducing bacteria and C4 acylcarnitine (butyrylcarnitine), a metabolite deriving from the carnitine metabolism, suggesting an intimate connection between dietary inputs into $H_2S$ signalling in the gut [49,50]. Further validation studies are therefore needed to elucidate the interaction between nutrition and sulphur homeostasis.

Next, we applied CIBERSORTx to reconstruct immune infiltrates from tumour RNAseq data, estimating the abundance of 22 immune cell types in TCGA samples. Our results pointed at four immune cell types following a continuous behaviour (Fig. 4A, Supplementary Table 1), namely T cells CD8+, T cells CD4 memory resting, and T cells follicular helper, which showed higher abundance in the proximal colon, while macrophages M0 followed an opposite trend and had higher abundance in the rectum. None of the retrieved immune cell types followed a sigmoid abundance pattern along the colon.

### 3.4. Continuous genes and metabolites colocalise in CRC-related pathways

Results show that bacterial, immune cell and metabolite abundances of tumours along the colon can be best described by continuous models. It is tempting to speculate that the high fraction of continuous molecular alterations in tumours is indicative of gradual ecosystem changes in which tumours evolve. To verify this hypothesis, we examined whether genes with continuous expression profiles were closer to continuous or cancer-related metabolites in cellular pathways than expected by chance.

We assembled a list of metabolites relevant to our study, including those with a continuous profile that we detected experimentally and others linked to cancer found in the literature [24–26]. Then, we implemented a network algorithm that, given the set of continuous genes and the list of assembled metabolites, searched for Reactome pathways in which continuous genes and the compiled metabolites were significantly closer in proximity than expected by chance (Fig. 5A). When we run the algorithm on genes with a continuous expression profile, this process identified 16 pathways (Table 1, Supplementary

