



OPEN

Colorectal cancer and inflammatory bowel diseases share common salivary proteomic pathways

Nicole Contran¹✉, Giorgio Arrigoni², Ilaria Battisti², Renata D'Incà³, Imerio Angriman³, Cinzia Franchin², Maria L. Scapellato⁴, Andrea Padoan¹, Stefania Moz¹, Ada Aita¹, Edoardo Savarino³, Greta Lorenzon³, Fabiana Zingone³, Gaya Spolverato³, Salvatore Pucciarelli³, Evelyn Nordi¹, Paola Galozzi¹ & Daniela Basso¹

Inflammatory bowel diseases (IBD) are high risk conditions for colorectal cancer (CRC). The discovery of IBD and CRC noninvasive protein/peptide biomarkers using saliva and feces was the aim of this study involving 20 controls, 25 IBD (12 Crohn's Disease-CD), 37 CRC. By untargeted proteomic (LTQ-Orbitrap/MS), a total of 152 proteins were identified in saliva. Absent in controls, 73 proteins were present in both IBD and CRC, being mainly related to cell-adhesion, cadherin-binding and enzyme activity regulation (g-Profiler). Among the remaining 79 proteins, 14 were highly expressed in CD and 11 in CRC. These proteins clustered in DNA replication/expression and innate/adaptive immunity. In stool, endogenous peptides from 30 different proteins were identified, two being salivary and CD-associated: Basic Proline-rich Protein 1 (PRBs) and Acidic Proline-rich Phosphoprotein. Biological effects of the PRBs-related peptides GQ-15 and GG-17 found in CD stool were evaluated using CRC cell lines. These peptides induced cell proliferation and activated Erk1/2, Akt and p38 pathways. In conclusion, the salivary proteome unveiled DNA stability and immunity clusters shared between IBD and CRC. Salivary PRB-derived peptides, enriched in CD stool, stimulate CRC cell proliferation and the pro-oncogenic RAS/RAF/MEK/ERK and PI3K/AKT/mTOR pathways suggesting a potential involvement of PRBs in IBD and cancer pathogenesis.

Colorectal cancer (CRC) and inflammatory bowel diseases (IBD) exhibit shared pathogenetic mechanisms, including chronic inflammation and altered microbiome-mucosa interaction, while lacking sensitive and specific diagnostic biomarkers. Patients with IBD appear to be at notably increased risk of CRC, estimated at 1%, 2% and 5% at 10, 20, and more than 20 years after the initial IBD diagnosis, respectively¹. This association is not unexpected, given the persistent hallmarks of cancer within the inflamed intestinal mucosa of IBD patients. These hallmarks involve chronic activation of pro-inflammatory signaling pathways, such as NF- κ B, with increased pro-inflammatory and decreased anti-inflammatory cytokines, alterations in immune cell response, including expansion of immunosuppressive T regulatory cells in the inflamed compared to non-inflamed mucosa, and effects of the altered microbiome, alongside changes in apoptotic and autophagic signaling pathways².

CRC ranks as the third most common cancer in industrialized countries and the fifth in terms of incidence and mortality worldwide^{3,4}. Despite decreasing trends observed in industrialized nations due to screening and improved treatment, developing countries are experiencing a rapid rise attributed to dietary and lifestyle⁵. Annual CRC detection using fecal occult blood (FOB) in average-risk asymptomatic subjects aged 50–75 years is recommended for reducing mortality⁶. However, FOB sensitivity and specificity remain suboptimal, even with the more advanced fecal immunochemical test (FIT) which outperforms classical guaiac-based testing (gFOBT)^{7,8}. Enhanced testing methods are crucial for preventing delayed diagnosis and unnecessary endoscopy, and for improving clinical care and resource conservation.

Regular endoscopic surveillance in IBD may facilitate early CRC detection, but compliance issues among patients can impede early CRC diagnosis rate^{1,9}. Diagnosing IBD at onset poses challenges due to its non-specific symptoms, often resembling functional intestinal disorders, hindering effective diagnostic work-up with imaging technology, endoscopy and histology. Although fecal calprotectin might support rule-out and rule-in strategies

¹Department of Medicine (DIMED), University of Padova, 35128 Padova, Italy. ²Department of Biomedical Sciences (DBS), University of Padova, 35128 Padova, Italy. ³Department of Surgery, Oncology and Gastroenterology (DiSCOG), University of Padova, 35128 Padova, Italy. ⁴Department of Cardiac, Thoracic and Vascular Sciences and Public Health, University of Padova, 35128 Padova, Italy. ✉email: nicole.contran@studenti.unipd.it

for further investigations in patients with suspected IBD, its sensitivity and specificity are limited up to 80%¹⁰. More sensitive and specific IBD and CRC biomarkers are under evaluation, primarily through genetic, epigenetic, metabolomic, proteomic, peptidomic and metagenomic studies^{11–13}.

Although serum is the preferred biological matrix, stools and saliva are also investigated. Saliva components not only reflect oral health, but may also mirror the presence of disease, including cancer, at distant sites^{14–16}. Moreover, saliva's non-invasive collection makes it ideal for repeat sampling, especially in vulnerable populations, such as children and the elderly.

This study aims to assess the proteomic and peptidomic profiles of saliva and stool to identify potential CRC and IBD biomarkers. In IBD patients stools, peptides derived from Basic salivary proline-rich protein 1 (PRB1, UniProt: P04280) were identified. These peptides were found to influence proliferation signaling pathways of colorectal cancer cells, highlighting the potential role of salivary proteins in CRC and IBD pathogenesis and diagnosis.

Results

Salivary peptidomic analysis

Peptides in biological fluids could be rapidly and reliably identified by high throughput MALDI-TOF/MS instruments, based on laser radiation and soft ionization of crystallized molecules on a plate. Since MALDI process generates almost exclusively singly charged molecules (features), their mass-to-charge ratio (m/z) corresponds to the mass. MALDI-TOF/MS was employed to analyze saliva samples, and a total of 668 salivary features were identified, of which 74 were significantly different between patient and control groups (Supplementary Table 1). Representative MALDI-TOF/MS spectra found in healthy controls (CS), Crohn's disease (CD), Ulcerative colitis (UC) and CRC are shown in Supplementary Fig. 1. These spectra depict identified peptides within the mass range of 1000 to 4000 m/z . Each peptide appears as a feature that fall on the x-axis according to its unique m/z value, while the height or intensity of these features on the y-axis corresponds to their relative abundances. The frequency of the 18 most significant ($p < 0.0005$) features in the different groups is shown in Fig. 1. Among these features, fourteen were observed with greater frequency in patients than in CS, while the remaining four exhibited converse tendencies. Notably, no specific CRC-associated features were identified, whereas peptides 1096 m/z , 1280 m/z , 1480 m/z , 1612 m/z , which were markedly rare in CS, demonstrated apparent correlations with IBD.

These data highlight the marked enrichment of salivary peptides, which differ in their profiles primarily between IBD and CS cohorts. It should be noted, however, that MALDI-TOF/MS analysis does not allow to discern whether the identified peptides originate from human or non-human sources, particularly those from the oral microbiome. To focus solely on human-derived proteins, subsequent analysis employing LTQ-Orbitrap/MS was conducted.

