DM

Discrimination Between Colorectal Cancer Patients and Healthy Individuals Based on the Urinary Volatilomic Profile

MASTER DISSERTATION

Francisca Silva Viveiros
MASTER IN APPLIED BIOCHEMISTRY

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

September | 2025

Discrimination Between Colorectal Cancer Patients and Healthy Individuals Based on the Urinary Volatilomic Profile

MASTER DISSERTATION

Francisca Silva Viveiros
MASTER IN APPLIED BIOCHEMISTRY

SUPERVISOR
José de Sousa Câmara

Discrimination Between Colorectal Cancer Patients and Healthy Individuals Based on the Urinary Volatilomic Profile

Francisca Silva Viveiros

Faculdade de Ciências Exatas e de Engenharia, Universidade da Madeira

Centro de Química da Madeira

Setembro 2025

Funchal – Portugal

AGRADECIMENTOS

Ao longo deste percurso, contei com o apoio, orientação e disponibilidade de diversas pessoas e, a todas, expresso aqui o meu sincero reconhecimento e agradecimento.

Ao meu orientador, Professor Doutor José de Sousa Câmara pela competência profissional, dedicação, disponibilidade e orientação ao longo de todo este processo.

Ao Centro de Química da Madeira (CQM) por fornecer todas as condições necessárias para o desenvolvimento desta dissertação. Também gostaria de agradecer às instituições financiadoras envolvidas: **FCT** – Fundação para a Ciência e a Tecnologia, através dos projetos CQM *Base Fund* (UIDB/00674/2020) (https://doi.org/10.54499/UID/00674/2025) e *Programmatic Fund* (UIDO/00674/2020); **ARDITI**- Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação, através dos projetos M1420-01-0145-FEDER-000005- Centro de Química da Madeira – CQM^+ (Programa Madeira 14-20); **FCT** e **programa Madeira 14-20** para a Rede Nacional de Espectrometria de Massa (RNEM) através do projeto PROEQUIPRAM (M14-20 M1420-01-0145-FEDER-000008).

Ao Centro de Investigação Doutora Maria Isabel Mendonça SESARAM, EPERAM, nomeadamente, Doutora Isabel Jardim, Professora Doutora Ana Célia, Doutor Pedro Berenguer, pela cedência de amostras fundamentais para a realização desta dissertação.

Aos meus amigos e colegas do CQM que me ajudaram neste processo: Telma Gomes, Teresa Abreu, Patrícia Sousa e Rosa Perestrelo. À minha colega de mestrado, Mariana Martins, que me ajudou e apoiou em todos os momentos, pela amizade e por tornar todo este percurso mais leve e risonho.

À minha família e amigos, que me ajudaram a ultrapassar as maiores dificuldades tanto académicas como pessoais. Em especial, à minha mãe, que é um apoio incondicional na minha vida, aconselhou-me e ajudou-me nos momentos mais difíceis e, pela força que transmite, mostrando-me sempre que há luz ao fundo do túnel.

Agradeço a Deus por me acompanhar nesta caminhada e nunca permitir que eu desistisse.

Por fim, presto uma homenagem especial à memória do meu tio José Teixeira, que partiu durante esta jornada, mas cuja importância e influência neste trabalho estarão para sempre na minha vida.

Muito obrigada a todos.

ABSTRACT

Colorectal cancer (CRC) is the third most common and second most lethal cancer worldwide. The colon and rectum are both components of the gastrointestinal system, extending from the mouth to the anus. They play a crucial role in the digestion of food and beverages, responsible for the breakdown of consumed substances, nutrient absorption, and waste elimination.

The objective of this thesis was to determine the urinary volatilomic profile of CRC patients and healthy individuals (HCs) through a non-invasive approach, aiming to identify potential biomarkers for diagnosis and treatment. Urine samples from HC subjects (n = 17) and CRC patients (n = 19) were analyzed using solid-phase microextraction combined with gas chromatography mass spectrometry (HS-SPME/GC-MS), a sensitive and reliable methodology. A total of 68 volatile organic metabolites (VOMs) from both study groups were identified in this study. The chemical families that contributed the most were ketones, phenolic compounds, and norisoprenoids. Data were normalized using multivariate ANOVA followed by Turkey´s test ($p < 0.05$), which identified 30 VOMs with significant differences between groups, including several associated with oncological conditions and processes such a microbial dysbiosis, lipid peroxidation, and chronic intestinal inflammation. Subsequently, multivariate OPLS-DA analysis revealed a clear separation between groups, which highlighted important VOM. 2-Pentanone and 2,5-dimethylfuran were more associated with CRC group, and naphthalene, 2-acetylfuran, p-xylene, theaspirane, 2-ethyl-1-hexanol, hexanal, hemimelitene, and 1,2-dihyfro-1,1,6-trimethylnaphthalene with HCs. An important finding was the early detection of an individual initially classified as a control who was later diagnosed with metastatic cancer. This result highlights the strong discriminatory power of the urinary volatilomic profile developed in this study and confirms the hypothesis that metabolic alterations related to CRC can be detected in urine even before clinical diagnosis. However, to increase the robustness of this study, it would be necessary to increase the number of participants and obtain more relevant information such as diet, disease stage, and investigate other profiles such as proteomics and genomics.

Keywords: Colorectal cancer; Urine; Biomarkers; SPME/GC-MS; Statistical analysis

RESUMO

O cancro colorretal (CRC) é o terceiro cancro mais comum e o segundo mais letal no mundo. O cólon e o reto, são ambos integrantes do sistema gastrointestinal, que se estende desde a boca ao ânus e exercem uma função importante na digestão de alimentos e bebidas, sendo o sistema responsável pela digestão das substâncias consumidas, assimilação de nutrientes e eliminação de resíduos.

O objetivo desta tese foi determinar o perfil volatilómico urinário de doentes com CRC e indivíduos saudáveis (HCs) através de uma abordagem não invasiva, pretendendo a identificação de potenciais biomarcadores para o diagnóstico e tratamento. As amostras de urina dos HCs (n = 17) e de doentes com CRC (n = 19) foram analisadas utilizando microextração em fase sólida acoplada a cromatografia em fase gasosa – espectrometria de massa (HS-SPME/GC-MS), uma metodologia sensível e fiável. Foram identificados 68 metabolitos orgânicos voláteis (VOMs) em ambos os grupos estudados. As famílias químicas que mais contribuíram foram as cetonas, os compostos fenólicos e os norisoprenóides. Os dados foram normalizados utilizando a análise multivariada ANOVA seguida de teste de Turkey ($p < 0.05$), que identificou 30 dos VOMs com diferenças significativas entre os grupos, entre os quais vários associados a condições oncológicas e a processos como disbiose microbiana, peroxidação lipídica e inflamação intestinal crónica. De seguida, análise multivariada OPLS-DA revelou uma separação clara entre os grupos, confirmada pelo teste de permutação e pelas VIP scores, que realçaram VOM importantes. O 2-pentanone e o 2,5-dimetilfurano foram mais associados ao grupo CRC e o naftaleno, 2-acetilfurano, p-xileno, teaspirano, 2-etil-1-hexanol, hexanal, hemimeliteno e o 1,2-dihidro-1,1,6-trimetil-naftaleno aos HCs. Uma descoberta importante foi a deteção precoce de um indivíduo que inicialmente foi classificado como controlo, que mais tarde recebeu um diagnóstico de cancro metastático. Este resultado realça o forte poder discriminativo do perfil volatilómico urinário desenvolvido neste estudo e confirma a hipótese de que as alterações metabólicas relacionadas com o CRC podem ser detectadas na urina mesma antes do diagnóstico clínico. Contudo, para aumentar a robustez deste estudo, seria necessário aumentar o número de participantes e obter mais informações relevantes como a alimentação, estádio da doença e investigar outros perfis como proteómica e genómica.

Palavras-chave: Cancro colorretal; Urina; Biomarcadores; HS-SPME/GC-MS; Análise estatística

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AJCC	American Joint Committee on Cancer
ANOVA	Analysis of Variance
APC	Adenomatous Polyposis Coli
ASR	Age Standardized Rate
AUROC	Area Under the Receiver Operating Characteristic Curve
BRAF	B-Rapidly Accelerated Fibrosarcoma
BMI	Body Mass Index
BSTFA	N,O-bis(trimethylsilyl)trifluoracetamide
CAR/PMDS	Carboxen/Polydimethylsiloxane
CEA	Carcinoembryonic Antigen
CCND1	Cyclin D1 gene
CIN	Chromosomal Instability Pathway
CIMP-high	CpG Island Methylator Phenotype-high
CMS1/3/4	Consensus Molecular Subtype
COX-2	Cyclooxygenase-2
CRC	Colorectal Cancer
CT	Computed Tomography
dHS-SPME	Dynamic Solid-Phase Microextraction
DVB/CAR/PDMS	Divinilbenzene/Carboxen/Polydimethylsiloxane
ECF	Ethyl Chloroformate
EGFR	Epidermal Growth Factor Receptor
ERK	Raf- MEK-Extracellular Signal-regulated Kinase
FAIMS	Field Asymmetric Ion Mobility Spectrometer
FAP	Familial Adenomatous Polyposis
FDA	Fisher Discriminant Analysis
FIT	Fecal Immunochemical Tests
FOBT	Fecal Occult Blood Testing
5-FU	5-Flurouracil
GC	Gas Chromatography
GC-IMS	Gas Chromatography Coupled with Ion Mobility Spectrometry

GC-MS	Gas Chromatography-Mass Spectrometry
gFOBT	Fecal Occult Blood Test
hbA1c	Hemoglobin A1c
HNPCC	Hereditary Non-polyposis Colorectal Cancer
HS-SPME	Headspace Solid-phase Microextraction
IBD	Inflammatory Bowel Disease
IS	Internal Standard
ITEX-GC-MS	Gas Chromatography in Tube Extraction
KNN	K-nearest Neighbors' Algorithm
KRAS	Kristen Rat Sarcoma Viral Oncogene Homolog
lncRNAs	Long non-coding RNAs
LSD	Least Significant Difference
miRNAs	MicroRNAs
MRI	Magnetic Resonance Imaging
MSI	Microsatellite Instability Pathway
MSI-high	Microsatellite Instability-high
mtsDNA	Multitarget stool DNA test
MUTYH	mutY DNA glycosylase gene
NaCl	Sodium Chloride
NF-kB	Nuclear Factor kB
NIST	National Institute of Standards and Technology
NSAIDSs	Nonsteroidal Anti-inflammatory Drugs
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
PCA	Principal Component Analysis
PDMS/DVB	Polydimethylsiloxane/Divinylbenzene
Pi3K	Phospoinositide 3-kinase
PLS-DA	Partial Least Squares Discriminant Analysis
PPARδ	Peroxisome Proliferator-activated Receptor delta
ppm	Parts per million
PVC	Polyvinyl chloride
RAM	Autonomous Regions of Madeira
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RT	Retention Time

SNM	Screening of Nonvolatile Metabolites
SVM	Screening of Volatile Metabolites
TMCS	Trimethyl chloride
TNM	Tumor Node Metastasis
TOF-MS	Gas Chromatography time-of-flight mass spectrometry
TP53	Tumor protein 53
VEGF	Vascular Endothelial Growth Factor
VIP	Variable Importance in Projection
VOMs	Volatile Organic Metabolites
wASR	World Age-Standardized Rate

Chapter I. Introduction

Cancer is a complex process of cellular evolution that begins with a genetic mutation that gives the cell a selective advantage. That mutant cell multiplies faster than the normal cells around it, forming a group of mutant cells, or a clone, that grows and develops abnormally (1). Carcinogenic cells manage to bypass the natural restrictions on cell division, ignoring the signals that would normally prevent uncontrolled growth, thanks to redundant regulatory systems present in cells (2). Furthermore, to survive and adapt to the different environments they encounter as they invade and colonize the body's tissues, cancer cells need to acquire new mutations (2). Therefore, tumor development occurs through a series of mutation and natural selection cycles, as carcinogenic cells adjust and evolve to survive in adverse conditions (1).

According to global estimates from 2022, cancer continuous to be a major public health concern, with over 14 million cases across all continents and both sexes (Figure 1) (3). Breast cancer emerged as the most prevalent type, accounting for 13.3% of all cases (1 871 979), followed by colorectal (10.5%; 1 477 839), lung cancer (9.1%; 1 277 987), and prostate cancer (8.6%; 1 212 426) (3). Thyroid (4,6%; 652 935) and stomach cancer (3,8%; 542 618), also showed significant incidence rates. These six types account for almost half of the global cancer burden, while the remaining 50.1% (7 072 257 cases) are attributed to other malignancies (3).

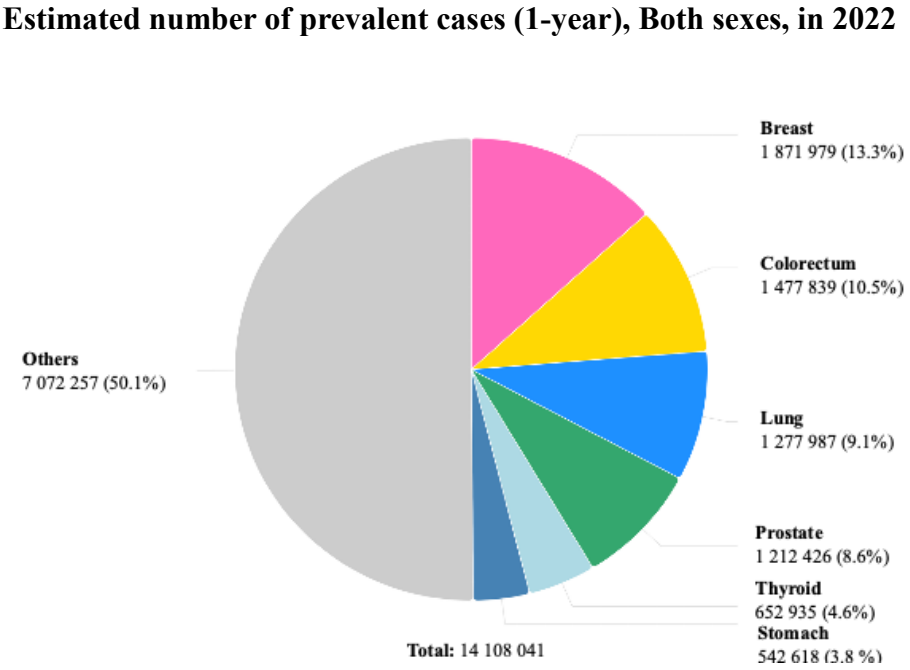


Figure 1 - Estimated number of prevalente cases (1-year), Both sexes, in 2022 Continents, all cancers. Source: GLOBOCAN, 2022 (3)

1.1 Colorectal cancer

Colorectal cancer (CRC) was in the past underdiagnosed, but nowadays it is one of Western countries' most prevalent types of cancer, with a high morbidity and fatality level (4). Globocan (2022) statistics reveal that CRC, in both males and females, is globally the most common form of cancer in terms of new cases, with a total of 1 926 425 new cases, and its terms of deaths, CRC is the third most common and most deadly, with 904 019 deaths (3).

The risk factors for developing such a kind of cancer include older age, a high-fat intake, inactivity, and a state of inflammatory bowel disease (5). However, some factors can reduce such a risk, such as a balanced diet with a high intake of fruits and vegetables and a healthy exercise level (6).