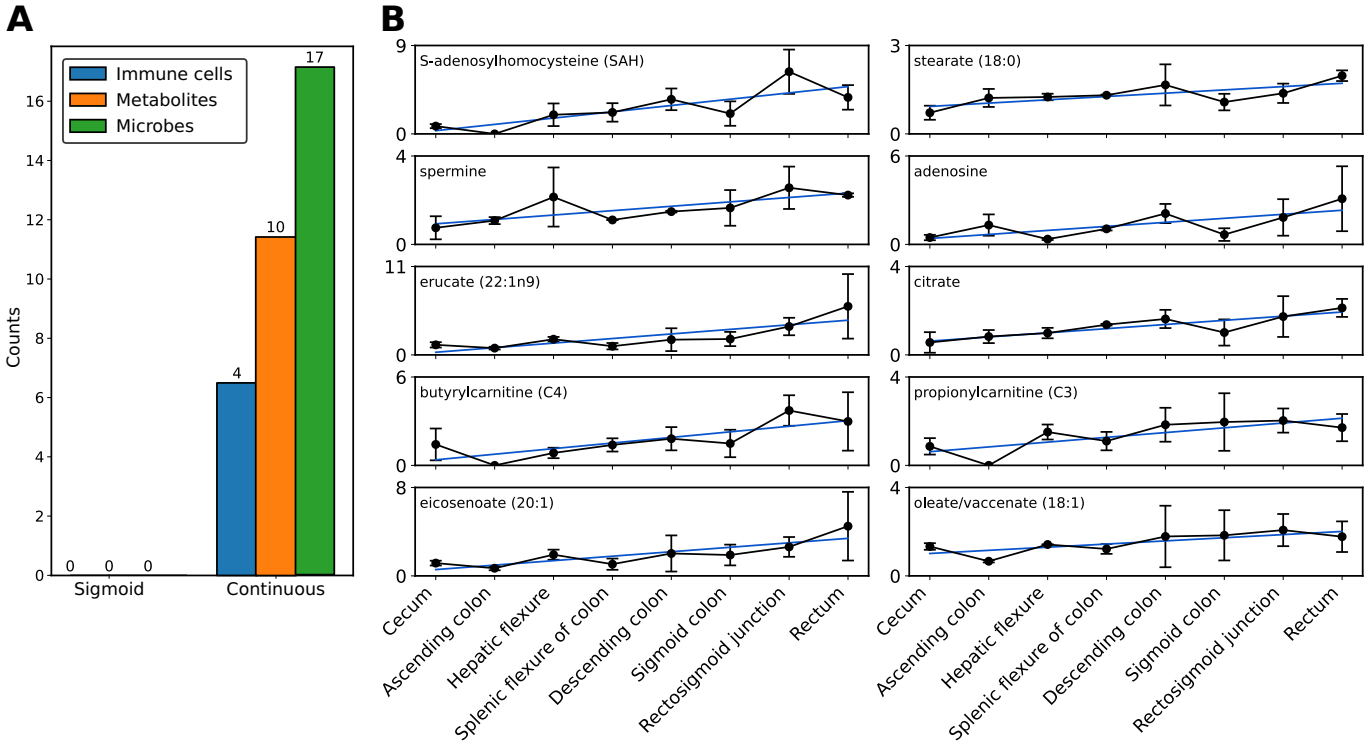


**Fig. 4.** Tumours environmental factors follow prevalently continuous trends along the colon. (A) Number of sigmoid and continuous features identified in immune cells, metabolites, and microbes datasets. (B) Plots of metabolites following a continuous behaviour along the proximal to distal axis of the colon. Dots indicate the median abundance per section, while error bars are the median absolute deviation. On top of each graph is the regression line from the best-fitting model.