Salivary proteomic analysis

The salivary proteomic analysis of tryptic digested peptides was performed by LTQ-Orbitrap/MS, which returns their amino acid sequences. A bioinformatic analysis of these sequences is made to match them with the known

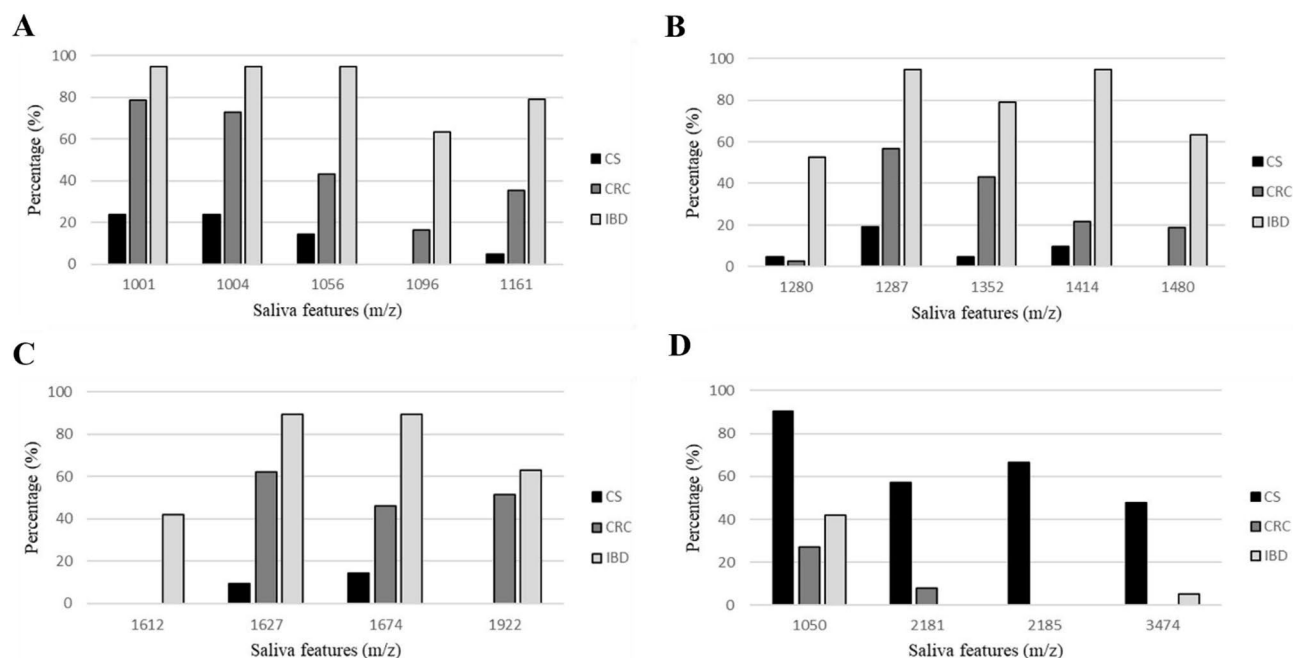


Figure 1. Most significant saliva features ($p < 0.0005$). The frequencies of positive findings among CRC and IBD patients in comparison with controls (CS) are shown. Panels (A, B, C) represent features found more frequently in CRC and IBD patients with respect to controls. Panel (D) represents features found more frequently in controls with respect to CRC and IBD patients.

human proteins. The increasing number of matching peptides covering multiple sequences of any single protein empowers its correct identification. The LTQ-Orbitrap/MS proteomic analysis of the CS, CD, UC and CRC saliva pools allowed the identification of a total of 225 proteins. To enhance the robustness, we selected those proteins identified by means of at least four peptides. This left 152 proteins (Supplementary Table 2), among which 73 were commonly found in CD, UC and CRC patients, but absent in CS. A high percentage (64%, 47/73) of these intestinal disease-associated proteins were not found in our previous salivary analysis of controls and SARS-CoV-2 infected subjects¹⁸. These 73 proteins were analyzed by *g*-Profiler software (p -value < 0.0001) to identify enriched GO-Terms corresponding to Molecular Function (GO:MF), Biological Process (GO:BP) and Cellular Components (GO:CC) (Supplementary Fig. 2). Molecular Functions were mainly related to cell adhesion and cadherin binding, inhibitor/regulator activity of peptidases and endopeptidases. The enriched Cellular Components included exosomes, vesicles, membrane bounded organelle and anchoring junctions. Enriched Biological Process were mainly related to negative regulation of enzymatic activity and to glucose and nucleic acid metabolism.

To identify which of the remaining 79/152 proteins were differently associated with diseases, the abundance ratios found in UC, CD and CRC with respect to controls were calculated and subdivided in five categories based on their belonging to the following percentiles ranges: (1) below 50th (median); (2) 51th–75th; (3) 76th–90th; (4) 91st–99th; (5) above 99th percentile. Fifty-six proteins were classified in the same category in UC, CD and CRC. On the contrary 33 proteins belonged to different categories in different diseases, as detailed in Table 1. Seven proteins were highly abundant in saliva of both CRC and IBD patients, 11 proteins were mainly increased in CRC and 14 in CD patients' saliva. One protein, Cystatin-SA, increased mainly in the two IBD types, CD and UC.

To identify whether the disease-associated proteins exhibit specific biological alterations, we selected those with an abundance in CRC and CD of at least five-fold with respect to controls ($CRC/CS \geq 5$ and $CD/CS \geq 5$) and verified the biological process they clustered in using DAVID tool (Fig. 2). For UC, only five proteins were at least five-fold higher than in controls and no cluster for them was identified.

Overall, these data provide evidence that CD, more than UC and CRC, alters the salivary proteome in agreement with the assumption that this disease is not confined to the intestine. Moreover CD, more than UC, shares with CRC overexpressed salivary proteins involved in critical carcinogenetic pathways, supporting the link between chronic inflammatory diseases and cancer.

Faecal endogenous peptides analysis

To identify potential IBD biomarkers able to distinguish CD from UC, stool endogenous peptides were analyzed by LTQ-Orbitrap LC-MS/MS. The analyses were made using the CS, CD and UC stool pools prepared using individual samples from the second cohort that appeared more enriched in protein bands at SDS-PAGE electrophoresis. The analyzed peptides were endogenous because stool pools were analyzed without previous tryptic digestion and were of human origin, since the bioinformatic analysis was carried out by referring to the human proteome with the exclusion of all non-human sequences.

This analysis revealed the presence of peptides belonging to 30 different proteins. Twenty-six proteins were selected on the basis of a number of matching peptides ≥ 4 (Supplementary Table 3). Four of the 26 proteins were detected in CD stool only and two of them were of salivary origin: Basic Salivary Proline-rich Protein 1, Salivary Acidic Proline-rich Phosphoprotein 1/2 (Fragment), Carboxypeptidase A1 and Immunoglobulin kappa constant. We focused on Basic Salivary Proline-Rich Proteins (PRBs) for which four isoforms are recognized (PRB1, UniProt: P04280; PRB2, UniProt: P02812; PRB3, UniProt: Q04118; PRB4, UniProt: P10163). In both CS and patients' saliva pools the PRB2 isoform was identified. PRB2 was not included among the detected salivary proteins (Supplementary Table 2) because only two instead of four trypsin digested peptides were identified, despite the high Peptide Spectrum Match value of 85. One of the two PRB2 trypsin digested peptides found in saliva matched exactly with the GQ-15 and partially with the GG-17 unique endogenous peptides belonging to the PRBs family found in stool (Table 2).