CRC is often developed from adenomas that have accumulated secondary mutations, becoming invasive carcinomas, with a significant portion of such cases being adenocarcinomas (7). Symptoms of CRC include blood in stool, tarry stools, abdominal pain, and iron-deficiency anemia (6). In a small proportion of cases, CRC symptoms become apparent for the first time when dissemination to distant regions of the body has already occurred (7).

The main treatment options include surgery, chemotherapy, and radiotherapy, especially for cases of rectal cancer. More than half of patients with CRC develop metastases, and most of these cases are considered unresectable (7). Despite advances in the treatment of metastatic CRC (mCRC), many patients eventually develop resistance to conventional therapies (7). After surgery, CRC is staged using Classification of Malignant Tumors (TNM), which is a globally recognized standard for classifying the anatomical extent of the spread of malignant tumors (7).

CRC and follow-up diagnostic modalities have a significant bearing on the psychological and physical state of a patient, and in most cases, impact overall life quality (8). Existing diagnostic and screening techniques include colonoscopy, flexible sigmoidoscopy, stool tests such as fecal immunochemical tests (FIT) and fecal occult blood tests (gFOBT), and imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) (9). Despite current techniques, factors such as invasiveness, patient comfort, and the need for increased accuracy mean that there is a need for non-invasive, cost-effective, and patient-friendly tools to aid early detection, increase compliance, and reduce overall CRC incidence and mortality (10).

1.2 Anatomy of the gastrointestinal system and colorectal cancer

The colon and rectum, both parts of the gastrointestinal system, extend from the mouth to the anus and play an important role in digesting food and beverages (11). The system is responsible for digesting consumed matter, absorbing nutrients and eliminating waste, given that the function of the colon and rectum is associated with the elimination phase of these processes (12).

The colon originates from the cecum and is divided into four sections: the ascending colon, transverse colon, descending colon and sigmoid colon (Figure 2) (11). All segments serve a role in absorbing water and electrolytes, and in storing and transporting stool to the rectum for expulsion (12). CRCs are most frequently differentiated according to location in the colon or rectum in an anatomic manner (11). CRCs are typically classified based on their anatomical location within the colon or rectum (12). Tumors in the ascending or transverse colon are called proximal or right-sided cancers, while those in the descending or sigmoid colon or rectum are categorized as distal or left-sided cancers. (12). It is important to remember that appendiceal and anorectal cancers, despite location near the anus and rectum, and near and part of the

colon, respectively, are not part of CRC statistics (11). Cancers in these locations occur in a variety of cell types and have a pathologic and clinical behavior that is different from conventional CRC (12). Understanding the anatomy of the gastrointestinal system and the specific locations of colorectal cancers is essential for diagnosis, treatment planning, and continued research to improve outcomes for affected patients (12).

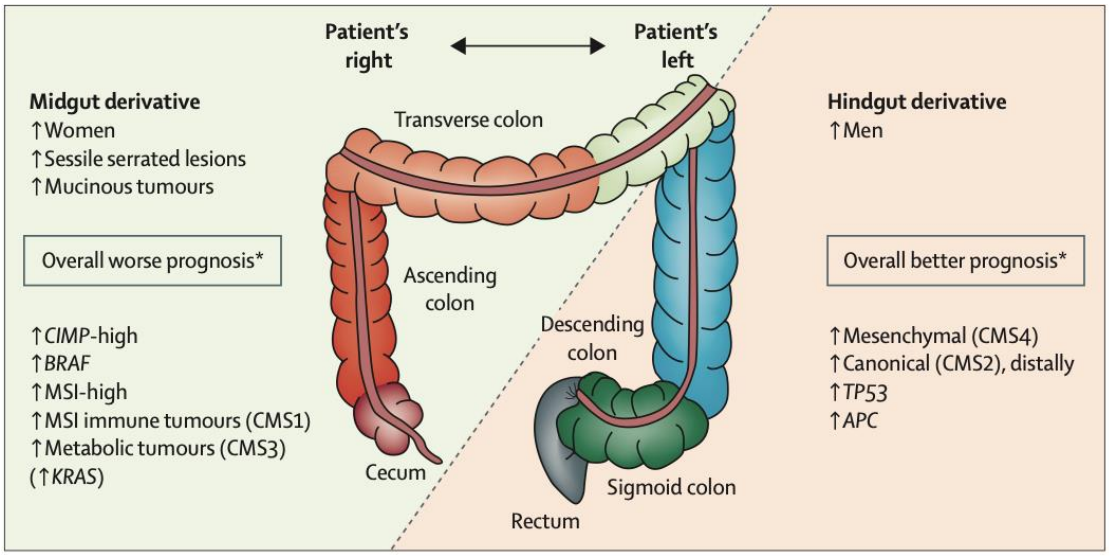


Figure 2 - Differences in right-sided versus left -sided colon and rectum (12). CIMP-high- CpG Island Methylator Phenotype-high; BRAF- B-Rapidly Accelerated Fibrosarcoma; MSI-high- Microsatellite Instability-high; CMS1/3/4- Consensus Molecular Subtype; KRAS- Kirsten Rat Sarcoma Viral Oncogene Homolog; TP53- Tumor Protein 53; APC- Adenomatous Polyposis Coli.

1.3 Incidence and mortality rates worldwide

CRC represents a global disease problem, and it is the third most common and second most lethal cause of cancer in 2022 (3). There were approximately 1 926 425 new cases of CRC worldwide, accounting for 9,6% of all cancer cases, and 904 019 deaths, accounting for 9,3% of all cancer deaths (3).

The incidence of CRC varies geographically, with high age-standardized incidence in high-income countries, including European, North American, and Australian/New Zealand countries (Figure 3) (3). However, high mortality is disproportionately high in low- and middle-income nations, a reflection of inequality in access to early detection and care (4). Asia carries the largest proportion of new cases and deaths with 27,6% of new cases and 27.5% of deaths worldwide (3). As shown in the figure, colorectal cancer is slightly more common in men than in women (Figure 3) (3). This distribution is seen in most parts of the world, with men tending to suffer most from

the disease (3). The literature proposes that life factors, including diet, exercise, alcohol consumption, and tobacco use, could contribute a lot towards such a gender variation (13).

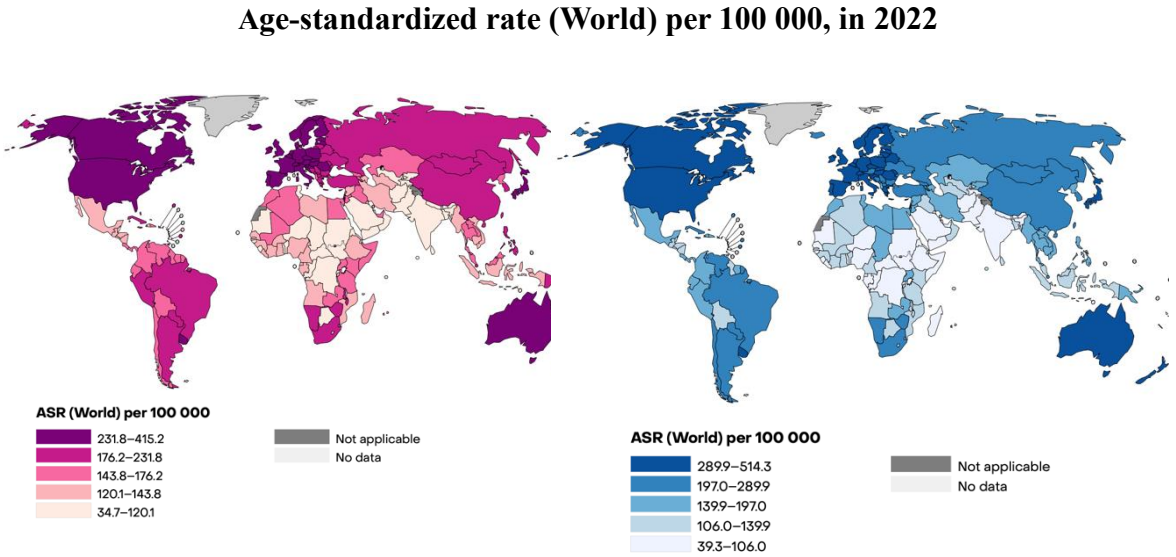


Figure 3- Age-standardized rate (World) per 100 000, Incidence, Females and Males, in 2022. Source: GLOBOCAN, 2022 (3).

Five-year prevalence estimates have proven that CRC affects more than 5.7 million lives worldwide, heightening long-term disease burden for both care providers and healthcare systems (14). Regarding advances in CRC therapy and screening, such statistics serve to present a picture of the necessity for effective prevention, increased access to care, and public policies to address the global burden of CRC (13).

CRC is forecasted to exhibit a significant rise in new cases and deaths in 2030, according to Globocan statistics (14). The incidence of CRC will increase worldwide, with new cases predicted to rise in all regions (Figure 4) (14).

Asia will have a sharp rise in new cases and deaths, in proportion to its high and growing population (14). By 2030, new cases will rise to more than 1.2 million, a significant increase compared with previous years (14). Deaths in Asia will increase, contributing to the increased burden in Asia (3). In Europe, despite new cases rising moderately compared with Asia, new cases in 2030 will nevertheless be significant, with near 600 000 cases predicted (14). Deaths will increase in Europe, but improvements in healthcare infrastructure, early detection, and care will counteract its impact somewhat compared with less developed countries (3). North America will have sustained high cases of colorectal cancer, with more than 200 000 new cases predicted in 2030 (14).

Deaths will increase at a similar rate, but improvements in screening, early detection, and care will reduce the impact compared with less developed countries (3). In Latin America, new cases and deaths will increase significantly, in proportion to changing demographics and increasing burden of factors such as poor nutrition and inactivity (14). By 2030, the number of new cases will rise to over 177 077, with a similar increase in deaths (3). Africa, despite having least in terms of actual cases, will see a significant climb in new cases and deaths in 2030 (14). This increase is a concern, with many African countries having a lack of care infrastructure and access to early detection and care which could exacerbate its impact (3). In Oceania, the number of deaths and cases from CRC will continue to rise, but will remain relatively low compared to other regions (3).

These projections work to highlight the growing global burden of CRC, and underline the necessity for enhanced prevention, early detection, and therapy efforts (4). As cases of CRC rise in all parts of the world, global health systems need to extend access to medical care and screening in emerging economies such as Asia, Latin America and Africa (14).

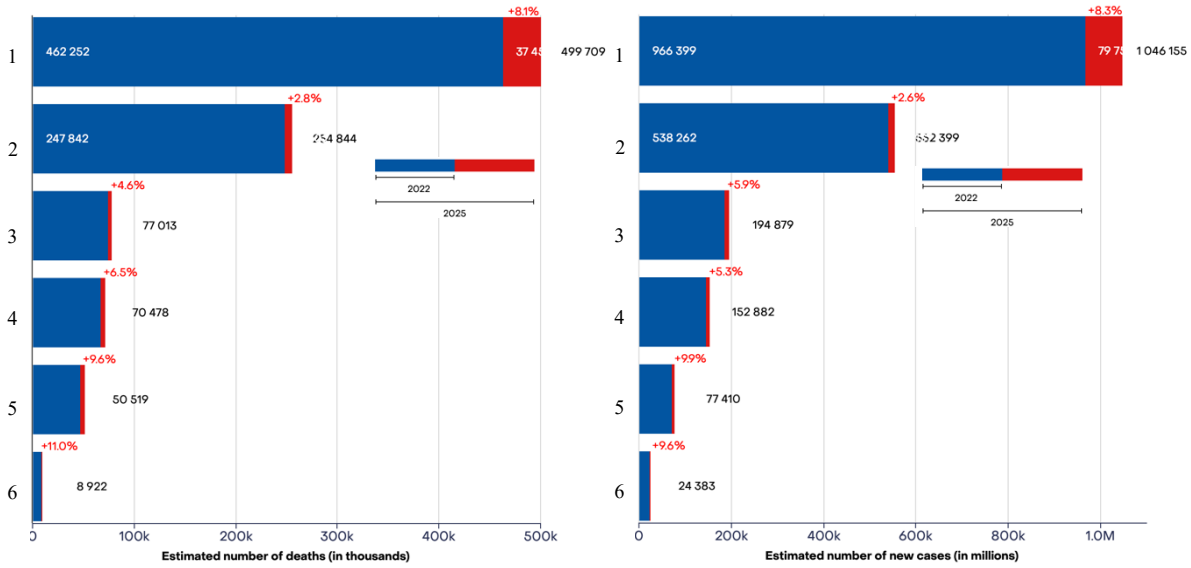


Figure 4 - Estimated number of new cases and deaths from CRC from 2022 to 2030 (1- Asia; 2- Europe; 3- Latin America and the Caribbean; 4- Northen America; 5- Africa; 6- Oceania). Source: GLOBOCAN, 2022 (3).

1.3.1 The impact of colorectal cancer in Portugal

In Portugal, CRC is responsible for a high proportion of deaths caused by cancer, approximately 15.2% of new cases of cancer detected annually (15).

As for the geographical distribution of the incidence rate in 2020 per 100 000 person-years standardized to the European population in Portugal, colon cancer ranked second with 4 689 new cases (26.1/100,000 person-years) (15). The southern region was the most recurrent, in Portoalegre (33.3/100 000 person-years), in Beja (32.3/100 000 person-years) and in the Coimbra district (31.4/100 000 person-years) (15). Rectal cancer ranked sixth, with 1 991 new cases at a rate of 11.6/100 000 person-years (15). It was most frequent in the Coimbra district (15.7/100 000 person-years) and in the Autonomous Region of Madeira (RAM, 14.9/100 000 person-years) (15).

Regarding the geographical distribution of the incidence rate in 2020 per 100 000 person-years and standardized to the European population in Portugal, for man, colon cancer was the third most common cancer with 3 682 new cases (33.6/100,000 person-years) (15). The districts with the highest rates were Beja (43.7/100 000 person-years), Portoalegre (42.0/100 000 person-years), and Coimbra (38.2/100 000 person-years) (15). In men, rectal cancer was the seventh most common, with 1 242 new cases (16.2/100 000 person-years) (15). The districts with the highest incidence were Coimbra (23.1/100 000 person-years), Portoalegre (22.1/100 000 person-years) and Madeira (21.3/100 000 person-years) (15).

According to the standardized geographical distribution of the incidence rate in 2020 (/100 000 person-years) (European population) in Portugal, colon cancer was the second most common type of cancer with 2 008 new cases (19.9/100 000 person-years) (15). The districts of Coimbra (26.0/100 000 person-years), Portoalegre (25.4/100 000 person-years) and Leiria (23.9/100 000 person-years) showed the highest incidence rates (15). Rectal cancer ranked ninth with 749 new cases (7.9/ 100 000 person-years) (15). The regions with the highest incidence were the RAM (10.4/100 000 person-years) and the districts of Faro (10.3/100 000 person-years) and Aveiro (9.9/100 000 person-years) (15).

