Identification of Potential Urinary Protein Biomarkers in Colorectal Cancer: A Pilot Study Using a Proteomic Approach

Bruno Augusto Alves Martins1, Raphaela Menezes de Oliveira2, Mickaella Michelson Martins2, Paulo Gonçalves de Oliveira1, Aline Maria Araújo Martins3

1 Department of Colorectal Surgery, School of Medicine, University of Brasilia, Brasilia, Brazil
2 Cell Biology Department, Institute of Biological Sciences, University of Brasilia, Brasilia, Brazil
3 Medical Science Graduate Program, School of Medicine, University of Brasilia, Brasilia, Brazil

Address for correspondence Bruno Augusto Alves Martins, MSc, SQN 212, block B, ap 205, Brasilia, Federal District, 70864-020, Brazil (e-mail: brunomartins.coloprocto@gmail.com).


Graphical Abstract

Control

CRC

Underwent colonoscopy
No abnormalities

CTRL

CRC

Protein quantification
Qubit

Underwent CRC surgery

Mass spectrometry-based proteomics
(uHPLC- MSMS Orbitrap)

m/z

Function Proteomics
Gene Ontology Analysis (GO)

Received
March 13, 2023


Accepted after revision
June 21, 2023

ISSN 2237-9363.

© 2023. Sociedade Brasileira de Coloproctologia. All rights reserved.

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil
Colorectal cancer (CRC) is the third most diagnosed malignancy worldwide, and it is also the second leading cause of cancer-related deaths. Despite recent progress in screening programs, noninvasive accurate biomarkers are still needed in the CRC field. In this study, we evaluated and compared the urinary proteomic profiles of patients with colorectal adenocarcinoma and patients without cancer, aiming to identify potential biomarker proteins. Urine samples were collected from 9 patients with CRC and 9 patients with normal colonoscopy results. Mass spectrometry (label-free LC—MS/MS) was used to characterize the proteomic profile of the groups. Ten proteins that were differentially regulated were identified between patients in the experimental group and in the control group, with statistical significance with a p value ≤ 0.05. The only protein that presented upregulation in the CRC group was beta-2-microglobulin (B2M). Subsequent studies are needed to evaluate patients through different analysis approaches to independently verify and validate these biomarker candidates in a larger cohort sample.

In this study, we evaluated and compared the urinary proteomic profiles of patients with colorectal adenocarcinoma and patients without cancer, aiming to identify potential biomarker proteins.

**Material and Methods**

**Patients and Study Design**

This study is characterized as a pilot, prospective and translational experiment in a real-life context. Individuals who were referred to the Department of Colorectal Surgery of Brasilia University Hospital for colonoscopy or for surgical treatment of colorectal adenocarcinomas were invited to participate in the study. All samples were collected from March 2018 to December 2018.

Patients who met the following inclusion criteria were enrolled in the CRC group: age between 20 and 90 years and histologically confirmed diagnosis of colorectal adenocarcinoma. The healthy control group comprised individuals who were asymptomatic and had no evidence of neoplasms at their screening colonoscopy.

Patients with at least one of the following characteristics were excluded from enrolment: individuals with inflammatory bowel disease, confirmed diagnosis or clinical suspicion of genetic colorectal cancer syndromes (familial adenomatous polyposis, hereditary nonpolyposis colorectal cancer or other hereditary CRC syndrome), patients who underwent chemotherapy and/or neoadjuvant radiotherapy and those with synchronous and metachronous colonic tumors.

Study protocols and procedures were approved by the ethics committee (protocol number 83200917.9.0000.5558), and analyses were carried out in agreement with the Declaration of Helsinki. Written informed consent was obtained from all the participants.

**Sample Collection**

For the CRC group, urine samples were collected on the day of surgery after urinary catheterization (50 ml). In the nonpathological control group, midstream urine samples
(50 ml) were collected on the day of colonoscopy. Samples were stored as 10 ml aliquots in Falcon tubes at −80 °C until assayed within 1 hour of sample withdrawal.