GQ-15 (1495.66 m/z) and GG-17 (1578.75 m/z) were synthesized to enable in vitro studies to verify whether they affect cell proliferation and the related signaling pathways Akt, Erk 1/2 and p38, using the colorectal cancer cells lines HT-29 and HCT-116. Initial exposure of HT-29 and HCT-116 cells to varying concentrations (0.1 and 100 nM) of GQ-15 and GG-17 was conducted to assess potential dose-dependent effects on cell growth. The percentage of cells/mL counted after 48 h of stimulation relative to the number of unexposed cells/mL counted at 24 h of culture was depicted (Fig. 3). Data were compared with epidermal growth factor (EGF) stimulation. Notably, differential sensitivities were observed in the two cell lines; while HCT-116 cells displayed no significant response, HT-29 cells exhibited significant dose-dependent stimulation of cell growth, particularly evident with GG-17. Subsequent experiments were performed using the higher peptides concentration at 100 nM. GG-17 emerged as a powerful stimulant of HT-29 proliferation, being even more effective than EGF. This pro-proliferative effect of GG-17 was observed both when the peptide was added alone (Fig. 4A) or in combination with EGF (Fig. 4B). Conversely, the proliferation of HCT-116 cell line remained unaffected by EGF, GQ-15, or GG-17 (Fig. 4C and 4D).

To further elucidate the underlying mechanisms, the RAS/RAF/MEK/ERK (MAPK) and PI3K/AKT/mTOR signaling pathways were evaluated via Western blot analysis of phosphorylated Erk 1/2, p38 and Akt of HT-29 and HCT-116 cells treated or untreated with GQ-15 or GG-17 peptides (Fig. 5). In HT-29 cells, both GQ-15 and GG-17, similarly to EGF, induced Erk 1/2 phosphorylation. The p38 phosphorylation was induced by the GG-17 peptide, not GQ-15, and by EGF alone or combined. Akt phosphorylation remained unaffected by the two peptides at both phosphorylation sites, whereas EGF treatment caused a reduced Akt³⁰⁸ phosphorylation. In HCT-116 cells GQ-15 induced Erk 1/2 and p38, while it reduced Akt phosphorylation; in this cell line GG-17 exerted an opposite behavior, inducing Akt but not Erk 1/2 and p38 phosphorylation. Co-stimulation with EGF

	UniProt Accession Nr	Description	DAVID GAD classification	DAVID DisGeNET classification	UC/CS	CD/CS	CRC/CS
Increased mainly in UC and CRC	A8K2U0	Alpha-2-macroglobulin-like protein 1	Neurological		33.74	14.49	24.86
	P04083	Annexin A1	Cancer, renal, cardiovascular, hematological, infection, metabolic	Neoplasm metastasis and invasiveness, prostatic neoplasm, gastric/stomach disease, lung adenocarcinoma, mouth neoplasm, squamous cell esophageal carcinoma, kidney disease, lymphatic metastasis	3.06	0.00	2.45
Increased mainly in CD and CRC	P01859	Ig gamma-2 chain C region	Immune	Idiopathic inflammatory myopathies	17.86	22.51	31.44
	P01834	Immunoglobulin kappa constant	Immune, infection	Hepatomegaly, dysgammaglobulinemia, idiopathic inflammatory myopathies	4.19	7.54	6.20
	P62805	Histone H4	Aging	Malignant breast neoplasm, prostatic neoplasm, neurodegenerative diseases	2.39	21.05	23.27
	P01040	Cystatin-A	Immune, metabolic	Neoplasm metastasis, glioblastoma, nasopharyngeal carcinoma, ichthyosis	1.03	2.07	2.43
	P02790	Hemopexin		Autoimmune diseases, non-insulin-dependent diabetes mellitus, drug toxicity, drug induced diseases, acute renal insufficiency	0.99	4.15	4.71
Increased mainly in UC and CD	P09228	Cystatin-SA	Cardiovascular, immune		2.35	5.05	1.36
Increased mainly in CRC	P01861	Ig gamma-4 chain C region		Alveolar and embryonal rhabdomyosarcoma	15.5	20.29	27.34
	P02787	Serotransferrin	Aging, immune, infection, hematological, metabolic, neurological, cardiovascular	Carcinoma, neurodegenerative diseases and others	12.08	18.06	24.23
	P05109	Protein S100-A8	Immune, renal, other	Dermatitis, leukemia	11.13	25.18	40.68
	E7EQB2	Kaliocin-1 (Fragment)	Aging, immune, infection, metabolic, neurological, renal	Malignant colon tumor, ulcerative colitis, enterocolitis, yeast and other fungal infection, others	3.78	10.11	23.93
	P59665	Neutrophil defensin 1	Infection		2.75	4.15	7.69
	U3KPS2	Myeloblastin	Cardiovascular	Myeloid leukemia	1.77	3.63	16.24
	Q8N4F0	BPI fold-containing family B member 2		Colorectal carcinoma	1.03	0.48	2.70
	P31025	Lipocalin-1	Neurological		1.02	0.33	2.89
	P08246	Neutrophil elastase	Cancer, cardiovascular, hematological	Neutropenia	0.66	1.73	7.70
	P20160	Azurocidin	Cardiovascular, infection		0.34	0.50	4.19
	P02671-2	Isoform 2 of Fibrinogen alpha chain		Fibrinogenemia, thrombophilia, kidney disease, mouth neoplasm, diabetes mellitus, colorectal and gastric carcinoma	0.00	1.75	2.57
Continued							

	UniProt Accession Nr	Description	DAVID GAD classification	DAVID DisGeNET classification	UC/CS	CD/CS	CRC/CS
Increased mainly in CD	Q02413	Desmoglein-1	Immune, other	Metabolic disease, dermatitis, keratoderma	1.94	7.95	3.69
	P01037	Cystatin-SN	Metabolic	Malignant neoplasm of stomach, prostate, hereditary diffuse gastric cancer	1.62	3.48	0.87
	P01036	Cystatin-S		Malignant breast neoplasm	1.42	2.93	0.77
	P32119	Peroxisredoxin-2	Aging, neurological	Neoplasm invasiveness, carcinomatosis, leukemia, melanoma, esophageal neoplasm, osteosarcoma, prostatic neoplasm, liver carcinoma	1.26	5.37	0.98
	P15924-2	Isoform DPII of Desmoplakin		Cardiomyopathies, skin disease	1.09	7.87	0.22
	P06702	Protein S100-A9	Immune, renal	Malignant colonic tumor, celiac disease	1.01	28.05	10.89
	Q86YZ3	Hornerin	Immune		0.56	12.70	0.33
	Q5D862	Filaggrin-2	Immune, hematological		0.44	2.55	0.00
	Q08380	Galectin-3-binding protein	Cancer	Liver cirrhosis and fibrosis	0.20	2.58	1.53
	P69905	Hemoglobin subunit alpha	Immune, hematological, cardiovascular, infection, metabolic, other	Carcinomatosis, non-insulin dependent diabetes mellitus, thalassemia, anemia, hemoglobinopathies polycythemia, erythrocytosis,	0.13	20.12	0.06
	P02042	Hemoglobin subunit delta	Cardiovascular, hematological	Thalassemia	0.12	14.66	0.05
	P69891	Hemoglobin subunit gamma-1	Hematological	Hemoglobin F disease, Parkinson disease, prostatic neoplasm, thalassemia	0.09	3.02	0.03
	P68871	Hemoglobin subunit beta		Anemia, hemoglobinopathies, thalassemia	0.09	9.70	0.03
	Q06830	Peroxisredoxin-1	Aging, infection, neurological	Non-small cell lung carcinoma, hepatoma, esophageal, liver and lung neoplasm	1.15	5.02	2.08

Table 1. List of 33 salivary proteins with differences in abundance in UC, CD and CRC with respect to CS. The abundance ratios (UC/CS, CD/CS and CRC/CS) are reported. The DAVID (Database for Annotation, Visualization and Integrated Discovery) functional annotation tool was used to define the disease annotation category DisGeNET and/or GAD (Gene-Disease Associations) classification.

modulated the signaling effects of both GQ-15 and GG-17 peptides, reducing the effect of GQ-15 while enhancing those of GG-17 on Erk 1/2, and amplifying the effect of GQ-15 on Akt.