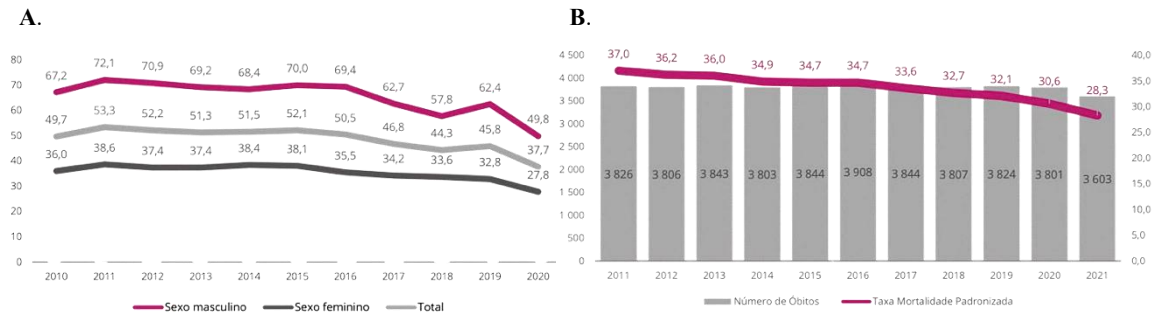
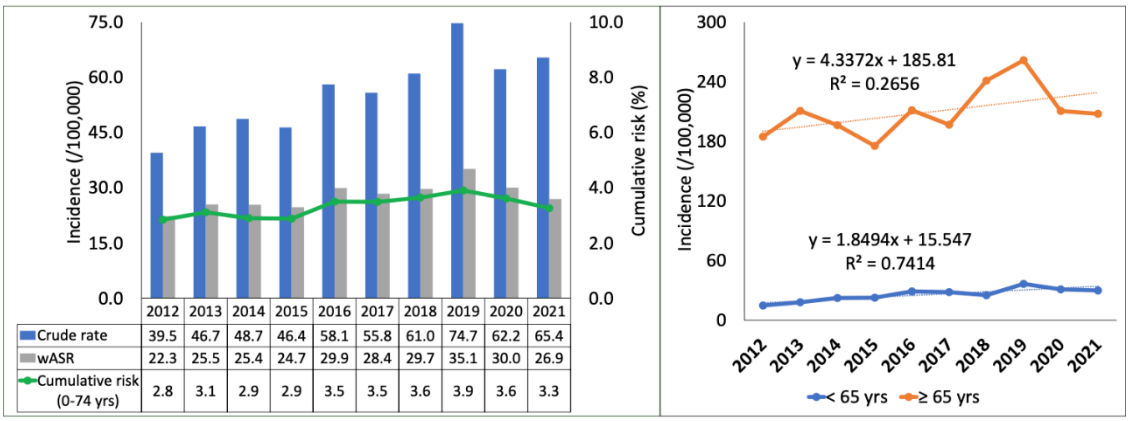


Figure 5 - A. Standardized Incidence Rate 2010-2020. (Source: Evaluation and Monitoring of Population-Based Cancer Screening, data published by the National Cancer Registry. **B.** Standardized Mortality Rate and Number of Deaths from Colon and Rectal Cancer 2011-2021 (Source: Evaluation and Monitoring of Population-Based Cancer Screening, Deaths by cause of Death (105). Standardized rates calculated by DSIA/DGS, based on the European standard population (2013 version) defined by EUROSTAT and using the direct standardization method and five-year age groups (106).

The most updated statistics register 6 092 new cases (16.1% of all malignant tumors) in men, and 4 483 new cases (14.1% of all malignant tumors) in females, in 2022 (3). Deaths have fallen steadily with advances in screening techniques and therapies, but CRC is still a significant public health issue (Figure 5) (16). All of these programs aim to detect colorectal cases at an early stage, when effective therapies are available, and a greater proportion of cases have a high chance of being cured (16). CRC screening in Portugal involves the use of screening tests, such as the immunologic test for occult blood in the stool, for individuals between 50 and 74 years of age (15). Periodical tests are proposed for individuals in the high-risk age group, with a view to finding symptoms of cancer when these have not yet become clinically apparent (15).

In the RAM (2 populated Atlantic islands: Madeira and Porto Santo, total area= 801 km^2, population= 250 744 – census 2021) a total of 1 428 cases of CRC were diagnosed among both sexes between 2012-2021 (Figure 6) (17). CRC was the third most frequent cancer overall and the second among each sex. A median age of 68 years (IQR= 77-59) was obtained, and 32.1% were diagnosed as stage III and 26% with metastatic disease (17). The World Age-Standardized Rate (wASR) varied from 22.3 in 2012 to 35.2/100 000 in 2019 (17). In the RAM, in 2020, wASR was 30.0/100 000 comparable to Southern Europe = 31.9 (World= 19.5 and Portugal= 39.4/100 000) (17). The minimum cumulative risk (0-74 years) was obtained in 2012 (2.8%) and the maximum in 2019 (3.9%) (Figure 4) (17). A tendency for a more "stable" increase in the age-specific rate among young group (<65 years; Pearson's R= 0.86) was found compared to elderly (≥ 65 years; Pearson's R= 0.52) (17). Its incidence continues to increase over the years, possibly

due to the risk factor exposure including animal-source intake, sedentarism, increased body mass index (BMI), alcohol, tobacco consumption, among others (18). In 2021, started the population based colorectal screening (50-74 years) with a simple fecal occult test before colonoscopy (17).



	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Crude rate	39.5	46.7	48.7	46.4	58.1	55.8	61.0	74.7	62.2	65.4
wASR	22.3	25.5	25.4	24.7	29.9	28.4	29.7	35.1	30.0	26.9
Cumulative risk (0-74 yrs)	2.8	3.1	2.9	2.9	3.5	3.5	3.6	3.9	3.6	3.3

Figure 6 - Crude, age-specific and world age-standardized incidence rate (wASR), and cumulative risk. 2012-2021 (17).

1.4 Colorectal cancer risk factors

CRC is a disease with both modifiable and non-modifiable etiological factors responsible for its causation (5). There are factors such as age and family background, whose role cannot be changed, but modifiable factors such as exercise and nutrition can have a significant role in changing susceptibility towards developing such as type of malignancy (Figure 7) (6). Identification of such factors holds key in its prevention, early detection, and effective therapy (7).

1.4.1 Modifiable risk factors

Several causes, such as lifestyle have been strongly associated with colorectal cancer (13). Most colorectal cancers are associated with risk factors that can be changed, providing an opportunity for prevention (18).

1.4.1.1 Being overweight or obese

Being overweight and obesity have a strong potential for developing CRC (18). In studies, one-third of cases of CRC have been linked with obesity, high consumption of alcohol, and high consumption of red meat (19). There is a strong relation between a high

BMI and high risk of developing colorectal cancer, with high impact in males through high fat in the abdominal cavity, high inflammation and high susceptibility towards cancer (18,19).

1.4.1.2 Physical Inactivity

Lack of exercise raises colorectal cancer risk through increased insulin sensitivity, lowering insulin demand for production (20). Reduced insulin blood is linked with increased CRC risk (20). Physical exercise also averts obesity, a key colorectal cancer risk, through averted chronic inflammation (20). The American Cancer Society recommends adults exercise 150 to 300 min per week at moderate intensity, which is 30 to 60 min per day, five days per week (12). Adults can also benefit from 75 to 150 min of vigorous intensity activity per week, which is about 10 to 20 min per day, five days a week (18).

1.4.1.3 Diet

Diet plays a significant role in CRC (21). Consumption of high amounts of red (e.g., pork, beef, and lamb) and processed (e.g., hot dogs and processed luncheon meats) foods increases the risk (21). High-temperature cooking (frying, broiling, grilling) of foods can produce unhealthy chemicals that can promote the development of cancer (21). On the other hand, a healthy model of diet with high consumption of fruits, vegetables, whole grain foods, and avoidance of red and processed foods can reduce the risk (18).

1.4.1.4 Smoking

Although smoking is most commonly associated with lung cancer, it is responsible for a variety of other types of cancer, including CRC (22). Cigarette smoke contains many carcinogens, which can cause mutations in DNA and lead to irreparable damage to the colorectal mucosa (22). The colon and rectum are exposed to these compounds through the circulation or direct contact after ingestion of saliva (22). The results of an animal study in which mice with induced colonic inflammation were exposed to tobacco derived carcinogens suggested that smoking could increase the risk of colon cancer development (22).

1.4.1.5 Alcohol Consumption

There is a widespread scientific consensus that consumption of alcohol can lead to a range of types of cancer (23). According to National Toxicology Program's Report on Carcinogens, published by the US Department of Health and Human Services, consumption of alcoholic beverages is a confirmed human carcinogen (23). There have

been a range of proposed processes through which consumption of alcohol can lead to an increased likelihood of developing cancer (23). One such mechanism is through the metabolism of ethanol, an active constituent in alcoholic beverages (23). In the organism, when ethanol is metabolized, it is metabolized to form acetaldehyde, a toxic compound and suspected human carcinogen (24). Acetaldehyde can cause both DNA and protein damage and can induce such changes in cells that can stimulate the development of cancer (24). Besides acetaldehyde, consumption of alcohol can generate reactive oxygen species (ROS), chemically reactive species with oxygen in them (24). ROS can induce oxidative damage to DNA, proteins, and fats in an organism, providing an opportunity for mutations in cells and the development of cancer (23).

1.4.2 Non- modifiable risk factors

1.4.2.1 Age

The trends in CRC cases and deaths are partly fueled by the older age group, with most disease burden experienced in this group (25). Overall, such a trend, nevertheless, covers changing trends in younger age groups (25). In older adults, CRC cases have decreased in the past decade, but in adults under 50, CRC cases have increased at a pace of about 2% per year over the same decade (12). Besides, CRC cases have leveled off in adults 50-64 years of age (12). Such changing trends have seen an increasingly younger age group of patients emerging (12). In 2019, 20% of all CRC cases in adults 54 years and below, a significant rise over 11% in 1995, have been witnessed (12). Despite a decline in the proportion of the overall population under 50 years, down from 80% in 1995 to 71% in 2019, such an increase in rise is witnessed (12). This rise is occurring despite a decline in the proportion of the overall population that is younger than 50, which has decreased from 80% in 1995 to 71% in 2019 (12). Over 2011 and 2020, CRC deaths have increased at a pace of about 1% per year in adults under 50, in contrast with a drop in older age groups (12). Specifically, deaths have reduced at a pace of 0.5% per year in adults under 50, in contrast with a drop in older age groups (12).

1.4.2.2 Personal history of colorectal polyps or cancer

A history of adenomatous polyps places a patient at a high risk for developing colorectal carcinoma (26). It is specifically high when such a polyp is large, numerous, or with dysplasia, a name for aberrant cellular abnormalities presents in a precancerous

state (26). Patients with such a characteristic of a polyp have a high chance of developing colorectal carcinoma through the opportunity for such a lesion to become malignant (6,26).

Furthermore, individuals with a history of CRC, even when initially resected, have a heightened risk for developing secondary malignancies in any part of the colon or rectum (26). That heightened risk is even larger when the first malignancy is one with a more virulent disease and with a greater opportunity for recurrence and development of new malignancies in the gastrointestinal tract (26). For sure an individual, follow-up and active screening become a necessity for controlling and minimizing future increased disease (6,8).

1.4.2.3 Personal history of inflammatory bowel disease

Individuals with a history of inflammatory bowel disease (IBD), such as disease in the form of ulcerate colitis and Crohn's disease, have a high probability of developing colorectal cancer (27). IBD is a disease with a character of ongoing colonic inflammation, over a long duration of years (27). Chronic inflammation, most prominently in untreated disease and poorly managed disease, can result in abnormalities in cells in both the colon and the rectum, with a high probability of developing dysplasia (27). Dysplasia is an atypical, but not malignant, condition with a probability of developing malignant tumors in the long-term (6).

Given the heightened CRC risk with IBD, individuals with such a disease will have to begin colorectal cancer screening at an early age compared to the general population (27). Besides, the heightened frequency of screening will vary with disease duration and severity (16). Inflammatory bowel disease can be differentiated from irritable bowel syndrome, with the latter not being a cause of heightened CRC risk (27). For individuals with IBD, early intervention and continued follow-up will go a long distance in counterbalancing heightened CRC risk, most notably for long-standing and untreated disease (6,8).

1.4.2.4 Family history of colorectal cancer or adenomatous polyps and colorectal cancer risk

While most CRC cases occur in individuals with no family disease, approximately one in three individuals with CRC has a family member with a similar disease (28). CRC in a first-degree family member (parent, sibling, or offspring) vastly increases an individual's risk for developing the disease (28). Even a larger increased risk is seen when

the family member develops disease at an age less than 50 years, or when a family group of several first-degree family susceptibility, a mix of genetic inherited mutations may predispose an individual to CRC, and familial aggregation of disease implicates a genetic basis for disease susceptibility (28). In addition, shared life experiences, such as diet and activity level, could cause familial clustering (6).

The presence of family members with adenomatous polyps is also associated with a rise in CRC risk, in that such polyps, having a malignant potential, predispose to malignancy in the colon (29). For individuals with a family background of adenomatous polyps or colorectal carcinoma, early screening can sometimes even be recommended, even at a younger age than the general age recommendation (6).

1.4.2.5 Inherited syndromes and colorectal cancer risk

Approximately 5% of CRCs occur in individuals with inheritable genetic mutations predisposing them to CRC. Inherited mutations have been linked with a variety of family syndromes, two of them being Lynch syndrome (formerly hereditary non-polyposis CRC, HNPCC) and familial adenomatous polyposis (FAP) (30). There are a variety of less prevalent genetic syndromes with a heightened risk for CRC, and such an association underlines the utility of genetic testing and early detection in such individuals (30).

Lynch syndrome is the most prevalent hereditary colorectal syndrome of cancer, and constitutes approximately 2% to 4% of all CRC cases (30). It is most frequently caused by mutations in DNA mismatch repair genes, such as MLH1, MSH2, MSH6, PMS2, and EPCAM, that have been inherited (30). All of these genes have an important function in repairing errors in DNA, and when not functioning, they predispose one to an accumulation of genetic errors that can result in cancer (30). In individuals with Lynch syndrome, CRCs occur at a relatively early age, most frequently in the right side of the colon (6,31). Polyps in individuals with Lynch syndrome occur less frequently in comparison with individuals not having the syndrome (32). In individuals with Lynch syndrome, a 50% chance of developing CRC during a life span but varies with the specific gene mutated (32). In addition to CRC, a predisposition for developing other types of cancer, most prominently in females, including carcinoma of the endometrium (the lining of the uterus), ovary, stomach, small intestine, pancreas, kidneys, prostate, and breast, in persons with Lynch syndrome (32). Turcot syndrome, a disease with a mutation in one of

Lynch syndrome genes, increases the likelihood???? of colorectal and brain malignancies, especially a brain malignancies, glioblastomas (6,32).

Familial Adenomatous Polyposis is a genetic disorder that increases a person´s likelihood of developing CRC (30). FAP is a result of mutations in the APC gene, and it is an autosomal dominant disorder (30). Approximately 1% of all colorectal malignancies may be associated with FAP (30). In most cases of FAP, individuals develop hundreds or even thousands of adenomatous polyps in the colon or rectum, beginning in early teens (10 to 12 years of age) (30). Unless surgically removed, these polyps will become malignant in approximately 40 years, with nearly all individuals with FAP developing colorectal carcinoma, excluding surgical resection of the colon for preventive indications (30). There is a heightened susceptibility to developing malignancies in other organs, including gastric, small intestine, pancreatic, hepatic and other organ malignancies (6).

1.4.2.6 Type 2 diabetes

Type 2 diabetes is a long-standing disease with insidiously progressive insulin derangement, often with hyperglycemia and hyperinsulinemia (33). In the United States, about 37.3 million individuals have diabetes, with documentation of abnormal fasting plasma glucose or hemoglobin A1c (hbA1c) (33). Black adults disproportionately have diabetes, with an age-adjusted prevalence of 16.8%, compared with 11.2% in White adults (33). Incidence and prevalence of diabetes have increased at a larger proportion in Black adults, compared with White adults, and widening racial gaps in disease burden have become apparent (33). Epidemiologic and biologic studies have postulated that individuals with a diagnosis of type 2 diabetes have an increased risk for developing CRC (33). The insulin and derangement in metabolism in diabetes, with ongoing oxidative and inflammatory processes, could underlie CRC development (33). Alteration in networks of cellular signaling with increased proliferation and survival of malignant cells can occur through such processes (33).