Sample Preparation
The quantification of the protein concentration of the sample was performed from fluorescence detection by the Qubit™ technique (Invitrogen). First, the fluorescent reagent was prepared from the junction between fluorophore and buffer solutions at a ratio of 1:200. Then, 5 μl aliquots of sample diluted in Milli-Q® water (1:3) were added to the reagent to reach a final volume of 200 μl. The mixtures were incubated for 15 min each, sequentially, and then analyzed in portable Qubit™ equipment. After reading each sample, the final concentration was obtained by multiplying the reading value by the chosen dilution factor.

Thirty micrograms of protein were aliquoted into low adsorption Eppendorf tubes for proteins and peptides (LoBind) for the digestion process. Lyophilized in a SpeedVac™ SC100 rotary concentrator (Savant™), the protein extracts were initially resuspended in a solution of 20 mM TEAB, 8 M urea and 50 mM DTT (pH 7.9) and incubated for 25 min at 55 °C and 400 rpm. Under shelter from light and after cooling, enough IAA solution was added to reach a final concentration of 14 mM and incubated again for 40 min at 21 °C and 400 rpm. Subsequently, a DTT solution at a final concentration of 5 mM was added to stop the alkylation reaction.

The samples were then diluted 1:5 with a 20 mM TEAB solution (pH 7.9), as trypsin enzyme is intolerant to high urea concentrations, by adding a CaCl2 solution sufficient to reach a final concentration of 1 mM and trypsin (Promega) in a 1:50 ratio. Afterwards, the samples were incubated for 13 hours at 37 °C and 300 rpm, and after the digestion period, TFA was added at a final concentration of 1% to prevent nonspecific cleavages and to stop the reaction. Protein digests were desalted immediately.

Tryptic peptides were desalted in reversed-phase home-made microcolumns. Constructed from Empore™ SPE discs (Sigma—Aldrich, USA) with hydrophobic C18 particles. Through microcolumns, peptides are purified and enriched by removing salt and performing subsequent elution. To prepare the microcolumns constructed in P200 tips, centrifugation sequences were performed at 1000 × g for 3 min with 100 μl of 100% MeOH, followed by 100 μl of acetonitrile 80% (v/v) and 0.5% acetic acid solution (v/v), and finally 100 μl of 0.5% acetic acid solution (v/v). Finally, protein digests were added to the columns, centrifuged at 900 × g for 4 min and desalted twice with 100 μl of 0.5% acetic acid solution (v/v) at 1000 × g for 3 min. Peptides were eluted with increasing concentrations of acetonitrile (25%, 50%, 80% and 100%) while the acetic acid concentration was maintained at 0.5% in the solutions, and slow centrifugation was performed at 600 × g for 3 min. Fractions of 20 μl (v/v) each were collected in Eppendorf LoBind tubes. The eluted peptides were lyophilized in a SpeedVac™ SC100 rotary concentrator (Savant™) and stored at −80 °C until quantification, which was also performed using the Qubit™ platform.

LC—MS/MS Analysis and Bioinformatics
The samples were analysed with a UHPLC-nano system (Dionex) coupled online with an LTQ-Orbitrap Elite mass spectrometer (ThermoScientific). Precisely 6 μg of total protein extracted from the initial sample volume was loaded onto a 5 cm PepSwift Monolithic Trap Column column (200 μm internal diameter, Dionex-nanoViper) and separated onto a PepSwift Monolithic Nano Column high resolution analytical Column 25 cm (internal diameter 100 μm, Dionex-nanoViper) and eluted using a gradient from 100% phase A (0.1% formic acid) to 26% phase B (0.1% formic acid, 95% acetonitrile) for 180 min, 26% to 100% phase B for 5 min and 100% phase B for 8 min (a total of 193 min at 200 nL/min). After each run, the column was washed with 90% phase B and re-equilibrated with phase A.