Peptides derived from salivary PRBs detectable in stool are biologically active, able to stimulate colorectal cancer cell growth and the key pro-oncogenic signaling pathways, namely RAS/RAF/MEK/ERK (MAPK) and PI3K/AKT/mTOR.

Discussion

The primary objective of this study was to identify novel non-invasive biomarkers potentially applicable to the diagnosis of CRC and IBD by analyzing salivary and fecal proteins/peptides. Fecal samples are already used as diagnostic tool for CRC and IBD via fecal occult blood and calprotectin assays. The selection of saliva as a potential source for biomarker exploration stemmed from its non-invasive collection process, which can be self-collected, thereby fostering patients' compliance¹⁹. Moreover, in contrast to serum or plasma, saliva presents a lower concentration of abundant proteins like albumin and immunoglobulins, that might mask less abundant biomarkers and thus facilitating the proteomic analysis²⁰. Saliva has been extensively investigated in various cancer contexts, including breast cancer, culminating in the concept of "salivaomics" encompassing proteomics, metabolomics, and transcriptomics, yielding promising outcomes¹⁹. Saliva, being continuously generated and traversing the entire gastrointestinal tract²¹, might harbor components pertinent to the initiation or progression of gastrointestinal tract inflammation and/or cancer.

In the present study, a substantial abundance of salivary peptides were identified by MALDI-TOF/MS in both patients and controls, in agreement with findings previously reported by Tong et al.²². Nevertheless, within these highly enriched peptides profiling, no discernible feature specifically associated with CRC was delineated.

Proteomic analysis by LTQ-Orbitrap/MS successfully identified a high number of human proteins in saliva, as expected²⁰. Non-human proteins, from the microbiome, were excluded from the bioinformatic analysis. A high number of proteins (n = 73) were correlated with colorectal diseases, being absent in controls, but present in IBD and CRC saliva. This correlation was further supported by the noteworthy observation that 64% (47/73) of these proteins were absent in saliva from a cohort of previously investigated patients devoid of intestinal diseases¹⁸. Notable proteins linked to colorectal diseases encompassed Annexins, Heat shock proteins, Histones

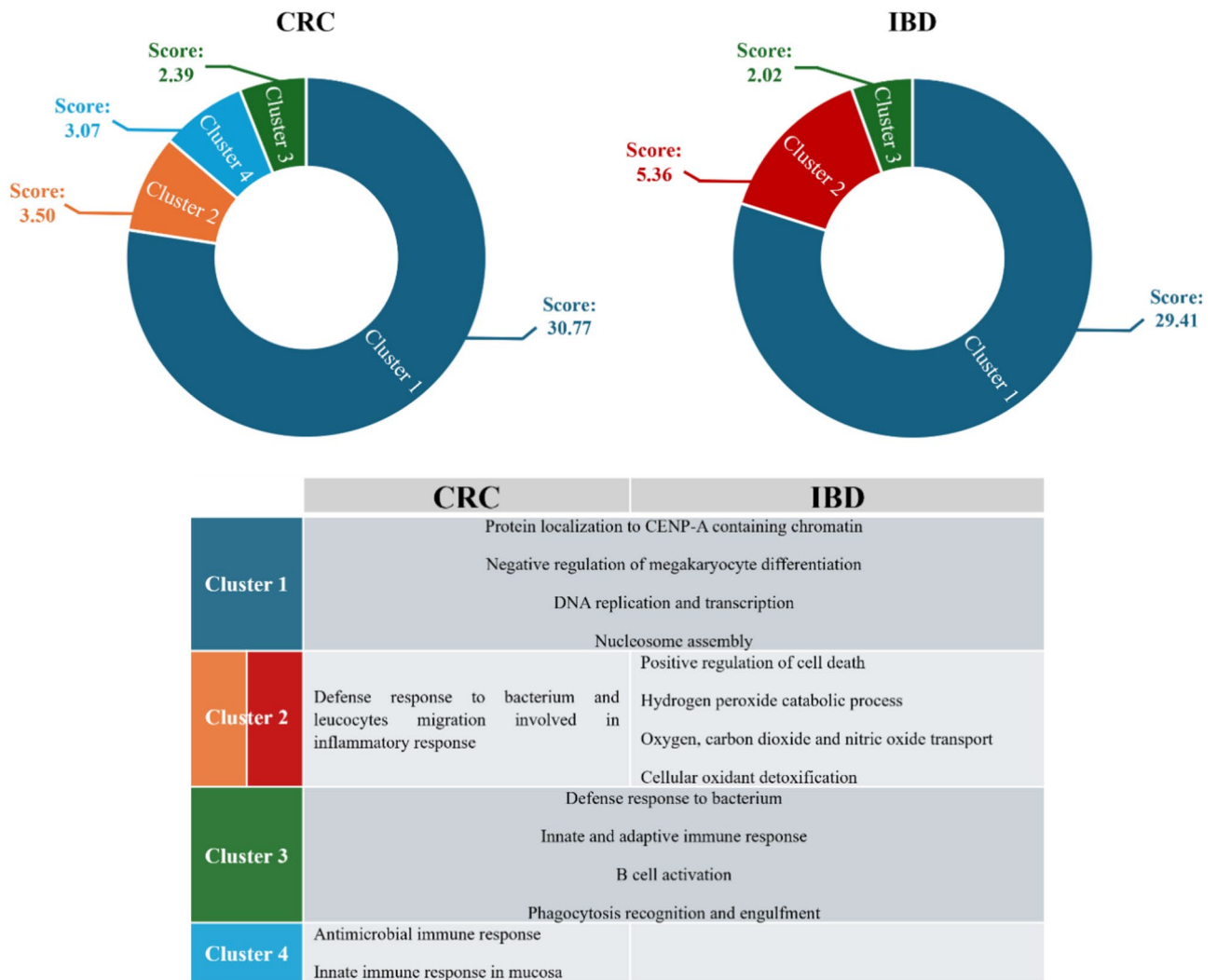


Figure 2. Clusters of biological processes in proteins highly represented in CRC with respect to controls (Left) and highly represented in CD with respect to controls (Right), using DAVID tool.

Biological matrix	Salivary PRB2 tryptic peptides	Stool Endogenous peptides
Saliva PRB2 peptide 2	SPPGKPQGPPPQGGNQPQGGPPPPGKPKQGGPPQGGNK	
Stool PRB1 GQ-15		<u>GNQPQGGPPPPGKPKQ</u>
Stool PRB1 GG-17		GPPPQGGNKPKQGGPPPG

Table 2. Basic salivary proline-rich protein 2 (PRB2, UniProt: P02812) unique peptide 2 identified in saliva and stool. Salivary peptides were obtained after trypsin digestion. Stool peptides PRB1 GQ-15 (underlined italics) and GG-17 (bold) were identified in Crohn’s disease as endogenous peptides. The saliva PRB2 peptide 2 matched with both PRB1 GQ-15 (underlined italics) and with PRB1 GG-17 (bold).

and glycolytic enzymes. These findings are congruent with the documented association with glycolytic process, stress response, and cellular metabolism, as elucidated through the GO-Terms analysis of the biological processes (see Suppl. Figure 2), and further supported by the fact that aerobic glycolysis is a hallmark of cancer²³. Among the identified salivary proteins, a subset of 79 found in both patients and controls, varied in abundance. Remarkably, within this subset, certain proteins exhibited significantly elevated levels in CRC cases. Notably, some of these proteins, including BPI-fold containing family B member 2 (BPIFB2; UniProt: Q8N4F0) and Kaliocin-1 (Fragment, UniProt: E7EQB2) have been previously implicated in CRC. Namely, BPIFB2 protein, highly enriched in salivary glands, demonstrates expression across different tumor types, including colorectal cancer (www.proteinatlas.org/ENSG00000078898-PIFB2/pathology/colorectal+cancer), and Kaliocin-1, also known as Lactotransferrin-1 (LTF), produced by exocrine glands, exerts an anti-bacterial and anti-cancer activity through cell cycle arrest, induction of apoptosis, immune modulation and inhibition of migration and invasiveness²⁴.