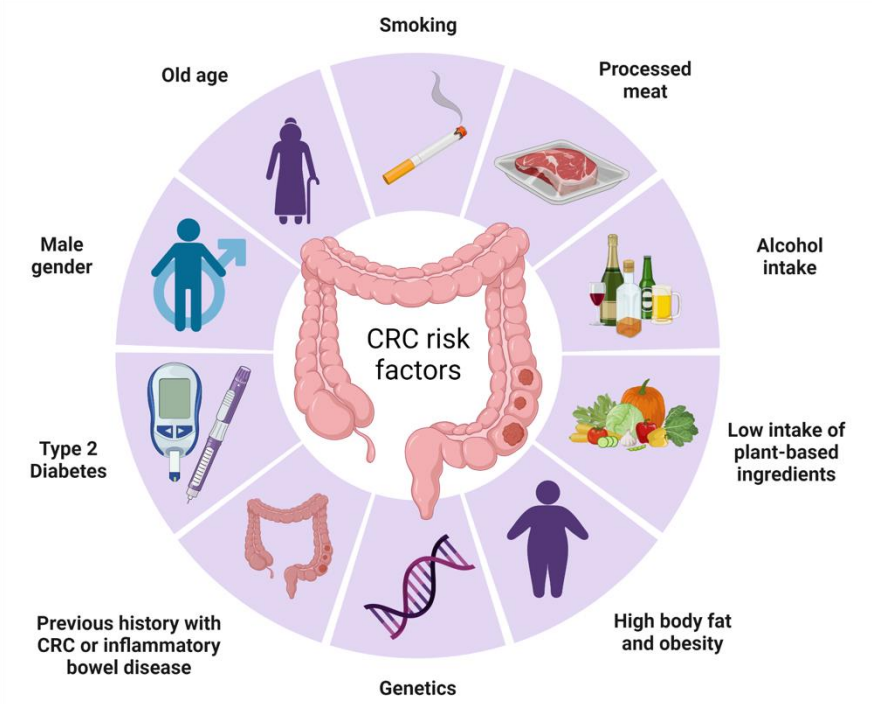


Figure 7 - Modifiable and unmodifiable colorectal risk factors (BioRender) (107).

1.5 Primary and secondary prevention of colorectal cancer

Primary and secondary prevention both contribute to a considerable reduction in CRC cases and deaths (34). Primary prevention targets minimizing overall CRC development risk through medical interventions, lifestyle modifications, and early detection of high-risk groups (34). On the other hand, secondary prevention aims at early CRC detection when its treatment is most effective and precancerous lesions removed even before becoming malignant. (34).

1.5.1 Primary prevention

Primary CRC prevention involves intervention in a range of factors involved in disease development (16). These risk factors are primarily related to lifestyle modifications, including smoking, diet, physical activity, and alcohol consumption, as well as medical interventions and genetic predisposition (34).

Perhaps one of the most significant modifiable CRC susceptibility factors is cigarette smoke use (34). In addition, cigarette smoke use is associated with a range of additional types of cancer (18). Tobacco smoke carries carcinogens, whose mutations in colon cell DNA can result in the development of cancer (35). To inhibit cases of CRC in

relation to cigarette use bans in public, increased taxes for cigarette sales, and increased cigarette-smoking cessation programs (34). All these interventions have effectively reduced cigarette use, in turn curbing CRC and additional cigarette-related disease (34). Diet and exercise form additional integral parts of CRC prevention. Involvement in a high intake of red and processed meat and low-fiber diet is linked with increased CRC susceptibility (16). In contrast, a high intake of fruits, vegetables, and whole grain foods, foods rich in fiber, confers protective factors (16). In addition, a high level of exercise is linked with a reduced CRC risk, possibly through its beneficial effects in enhancing digestion and minimizing inflammation (34). Promotion of healthy diets and an active life can go a long way in CRC prevention (34). Excessive use of alcohol is a well-established CRC risk factor (34). Alcohol is metabolized to acetaldehyde, a compound with the potential to cause DNA damage in colon cells and contribute to cancer development (23). Public awareness about moderation in alcohol consumption can contribute to a reduced CRC burden (34). Chemoprevention, in addition to lifestyle changes, is an alternative to reducing CRC risk that involves the use of drugs and chemicals (36). Anti-inflammatory drugs, such as aspirin, have been shown to inhibit colorectal adenomas, benign growths with a potential to become CRC, and have a protective role in CRC development (36). Inhibition of an enzyme, cyclooxygenase-2 (COX-2), responsible for inflammation and for the growth of cancer cells, is postulated to produce such protective actions, but use of such drugs is not free of danger and must therefore be under medical guidance (34). Folic acid, a B vitamin involved in DNA synthesis and its repair, has been touted as a preventive for CRC (37). Long-term folic acid supplements have been proposed to inhibit adenomas and CRC through inhibition of DNA damage, but studies have produced conflicting results, and additional studies must occur before folic acid can become a general recommendation for CRC prevention (37). Likewise, a protective role for CRC has been proposed for calcium, with its proposed mechanism of protecting CRC development through its chelation with bile acids in the colon, possibly lessening their carcinogenicity (38). However, research studies about calcium supplements have conflicting findings, and its role in CRC prevention must be confirmed through future studies (34,38). Genetic testing is an important part of primary prevention, specifically in individuals with a family history of CRC or familial syndromes such as Lynch syndrome or FAP (32). For individuals at high genetic risk, early interventions such as increased follow-up or preventive operations can effectively lessen the chance of developing CRC (34).

1.5.2 Secondary prevention

Secondary prevention is focused on early detection of CRC when in an early stage and not yet in an advanced stage (34). The purpose is to detect and remove precancerous lesions, such as adenomas, before becoming an invasive form of cancer (34). It has been proven to maximize survival, with early CRC benign easier to manage (34).

The most important part of secondary prevention is screening. There are a variety of tests for CRC, and one such common one is fecal occult blood testing (FOBT), a noninvasive testing that detects concealed blood in stool, a sign that could mean CRC, including precancerous lesions, is present (34). It is recommended for persons over 50 years with an average CRC level of risk (34). Despite its effectiveness and economy, its use is not ideal, with no detectability for nonbleeding polyps and no direct visualization of the colon (34). The flexible sigmoidoscopy is a kind of screening that allows physicians to visualize the descending part of the colon with a flexible tube (18). It identifies polyps and tumors in the sigmoid colon and rectum and is recommended for persons over 50 years with no additional CRC level of risk, at five-year intervals (18). While less uncomfortable, sigmoidoscopy sees only part of the colon, and its success in discovering polyps or CRC in the upper portion of the colon is, therefore, compromised (34). The gold standard for CRC screening is regarded to be colonoscopy, with its full view of the entire colon (16). Colonoscopy allows both visualization and resection of polyps, and in the process, can prevent them becoming CRCs (16). Colonoscopy is recommended at 10-year intervals for individuals with no family history of CRC and no other risk factors (16). Double-contrast barium enema and CT colonography can serve as alternatives for individuals in whom colonoscopy cannot be conducted, but these tests have less success in discovering small polyps (34). For high-risk persons, such as those with family history of CRC, a history of IBD, or with a genetic syndrome such as Lynch Syndrome or FAP, screening must commence early and at shorter intervals (34). For individuals with Lynch Syndrome, colonoscopy must commence at 20-25 years, and for individuals with FAP, annual sigmoidoscopy must begin during adolescence (34). Polyp removal through colonoscopy is a key aspect of secondary prevention, too (34). Patients with a history of adenomatous polyps have an increased probability of developing additional polyps and CRC (34). Periodic colonoscopy, typically every 3-5 years, is advised for individuals with a history of having a removed polyp (34).

1.6 Colorectal cancer carcinogenesis

1.6.1 Adenoma-carcinoma sequence

Most colorectal malignancies arise from precancerous adenomas, and these can broadly be grouped into two categories: traditional tubular adenomas and serrated polyps (Figure 8) (39). Adenoma development arises through a derangement in the mechanism responsible for regulating proliferation and DNA repair in cells (39). In the colon, these processes are important, with continuous regeneration of cells in the intestinal mucosa, with proliferation taking place at the base of the crypts and cells moving towards the colonic lumen, with loss of processes of terminal differentiation and programmed cell death, and with the development of increasingly dysplastic histologic features and, in time, gain invasive capabilities (39).

This transformation involves a cascade of mutations in important growth-regulating genes, indicative of development towards malignancy from normal tissue (39). Constitutive activation of Wnt-regulated genes, most of which have a high tumorigenesis association, is a result of these mutations. Some of these include MYC, cyclin D1 gene (CCND1), vascular endothelial growth factor (VEGF), and peroxisome proliferator-activated receptor delta (PPARδ) (39). In addition, mutations in Wnt-regulated genes, including AXINI1, AXIN2, and CTNNB1, can amplify Wnt signaling in the absence of an APC mutation (39). This confirms Wnt pathway's important role in the proliferation of intestinal epithelial cells and its role in contributing towards tumorigenesis through increased expression of a variety of growth and survival-related genes (39). The Wnt pathway is activated in nearly all CIN-containing tumors, and APC mutations have been identified in approximately 80% of them (39). To synthesize findings in more than 30 gene expression studies conducted with differing techniques and platforms, a consensus model for CRC subtyping, wherein colorectal tumors segregate into four types based on robust, processed gene expression profiles, was put forward by the CRC Subtyping Consortium. CMS2 tumors have Wnt and MYC pathway activation, with SCNAs, and exhibit a CIN phenotype (39).

MYC, a gene not commonly in most colorectal malignancies, but MYC amplifications occur with high frequency. It has been established that expression of MYC can be increased with Wnt pathway activation (39). On the other hand, meta-analyses have not confirmed a definite association between MYC expression and overall disease survival in cases of CRC (39). Apart from mutations in APC, autosomal dominant

mutations in the APC gene cause familial adenomatous polyposis (FAP), a disease with many adenomatous polyps (39). KRAS mutations, an activating mutation, occur with high frequency following mutations in APC and in approximately 40% of colorectal malignancies (39). KRAS is a part of variety of growth factor signaling networks, including epidermal growth factor receptor (EGFR) signaling. In the EGFR pathway, KRAS activation induces constitutive activation of the Raf-MEK-extracellular signal-regulated kinase (ERK) cascade, phosphoinositide 3-kinase (pi3K) signaling through MTOR, and nuclear factor kB (NF-kB) transcription factors (39). Progress through the cell cycle is facilitated through such a cascade and is most activated in colorectal malignancies (39). Targeting such a pathway with therapeutic drugs, most prominently in metastatic CRC, is a critical clinic practice (39). However, colorectal malignancies with mutations in KRAS and BRAF are not sensitive to inhibition with drugs for EGFR, as such malignancies have activity in the Raf-MEK-ERK and PI3K pathways (39).

1.6.1.1 Chromosomal instability (CIN) pathway

CIN accounts for 65-70% of cases of sporadic CRC and is characterized by aneuploidy, deletion, amplification, and loss of heterozygosity in its chromosome abnormalities (40). CIN tumors have fewer mutations in coding regions but have defects in chromosome segregation and in the repair of DNA damage (40). Loss of function of APC is the initial genetic alteration in this pathway, and it induces Wnt signaling activation and induces proliferation of cells (40). Other mutations, including TP53, KRAS, and PIK3CA promote growth and therapy resistance (39).

1.6.1.2 KRAS mutations and metabolic alterations

KRAS mutations often coexist with metabolic shifts, most prominently through PIK/AKT activation, with a function in survival and proliferation of cells (41). CMS3 (with CIN) tumors have characteristically aberrant amino acid and glucose metabolism, and these can serve as therapeutic targets (39,41).

1.6.1.3 COX-2

COX-2 expression in CRCs and adenomas is heightened and induces progression of the tumor through its conversion of arachidonic acid to proliferation-supportive prostanoids (42). Inhibition of COX, via drugs including nonsteroidal anti-inflammatory drugs (NSAIDSs), lessens CRC risk and confirms use in subjects at heightened risk (39,42).

1.6.1.4 TP53 mutations

TP53 mutations are the most common genetic mutations in cancer and play a critical role in colorectal carcinogenesis in its later stages (43). Loss of function of TP53 allows unregulated growth of cells (43). Approximately 60% of CIN tumors have inactivating mutations in TP53, and loss of 18q arm heterozygosity (DCC and SMAD2/SMAD4 genes included) is a marker for poor prognosis (39,43).

1.6.1.5 Microsatellite instability (MSI) pathway

The MSI pathway is characterized by high-frequency mutations in microsatellite regions secondary to an impairment in the DNA mismatch repair system (44). MSI tumors develop mutations in MLH1, MSH2, MSH6, PMS2, or EPCAM and enter a hypermutable state conductive to tumorigenesis (44). MSI arises in approximately 15% of sporadic colorectal malignancies, and most MSI malignancies arise secondary to epigenetic inactivation of MLH1 (44). MSI malignancies often develop mutations in TGFBR2M that promote unregulated proliferation of cells (44). Despite a relatively good prognosis in comparison with CIN malignancies, MSI malignancies respond to immunotherapies, such as PD-1 blockers (39,44).

1.6.1.6 Serrated neoplasia pathway

The serrated neoplasia pathway accounts for about 15% of CRCs and is associated with serrated polyps, extending from benign lesions to precancerous adenomas (45). Serrated neoplasia is distinguished through early BRAF V600E mutation activation, with constitutive activation of the MAPK pathway and unregulated proliferation of cells (45). Serrated tumors exhibit high CpG island methylation (CIMP), and it inactivates tumor-suppressor genes and induces carcinogenesis. BRAF mutations and CIMP occur in both MSI-high and MSS serrated tumors (39,45).

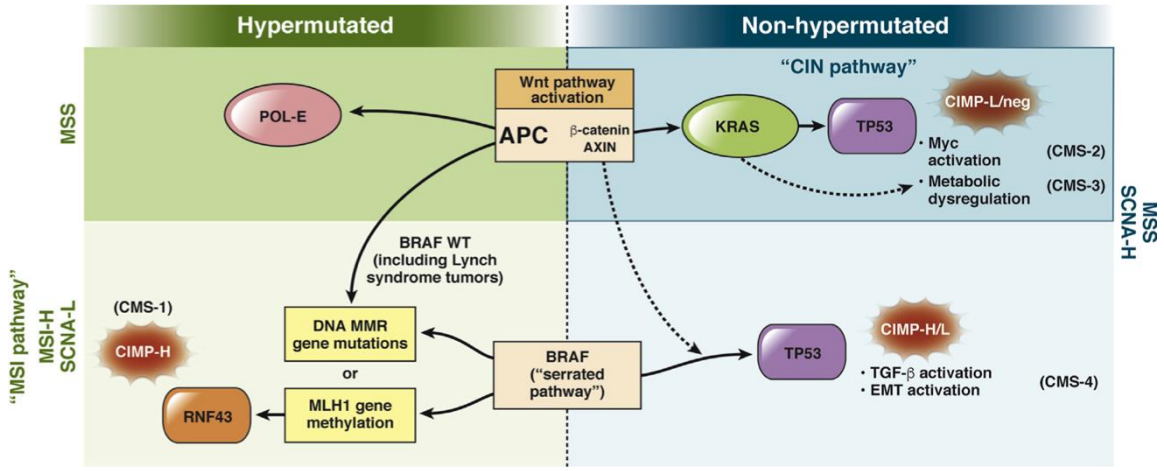


Figure 8 - Pathways of colorectal carcinogenesis. Activation of the Wnt pathway (primarily via APC mutations) or a mutation in BRAF can initiate colorectal tumorigenesis. BRAF mutations promote tumorigenesis via the serrated neoplasia pathway, leading to MSI with hypermutation or MSS without hypermutation 8indicated in the figure). Colorectal tumor classifications include CIN, MSI, and the serrated pathway (see CMS). EMT, epithelial to mesenchymal transition (39).