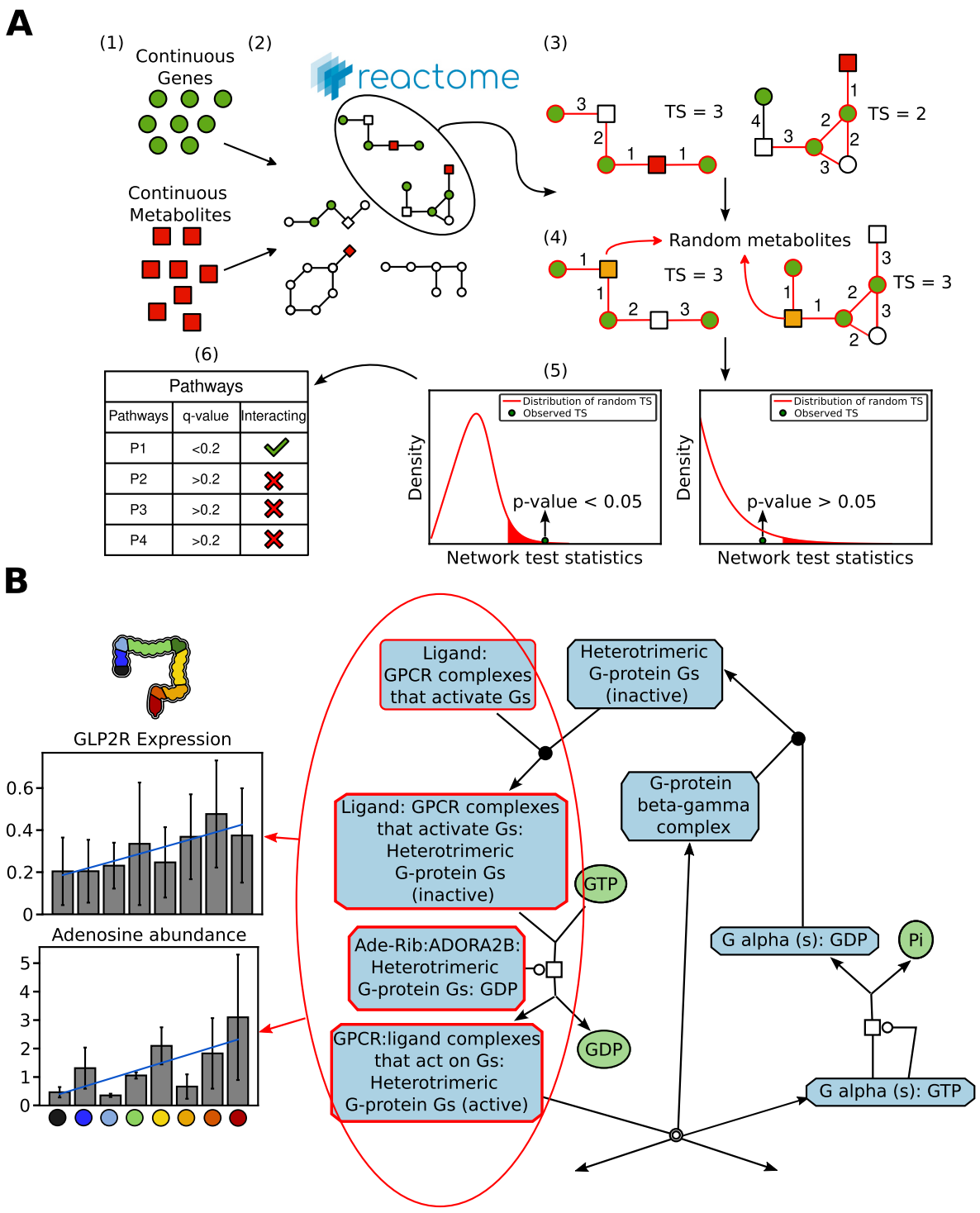


**Fig. 5.** Network analysis methods show that tumour continuous molecular and environmental features closely interact in cancer-related pathways. (A) Graphical scheme of the network search algorithm developed to identify Reactome pathways containing close interactions between genes with a continuous expression profile, and a set of target metabolites composed of our experimentally detected continuous metabolites and other cancer-relevant metabolites. (B) A subset of the G alpha (s) signalling events pathway retrieved from Reactome (https://reactome.org/, 'R-HSA-418555'). Genes with a continuous expression profile are found in complexes with metabolites that share the same type of profile. Expression is measured in transcripts per million, while metabolite abundance is reported as scaled intensity.

Table 3). The same process was repeated on genes with continuous mutation frequency and promoter methylation, identifying 8 and 10 pathways respectively (Supplementary Table 3). These pathways were mainly involved in cellular functions such as metabolism (e.g. glycosaminoglycan, carbohydrate metabolism), signal transduction (e.g. G alpha (s) signalling events, Fig. 5B), gene expression (e.g. Epigenetic regulation of gene expression), organisation of extracellular matrix, and immune system (e.g. Class I MHC mediated antigen processing & presentation).

We repeated this network analysis approach by using the MetaboAnalyst tool [51], which given a list of genes and metabolites allows to perform enrichment analysis on KEGG pathways. 2 out of our 16 previously identified pathways were also identified with MetaboAnalyst (Supplementary Table 3) as significant pathways, confirming that

**Table 1**
List of Pathways containing close interactions between genes with continuous expression profile and metabolites with either continuous abundance profile, associated with CRC or differentially abundant in left vs right colon.

| Superpathway | Pathway | q-Value |
|---|---|---|
| Signal transduction | G alpha (s) signalling events | 0.03 |
| Metabolism of RNA | Metabolism of RNA | 0.12 |
| Metabolism of proteins | Post-translational protein modification | 0.09 |
| | Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO) and transfer to a nascent protein | 0.04 |
| | Metabolism of proteins | 0.19 |
| Metabolism | Biological oxidations | 0.07 |
| | Fatty acid metabolism | 0.10 |
| | Glycosaminoglycan metabolism | 0.03 |
| | Metabolism of carbohydrates | 0.07 |
| | Metabolism of vitamins and cofactors | 0.07 |
| Immune system | Class I MHC mediated antigen processing & presentation | 0.00 |