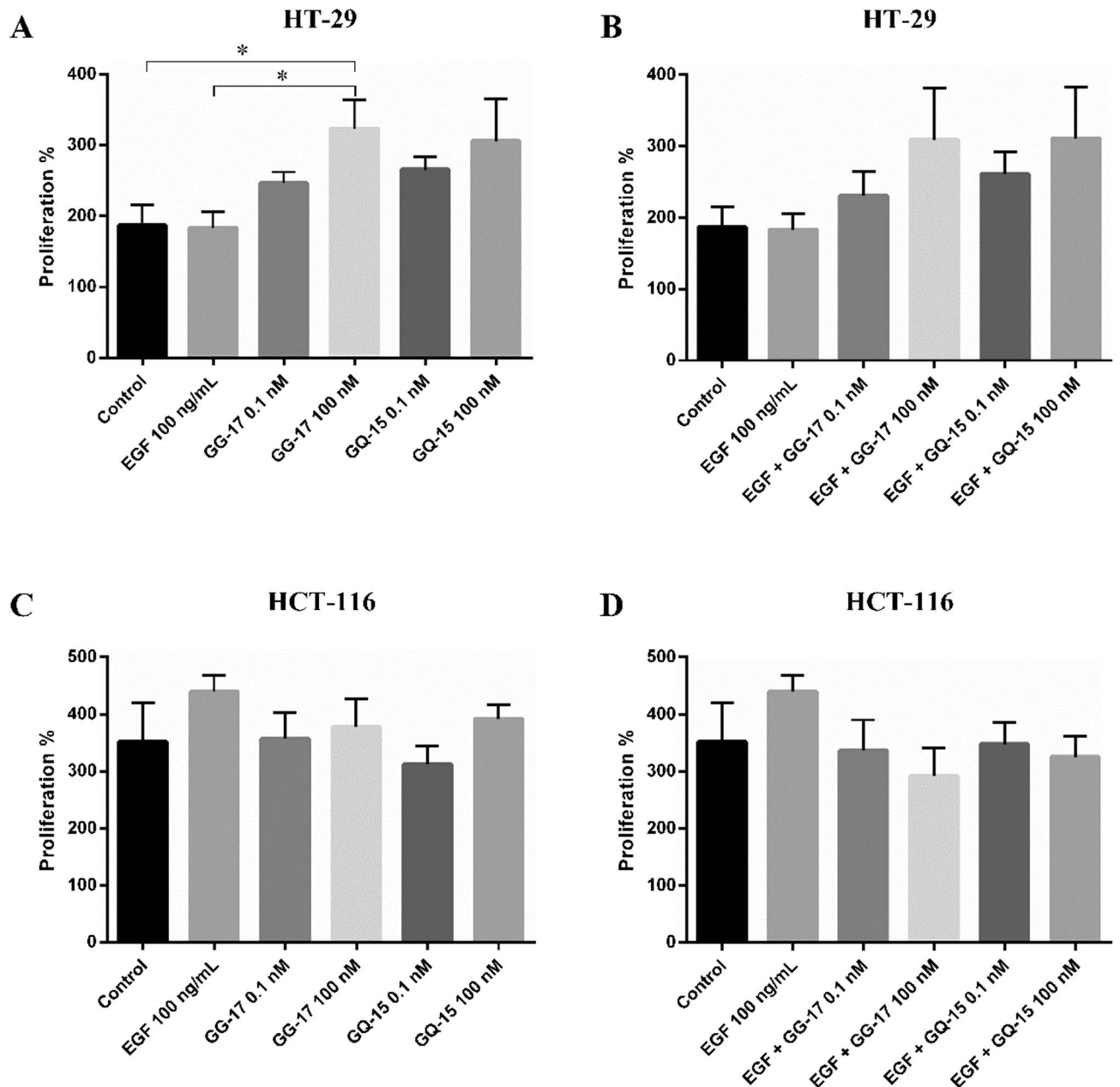


Figure 3. Dose-dependent proliferation experiments with HT-29 (A and B) and HCT-116 (C and D) cell lines. Cells were stimulated with EGF 100 ng/mL, GQ-15 at 0.1 nM and 100 nM, and GG-17 at 0.1 nM and 100 nM synthetic peptides both added alone (panels A and C) or combined (panels B and D) and counted after 24, 48 and 72 h. The percentage of cells/mL counted after 48 h of stimulation relative to the number of unexposed cells/mL counted at 24 h of culture is shown. One-way ANOVA: Panel A, $F = 5.163$, $p = 0.010$; Panel B, $F = 0.840$, $p = 0.484$; Panel C, $F = 0.642$, $p = 0.597$; Panel D, $F = 1.441$, $p = 0.258$. * = $p < 0.05$ by Tukey's multiple comparison test.

Overall, the analysis revealed that proteins associated with colorectal cancer predominantly clustered within processes involved in DNA stability maintenance. However, this main cluster was not CRC exclusive, as it also overlapped with the main protein cluster associated with CD. Additionally, minor clusters of inflammation, innate and adaptive immunity exhibited shared characteristics between CRC and CD. These results imply shared biological alterations, mainly inflammation, immunity, and control of DNA replication between cancer and chronic inflammatory disorders. Such findings are in line with those reported by Zheng et al.²⁵, who analyzed salivary exosomal proteins in IBD and identified altered processes encompassing inflammation, innate and adaptive immune response, DNA binding and proteasomes activity when compared to controls. Two CD-associated proteins identified in this study, Peroxiredoxin-1 and Peroxiredoxin-2, appear of interest for their primarily involvement in oxidation processes. A differential expression of proteins involved in oxidation/reduction have been documented among IBD patients, including an upregulation of Peroxiredoxin-2 in blood mononuclear cells,