1.7 Colorectal cancer clinical manifestation

CRC typically presents with symptoms that occur in its later stages and can vary according to the location and size of the tumor (6). CRC in its early stages is often asymptomatic, and symptoms develop with progression of the disease (Figure 9) (46).

One of its early symptoms is hematochezia, during which small amounts of blood can be detected in the stool (6). In such a case, no apparent change can be detected, but a positive reading can result in a fecal occult blood test (6). In cases in which a significant amount of blood is involved, stool can become mucus-containing or jellylike (6). With progression of disease, one of its common symptoms is obstruction in the intestines, secondary to growth in the tumor bulk (6). Abdominal pain, bloating, nausea, vomiting, weakness, and difficulty in passing stool can result in obstruction. Involvement of an abdominal mass is a common feature of carcinoma of the right-sided colon, in which a palpable tumor develops (46).

Systemic symptoms, such as cachexia (significant loss of weight and muscle tissue), anemia and general loss of weight, can develop in later phases of disease (6). In early phases, CRC will not have apparent symptoms, and a delayed diagnosis can then develop when the tumor is a significant size (6). Symptoms of a tumor vary according to location in anatomic site (6). For carcinoma of the right side of the colon, blood in stool and obstruction of the gut develop. For carcinoma of the rectum, changed bowel habits

are most common symptoms (6). Symptoms in early phases of CRC are not specific and therefore contribute to delayed disease recognition and therapy. (46).

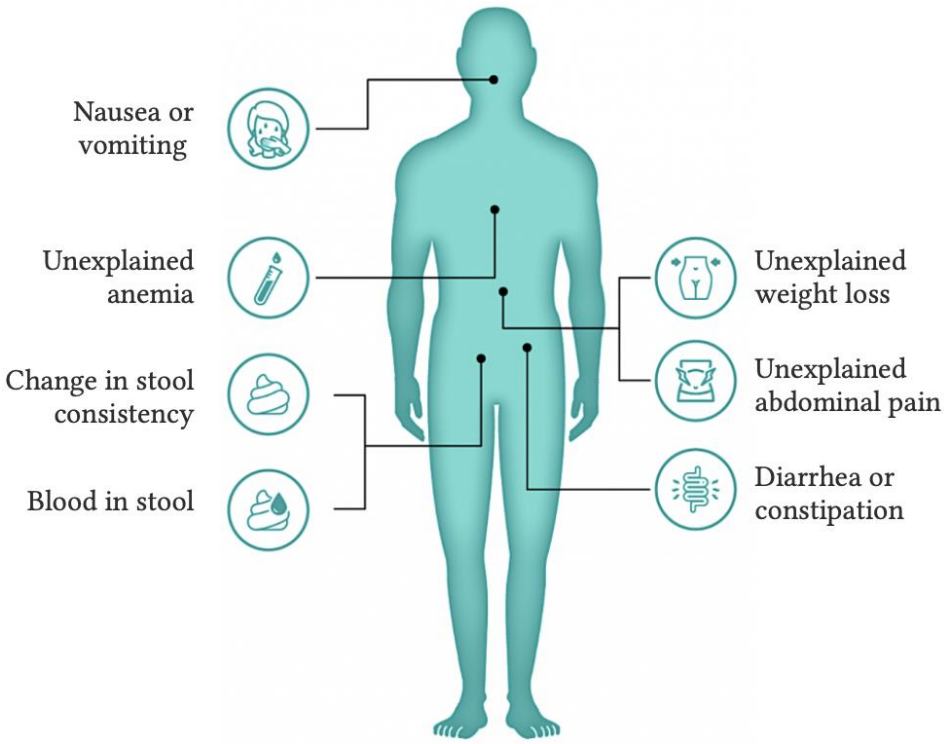


Figure 9 - Clinical manifestations of colorectal cancer (Fonte: Cancro Online) (108).

1.8 Colorectal cancer screening

CRC is usually asymptomatic at its initial stages, and thus, there is a low level of early detection (16). This delayed identification of the disease translates to the fact that most patients present themselves for diagnosis when the disease has reached late stages, where the possibility of treatment is slim, and the chance of curative therapy is usually lost (16). The key to successful CRC treatment is early identification via effective screening programs (9).

An optimal CRC screening test would fulfill a number of important criteria: it would be non-invasive, very sensitive, safe, readily accessible, and cost-effective (9). Along with these attributes, for successful implementation, high levels of adherence and ongoing monitoring to ascertain long-term effects are also necessary (9). The objective of CRC screening is to reduce the incidence and mortality of the disease sustainably by detecting it at a phase where interventions can avert progression or provide curative

results (9). Realization of these objectives demands not just improvement in screening technologies but also means to enhance public participation and streamline health service delivery mechanisms (16).

1.8.1 Stool-based tests

1.8.1.1 Fecal immunochemical test (FIT)

The FIT is a very practical and widely used tool, it detects human hemoglobin in stool via immunochemical antibodies and is thus more specific and trustworthy than the older guaiac-based fecal occult blood tests (gFOBT) (Figure 10) (47). FIT involves only a single stool sample, in contrast to gFOBT, which requires three samples; thus it is more patient-friendly and is associated with improved participation rates (9). FIT, in contrast to gFOBT, is not subject to dietary or medication-related interference, which further increases its reliability (47). Sensitivity for the detection of CRC depends on the threshold for hemoglobin detection, with 79% sensitivity at 20 µg/g and 91% at 10 µg/g (47). Specificity, on the other hand, ranges from 94% to 90%, respectively (47). Although FIT is less efficient at detecting advanced adenomas (with a sensitivity from 15% to 40%), its simplicity and low cost make it the backbone of the population-based CRC screening programs (9). Nevertheless, it needs to be repeated annually or biennially to reach cumulative detection rates that are comparable to a single decennial colonoscopy (47).

1.8.1.2 Multitarget stool DNA test (mtsDNA)

The mtsDNA adds detection of tumor-derived DNA, such as methylated DNA and gene mutations, to FIT to increase its diagnostic yield (16). In a large study of more than 10 000 participants, mtsDNA showed 92% sensitivity for CRC and 42% for advanced adenomas, whereas FIT alone showed 74% and 24%, respectively. Although more sensitive, mtsDNA is less specific (87% compared with 95% for FIT) and is considerably more expensive (~$600 per test) (9). Its complicated stool collection procedure and logistic issues also preclude its widespread implementation (9). False positives are also a problem, which can result in unnecessary follow-up tests (9). Cost-effectiveness studies suggest that FIT and colonoscopy are still more cost-effective options unless the cost of mtsDNA is reduced significantly, or participation rates are increased considerably (9).

1.8.2 Direct visualization tests

1.8.2.1 Colonoscopy

Colonoscopy is the gold standard for CRC detection offering both diagnostic and therapeutic capabilities (Figure 10) (9). It allows for complete visualization of the colon, enabling the removal of precancerous lesions during the procedure (9). Numerous studies have demonstrated its effectiveness, with reductions in CRC incidence and mortality ranging from 29% to 68% (9). However, colonoscopy is invasive, requires extensive bowel preparation, and is associated with risks such as bleeding and perforation (9). These factors, combined with logistical and financial barriers, contribute to lower adherence rates compared to non-invasive tests (9). While it is primarily used as a follow-up procedure after positive stool-based tests, it is also a primary screening tool in the United States, where it is recommended every 10 years for individuals at average risk (9).

1.8.2.2 Flexible Sigmoidoscopy

Flexible sigmoidoscopy, which inspects only the distal colon, provides a less invasive option compared to colonoscopy (9). Large-scale trials have demonstrated reductions in CRC incidence (18-23%) and mortality (22-31%) using this approach (9). It fails to find lesions in the proximal colon and necessitates a follow-up colonoscopy for positive tests (16). The requirement for bowel preparation and the absence of sedation additionally hampers patient compliance (16). As a result, its application has decreased, with nations such as the UK moving to FIT-based screening (9).

1.8.2.3 Computed tomographic colonography (CTC)

The CTC offers 3D or 4D reconstructions for close imaging of the colon (Figure 10) (9). Sensitivity for adenoma detection ≥10 mm is 84% to 92%, with 85% to 96% specificity (9). Although it is less invasive with no sedation requirements, it still requires bowel preparation and subjects patients to radiation (9). Extracolonic finding, which occurs in up 66% of individuals, usually results in further testing, which complicates care and adds expense (9). Although promising, the wide availability of trained radiologists and special equipment has limited its uptake (9).

1.8.2.4 Colon capsule

The colon capsule, which is a wireless camera capsule that the patient swallows, provides a non-invasive way to visualize the colon (48). Technological advancements have enhanced its sensitivity to 88% for adenomas ≥6 mm, with 82% specificity (9). However, difficulties with bowel preparation and the need for expert readers make it

difficult to use in practice (48). Advances in the future, such as artificial intelligence to automate the analysis, may make it easier to use for population-based screening (9).

1.8.3 Blood-based tests

Blood-based tests, such as the SEPT9 assay, represent a less invasive approach to CRC screening (Figure 10) (9). The SEPT9 test detects methylated DNA from the SEPT9 gene, a biomarker for CRC (9). It is particularly useful for individuals who are unable or unwilling to undergo stool-based or imaging tests (9). However, its sensitivity (68%) and specificity (80%) are lower than those of FIT and mtsDNA. The SEPT9 test is FDA-approved but not widely reimbursed or recommended in major guidelines due to its limited efficacy (9). Its role in screening is thus restricted to cases where other methods are not feasible (9).

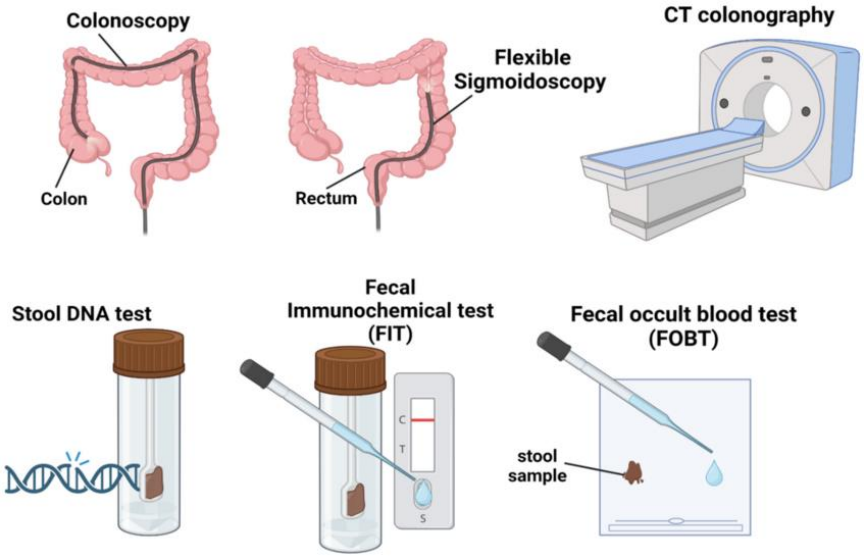


Figure 10 - Types of colorectal cancer diagnosis *(109)*.

Table 1- Adapted summary table from the American Cancer Society of the characteristics of recommended colorectal cancer screening tests (12).

	Benefits	Limitations	Timing
Visual examination			
Colonoscopy	• Examines the entire colon • Can biopsy and remove polyps • Can diagnose other diseases • Required for abnormal results from all other tests	• Full bowel cleansing • Can be expensive • Sedation is usually needed, necessitating a chaperone to return home • Patient may miss a day of work • Highest risk of bowel tears or infections compared with other tests	10 years
Computed tomographic colonography (CTC)	• Examines the entire colon • Fairly quick • Few complications • No sedation needed • Noninvasive	• Full bowel cleansing • Cannot remove polyps or perform biopsies • Exposure to low-dose radiation • Colonoscopy necessary if positive • Not covered by all insurance plans	5 years
Flexible sigmoidoscopy	• Fairly quick • Few complications • Minimal bowel preparation • Does not require sedation or a specialist	• Partial bowel cleansing • Views only one-third of the colon • Cannot remove large polyps • Small risk of infection or bowel tear • Slightly more effective when combined with annual fecal occult blood testing • Colonoscopy necessary if positive • Limited availability	5 years

Table 1 – Continuation.

	Benefits	Limitations	Timing
Stool testes (low-sensitivity stool tests, such as single-sample FOBT done in the doctor´s office or toilet bowl tests, are not recommended)			
Fecal immunochemical test	• No bowel cleansing or sedation • Performed at home • Low cost • Noninvasive	• Requires multiple stool samples • Will miss most polyps • May produce false-positive test results • Slightly more effective when combined with a flexible sigmoidoscopy every 5 years • Colonoscopy necessary if positive	Annual
High-sensitivity guaiac-based fecal occult blood test (gFOBT)	• No bowel cleansing • Can be performed at home • Requires only a single stool sample • Noninvasive	• Will miss most polyps • More false-positive results than other tests • Higher cost than gFOBT and RT • Colonoscopy necessary if positive	Annual
FIT-DNA test (Cologuard®)	• No bowel cleansing • Can be performed at home • Requires only a single stool • Noninvasive	• Will miss most polyps • More false-positive results than other tests • Higher cost than gFOBT and RT • Colonoscopy necessary if positive	3 years

1.8.4 Tumor biomarkers

The exploration of biomarkers for CRC diagnosis has paved the way for more non-invasive, precise, and accessible diagnostic methods (49). Biomarkers are biological molecules found in blood, urine, or other body fluids and tissues that act as indicators of normal or pathological biological processes, such as cancer or disease progression (50). These changes may involve gene expression, proteomic, and metabolomic characteristics specific to factors such as genetic mutations, transcription alterations, or post-translational modifications (51). Various types of biomarkers exist, including proteins, nucleic acids (such as microRNAs), antibodies, peptides, and volatile organic metabolites (VOMs), which can be detected in bodily fluids such as blood, urine, saliva, and feces, enabling non-invasive and simple assessments (50). Biomarkers are used in different clinical contexts, including predicting disease development, detecting hidden cancers, distinguishing between benign and malignant lesions, assessing the prognosis of cancer patients, and monitoring recurrences and therapeutic responses (50). A common approach in the search for cancer markers is identifying substances that are absent or found in abnormal concentrations in cancer patients, aiding in the identification of useful biomarkers for preventive medicine, early diagnosis, and treatment selection (50).

1.8.4.1 MicroRNAs and colorectal adenoma

MicroRNAs (miRNAs) are small non-coding RNA molecules, about 18-25 base pairs, that control gene expression by binding to the 3' untranslated region (of messenger RNA (mRNA), preventing translation (52). These molecules affect cell growth, differentiation, and apoptosis and are part of various cell processes (50). In cancer, miRNAs control oncogenic and tumor-suppressor pathways, affecting cell migration, invasion, metastasis, immune system interactions, and angiogenesis (52). In CRC miRNAs are involved in disease progression, including the activation of epithelial-mesenchymal transition (52). One of the strengths of miRNAs is their stability in biological fluids like plasma, stool, urine, saliva, and tissue samples, being resistant to degradation by endogenous RNases (52). This makes miRNAs valuable candidates for diagnostic, prognostic, and predictive biomarkers in CRC (50).