| | Immune System | 0.05 |
| Gene expression (transcription) | Epigenetic regulation of gene expression | 0.10 |
| | Gene expression (transcription) | 0.07 |
| | Generic transcription pathway | 0.03 |
| Disease | Diseases of metabolism | 0.03 |

Glycosaminoglycan biosynthesis and N-Glycan biosynthesis indeed contains continuous gene in close proximity with either continuous or cancer associated metabolites.

To test if the observed number of pathways with high proximity between continuous genes (expressed, methylated and mutated) and metabolites was larger than expected by chance, we repeatedly randomised the identity of the input metabolites and found a significantly higher number for the original set ($p \leq 0.0001$; randomisation test; Fig. 5C, Supplementary Fig. 2C).

We wondered if we would also find close interactions between sigmoid genes and metabolites. For gene expression, we identified only two pathways enriched in close interactions: Metabolism of RNA and Post-translational protein modification.

While more analysis is necessary to determine causality and directionality, our findings suggest that environmental factors, such as diet and microbiome, may impact tumour molecular properties in a gradually changing manner along the colon. Gradients of selective forces might result in the development of heterogeneous cancer types and their anatomic biases within the colon.

## 4. Discussion

In this study, we first demonstrate that many cancer types show high anatomic location-specificity in their driver gene mutation frequencies. One of the tissues with the largest differences between tumours from different anatomic positions is the colon.

We therefore address the long-standing question of whether a left-right or a continuous model better describes the biological and clinical differences observed in tumours along the proximal-to-distal axis of the colon. Here we describe a comprehensive analysis of CRC's molecular and environmental features, by combining publicly available and newly experimentally generated omics datasets. Our results demonstrated that while the two models are not mutually exclusive for CRC molecular features, specific features follow a continuous behaviour. Genes involved in cell cycle, DNA replication and chromatin maintenance can be best described by a left-right model. Instead, genes related to metabolism and adaptation to the local tissue-environment show continuous changes in their molecular properties along the colon. Intriguingly, we observed a possible link between alterations in those genes and environmental factors, with the microbiome likely playing a central role mediated by the local metabolome.