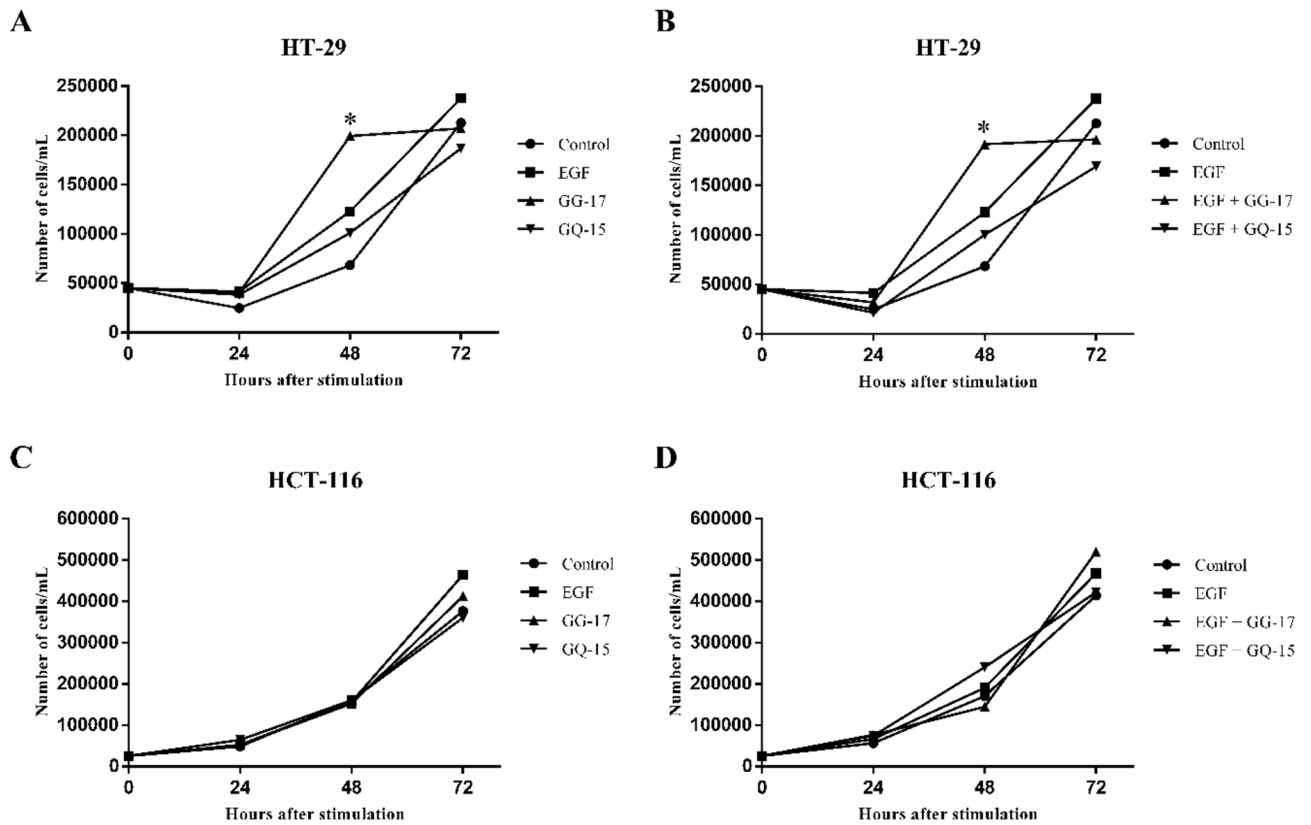


Figure 4. Proliferation experiments with HT-29 (A and B) and HCT-116 (C and D) human colorectal cancer cell lines. Cells were stimulated with EGF 100 ng/mL, GQ-15 100 nM and GG-17 100 nM synthetic peptides both added alone (panels A and C) or combined (Panels B and D) and counted after 24, 48 and 72 h. Repeated measures analysis of variance: (A) time $F = 51.63$, $p < 0.0001$; treatment: $F = 1.54$, $p = 0.221$; interaction $F = 1.63$, $p = 0.115$. (B) Time $F = 66.33$, $p < 0.0001$; treatment: $F = 2.16$, $p = 0.109$; interaction $F = 2.22$, $p = 0.026$. (C) Time $F = 141.60$, $p < 0.0001$; treatment: $F = 0.042$, $p = 0.988$; interaction $F = 0.297$, $p = 0.974$. (D) Time $F = 126.80$, $p < 0.0001$; treatment: $F = 0.088$, $p = 0.966$; interaction $F = 0.311$, $p = 0.970$. Tukey's multiple comparison test: * $p < 0.001$ with respect to control (panel A and B) and $p < 0.05$ with respect to GQ-15 (panel A) or EGF + GQ-15 (panel B) at 48 h. Each point and line report mean values of ten replicated measures. SD for graphical clarity is not shown.

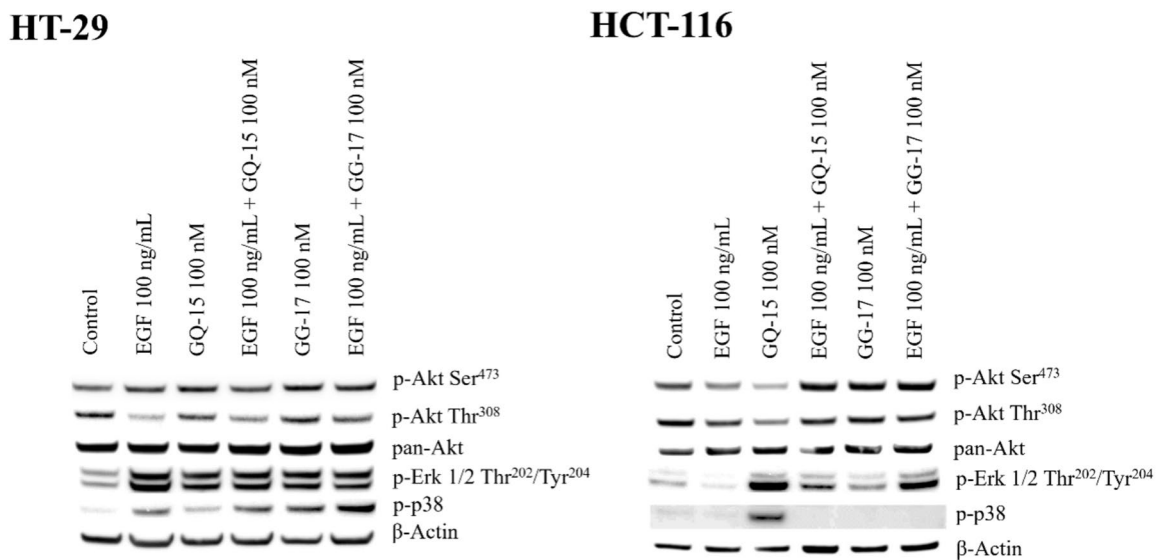


Figure 5. Western blot analyses of HT-29 and HCT-116 cell lines. Both cell lines were stimulated with EGF 100 ng/mL, GQ-15 100 nM, GQ-15 100 nM + EGF 100 ng/mL, GG-17 100 nM, GG-17 100 nM + EGF 100 ng/mL. Proliferation-related targets Akt, Erk 1/2 and p38 are shown. The results of any target were cropped from different blots. Full-length blots are included in the Supplementary Information file.

which imply a potential migration of these cells to the oral mucosa, as indicated by Hatsugai et al.²⁶. Moreover, heightened antioxidant activity alongside increased oxidative damage to proteins and lipids has been described in the saliva of patients with *BRCA1* mutations and breast cancer, suggesting a potential association between altered oxidative-antioxidative balance and oral manifestations²⁷.

Furthermore, a comprehensive assessment of proline-rich proteins (PRPs), constituting the primary components of human saliva²⁸, was undertaken. This evaluation was prompted by the indication of endogenous peptides derived from basic salivary proline-rich protein 1 (PRB1) in stool samples from CD patients. The presence of PRB1 peptides in stool may stem from salivary protein fragmentation during transit through the gastrointestinal tract. However, an alternative possibility exists wherein these peptides originate directly from saliva and traverse the gastrointestinal tract without undergoing further degradation. This hypothesis is supported by the observation that proteolysis of PRPs can occur during or prior to granule maturation, implying that resulting peptides are inherent constituents of saliva²⁹. An increase in PRPs-derived peptides, observed in the saliva of patients with different diseases such as asthma³⁰, appears to exert significant biological effects. Pinto da Costa et al. demonstrated that a PRPs-derived peptide (GPPPQGGPRQG) compromises the stabilization of adherent junctions, such as E-cadherin, thereby inducing apoptosis³¹. Interestingly, this peptide sequence shared the first seven amino acids with the GG-17 peptide identified by us in the stool of CD patients. The behavior of PRPs in the context of CRC and IBD in saliva was further investigated. Basic salivary proline-rich protein 2 was found to be less abundant in IBD and CRC patients than in controls, suggesting an elevated level of degradation and peptide release within the oral cavity of patients with intestinal diseases. In agreement, lower levels of a salivary peptide at 1226.7 m/z from PRPs have been described in patients undergoing hemodialysis for chronic kidney disease compared to controls²².