1.8.4.2 MicroRNAs and colorectal cancer

Several studies have investigated miRNA expression profiles at different stages of CRC (53). For example, high plasma levels of miR-92a and linked to advanced adenomas with sensitivities of 62.2.% and 64.9% and specificities of 84.7% and 81.4%, respectively (50). Their combination enhances diagnostic performance beyond single marker's sensitivities and specificities (50). In contrast, miRNAs like miR-760 and miR-601 are downregulated in advanced adenomas. miRNAs panels comprising miR-532-3p, miR-195, miR-331, miR-17, miR-142-3p, miR-15b, miR-532, and miR-652 have shown 88% sensitivity and 64% specificity in differentiating advanced adenomas (50). Likewise, high levels of miR-21, miR-29a, and miR-125b have been linked to early CRC and can differentiate advanced from non-advanced disease stages (50). Specific miRNAs like miR-141 are linked to advanced-stage CRC (stage IV) with a sensitivity of 66.7% and specificity of 84% for discriminating stage IV from stages I-II (50). High levels of miR-29a have also been implicated in the early detection of liver metastases (50). Prognostic miRNAs like miR-221 are linked to unfavorable survival outcomes and are associated with p53 expression (50). High miR-203 levels are associated with tumor stage, lymph node metastasis, peritoneal and distant metastases, and overall poorer survival (50). Other miRNAs like miR-19A are linked to poor response to chemotherapy, in particular, resistance to FOLFOX regimen (50). In contrast, high miR-204-5p levels are associated with enhanced sensitivity to 5-fluorouracil (5-FU) due to the downregulation of RABB22A, a RAS oncogene family member (53). Some miRNAs are also predictive of resistance to targeted therapies (53). For instance, downregulation of miR-20a, miR-130, miR-145, miR-216, and miR-372 is associated with resistance to oxaliplatin, with sensitivity and specificity of 92% and 88%, respectively (50). Elevated miR-126 levels predict poor response to bevacizumab, and a 30% increase in miR-155-5p during treatment is linked to improved progression-free and overall survival (50). These findings illustrate the promise of miRNAs as diagnostic, prognostic, and predictive biomarkers (53). However, existing evidence is limited by heterogeneity of study designs, lack of standardization, and lack of large-scale validation (50).

1.8.4.3 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are RNA transcripts exceeding 200 nucleotides in length, accounting for 68% of the human transcriptome (50). They are also involved in diverse cellular processes, including transcriptional regulation, post-

transcriptional mRNA control, protein stability, subcellular organization, and epigenetic regulation (50). These processes play critical roles in cell proliferation, migration, and survival, and lncRNAs show differential expression across tissues and cell types (50). Several lncRNAs have been identified as either oncogenic or tumor suppressors in CRC (50). Tumor suppressive lncRNAS often interact with p53, a key tumor suppressor gene, for example, the lncRNA LED is upregulated by p53, and its downregulation is associated with CRC and other cancers (50). Similarly, lincRNA-p21 enhances radiotherapy sensitivity and promotes apoptosis in CRC, while its depletion accelerates disease progression (50). Another lncRNA, DINO, is involved in the p53-dependent DNA-damage response and exhibits low expression in CRC (50). On the oncogenic side, lncRNAs like MALATI and HOTAIR are implicated in CRC tumor progression and metastasis. Increase MALATI expression correlates with poor prognosis, while HOTAIR is associated with advanced tumor stages and metastasis (50).

1.8.4.4 Volatile organic metabolites

Low molecular weight VOMs show promise as biomarkers of cancer (54). VOMs are indicators of inflammatory and oxidative processes and are classified as volatile based on their physical and chemical properties (54). They include a wide variety of compounds such as alkanes, alkenes, ketones, and aldehydes. Although the metabolic origin and function of most VOMs remain unclear, they are the products of both endogenous metabolic processes and exogenous sources, i.e., environment and diet (54). Only VOMs of endogenous origin can be considered as indicators of pathologies (55). Analysis of cancer-related compounds is a new, promising medical diagnostic area and is gaining more scientific interest since it is not invasive, low-cost, and reproducible (55). It is based on the alteration of cells associated with tumor development, reflecting genetic and protein changes causing VOM emission (55).

Urinary biomarkers have been useful indicators of health throughout history (55). Nowadays, urine analysis is an essential diagnostic tool, applied in physical, chemical, and microscopic examinations (55). Urine is a promising source of volatile biomarkers, whose profile depends on internal and external factors such as genetics, diet, fluid intake, daily habits, and collection conditions (55). The majority of compounds in urine are endogenous metabolites, such as sulfur volatile compounds, which can be produced in tissues or by bacteria in the gastrointestinal tract (55). External-origin VOMs, such as environmental pollutants and food- and hygiene product-related compounds, are also

present in urine (55). Urine analysis can provide valuable information on metabolite levels, metabolic defects, drug action, and chemical exposure (55). Urine also utilized to track drug use, doping, diet, and food contaminant intake, as urine contains both the target substance and its metabolites (55). Through urine metabolomic profiling, one can obtain an "endogenous marker fingerprint" for every individual, representing such factors as age, sex, health status, and diet (55).

The identification of volatile biomarkers in urine for disease diagnosis is a promising but still developing area. One of the advantages of this approach is the ease of collection, which is a painless, non-invasive procedure that can be repeated as often as necessary (56). Urine also offers advantages over other body fluids, like exhaled air or serum, as it contains numerous important metabolites in concentrations similar to those in plasma but with less protein complexity and less requirement for pre-treatment (56). However, one of the challenges of using urine in metabolic studies is the huge biological variation in its composition, and thus, one must define intra- and interindividual variation in metabolites in the control population (56). Though research in this area has advanced, additional studies are needed to standardize urine metabolomics and confirm the findings obtained so far (56). Urinary biomarkers for CRC are mainly comprised of volatile organic compounds and metabolites excreted in urine (56). Analysis of such compounds can identify specific patterns for cancer presence, and thus, patients with the disease can be distinguished from healthy controls (57). The presence of urinary biomarkers offers a non-invasive strategy for screening and early detection (56). Furthermore, urinary biomarkers can be utilized to follow up the response to treatment as well as recurrence of the disease, providing a valuable tool for clinical management of this disease (56). Techniques like headspace solid-phase microextraction (HS-SPME) combined with gas chromatography-mass spectrometry (GC-MS) have been utilized to identify VOMs as potential volatile biomarkers of cancer (55,58).

Qiu et al. (59) conducted a profiling analysis of urine samples from 60 CRC patients at various stages of the disease and 63 HCs (59). The patients had previously undergone serum experiments (59). The study included preoperative and postoperative urine samples collected on the seventh day after surgery to measure changes in volatilomic profile (59). The analysis was performed using GC-MS with solvent extraction and ethyl chloroformate derivatization (59). A total of 187 VOMs were identified that separated CRC patients from HCs in 90% of samples (59). Sixteen possible biomarkers were identified between preoperative CRC patients and HCs, with some

metabolites decreasing, such as 3-methylhistidine, histidine, and citric acid, and others increasing, such as 5-hydroxytryptophan, 5-oxoproline, p-cresol, and phenyl acetate (59). After surgery, there was a return to healthy levels of VOMs, such as 5-hydroxytryptophan, 2-hydroxyhippuric acid, succinic acid, and phenylacetylglutamine (59). In addition, six metabolites exhibited characteristic expression levels that could distinguish different stages of CRC (59). For instance, elevated levels of indoleacetic acid were observed in stage I, p-hydroxyphenylacetic acid in stage II, and 5-hydroxyindoleacetic acid in stage III (59). The maximum level of 2-methylpropanoic acid was observed in stage II, with a drastic fall in stage IV patients (59). In addition, 21 VOMs predominantly amino acids and phenyl-containing compounds, facilitated the separation of preoperative and postoperative CRC patients, likely reflecting metabolic changes due to the surgical procedure (59).

In a study conducted by Silva et al. (60), the urinary volatilomic profile of 54 subjects was established and explored using HS-SPME/GC-MS. The study included 33 cancer patients – 14 with leukemia, 12 with CRC, 7 with lymphoma, and 21 HCs (60). A total of 82 VOMs from different chemical families were identified (60). Preoperative CRC patients were characterized by the presence of benzene derivatives, terpenoids, and phenolic compounds (60). High levels of VOMs such as p-cymene, anisole, γ-terpinene, bornylene, dimethyl disulfide, 4-methylphenol, 1,2-dihydro-1,1,6-trimethylnaphthalene, 1,4,5-trimethylnaphthalene, and 2,7-dimethylquinoline were observed, and reduced levels of VOMs such as 1-octanol, heptanal, hexanal, and dimethyl disulfide when compared to HCs (60). Notably, 4-methyl-2-heptanone was exclusively detected in CRC patients (60).

Cheng et al. (61) employed solvent extraction and derivatization followed by gas chromatrogray-time-of-flight mass spectrometry (GC-TOF-MS) to identify metabolite markers for CRC (61). The study involved 103 patients with CRC and 101 HCs, and 163 VOMS were detected (61). Statistical modeling through univariate and multivariate analysis led to the identification of 19 VOMs as potential biomarkers (61). A predictive model composed seven metabolites: citric acid, hippuric acid, p-cresol, 2-aminobutyric acid, myristic acid, putrescine, and kynurenate was able to distinguish CRC patients from HCs (61). The area under the receiver operating characteristic curve (AUROC) was 0.993, while the sensitivity and specificity for the training set were 97.5% and 97.6 %, respectively (61). The model also performed well with a test set, with an AUROC of 0.998, sensitivity of 97.5%, and specificity of 100% (61).

Arasaradnam et al. (62) used a field asymmetric ion mobility spectrometer (FAIMS) to examine urine samples from 83 CRC patients and 50 HCs (62). Fisher discriminant analysis of FAIMS data revealed that the technique had 88% sensitivity and 60% specificity in the discrimination of CRC patients from HCs (62). Concurrently, an in-tube extraction (ITEX)-GC-MS experiment was also conducted, but no specific chemical markers in CRC patients were identified compared to HCs (62). However, nine volatile compounds associated with CRC were identified, and their peaks were matched to 26 NIST targets (62).

Liesenfeld et al. (63) analyzed urine samples of 170 samples, grouped into four groups: pre-surgery (79), post-surgery (within a few days) (9), six months after surgery (46), and 12 months after surgery (36) CRC patients (63). GC-MS detected a total of 82 VOMs, and these metabolites were utilized to separate pre-surgery from post-surgery CRC patients (63). Several important metabolites have been attributed to changes in the gut microbiome due to CRC surgery, such as compounds like 2,3-butanediol, pyrogallol, hydroquinone, and maleamic acid (63). Ten metabolites, such as hydroxyproline dipeptide (Hyp-Hyp9), p-cresol-β-O-glucuronide, and hippuric acid, were found to differentiate CRC patients by cancer stage (63). Higher concentrations of Hyp-Hyp and p-cresol-β-O-glucuronide were observed in patients with intermediate stage disease (II-III), while these metabolites, along with hippuric acid and p-cresol, were higher in patients with early stage (I) (63). In contrast, these metabolites were less concentrated in patients with advanced-stage CRC (63). A multivariate model using 20 metabolites was constructed to separate pre-surgery from post-surgery CRC patients, with an AUROC of 0.89 (95% Cl: 0.84-0.95) (63).

Mozdiak et al. (64) utilized FAIMS in the creation of a VOM-based screening assay for CRC and adenomas, as well as gas chromatography coupled with ion mobility spectrometry (GC-IMS) (64). The study involved a total of 163 patients. FAIMS analysis revealed high sensitivity (1.9) and specificity (0.92) in the discrimination between CRC patients and HCs (64). When CRC cases and adenomas were included, however, the accuracy of the test was significantly reduced (64). No characteristic VOM biomarkers were identified using GC-IMS (64). In summary, VOM signatures facilitated efficient discrimination between premalignant and malignant patient, with greater sensitivity and uptake than the fecal occult blood test (FOBT) in bowel cancer screening (64).

Table 2 - Adapted summary table from a review of all studies investigation of CRC in urine *(65)*.

Subjects	Analytical Technique	Main Analytes	GC Column	Statistical Approach	Ref
• 60 CRC: Stage I: 7 Stage II: 23 Stage III: 21 Stage IV: 9 • 63 HCs	Solvent extraction with chloroform and derivatization with ECF + GC-MS	SNM: amino acids, organic acids	DB-SMS (30 m × 250 µm i.d., 0.25 µm film thickness)	PCA, OPLS-DA	(59)
• 12 CRC • 21 HCs	HS-SPME with CAR/PMDS (75 µm) + GC-MS	SVM: hidrocarbons; aldehydes; sulfur compounds	BP-20 (30 m × 0.25 mm ID, 0.25 µm film thickness)	one way ANOVA, LSD, PCA	(60)
• 103 CRC Stage I: 24 Stage II: 45 Stage III: 27 Stage IV: 5 • 101 HCs	Solvent extraction with methanol and derivatization with methoxamine (in pyridine) and BSTFA (1% TMCS) + GC-TOFMS	SNM: amino acids; organic acids; saccharides	DB-5MS (30 m × 250 µm I.D., 0.25 µm film thickness;	PCA, OPLS-DA, ROC curve, Student´s *t*-test, Wilcoxon-Mann-Whitne test	(61)
• 83 CRC • 50 HCs	ITEX + GC-MS	SVM: ketones; aldehydes; nitrogen compounds	Rxi-624Sil c(20 m length × 0.18 mm ID, 1.0 µm film thickness)	FDA, KNN method	(62)
Total for GC-MS and ^{1}H-NMR is 199 CRC: • CRC pre-surgery: 97 • CRC post-surgery: 12 • CRC follow up: 90	Solvent extraction with methanol and derivatization with methoxyamine (in pyridine) and BSTFA (1% TMCS) + GC-MS	SNM: alcohols; amino acids; organic acids; saccharides	HP-5-MS (30 m film thickness 0.25 mm; 0.25 µm film thickness)	Wilcoxon-Mann-Whitney tests, PLS-DA, one-way ANOVA, ROC curve	(63)

Table 2 - Continuation.

Subjects	Analytical Technique	Main Analytes	GC Column	Statistical Approach	Ref
• CRC pre-surgery: 163 • CRC follow-up: 137	Solvent extraction with methanol and derivatization with methoxyamine (in pyridine) and BSTFA (1% TMCS) + GC-MS	SNM: amino acids	HP-5MS (30 m film thickness 0.25 mm; 0.25 µm film thickness)	One-way ANOVA, Pearson Chi-squared test, Pearson´s partial correlation coefficients, Cox proportional hazard models	(66)
• 12 CRC • 80 adenomas • 14 diverticular diseases • 5 hemorrhoids • 14 inflammatory bowel disease • 1 excluded • 37 HCs	Not specified + GC-IMS	Undetermined	Not specified	ROC curve, Sparse logistic regression, Random Forest, Gaussian process classifier, Support vector machine, Neural network	(64)

Abbreviations: ANOVA – Analysis of variance; BSTFA – N,O-bis(treimethylsily)trifluoroacetamide; CAR/PMDS – Carboxen/Polydimethylsiloxane; CRC – Colorecetal Cancer; ECF – Ethyl Chloroformate; FDA – Fisher Discriminant Analysis; GC – Gas Chromatography; GC-IMS – Gas Chromatography Coupled with Ion Mobility Spectrometry; GC-MS – Gas Cromatography-Mass Spectrometry; HCs – Healthy Controls; HS-SPME – Headspace Solid-phase Microextraction; KNN – K-nearest neighbors algorithm; LSD – Least Significant Difference; OPLS-DA – Orthogonal partial least squares discriminant analysis; PCA – Principal Component Analysis; PMDS/DVB – Polydimethylsiloxane/Divinylbenzene; PLS-DA – Partial least squares discriminant analysis; ROC – Receiver operating characteristic; SNM – Screening of nonvolatile metabolites; SVM – Sreening of volatile metabolites; TMCS – Trimethyl chloride; 1H-NMR – Proton Nuclear Magnetic Resonance.