A main novelty of our work is that we analysed molecular changes in light of environmental selective forces. This allowed us to identify small subnetworks of interactions between genes and the tumour ecosystem. Most of these interactions were identified in pathways closely related to the establishment of the tumour microenvironment or involved in CRC development.

A significant role in CRC development has been attributed to a group of pathways involving continuous genes and metabolites located close to each other: G alpha (s) signalling events. The activation of G protein-coupled receptors, which respond to various luminal metabolites, has gained attention due to its association with the regulation of inflammation and immune response [52–54]. In these pathways, we identified close interaction between G-protein-coupled receptors (and phosphodiesterases) with nucleosides and amine (Supplementary Table 3). Our profiling analysis shows that adenosine and G protein-coupled receptors in this pathway follow a linear trend increasing from the cecum to the rectum (Fig. 5B). In contrast, the abundance of immune T cells shows the opposite trend (Supplementary Fig. 2D). Previous studies have shown that adverse conditions like inflammation or hypoxia can lead to the accumulation of extracellular adenosine, which activates cellular signalling pathways through G-protein-coupled receptors, leading to immunosuppressive effects that can be beneficial to colorectal tumorigenesis [55]. The inverse correlation between adenosine abundance and T cells has been observed before [56]. Interestingly gene-metabolites pairs with continuous gradients have been found in immune system pathways like the Class I MHC mediated antigen processing & presentation pathway, which is a crucial pathway in the activation of CD8 T cells through the recognition of cancer cell antigens [57]. While our models did not allow for interpreting directionality, it is tempting to speculate about the existence of causal links between the tumour and its environment being responsible for the observed sigmoid and continuous trends of the tumours. Future experimentation (e.g. controlled manipulation of the environment in vitro tumour evolution experiments or longitudinal studies) will have to clarify causality and directionality. Discerning the biological features influencing the tissue microenvironment is a challenging hot topic in CRC. Thus our multi-omics approach on large datasets provides a proxy for further experimental validations to investigate a possible causality between microbiome, metabolites abundance and gene expression.

Obviously our results could be confounded by biases in the composition of the patient cohorts used in this study. While it would be very interesting to see if our observations hold for cohorts from currently geographically underrepresented locations, we confirmed that ethnicity and gender do likely not confound our observations as they are not associated with position along the proximal-to-distal axis of the colon (ordinary least square regression, $p > 0.05$).

Interestingly, we find several lines of evidence that gradients of molecular alteration profiles affecting cancer driver genes and CMS could be favoured through specific environmental properties of the tumour ecosystem: 1. We find that FBXW7 (the fifth most frequently mutated colorectal cancer gene [29]) follows a continuous promoter methylation trend. This gene has been identified in the pathway analyses to closely interact with continuously abundant metabolites in the Immune system pathway. 2. SBS88 follows a continuous trend along the colon. This SBS is caused by colibactin and has been recently described to contribute to mutations in colon cancer driver genes such as APC [58]. Interestingly, our analysis of microbiome composition along the colon revealed an increase of colibactin-positive patients towards the distal parts of the colon (Supplementary Fig. 2E). Ultimately, APC, which contains mutations that are in agreement with the SBS88 signature, shows an increase of mutations towards the rectum. 3. CMS2 follows a continuous trend along the colon. One of the genes (TP53RK) characteristic of the CMS2 expression signature, as defined in the work of S. Buechler et al. [59], follows a continuous expression and methylation trend and comes up in the respective network analyses as forming a subnetwork with continuously abundant metabolites in metabolism of RNA and tRNA modification in the nucleus and cytosol pathways.

The intricate link between a tumour and its environment has important clinical applications: first, metabolic or bacterial profiles could serve as biomarkers for tumour presence and position. Second, more precise localisation information along the colon might hold diagnostic or prognostic potential and, consequently, guide tailored therapeutic interventions. In agreement, a previous study established a link between tumour localisation in the colon and prognosis [8].

Beyond the biological and clinical implications, this study marks an important step in the field about the existence of gradient selection. Future methodological work should be dedicated to developing population genomics methods to understand how selection is constrained in space and will have to develop models capable of identifying genes under location-specific selection.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbadis.2024.167311.