The mass spectra were acquired in positive mode by applying data-dependent analysis in tandem mass spectra acquisition (MS/MS). Each MS in Orbitrap (mass range: m/z 350–1800 and resolution: 120000) was followed by MS/MS of the fifteen most intense ions in the LTQ. Fragmentation in the QTL occurred by collision-induced high-energy dissociation, and selected ion sequences were dynamically excluded for 15 seconds.

Data processing was performed with ProteomeDiscoverer v.1.3 beta (Thermo Scientific). The search and identification of proteins was also carried out with the ProteomeDiscoverer program and Peaks software, with the Mascot v.2.3 algorithm against a Homo sapiens database installed on the laboratory server, using the Database on Demand tool containing the proteins found in UniProt/SWISS-PROT and UniProt/TrEMBL. Contaminant proteins (various types of albumins, human keratins, BSA and porcine trypsin) were added to the database and manually removed from the identification lists. Searches were performed with the following parameters: MS precision of 10 ppm, MS/MS of 0.05 Da, up to 2 missing cleavage sites, carbamidomethylation of cysteines as a modification and oxidation of methionine, and N-terminal acetylation of protein as variable modifications. The number of proteins, protein group and number of peptides were filtered with a false-positive detection rate less than 1%, and peptides with rank 1 and a minimum of 2 peptides per protein were accepted for identification with Proteome Discoverer.

Progenesis QI software (http://www.nonlinear.com/progenesis/qi) (Nonlinear Dynamics©) was used to process the spectra and analyze and interpret the data related to the comparison of the proposed biological scenarios. ANOVA (p<0.05) and fold change (≥2) filters were applied to determine statistical significance.

For peptide identification, the Peaks® Studio 7.0 platform (http://www.bioinfor.com/peaks-studio) (Bioinformatics Solutions, Inc.) was used.

To evaluate the functional annotations (molecular function, cellular component, and biological process) from the
categorization by Gene Ontology (GO), Strap software (http://www.bumc.bu.edu/cardiovascularproteomics/cpctools) was used.

**Results**

Materials were collected from fifteen individuals in the CRC group; however, 6 were excluded from the analysis for the following reasons: four due to insufficient or inadequate samples for evaluation, one for having a history of chemotherapy and one for showing insufficient protein quantification after sample processing. In the control group, materials were collected from 14 patients; however, five were excluded from the analysis due to insufficient or inadequate samples for evaluation (Figure 1). Clinical-pathological characteristics of the study participants are described in Table 1.

Ten proteins with differentiated regulation were identified between patients in the experimental group and in the control group, with statistical significance with a p value ≤ 0.05 (Table 2). The only protein that presented upregulation in the CRC group was beta-2-microglobulin. The annotations of the term GO regarding the cellular components (Figure 2) showed that the regulated proteins are predominantly located in the extracellular environment (23%) and in the plasma membrane (14%). For the biological processes (Figure 3) in which the regulated proteins are involved, the annotations in GO terms revealed prevalence in development (23%), cellular processes (14%), metabolic processes (14%) and regulation (14%). In relation to molecular functions (Figure 4), there was evidence of major involvement with binding (53%) and catalytic activity (24%).

**Discussion**

Through mass spectrometry (MS), protein expression can be identified and quantified in an extremely sensitive manner, even for molecules that are present in low abundances in biological samples. Furthermore, this method is effective in detecting posttranslational modifications, functionalities, localization, and interactions of proteins, helping clarify cell signalling pathways. Due to these characteristics, MS is the main technique of performing translational proteomics in research for prospecting biomarkers, especially in the field of oncology.

Proteomic studies for prospection of diagnostic biomarkers encompass a wide variety of approaches and matrix types. The main matrices used are blood-based samples (serum and plasma), tumour tissue samples, urine samples, stools and samples from colorectal neoplasia models (animal models or organoid cultures).