Table 3 - Summary of potential urinary small organic molecules biomarkers for colorectal cancer (55,65).

Volatile organic metabolite	Formula	CAS Number	Reference
1-Octanol	$C_8H_{18}O$	111-87-5	(60)
Hexen-1-ol	$C_6H_{12}O$	928-95-0	(62)
p-cresol	C_7H_8O	106-44-5	(59–61,63)
Phenol	C_6H_5OH	108-44-5	(61)
Guaiacol	$C_7H_8O_2$	90-05-1	(63)
2,3-Butanediol	$C_4H_{10}O_2$	513-85-9	(63)
Acetaldehyde	C_2H_4O	75-07-0	(62)
Heptanal	$C_7H_{14}O$	111-71-7	(60)
Hexanal	$C_6H_{12}O$	66-25-1	(60,62)
3-Heptanone	$C_7H_{14}O$	106-35-4	(60,62)
4-Heptanone	$C_7H_{14}O$	123-19-3	(62)
2-Pentanone	$C_5H_{10}O$	107-87-9	(62)
2,3-Butanedione	$C_4H_6O_2$	431-03-8	(62)
Phenyl acetate	$C_8H_8O_2$	122-79-2	(59)
Anisole	C_7H_8O	100-66-3	(60)
p-cymene	$C_{10}H_{14}$	99-87-6	(60)
γ-terpinene	$C_{10}H_{16}$	99-85-4	(60)
Citric acid	$C_6H_8O_7$	77-92-9	(59)
2-hydroxyisobutyric acid	$C_4H_8O_3$	594-61-6	(63)
Isocitric acid	$C_6H_8O_7$	320-77-4	(59)
3-hydroxybutanoic acid	$C_4H_8O_3$	300-85-6	(63)
Latic acid	$C_3H_6O_3$	50-21-5	(63)
Myristic acid	$C_{14}H_{28}O_2$	544-63-8	(61,67)
Pyruvic acid	$C_3H_4O_3$	127-17-3	(61)
Succinic acid	$C_4H_6O_4$	110-15-6	(59)
Fumaric acid	$C_4H_4O_4$	110-17-8	(61)
2-Methoxythiophene	C_5H_6OS	16839-97-7	(60)
Dimethyl disulfide	$C_2H_6S_2$	624-92-0	(62)
Putrescine	$C_4H_{12}N_2$	110-60-1	(61)
Dimethyl-thiourea	$C_3H_8N_2S$	534-13-4	(62)

Table 3 – Continuation.

Volatile organic metabolite	Formula	CAS Number	Reference
Allyl isothiocynate	C_4H_5NS	57-06-7	(62)
1,2,4-Trimethylbenzene	C_9H_{12}	95-63-6	(60)
1,2-Dihydro-1,1,6-trimethyl-naphthalene	$C_{13}H_{16}$	30364-38-6	(60)
1,4,5-Trimethyl-naphthalene	$C_{13}H_{14}$	2131-41-1	(60)
2,7-Dimethyl-quinoline	$C_{11}H_{11}N$	93-37-8	(60)
2-Methyl-3-phenyl-2-propenal	$C_{10}H_{10}O$	101-39-3	(60)
Bornylene	$C_{10}H_{16}$	464-17-5	(60)
Furan	C_4H_4O	110-00-9	(68)
Carbon disulfide	CS_2	75-15-0	(68)
4-tert-butylphenol	$C_{10}H_{14}O$	98-54-4	(68)

1.9 Colorectal cancer staging

CRC staging is a key process in patient diagnosis and treatment planning because it determines the extent of cancer spread and provides a platform for prognosis evaluation (69). Staging accurately enables clinicians to tailor therapeutic plans and evaluates outcomes optimally (69). The most popular system applied in the staging of CRC is the American Joint Committee on Cancer (AJCC) TNM staging system, which analyzes three critical components: Tumor (T): this component analyzes the extent of tumor invasion into the colon or rectum layers (69). The wall contains various layers, including the mucosa (innermost layer), submucosa, muscularis propria, and serosa (outermost layer) (69). The tumor structures or organs; Nodes (N): this component analyzes whether cancer has extended to surrounding lymph nodes and the number of nodes involved (69). It ranges from N0 (no lymph node involvement) to N2 (involvement of four or more regional lymph nodes); Metastasis (M): the metastasis component analyzes whether the cancer has extended to distant organs, such as the liver or lungs (70). M0 denotes no distant metastasis, while M1 denotes metastatic disease (69). Integrating the TNM designations, CRC is classified into stages from 0 to IV: stage 0, or carcinoma in situ, is cancer that is confined to the innermost layer of the colon or rectum (70). It has not extended deeper into tissues or to lymph nodes or distant locations. Treatment at this stage typically consists of surgical resection locally; stage II is further classified into IIA, IIB and IIC (70). In IIA, cancer has spread through the wall of colon or rectum but has not invaded nearby tissues or lymph nodes (70). In IIB, cancer has grown through the wall and invaded nearby tissues (70). In IIC, cancer has spread through the serosa and invaded nearby structures (70). Treatment often involves surgery, and adjuvant chemotherapy may be indicated in more dangerous circumstances; stage III involves lymph node involvement with no distant metastasis and is subclassified into: IIIA, IIIB and IIIC (70). In IIIA, cancer has spread to 1-3 nearby lymph nodes but is still within the inner layers of the colon or rectum (70). In IIIB, cancer has spread through the wall of the colon and extended to 1-3 lymph nodes (70). In IIIC, cancer has spread to 4 or more lymph nodes regardless of tumor depth (70). Stage III requires a combination of surgery and adjuvant chemotherapy to improve prognosis; stage IV indicates distant metastasis and is subclassified into: IVA, and IVB (70). In IVA, cancer has spread to two or more distant organs or regions. Stage IV treatment often involves systemic chemotherapy, targeted therapies, and, in selected circumstances, surgical removal of metastatic sites (70).

The TNM staging system provides a universal system for the evaluation of CRC. Accurate staging involves an integration of physical examination, imaging studies (e.g., CT, MRI, or PET scans), and histopathological evaluation of biopsy or surgical specimens (69). Staging is very important for prognosis, with early-stage cancers (Stage 0-II) having significantly better survival rates compared to advanced stages (III and IV) (69). Emerging technologies, such as molecular profiling and biomarkers, are increasingly being integrated into staging to further define tumor biology and guide personalized treatment approaches (69).

1.10 Colorectal cancer conventional therapies

The treatment of CRC is multidisciplinary and is primarily founded on conventional treatment modalities, which include surgery, chemotherapy, radiotherapy, and the use of antiangiogenic agents (71,72). These treatments are personalized to the patient according to the stage of cancer, location of the cancer, and other clinical aspects. Each treatment modality plays a critical role in improving survival, preventing recurrence, and effectively controlling the disease (71).

1.10.1 Surgery

Surgery remains the cornerstone of treatment in colorectal cancer, particularly in early disease (73). The primary goal of surgery is the complete resection of the tumor with adequate margins, as well as the examination of regional lymph nodes to obtain an accurate pathological staging (73). There have been numerous advances in surgical technique over the years. Conventional open surgery, though still widely performed, has been supplemented by minimally invasive approaches, such as laparoscopic surgery (74). Laparoscopic surgery has a number of advantages over conventional open surgery, including smaller wounds, shorter hospital stays, less postoperative pain, and quicker recovery times (74). It has been shown that laparoscopic colectomy, when compared with open surgery, is associated with less blood loss, fewer complications, and quicker recovery (74). There have been some concerns, however, with the potential increased risk of tumor cell dissemination with minimally invasive approaches, although evidence for this is conflicting (71). The type of surgical procedure employed depends on the site of the tumor (74). For instance, right-sided colon cancers are typically managed with a right hemicolectomy, whereas left-sided cancers can be managed with a left hemicolectomy or

sigmoidectomy (74). Rectal cancers, due to their proximity to the anal sphincter, present more challenges (74). For these, the treatment may range from low anterior resection with primary anastomosis to abdominoperineal resection with permanent colostomy, depending on the proximity of the tumor to the anal sphincter and whether it is possible to preserve the sphincter (74).

1.10.2 Chemotherapy

Chemotherapy is one of the cornerstones in the management of CRC, particularly in advanced or metastatic disease (72). Chemotherapy is aimed at killing cancer cells that have escaped the primary tumor and reducing the risk of recurrence (72). In early-stage disease, chemotherapy is not routinely indicated unless the tumor has high-risk characteristics, such as poor differentiation, perineural invasion, or lymph node involvement (72). In these cases, adjuvant chemotherapy can eliminate residual cancer cells and dramatically improve survival (72). Some of the standard chemotherapy regimens for treating CRC, including 5-fluorouracil (5-FU, leucovorin, irinotecan), are particularly useful in patients with metastatic disease (72). Chemotherapy is also an important treatment in patients with stage III colon cancer, where it is administered after surgery to eliminate micrometastases (72). The use of chemotherapy in stage II colon cancer is still restricted, as the risk of recurrence is lower at this stage (72).

1.10.3 Radiotherapy

Radiotherapy is an important component in the treatment of rectal cancer, particularly in locally advanced disease (72). In contrast to colon cancer, which is not generally responsive to radiation, rectal cancer is often treated with radiotherapy to reduce the risk of local recurrence, particularly if the tumor is large or near vital structures such as the anal sphincter (72). In rectal cancer, radiotherapy is often used in the neoadjuvant setting, before surgery, to reduce the tumor size and facilitate surgical resection (72). This can make an inoperable tumor resectable or reduce the risk of recurrence by eliminating any residual cancer cells after surgery (72). Neoadjuvant chemoradiotherapy, the combination of chemotherapy and radiotherapy, has been shown to improve outcomes in patients with rectal cancer, particularly for those with locally advanced disease (T3 or T4 tumors) or lymph node involvement (72). The radiation dose and the type of chemotherapy drugs used are individualized according to the tumor size and location and

other factors (72). In some cases, radiotherapy may be used postoperatively, particularly if the tumor was not completely resected or if there is a high risk of local recurrence (72).

1.10.4 Antiangiogenic therapy

In the last several years, antiangiogenic therapies have emerged as a useful treatment approach in metastatic CRC (75). These therapies work by interrupting the blood supply to the tumor, inhibiting its growth and spread (75). One of the most common antiangiogenic agents is bevacizumab, a monoclonal antibody against vascular endothelial growth factor (VEGF), a protein that stimulates the growth of new blood vessels (75). Bevacizumab inhibits the blood supply to the tumor by blocking VEGF, limiting tumor growth and spread (75). Studies have shown that adding bevacizumab to chemotherapy can lead to significant improvements in survival and progression, particularly in patients with advanced disease (75). In addition, antiangiogenic therapies targeting other pathways, such as the epidermal growth factor receptor (EGFR), with monoclonal antibodies such as cetuximab and panitumumab, have also demonstrated benefits, particularly in tumors with wild-type KRAS genes (75).

1.11 Follow-up

Follow-up care is a fundamental aspect in the management of CRC, intending to improve survival, detecting recurrences, and monitoring for new primary cancers or metachronous polyps (76). The benefit of follow-up is that it enables early detection and intervention, which are crucial to improving long-term outcomes in patients with CRC (76). Vigilant monitoring guarantees that recurrence or a new lesion can be identified and treated at an early stage, when the possibilities for successful outcomes are highest (77). The stage of CRC at diagnosis has a significant impact on the determination of the patient's prognosis and the direction of follow-up care (77). Most recurrences occur in the first four years following surgical resection, and as such, the five-year survival rate is a reliable indicator of long-term success (77). The survival rates highly depend on the stage of the disease at diagnosis, ranging from over 90% survival for stage I, between 70 and 85 % for stage II, 25 to 80% for stage III, and less than 10% for stage IV. These figures highlight the critical need for stringent follow-up, particularly for patients presenting with advanced disease or for those with high-risk features (77). Advances in treatment, including refinement in surgical techniques and the introduction of targeted

therapies, have significantly improved the survival rates for patients with metastatic CRC (76). For example, in the past, the life expectancy of patients with small liver metastases was only 6 to 9 months (76). Presently, with the integration of advanced therapies and early detection, the median survival has improved to 24 to 30 months (76). In addition to disease stage, prognosis and follow-up care are also determined by specific molecular markers (76). Biomarkers such as microsatellite instability (MSI) and chromosomal deletions provide valuable information on tumor behavior and recurrence risk (76). These markers also direct individualized treatment and follow-up strategies, allowing clinicians to tailor care to the patient's unique needs (76). The primary objective of follow-up care is the early detection of recurrences, metachronous polyps, or new primary cancers. This is achieved through medical examination at regular intervals, laboratory tests, endoscopic procedures, and imaging studies (76). Colonoscopy is a cornerstone of follow-up care and is typically recommended one year after curative surgery (76). Repeat colonoscopies are performed at three- to five-year intervals if no significant findings are observed. In high-risk patients or those with features of concern, colonoscopy frequency can be increased to allow closer surveillance (76). Imaging studies, such as computed tomography (CT) scans of the abdomen and pelvis, are instrumental in monitoring for local recurrence or distant metastases (76). Laboratory tests, such as monitoring of carcinoembryonic antigen (CEA) levels, are also a standard component of follow-up care (76). Elevated CEA levels can be indicative of recurrence or metastasis, prompting further evaluation (76). Adherence to follow-up recommendations is often suboptimal, particularly in patients with logistical, financial, or psychological impediments (77).

1.12 Objective of the thesis and experimental process

The main goal of this thesis was to determine the urinary volatilomic profile of CRC patients and HC through a non-invasive approach, aiming to identify potential biomarkers for CRC diagnosis and management.

Urine samples from cancer-free individuals (HC, n= 17) and CRC patients (n= 19) were analyzed using HS-SPME/GC-MS, a sensitive and reliable methodology. The resulting data matrix was then subjected to advanced statistical analysis, employing multivariate techniques such as PCA and PLS-DA, to identify and define potential molecular biomarkers.

To achieve this objective, the specific aims are:

- Characterize the study population, including individuals with pathology, assessed from a clinical perspective, and the HC group.
- Apply the HS-SPME methodology followed by GC-MS analysis, previously optimized by the University of Madeira's research group for the study of volatile biomarkers in other cancer types.
- Establish a methodological, chemical, and biochemical foundation to support the development of a rapid, cost-effective, and non-invasive diagnostic tool for CRC detection.