## CRediT authorship contribution statement

**Tiziano Dallavilla:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Serena Galiè:** Writing – review & editing, Writing – original draft, Data curation. **Gaia Sambruni:** Writing – review & editing, Data curation. **Simona Borin:** Data curation. **Nicola Fazio:** Data curation. **Uberto Fumagalli-Romario:** Data curation. **Teresa Manzo:** Methodology, Data curation. **Luigi Nezi:** Methodology, Data curation. **Martin H. Schaefer:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Martin Schaefer reports financial support was provided by AIRC Italian Foundation for Cancer Research. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The molecular and clinical data are available in Supplementary Table 4. All the code used to generate the results of this study, is available at https://github.com/Dallavilla-Tiziano/profile_analysis/.

*Acknowledgements*

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

*Ethics approval and consent to participate*

The use of human samples was approved by the European Institute of Oncology Institutional Review Board (protocol n. R1083/19-IEO 1149). Samples were collected from patients diagnosed with CRC. All patients were non-metastatic and treatment-naïve at the time of surgical resection.

The IEO's Ethical Committee (registered as IEO 1781) has approved the collection of samples from healthy donors. All donors provided written informed consent to tissue collection, analysis and data publication according to the declaration of Helsinki.

*Fundings*

The work leading to this manuscript was supported by Fondazione AIRC (grant reference number MFAG 21791 and Bridge Grant n. 29162 to M.S., IG 26406 to L.N. and StartUp Grant 21474 to T.M.) and is partially supported by the Italian Ministry of Health with Ricerca Corrente and 5 x 1000 funds. S.G. is supported by Credit Agricole's through Fondazione IEO-CCM.

## References

[1] K.M. Haigis, K. Cichowski, S.J. Elledge, Tissue-specificity in cancer: the rule, not the exception, Science 363 (2019) 1150–1151.
[2] B. Baran, N. Mert Ozupek, N. Yerli Tetik, E. Acar, O. Bekcioglu, Y. Baskin, Difference between left-sided and right-sided colorectal cancer: a focused review of literature, Gastroenterol. Res. Pract. 11 (2018) 264–273.
[3] M. Yamauchi, P. Lochhead, T. Morikawa, C. Huttenhower, A.T. Chan, E. Giovannucci, C. Fuchs, S. Ogino, Colorectal cancer: a tale of two sides or a continuum? Gut 61 (2012) 794–797.
[4] M.S. Lee, D.G. Menter, S. Kopetz, Right versus left colon cancer biology: integrating the consensus molecular subtypes, J. Natl. Compr. Cancer Netw. 15 (2017) 411–419.
[5] P. Gervaz, P. Bucher, P. Morel, Two colons-two cancers: paradigm shift and clinical implications, J. Surg. Oncol. 88 (2004) 261–266.
[6] M. Yamauchi, T. Morikawa, A. Kuchiba, Y. Imamura, Z.R. Qian, R. Nishihara, X. Liao, L. Waldron, Y. Hoshida, C. Huttenhower, A.T. Chan, E. Giovannucci, C. Fuchs, S. Ogino, Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum, Gut 61 (2012) 847–854.
[7] J.M. Loree, A.A.L. Pereira, M. Lam, A.N. Willauer, K. Raghav, A. Dasari, V. K. Morris, S. Advani, D.G. Menter, C. Eng, K. Shaw, R. Broaddus, M.J. Routbort, Y. Liu, J.S. Morris, R. Luthra, F. Meric-Bernstam, M.J. Overman, D. Maru, S. Kopetz, Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and consensus molecular subtypes, Clin. Cancer Res. 24 (2018) 1062–1072.
[8] T. Ugai, N. Akimoto, K. Haruki, T.A. Harrison, Y. Cao, C. Qu, A.T. Chan, P. T. Campbell, S.I. Berndt, D.D. Buchanan, A.J. Cross, B. Diergaarde, S.J. Gallinger, M.J. Gunter, S. Harlid, A. Hidaka, M. Hoffmeister, H. Brenner, J. Chang-Claude, L. Hsu, et al., Prognostic role of detailed colorectal location and tumor molecular features: analyses of 13,101 colorectal cancer patients including 2994 early-onset cases, J. Gastroenterol. 58 (3) (2023) 229–245.
[9] D. Shalon, R.N. Culver, J.A. Grembi, J. Folz, P.V. Treit, H. Shi, F.A. Rosenberger, L. Dethlefsen, X. Meng, E. Yaffe, A. Aranda-Díaz, P.E. Geyer, J.B. Mueller-Reif, S. Spencer, A.D. Patterson, G. Triadafilopoulos, S.P. Holmes, M. Mann, O. Fiehn, D. A. Relman, et al., Profiling the human intestinal environment under physiological conditions, Nature 617 (2023) 581–591.
[10] A. Chikina, D. Matic Vignjevic, At the right time in the right place: how do luminal gradients position the microbiota along the gut? Cells Dev. 168 (2021) 203712.
[11] G.P. Donaldson, S. Melanie Lee, S.K. Mazmanian, Gut biogeography of the bacterial microbiota, Nat. Rev. Microbiol. 14 (2016) 20–32.
[12] O.O. Coker, C. Liu, W.K.K. Wu, S.H. Wong, W. Jia, J.J.Y. Sung, J. Yu, Altered gut metabolites and microbiota interactions are implicated in colorectal carcinogenesis and can be non-invasive diagnostic biomarkers, Microbiome 10 (2022) 35.
[13] G. Sambruni, A.D. Macandog, J. Wirbel, D. Cagnina, C. Catozzi, T. Dallavilla, F. Borgo, N. Fazio, U. Fumagalli-Romario, W.L. Petz, T. Manzo, S.P. Ravenda, G. Zeller, L. Nezi, M.H. Schaefer, Location and condition based reconstruction of colon cancer microbiome from human RNA sequencing data, Genome Med. 15 (2023) 32.
[14] G.H. Putri, S. Anders, P.T. Pyl, J.E. Pimanda, F. Zanini, Analysing high-throughput sequencing data in Python with HTSeq 20, Bioinformatics 38 (2022) 2943–2945.
[15] R. Heery, M.H. Schaefer, DNA methylation variation along the cancer epigenome and the identification of novel epigenetic driver events, Nucleic Acids Res. 49 (2021) 12692–12705.
[16] Y. Liu, N.S. Sethi, T. Hinoue, B.G. Schneider, A.D. Cherniack, F. Sanchez-Vega, J. A. Seoane, F. Farshidfar, R. Bowlby, M. Islam, J. Kim, W. Chatila, R. Akbani, R. S. Kanchi, C.S. Rabkin, J.E. Willis, K.K. Wang, S.J. McCall, L. Mishra, A.I. Ojesina, et al., Comparative molecular analysis of gastrointestinal adenocarcinomas, Cancer Cell 33 (2018) 721–735.e8.
[17] J. Guinney, R. Dienstmann, X. Wang, A. de Reyniès, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I. M. Simon, S. Gerster, E. Fessler, F. De Sousa, E. Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, et al., The consensus molecular subtypes of colorectal cancer, Nat. Med. 21 (2015) 1350–1356.
[18] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, M. S. Khodadoust, M.S. Esfahani, B.A. Luca, D. Steiner, M. Diehn, A.A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry, Nat. Biotechnol. 37 (2019) 773–782.
[19] A. Gihawi, C.S. Cooper, D.S. Brewer, Caution regarding the specificities of pan-cancer microbial structure, Microb. Genom. 9 (2023).
[20] A. Gihawi, Y. Ge, J. Lu, D. Puiu, A. Xu, C.S. Cooper, D.S. Brewer, M. Pertea, S. L. Salzberg, Major Data Analysis Errors Invalidate Cancer Microbiome Findings, bioRxiv, 2023.
[21] S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A.V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S.J. Hoyt, M. Diekhans, G.A. Logsdon, M. Alonge, S.E. Antonarakis, M. Borchers, G. G. Bouffard, S.Y. Brooks, G.V. Caldas, et al., The complete sequence of a human genome, Science 376 (2022) 44–53.