Although excreted proteins are not high in urine, their composition is significantly less complex than that of serum or plasma, and they can promptly reflect changes in the body. A small concentration of proteins may be advantageous in the prospection of reliable biomarkers because the urinary protein profile generated from glomerular filtering and tubular resorption does not contain large amounts of other nonrelevant plasma proteins, such as albumin. In addition, compared to stool, urine is less affected by microorganisms. Therefore, urine can be considered a good source for biomarker discovery using proteomic technologies.

This work used LC–MS/MS analysis to identify differential protein regulation in the urine of non-pathological individuals and patients with colorectal cancer. Ten proteins were differentially regulated in this proposed comparative scenario. The only protein that was upregulated in the CRC group was beta-2-microglobulin (B2M).

Beta-2-microglobulin (B2M) is a well-known housekeeping protein that is present on the surface of nuclear cells and in most body fluids and is a key component of the histocompatibility complex. Elevated levels of B2M are associated with various pathological conditions, such as kidney diseases, immunodeficiencies, autoimmune diseases, solid tumors and hematologic malignancies. In relation to CRC, there is a robust relationship between the upregulation of this protein B2M and the risk of developing the disease. Prinz and collaborators (2016) measured the B2M levels in serum stored samples from 12,300 individuals and found a hazard ratio of 2:21 for colorectal cancer risk. Mutations in the B2M gene still have value in determining the prognosis in CCR, particularly in microsatellite unstable tumours. Moreover, B2M may be involved in the process of growth, apoptosis and metastasis of neoplastic cells and is a possible focus for cancer target therapies.

Several results, i.e., the downregulation of uromodulin and EGF-containing fibulin-like extracellular matrix protein 1 (EFEMP1) in the CRC group, also emphasize the oncological context.

Uromodulin, or Tamm-Horsfall protein, is a glycoprotein encoded by the UMOD gene. Uromodulin is produced exclusively in the kidneys and is the most abundant protein in urine. It participates in ion transport processes, reduces the aggregation of calcium crystals and performs an immunomodulatory function in the urinary tract. This protein is a biomarker of acute and chronic kidney diseases, and its negative regulation is evidenced in cases of kidney neoplasms. Recently, Xin and collaborators (2022) demonstrated that alterations in the expression of uromodulin-like
1 (UMODL1) are associated with the survival and prognosis of CRC patients.

The fibulin family consists of glycoproteins that contain C-terminal domain and epidermal growth factor (EGF)-like modules with different characteristics. These proteins play a fundamental role in several biological processes, such as embryonic development, organogenesis, homeostasis, coagulation, and healing. Furthermore, these proteins are involved in the control of cell morphology, growth, adhesion, and motility. In the context of colorectal cancer, studies have reported the role of EGF-containing fibulin-like cell matrix protein 2 (EFEMP2) as a marker of early detection, recurrence and prognosis of CCR.\textsuperscript{25,26} For EFEMP1, also known as fibulin 3, the coding gene is related to Doyne honeycomb retinal dystrophy, and studies have shown that its negative regulation is also associated with advanced colon tumors; thus, EFEMP1 is a predictor of worse prognosis and lymph node metastases.\textsuperscript{27,28}

Another downregulated protein in the CRC group is collagen type I alpha 2 isoform 1 (COL1A2). During neoplastic progression, the balance between extracellular matrix formation and degradation is affected, with excessive collagen remodeling by metalloproteinases. As a result, small protein fragments of degraded collagens are released into the circulation, making them potential markers of carcinogenesis. Although most studies show increased expression of collagen proteins in CCR, including as a predictor of liver metastases, the specific role of COL1A2 is unknown. Apparently, COL1A2 is downregulated in CCR because its coding gene is hypermethylated. It was demonstrated in vitro that COL1A2 overexpression inhibits the proliferation, invasion and