To assess the potential biological effects of salivary-derived peptides, a series of in vitro studies were conducted by stimulating two human colorectal cancer cell lines with the PRBs-derived peptides, specifically GQ-15 and GG-17. Notably, the HT-29 cell line exhibited heightened responsiveness to GG-17 stimulation, eliciting dose-dependent cell proliferation. Furthermore, GG-17 induced phosphorylation of Erk 1/2 and p38, critical downstream targets of the pro-oncogenic RAS/RAF/MEK/ERK (MAPK) signaling pathway. GQ-15, similarly to GG-17 and EGF, triggered Erk 1/2, but not p38, phosphorylation, but it did not stimulate cell growth. In contrast, the response of HCT-116 cells to PRPs-derived peptides and EGF stimulation differed significantly from that of HT-29. Proliferation remained unaffected by any molecule or combination thereof. Moreover, while GQ-15, mainly if combined with EGF, induced Erk 1/2 and p38 phosphorylation, GG-17 solely activated Erk 1/2 without impacting p38. Moreover, PRBs peptides induced Akt phosphorylation at both 308 and 473 sites in HCT-116 cells, particularly in combination with EGF. The different behaviors observed in these two cell lines might depend on their distinct mutational statuses: HT-29 is wild-type for *KRAS* and harbors the p.R273H *p53* mutation, while HCT-116 carries the *KRAS* codon 13 mutation and is wild-type for *p53*³². Furthermore, the different sensitivity of these two cell lines to EGF stimulation might also be influenced by their differential expression levels of the EGF receptor, reported to be lower in HCT-116 compared to HT-29 cells³³.

These findings indicate that PRBs naturally derived salivary peptides have the potential to traverse the intestinal mucosa and stimulate cancer cell growth by activating pro-proliferative signaling pathways.

In conclusion, our examination of the salivary proteome unveiled a cluster of proteins shared between CRC and IBD, involved in DNA stability maintenance, innate, and adaptive immunity. Notably, reduced PRPs saliva levels in CRC and IBD coincide with the enrichment of two PRB-derived peptides are enriched in CD stool. These peptides stimulate in vitro CRC cell proliferation and activate signaling pathways involving Erk1/2, Akt and p38, thus suggesting a potential involvement of PRPs in the pathogenesis of IBD and cancer.

Methods

Studied cohorts

Two different cohorts of subjects were enrolled. The first cohort contributed saliva samples for peptidomic and proteomic analysis, while the second cohort provided stool samples for proteomic examination.

The first cohort included: 20 healthy controls (CS, 10 females, 10 males; mean age \pm SD: 40.0 \pm 10.2 years), 25 IBD (12 Crohn's Disease, CD; 13 Ulcerative Colitis, UC; 10 females, 15 males; mean age \pm SD: 47 \pm 12.9 years) and 37 CRC (20 females and 17 males; mean age \pm SD: 66.0 \pm 14.7 years). The second cohort included: 51 CS (19 females, 32 males; mean age \pm SD: 50.0 \pm 8.7 years) and 151 IBD (94 CD and 57 UC; 61 females, 90 males; mean age \pm SD: 48.1 \pm 15.8 years).

CS were healthcare personnel undergoing routine visits, without any sign of intestinal symptoms or a history of intestinal disease. CRC and IBD patients attended the Surgical Clinic and the Gastroenterology Unit of the University-Hospital of Padova.

This work was approved by our local ethics committee (University-Hospital of Padova, Prot. N. 3756/AO/16) and a fully informed consent was obtained in writing from all participants. The procedures used in this study adhere to the tenets of the Declaration of Helsinki. All methods took place in accordance with relevant guidelines and regulations.

Saliva and stool collection and processing

Saliva samples were collected in the morning before breakfast using Salivette® (Sarstedt AG & Co., Germany), centrifuged for 5 min at 4500 rpm and stored at -80 °C. Fecal samples, weighing 100 mg, were collected from the first bowel movement of the day, diluted in 100 μ L of ultrapure water, centrifuged for 15 min at 20,000 g, and the supernatants stored at -80 °C¹⁷.

Saliva MALDI-TOF/MS peptidomic analysis

Two hundred μL of saliva were prepared as detailed in Supplementary methods to perform peptidomic MALDI-TOF/MS analyses on a 4800 Plus MALDI-TOF/TOF Analyzer (AB Sciex, Canada) in reflector mode. Using MALDI-TOF/MS, crystallized samples are irradiated through pulsing laser beam that induces soft ionization of the molecules. The generated ions are accelerated in the flight tube under vacuum to reach the detector. The length of time to pass through the flight tube (Time of flight) is inversely correlated with the molecular mass of the ionized molecules and the generated ion signals are expressed as mass to charge ratio (m/z) with intensity correlated with abundance. Since generated ions are primarily single protonated molecules, the multiple charged molecules being very low in abundance, the m/z features of MALDI-TOF/MS spectra should be regarded as molecular weights³⁴.

Saliva and Stool LTQ-Orbitrap LC–MS/MS proteomic analysis

Saliva and fecal samples were subjected to SDS-PAGE electrophoresis to select samples which appeared more enriched in protein bands to prepare pools as detailed in Supplementary methods. Eight individual saliva samples from any patients' and control groups were selected and pooled to obtain the CS pool (0.5 $\mu\text{g}/\mu\text{L}$ final protein concentration), the UC pool (1.5 $\mu\text{g}/\mu\text{L}$), the CD pool (3.0 $\mu\text{g}/\mu\text{L}$) and the CRC pool (3.0 $\mu\text{g}/\mu\text{L}$). Similarly, three pools of fecal samples were prepared: the CS pool made with 6 samples (5 $\mu\text{g}/\mu\text{L}$), the UC pool made of 5 samples (4.4 $\mu\text{g}/\mu\text{L}$) and the CD pool made of 5 samples (4.6 $\mu\text{g}/\mu\text{L}$).

The proteomic analysis of saliva was performed using SDS-PAGE slices subjected to trypsin in-gel digestion. Stool endogenous peptides were directly extracted from stool pools following the FASP protocol for protein digestion. Both tryptic salivary peptides and stool endogenous peptides were analyzed with an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, US) after they have been treated as detailed in Supplementary methods. Data were analyzed with the Proteome Discoverer software (version 1.4, Thermo Fisher Scientific, US) interfaced to a Mascot server (version 2.2.4, Matrix Science, UK) and searched against the human section of the Uniprot Database (www.uniprot.org, version 20,150,401, 90,411 sequences), excluding all non-human proteins. Consequently, peptides matching with proteins of animal, plant, microbial and other non-human origins did not enter the analysis.

Peptides synthesis

The Basic salivary proline-rich protein 1 (UniProt: P04280) peptides GQ-15 (NH₂-GNQPQGPPPPGKPKQ-COOH, molecular weight: 1495.66 m/z) and GG-17 (NH₂-GPPPQGGNKPKQGPPPPG-COOH, molecular weight: 1578.75 m/z) were synthesized by Primm srl (Italy). Lyophilized powder peptides were resuspended in 2 mL of PBS 1X (concentration of 2 mg/mL). To avoid freeze–thaw cycles, 20 μL of aliquots were stored at $-80\text{ }^\circ\text{C}$.

In vitro cell proliferation and signaling experiments

Two human CRC cell lines, HT-29 and HCT-116, were used to study cell proliferation (Automated Cell Counter, Life Technologies, US) and the proliferation signaling pathways p38, Erk 1/2 and Akt by western blot analyses as detailed in Supplementary methods. Cells were left unstimulated or were stimulated with a positive control (EGF 100 ng/mL) or with the GQ-15 100 nM and GG-17 100 nM peptides.