[22] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[23] M.A. Walker, C.S. Pedamallu, A.I. Ojesina, S. Bullman, T. Sharpe, C.W. Whelan, M. Meyerson, GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts, Bioinformatics 34 (2018) 4287–4289.
[24] J. Chen, X. Liu, L. Shen, Y. Lin, B. Shen, CMBD: a manually curated cancer metabolic biomarker knowledge database, Database 2021 (2021) baaa094.
[25] Y. Cai, N.J.W. Rattray, Q. Zhang, V. Mironova, A. Santos-Neto, E. Muca, A.K. R. Vollmar, K.-S. Hsu, Z. Rattray, J.R. Cross, Y. Zhang, P.B. Paty, S.A. Khan, C. H. Johnson, Tumor tissue-specific biomarkers of colorectal cancer by anatomic location and stage, Metabolites 10 (2020).
[26] B.A. Baxter, K.D. Parker, M.J. Nosler, S. Rao, R. Craig, C. Seiler, E.P. Ryan, Metabolite profile comparisons between ascending and descending colon tissue in healthy adults, World J. Gastroenterol. 26 (2020) 335–352.
[27] A. Hagberg, P.J. Swart, D.A. Schult, Exploring Network Structure, Dynamics, and Function Using Networkx, 2008.
[28] A. Kamburov, K. Pentchev, H. Galicka, C. Wierling, H. Lehrach, R. Herwig, ConsensusPathDB: toward a more complete picture of cell biology, Nucleic Acids Res. 39 (2010) D712–D717.
[29] M.S. Lawrence, P. Stojanov, C.H. Mermel, J.T. Robinson, L.A. Garraway, T. R. Golub, M. Meyerson, S.B. Gabriel, E.S. Lander, G. Getz, Discovery and saturation analysis of cancer genes across 21 tumour types, Nature 505 (2014) 495–501.
[30] GTEx Consortium, The GTEx consortium atlas of genetic regulatory effects across human tissues, Science 369 (2020) 1318–1330.
[31] E. Nazemalhosseini Mojarad, P.J. Kuppen, H.A. Aghdaei, M.R. Zali, The CpG island methylator phenotype (CIMP) in colorectal cancer, Gastroenterol. Hepatol. Bed Bench 6 (2013) 120–128.
[32] H.T. Nguyen, H.-Q. Duong, The molecular characteristics of colorectal cancer: implications for diagnosis and therapy, Oncol. Lett. 16 (2018) 9–18.
[33] G. Weiting, C. Wen, B. Rui, H. Wangxiong, W. Dehao, Z. Shu, H. Hanguang, A novel 4-gene prognostic signature for hypermutated colorectal cancer (CRC), Ann. Oncol. 29 (2018) vi11.
[34] L. Zhang, J.W. Shay, Multiple roles of APC and its therapeutic implications in colorectal cancer, J. Natl. Cancer Inst. 109 (2017) djw332.
[35] A.P. Davis, T.C. Wiegers, R.J. Johnson, D. Sciaky, J. Wiegers, C.J. Mattingly, Comparative Toxicogenomics Database (CTD): update 2023, Nucleic Acids Res. 51 (2023) D1257–D1262.
[36] Y.-A. Moon, Emerging roles of polyunsaturated fatty acid synthesis pathway in colorectal cancer, Anim. Cells Syst. 27 (1) (2023) 61–71.
[37] A. Pakiet, J. Kobiela, P. Stepnowski, T. Sledzinski, A. Mika, Changes in lipids composition and metabolism in colorectal cancer: a review, Lipids Health Dis. 18 (2019) 1–21.
[38] P. Louis, G.L. Hold, H.J. Flint, The gut microbiota, bacterial metabolites and colorectal cancer, Nat. Rev. Microbiol. 12 (2014) 661–672.
[39] G.D. Poore, E. Kopylova, Q. Zhu, C. Carpenter, S. Fraraccio, S. Wandro, T. Kosciolek, S. Janssen, J. Metcalf, S.J. Song, J. Kanbar, S. Miller-Montgomery, R. Heaton, R. Mckay, S.P. Patel, A.D. Swafford, R. Knight, Microbiome analyses of blood and tissues suggest cancer diagnostic approach, Nature 579 (2020) 567–574.
[40] L. Yu, G. Zhao, L. Wang, X. Zhou, J. Sun, X. Li, Y. Zhu, Y. He, K. Kofonikolas, D. Bogaert, M. Dunlop, Y. Zhu, E. Theodoratou, X. Li, A systematic review of microbial markers for risk prediction of colorectal neoplasia, Br. J. Cancer 126 (2022) 1318–1328.
[41] S. Chen, L. Fan, Y. Lin, Y. Qi, C. Xu, Q. Ge, Y. Zhang, Q. Wang, D. Jia, L. Wang, J. Si, L. Wang, Bifidobacterium adolescentis orchestrates CD143 cancer-associated fibroblasts to suppress colorectal tumorigenesis by Wnt signaling-regulated GAS1, Cancer Commun. 43 (2023) 1027–1047.
[42] A.E.V. Quaglio, T.G. Grillo, E.C.S. De Oliveira, L.C. Di Stasi, L.Y. Sassaki, Gut microbiota, inflammatory bowel disease and colorectal cancer, World J. Gastroenterol. 28 (2022) 4053–4060.