---

**Table 1** Clinical-pathological characteristics of the study participants

<table>
<thead>
<tr>
<th>Gender</th>
<th>CRC group (n = 9)</th>
<th>Control group (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3 (33,3%)</td>
<td>3 (33,3%)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (66,7%)</td>
<td>6 (66,7%)</td>
</tr>
<tr>
<td>Age (years, median [variation])</td>
<td>62 [43-85]</td>
<td>63 [38-76]</td>
</tr>
<tr>
<td>BMI (kg/m(^2), median [variation])</td>
<td>23,7 [17,5-30,1]</td>
<td>27,04 [22,9-35,9]</td>
</tr>
<tr>
<td>Staging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1(11,1%)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>3 (33,3%)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>3 (33,3%)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>2 (22,2%)</td>
<td></td>
</tr>
<tr>
<td>Tumour location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right colon</td>
<td>3 (33,3%)</td>
<td></td>
</tr>
<tr>
<td>Left colon</td>
<td>2 (22,2%)</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>4 (44,4%)</td>
<td></td>
</tr>
<tr>
<td>CEA (median [ng/mL]/variation)</td>
<td>2,78 [1,17-180]</td>
<td></td>
</tr>
</tbody>
</table>

BMI: body mass index; CEA: carcinoembryonic antigen; CRC: colorectal cancer.

**Table 2** Urinary proteins identified by label-free mass spectrometry and different abundances in patients with colorectal cancer and in the control group.

<table>
<thead>
<tr>
<th>UniProt entry name</th>
<th>Protein</th>
<th>Fold</th>
<th>HNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0A0523H5</td>
<td>Collagen type I alpha 2 isoform 1</td>
<td>6,81</td>
<td>CTL</td>
</tr>
<tr>
<td>P05090</td>
<td>Apolipoprotein D</td>
<td>7,34</td>
<td>CTL</td>
</tr>
<tr>
<td>P07911</td>
<td>Uromodulin</td>
<td>4,34</td>
<td>CTL</td>
</tr>
<tr>
<td>Q05CF8</td>
<td>KNG1</td>
<td>10,60</td>
<td>CTL</td>
</tr>
<tr>
<td>H0YL3</td>
<td>Beta-2-microglobulin (Fragment)</td>
<td>2,62</td>
<td>CRC</td>
</tr>
<tr>
<td>B4DWH0</td>
<td>Highly similar to EGF-containing fibulin-like extracellularmatrix protein 1</td>
<td>10,71</td>
<td>CTL</td>
</tr>
<tr>
<td>Q6NSB3</td>
<td>Alpha-amylase (Fragment)</td>
<td>37,68</td>
<td>CTL</td>
</tr>
<tr>
<td>C9JMK5</td>
<td>Phosphoinositide-3-kinase-interacting protein 1 (Fragment)</td>
<td>6,57</td>
<td>CTL</td>
</tr>
<tr>
<td>C0JZY2</td>
<td>Titin</td>
<td>2,71</td>
<td>CTL</td>
</tr>
<tr>
<td>Q9HAU0</td>
<td>Pleckstrin homology domain-containing family A member 5</td>
<td>5,78</td>
<td>CTL</td>
</tr>
</tbody>
</table>

CRC: colorectal cancer. CTL: control group; HNA: higher normalized abundance.

---

2 (52x218) are associated with the survival and prognosis of CRC patients.

The fibulin family consists of glycoproteins that contain C-terminal domain and epidermal growth factor (EGF)-like modules with different characteristics. These proteins play a fundamental role in several biological processes, such as embryonic development, organogenesis, homeostasis, coagulation, and healing. Furthermore, these proteins are involved in the control of cell morphology, growth, adhesion, and motility. In the context of colorectal cancer, studies have reported the role of EGF-containing fibulin-like cell matrix protein 2 (EFEMP2) as a marker of early detection, recurrence and prognosis of CCR.\textsuperscript{25,26} For EFEMP1, also known as fibulin 3, the coding gene is related to Doyne honeycomb retinal dystrophy, and studies have shown that its negative regulation is also associated with advanced colon tumors; thus, EFEMP1 is a predictor of worse prognosis and lymph node metastases.\textsuperscript{27,28}