Statistical analysis

Statistical analyses were conducted utilizing STATA (US), and R software packages for statistical computing (US). For the evaluation of saliva peptidomic features, Fisher's exact test was used with Benjamini-Hockberg adjusted p -value. Differences between groups were assessed using the Z-test. Proliferation experiments were subjected to analysis using two-way repeated measures analysis of variance (ANOVA), with a significance threshold set at $p < 0.05$.

All methods were carried out in accordance with relevant guidelines.

Informed consent was obtained from all participants (University-Hospital of Padova, Prot. N. 3756/AO/16).

Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Received: 9 June 2023; Accepted: 22 July 2024

Published online: 31 July 2024

References

- Faye, A. S., Holmer, A. K. & Axelrad, J. E. Cancer in inflammatory bowel disease. *Gastroenterol. Clin. N. Am.* **51**, 649–666 (2022).
- Shao, B. Z. *et al.* The role of autophagy in inflammatory bowel disease. *Front. Physiol.* **12**, 621132 (2021).
- Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J. Clin.* **74**, 12–49 (2024).
- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Keum, N. & Giovannucci, E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).
- US Preventive Services Task Force *et al.* Screening for colorectal cancer: US preventive services task force recommendation statement. *JAMA* **325**, 1965–1977 (2021).
- de Wijkerslooth, T. R. *et al.* Immunochemical fecal occult blood testing is equally sensitive for proximal and distal advanced neoplasia. *Am. J. Gastroenterol.* **107**, 1570–1578 (2012).

8. Lee, J. K., Liles, E. G., Bent, S., Levin, T. R. & Corley, D. A. Accuracy of fecal immunochemical tests for colorectal cancer: Systematic review and meta-analysis. *Ann. Intern. Med.* **160**, 171 (2014).
9. Woo, D. H. *et al.* Predictors and clinical outcomes of follow-up loss in patients with inflammatory bowel disease. *J. Gastroenterol. Hepatol.* **33**, 1834–1838 (2018).
10. Lin, J. F. *et al.* Meta-analysis: Fecal calprotectin for assessment of inflammatory bowel disease activity. *Inflamm. Bowel. Dis.* **20**, 1407–1415 (2014).
11. Dragomir, M. P., Kopetz, S., Ajani, J. A. & Calin, G. A. Non-coding RNAs in GI cancers: From cancer hallmarks to clinical utility. *Gut.* **69**, 748–763 (2020).
12. Grady, W. M., Yu, M. & Markowitz, S. D. Epigenetic alterations in the gastrointestinal tract: Current and emerging use for biomarkers of cancer. *Gastroenterology* **160**, 690–709 (2021).
13. Bosch, L. J. W. *et al.* Novel stool-based protein biomarkers for improved colorectal cancer screening: A case-control study. *Ann. Intern. Med.* **167**, 855–866 (2017).
14. Koopaie, M., Kolahdooz, S., Fatahzadeh, M. & Manifar, S. Salivary biomarkers in breast cancer diagnosis: A systematic review and diagnostic meta-analysis. *Cancer Med.* **11**, 2644–2661 (2022).
15. Machida, T. *et al.* miR-1246 and miR-4644 in salivary exosome as potential biomarkers for pancreatobiliary tract cancer. *Oncol. Rep.* **36**, 2375–2381 (2016).
16. Mascitti, M. *et al.* Beyond head and neck cancer: The relationship between oral microbiota and tumour development in distant organs. *Front. Cell Infect. Microbiol.* **9**, 232 (2019).
17. Basso, D. *et al.* Peptidomic and proteomic analysis of stool for diagnosing IBD and deciphering disease pathogenesis. *Clin. Chem. Lab. Med.* **58**, 968–979 (2020).
18. Aita, A. *et al.* Salivary proteomic analysis in asymptomatic and symptomatic SARS-CoV-2 infection: Innate immunity, taste perception and FABP5 proteins make the difference. *Clin. Chim. Acta.* **537**, 26–37 (2022).
19. Kaczor-Urbanowicz, K. E. *et al.* Emerging technologies for salivaomics in cancer detection. *J. Cell. Mol. Med.* **21**, 640–647 (2017).
20. Loo, J. A., Yan, W., Ramachandran, P. & Wong, D. T. Comparative human salivary and plasma proteomes. *J. Dent. Res.* **89**, 1016–1023 (2010).
21. Marcuello, M. *et al.* Circulating biomarkers for early detection and clinical management of colorectal cancer. *Mol. Aspects Med.* **9**, 107–122 (2019).
22. Tong, P. *et al.* Identification of salivary peptidomic biomarkers in chronic kidney disease patients undergoing haemodialysis. *Clin. Chim. Acta.* **489**, 154–161 (2019).
23. Pavlova, N. N. & Thompson, C. B. The emerging hallmarks of cancer metabolism. *Cell. Metab.* **23**, 27–47 (2016).
24. Hassan, M. I., Waheed, A., Yadav, S., Singh, T. P. & Ahmad, F. Prolactin inducible protein in cancer, fertility and immunoregulation: Structure, function and its clinical implications. *Cell. Mol. Life Sci.* **66**, 447–459 (2009).
25. Zheng, X. *et al.* Salivary exosomal PSMA7: A promising biomarker of inflammatory bowel disease. *Protein Cell.* **8**, 686–695 (2017).
26. Hatsugai, M. *et al.* Protein profiles of peripheral blood mononuclear cells are useful for differential diagnosis of ulcerative colitis and Crohn's disease. *J. Gastroenterol.* **45**, 488–500 (2010).
27. Sawczuk, B. *et al.* Salivary gland function, antioxidant defence and oxidative damage in the saliva of patients with breast cancer: Does the BRCA1 mutation disturb the salivary redox profile?. *Cancers (Basel)* **11**, 1501 (2019).
28. Bennick, A. Salivary proline-rich proteins. *Mol. Cell. Biochem.* **45**, 83–99 (1982).
29. Vitorino, R. *et al.* Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics* **10**, 3732–3742 (2010).
30. Chen, F. *et al.* Association of increased basic salivary proline-rich protein 1 levels in induced sputum with type 2-high asthma. *Immun. Inflamm. Dis.* **10**, 602 (2022).
31. da Costa, J. P. *et al.* Anti-tumoral activity of human salivary peptides. *Peptides.* **71**, 170–178 (2015).
32. Ahmed, D. *et al.* Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis.* **2**, 71 (2013).
33. Unson, S. *et al.* Heteronemin and tetrac induce anti-proliferation by blocking EGFR-mediated signaling in colorectal cancer cells. *Mar Drugs.* **20**, 482 (2022).
34. Duncan, M. W. *et al.* Applications of MALDI mass spectrometry in clinical chemistry. *Clin. Chem.* **62**, 134–143 (2016).

Disclaimer

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Author contributions

Study concept and design: N.C. and D.B.; Conceived the experiments: N.C., G.A. and D.B.; Statistics and prepare figures and tables: N.C., I.B., C.F. and A.P.; Conducted the experiments: N.C., I.B., C.F., S.M., A.A., E.N. and P.G.; Drafting of the manuscript: N.C. and D.B.; Analyzed the results: N.C., G.A. and D.B.; Cohorts, patient characterization, data collection: R.D.I., I.A., M.L.S., E.S., G.L., F.Z., G.S. and S.P.; All authors reviewed the manuscript and approved the final version.

Funding

PNRR-POC-22-12375750, Diagnostic innovation for inflammatory bowel diseases in children and adults: non invasive stool and salivary testing with ecologic new methods and biomaterials (CUP I93C22000590006), Italian Ministry of Health. Funded by the European Union—NextGenerationEU – NRRP; M6C2—Investment: 2.1 “Enhancement and strengthening of biomedical research within the NHS”.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68400-z>.

Correspondence and requests for materials should be addressed to N.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024