[43] I. Sobhani, E. Bergsten, S. Couffin, A. Amiot, B. Nebbad, C. Barau, N. de'Angelis, S. Rabot, F. Canoui-Poitrine, D. Mestivier, T. Pédron, K. Khazaie, P.J. Sansonetti, Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures, Proc. Natl. Acad. Sci. USA 116 (2019) 24285–24295.
[44] H. Zhang, J. Wu, D. Ji, Y. Liu, S. Lu, Z. Lin, T. Chen, L. Ao, Microbiome analysis reveals universal diagnostic biomarkers for colorectal cancer across populations and technologies, Front. Microbiol. 13 (2022) 1005201.
[45] I. Aranda-Olmedo, L.A. Rubio, Dietary legumes, intestinal microbiota, inflammation and colorectal cancer, J. Funct. Foods 64 (2020) 103707.
[46] K. Hama, Y. Fujiwara, T. Hayama, T. Ozawa, K. Nozawa, K. Matsuda, Y. Hashiguchi, K. Yokoyama, Very long-chain fatty acids are accumulated in triacylglycerol and nonesterified forms in colorectal cancer tissues, Sci. Rep. 11 (2021) 6163.
[47] T. Manzo, B.M. Prentice, K.G. Anderson, A. Raman, A. Schalck, G.S. Codreanu, C. B. Nava Lauson, S. Tiberti, A. Raimondi, M.A. Jones, M. Reyzer, B.M. Bates, J. M. Spraggins, N.H. Patterson, J.A. McLean, K. Rai, C. Tacchetti, S. Tucci, J. A. Wargo, S. Rodighiero, et al., Accumulation of long-chain fatty acids in the tumor microenvironment drives dysfunction in intrapancreatic CD8+ T cells, J. Exp. Med. 217 (2020).
[48] M.A.B. Melone, A. Valentino, S. Margarucci, U. Galderisi, A. Giordano, G. Peluso, The carnitine system and cancer metabolic plasticity, Cell Death Dis. 9 (2018) 228.
[49] T. Hirano, D.A.C. Beck, D.R. Demuth, M. Hackett, R.J. Lamont, Deep sequencing of Porphyromonas gingivalis and comparative transcriptome analysis of a LuxS mutant, Front. Cell. Infect. Microbiol. 2 (2012) 79.
[50] O. Kabil, V. Vitvitsky, R. Banerjee, Sulfur as a signaling nutrient through hydrogen sulfide, Annu. Rev. Nutr. 34 (2014) 171–205.
[51] Z. Pang, Y. Lu, G. Zhou, F. Hui, L. Xu, C. Viau, A.F. Spigelman, P.E. MacDonald, D. S. Wishart, S. Li, J. Xia, MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation, Nucleic Acids Res. gkae253 (2024).
[52] S.A.K. Rasheed, L.V. Subramanyan, W.K. Lim, U.K. Udayappan, M. Wang, P. J. Casey, The emerging roles of Gα12/13 proteins on the hallmarks of cancer in solid tumors, Oncogene 41 (2022) 147–158.
[53] D. Jacenik, E.J. Beswick, W.M. Krajewska, E.R. Prossnitz, G protein-coupled estrogen receptor in colon function, immune regulation and carcinogenesis, World J. Gastroenterol. 25 (2019) 4092–4104.
[54] N.H. Moniri, Q. Farah, Short-chain free-fatty acid G protein-coupled receptors in colon cancer, Biochem. Pharmacol. 186 (2021) 114483.
[55] V. D'Antongiovanni, M. Fornai, C. Pellegrini, L. Benvenuti, C. Blandizzi, L. Antonioli, The adenosine system at the crossroads of intestinal inflammation and neoplasia, Int. J. Mol. Sci. 21 (2020).
[56] B. Mastelic-Gavillet, B. Navarro Rodrigo, L. Décombaz, H. Wang, G. Ercolano, R. Ahmed, L.E. Lozano, A. Ianaro, L. Derré, M. Valerio, T. Tawadros, P. Jichlinski, T. Nguyen-Ngoc, D.E. Speiser, G. Verdeil, N. Gestermann, O. Dormond, L. Kandalaft, G. Coukos, C. Jandus, et al., Adenosine mediates functional and metabolic suppression of peripheral and tumor-infiltrating CD8 T cells, J. Immunother. Cancer 7 (2019) 257.
[57] D.S. Chen, I. Mellman, Oncology meets immunology: the cancer-immunity cycle, Immunity 39 (2013) 1–10.
[58] B. Chen, D. Ramazzotti, T. Heide, I. Spiteri, J. Fernandez-Mateos, C. James, L. Magnani, T.A. Graham, A. Sottoriva, Contribution of pks E coli mutations to colorectal carcinogenesis, Nat. Commun. 14 (2023) 7827.
[59] S.A. Buechler, M.T. Stephens, A.B. Hummon, K. Ludwig, E. Cannon, T.C. Carter, J. Resnick, Y. Gökmen-Polar, S.S. Badve, ColoType: a forty gene signature for consensus molecular subtyping of colorectal cancer tumors using whole-genome assay or targeted RNA-sequencing, Sci. Rep. 10 (2020) 1–13.