Another downregulated protein in the CRC group is collagen type I alpha 2 isoform 1 (COL1A2). During neoplastic progression, the balance between extracellular matrix formation and degradation is affected, with excessive collagen remodeling by metalloproteinases. As a result, small protein fragments of degraded collagens are released into the circulation, making them potential markers of carcinogenesis. Although most studies show increased expression of collagen proteins in CCR, including as a predictor of liver metastases, the specific role of COL1A2 is unknown. Apparently, COL1A2 is downregulated in CCR because its coding gene is hypermethylated. It was demonstrated in vitro that COL1A2 overexpression inhibits the proliferation, invasion and
migration of cancer cells; thus, COL1A2 can be used as a prognostic biomarker in CCR. 29–33

Downregulation of apolipoprotein D (APOD) was also observed in the CRC group. Apolipoprotein D is a member of the lipocalin family that is primarily associated with high-density lipoproteins in plasma. This protein appears to play a multifunctional role and is associated with the cell cycle and proliferation. It has already been described that mRNA expression of APOD is downregulated in colorectal tumors, and diminished expression of APOD is related to lymph node metastasis, advanced stages, and lower overall survival. 34,35

This study has some limitations associated to the real-life experimental design, such as the low number of cases ascertained in each group. In addition, patients with different CRC stages were included indiscriminately in the analysis, which can generate some phenotyping bias, as patients with advanced disease could express a different proteomic profile from those with early lesions. The main strength of this study is the application of LC–MS/MS, a high-performance method for noninvasive matrix prospection to analyze potential biomarkers.

Conclusion

The LC–MS/MS analysis of urine samples from nonpathological individuals versus patients with colorectal cancer revealed ten proteins that were differentially regulated. The only protein that was upregulated in the colorectal cancer group was

Fig. 2 Cellular component Gene Ontology term annotation. Regulated proteins are predominantly located in the extracellular environment and in the plasma membrane.

Fig. 3 Biological process Gene Ontology term annotation. Regulated proteins are predominantly associated with development, cellular processes, metabolic processes, and regulation.
beta-2-microglobulin. Subsequent studies are needed to evaluate patients with different analysis approaches (i.e., using more shorter runs for analytical methods, and/or Data Independent Analyses [DIA] method), to independently verify and validate these biomarker candidates in a larger cohort sample.

Abbreviations

CRC colorectal cancer.
B2M beta-2-microglobulin.
FOBT faecal occult blood test.
CEA carcinoembryonic antigen.
GO Gene Ontology.
MS mass spectrometry.
EFEMP1 fibulin-like extracellular matrix protein 1.
UMODL1 uromodulin-like 1.
COL1A2 collagen type 1 alpha 2 isoform 1.
APOD apolipoprotein D.

Funding Statement
This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Conflicts of Interest
The authors declare no conflicts of interest related to this article.

Acknowledgements
We thank the team of researchers from the Laboratory of Biochemistry and Protein Chemistry (LBQP) - Biology Institute - Cell Biology Department - University of Brasilia - Brazil. We greatly thank and acknowledge the members of the Colorectal Surgery Department from the University of Brasilia.

References

Urinary Protein Biomarkers in Colorectal Cancer

Alves Martins et al.

30 Kehlet SN, Sanz-Pamplona R, Brix S, Leeming DJ, Karsdal MA, Moreno V. Excessive collagen turnover products are released during colorectal cancer progression and elevated in serum from metastatic colorectal cancer patients. Sci Rep 2016;6(01):30599. DOI: 10.1038/srep30599
33 van Huizen NA, Coebergh van den Braak RJJ, Doukas M, Dekker LJM, IJzermans JNM, Luider TM. Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue. J Biol Chem 2019;294(01):281–289. DOI: 10.1074/jbc.RA118.005087