2 Chapter II. Materials and methods

2.1 Extraction and analysis conditions

2.1.1 Materials and reagents

Sodium chloride (NaCl, 99.5%) was acquired from Panreac AppliChen ITW Reagents (Barcelona, Spain) to promote VOMs' salting-out. Hydrochloric acid (HCl, 37%) and 3-octanol (internal standard (IS), 99%) from Sigma-Aldrich (St. Louis, MO, USA) were used to prepare the solutions HCl 5 M and 3-octanol 5 parts per million (ppm), respectively, in ultrapure water obtained from a Mili-Q water purification system (Millipore, Bedford, PA, USA) (78). The helium of purity 5.0 (Air Liquide, Algés, Portugal) was used as a GC carrier gas. The glass vials, SPME holder for manual sampling, and fibre were purchased from Supelco (Merck KGaA, Darmstadt, Germany). The SPME device included a fused silica fibre coating partially cross-linked with 50/30 µm DVD/CAR/PMDS (divinylbenzene/carboxen/polydimethylsiloxane), which was conditioned at 270 ºC for 30 min before its use, according to the manufacturer's guidelines (78). Pure standards were used to confirm the VOMs identified through the National Institute of Standards and Technology (NIST) library and were acquired in their maximum purity available, which included octanoic acid (99 %), 2-ethyl-1-hexanol (98 %), and 2-pentanone (98 %) from Acros Organics (NJ, USA), dimethyl disulfide (> 98 %) and 4-heptanone (96 %) from Fluka Analytical (Honeywell Specialty Chemicals Seelze GmbH, Hannover, Germany), and 3-hexanone (98%) from Aldrich Chemistry (St. Louis, MO, USA) (78).

2.1.2 Subjects and urine sampling

This study involved 17 healthy individuals without any known pathology (HC group), and 19 CCR patients (CRC group) (Table 4). The HC group consists of men and women with no history of CRC. The urine samples from the HC group were collected from family members and friends to support the objectives of this thesis, with informed consent. The urine samples from oncological patients were collected in the Gastrointestinal Unit of SESARAM. The research protocol was approved by the local Ethics Committee (CES18/2022) and all participants were fully informed of the objectives of this study and signed an informed consent before being sampled.

Participants were instructed to collect the urine into a 100 mL sterile polyvinyl chloride (PVC) container that was provided to them. Upon collection, each sample was centrifuged at 4000 rpm for 20 min, aliquoted in 15 mL vials, and stored at -80 ºC to avoid any kind of degradation, such as enzymatic activity or oxidation of VOMs, until analysis. All the data collected from the participants was processed to respect confidentiality, privacy, and the ethical principles inherent to any research study involving human subjects.

Table 4 - Demographic and clinical characteristics of participants in the control and colorectal cancer (CRC) groups.

Variable	HC (n=17)	CRC (n=19)
Sex		
Male	13	15
Female	4	4
Mean Age ± SD (years)	68.7 ± 12.1	67.1 ± 11.7
Smokers		
Ever smokers	5	7
Never smokers	12	12
Current medication [n (%)]		
Yes	15 (88.2 %)	12 (63.2 %)
No	2 (11.8 %)	7 (36.9 %)
Tumor location [n (%)]		
Cecum	-	2 (10.5 %)
Ascending colon	-	2 (10.5 %)
Sigmoid colon	-	5 (26.3 %)
Recto-sigmoid transition	-	1 (5.2 %)
Rectum	-	9 (47.3 %)

Data are presented as absolute numbers and percentages [n (%)] or as mean ± standard deviation (SD), where applicable. Tumor location data are provided only for CRC patients.

2.1.3 HS-SPME procedure

The HS-SPME extraction was performed according to previously optimized conditions for the establishment of the urinary volatilomic profile (79). Instead of the CAR/PDMS fibre, a DVD/CAR/PDMS fibre was used to extract a wider range of VOMs. Briefly, 4 mL aliquots of urine adjusted to pH 1-2 with 500 µL of HCl (5 M) were transferred to an 8 mL sampling glass vial, and 0.8 g of NaCl and 10 µL of 3-octanol (5 ppm) were added, with stirring at 800 rpm (79). The vial was placed in a thermostat bath adjusted to 50.0 ± 0.1 ºC and then the SPME fibre was inserted in the vial for 60 min (79). After sampling, the SPME fibre was withdrawn into the needle, removed from the vial, and inserted in the injector port (250 ºC) of the GC-MS system for 6 min, for desorption of the analytes. Each sample was analyzed in triplicate (79).

2.1.4 GC-MS analysis

The GC-MS analysis was performed with a Agilent Technologies 6890N Network (Palo Alto, CA, USA) (79). The gas chromatograph was equipped with a 60 m × 0.25 mm ID × 0.25 µm film thickness, BP-20 (SGE, Dortmund, Germany) fused silica column. The chromatographic protocol selected for the separation of VOM prior to MS analysis consisted of setting the initial oven temperature at 40 ºC for 2 min, followed by a temperature gradient of 2.7 ºC/min up to 220 ºC, which was maintained for 5 min, for a total run time of 73.67 min. The column flow was kept constant at 1 mL/min, using helium as the carrier gas (mobile phase). The injection port operated in splitless mode and was maintained at 250 ºC. For the 5975 MS system, the transfer line, quadrupole, and source temperature were set at 270, 150, and 230 ºC, respectively. The electron impact mass spectrum was acquired at an ionization energy of 70 eV and an emission current of 10 µA. Data acquisition was performed in scan mode (30 - 300 m/z), and the electron multiplier was adjusted using the auto-tune mode. The identification of VOMs was obtained through manual interpretation of the spectra and comparison with data available in the Agilent MS ChemStation software, which includes the NIST05 mass spectral library, considering a minimum match of 70% between the experimental spectra and those in NIST05. The peak areas of the VOMs identified in this study were expressed relative area to the internal standard (IS, 3-octanol at 5 ppm).

2.2 Statistical Analysis

Statistical analysis was performed using MetaboAnalyst 6.0 (80). The workflow began with data pre-processing, which involved removing VOMs with missing values and normalizing the dataset. Normalization was carried out using cubic root transformation followed by auto-scaling. Subsequently, one-way ANOVA was applied, and Tukey's post-hoc test was used for multiple comparisons of means, considering p-values < 0.05 as statistically significant. The VOMs identified as significant were further analyzed through multivariate statistical methods. Principal component Analysis (PCA) was conducted to uncover trends and detect potential outliers, followed by Partial Least Squares Discriminant Analysis (PLS-DA). Hierarchical clustering analysis based on Pearson´s correlation was used to generate a heatmap, helping to identify clustering patterns among significantly altered VOMs across the study groups. Key variables in the PLS-DA model were assessed using Variable Importance in Projection (VIP) scores. Model validation was performed via 10-fold cross-validation and permutation testing (1000 random permutation of Y-observations). All sample triplicates were included in the data matrix and are displayed as individual points in the PCA score plots.

3 Chapter III. Results and discussion

3.1 Urinary volatilomic profile

The analysis of urinary volatilomic profile revealed clear differences between the study groups (HC vs CRC), as can be observed in Figure 11.

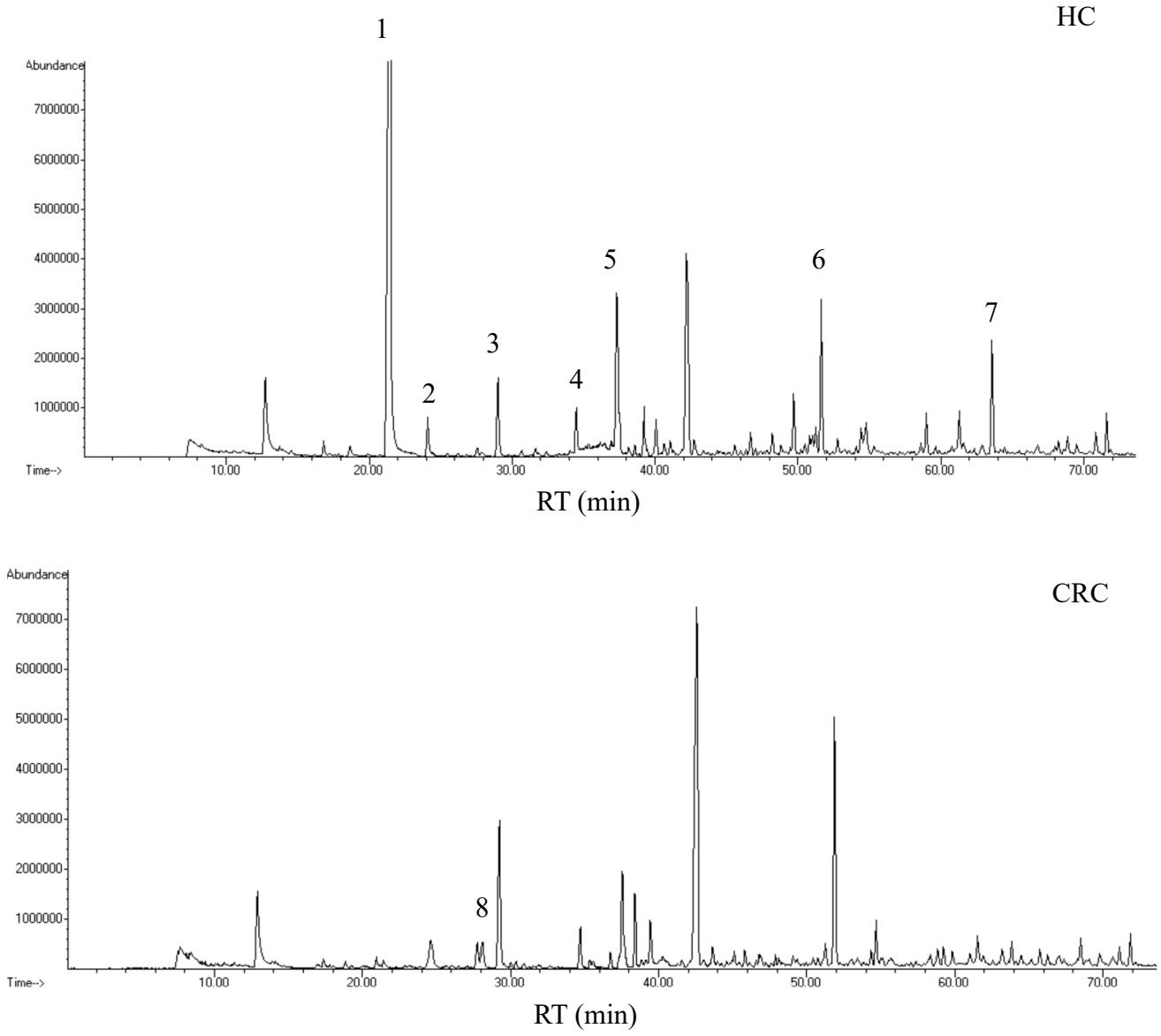


Figure 11 – Representative chromatograms of the groups analysed obtained using HS-SPME/GC-MS. Legend: HC: control group; CRC: colorectal group; 1: 4-heptanone; 2: 2-heptanone; 3: o-cymene; 4: 3-octanol (IS); 5: p-cymene; 6: TDN; 7: p-cresol; 8- γ-terpinene.

A total of 68 VOMs were identified in the urine analysed, comprising a diversity of chemical classes (Table 5). Among them 14 terpenes, six furanic compounds, seven norisoprenoids, four volatile phenols, seven ketones, four alcohols, three sulfur-containing compounds, and 22 others, comprising aldehydes, alkenes, benzene derivates, carboxylic acids and esters.

Table 5 – Identification, chemical family, possible origin, and mean relative peak area of the identified VOMs in the HC group and CCR patients (n=3; RSD < 20%).

RT (min)	VOMs	Formula	CAS nr	Possible Origin	Mean relative peak area	
					HC	CCR
Alcohols						
32.16	Tetrahydromyrcerol	$C_{10}H_{22}O$	18479-57-7	Exo	-	1.106
36.94	Tetrahydrolinalool	$C_{10}H_{22}O$	78-69-3	Exo	0.234	0.317
38.58	Dihydromyrcerol	$C_{10}H_{20}O$	18479-58-8	Exo	0.492	1.580
39.62	2-Ethyl-1-hexanol	$C_8H_{18}0$	104-76-7	Endo/Exo	0.870	1.313
Furanic Compounds						
11.52	2-Methylfuran	C_5H_6O	534-22-5	Endo/Exo	-	0.280
14.11	2,5-Dimethylfuran	C_6H_8O	625-86-5	Exo	-	0,363
17.25	2-Acetylfuran	$C_6H_6O_2$	1192-62-7	Exo	0.323	0.382
27.20	2-Pentylfuran	$C_9H_{14}O$	3777-69-3	Exo	0.064	0.039
27.76	Ocimene quintoxide	$C_{10}H_{18}O$	7416-35-5	Exo	0.260	0.243
50.21	4,7-Dimethylbenzofuran	$C_{10}H_{10}O$	28715-26-6	Exo	0.523	0.296
Ketones						
14.91	2-Pentanone	$C_5H_{10}O$	107-87-9	Exo	0.284	0.413
18.01	3-Hexanone	$C_6H_{12}O$	589-38-8	Endo/Exo	0.042	0.073
21.69	4-Heptanone	$C_7H_{14}O$	123-19-3	Endo/Exo	6.425	2.430
24.57	2-Heptanone	$C_7H_{14}O$	110-43-0	Endo/Exo	0.306	0.100
25.70	3-Methyl-2-heptanone	$C_8H_{16}O$	2371-19-9	Exo	-	0.027
28.29	3-Octanone	$C_8H_{16}O$	106-68-3	Exo	-	0.044

Table 5 – *Continuation*.

RT (min)	Metabolites	Formula	CAS nr	Possible Origin	Mean relative peak area	
					HC	CCR
Norisoprenoids						
41.50	Theaspirane	$C_{13}H_{22}O$	36431-72-8	Exo	0.308	0.170
43.83	α-Ionene	$C_{13}H_{18}$	79-77-6	Exo	-	0.164
46.02	β-Ionene	$C_{13}H_{20}O$	432-25-7	Exo	0.312	0.209
52.07	TDN	$C_{13}H_{16}$	30364-38-6	Exo	2.720	1.896
54.40	Damascenone	$C_{13}H_{18}O$	23726-93-4	Exo	0.793	0.423
61.77	Dihydro-β-ionene	$C_{13}H_{22}O$	1203-08-3	Exo	0.632	0.604
Sulphur-containing compounds						
10.03	Dimethyl dissulfide	$C_2H_6S_2$	624-92-0	Endo/Exo	0.170	0.399
29.83	1,3-Dithiane	$C_4H_8S_2$	505-23-7	Exo	-	0.022
31.15	2-Methoxythiophene	C_5H_6OS	16839-97-7	Exo	0.172	0.298
50.97	3-Sulfolene	$C_4H_6O_2S$	77-79-2	Exo	-	0.072
Terpenes						
25.44	Cosmene	$C_{10}H_{14}$	460-01-5	Exo	-	0.028
25.81	Limonene	$C_{10}H_{16}$	5989-27-5	Exo	-	0.058
28.10	*γ*-Terpinene	$C_{10}H_{16}$	99-85-4	Exo	-	0.408
29.48	*o*-Cymene	$C_{10}H_{14}$	527-84-4	Exo	1.950	1.377
30.16	Carvomenthene	$C_{10}H_{18}$	499-97-8	Exo	-	0.115
37.78	*p*-Cymene	$C_{10}H_{12}$	1195-32-0	Exo	6.811	2.150
39.05	Linalool oxide	$C_{10}H_{18}O_2$	60047-17-8	Exo	0.127	0.356
39.34	Nerol oxide	$C_{10}H_{16}O$	1786-08-9	Exo	-	0.122
42.30	Linalool	$C_{10}H_{18}O$	78-70-6	Exo	-	0.177
46.81	Menthol	$C_{10}H_{20}O$	15356-70-4	Exo	0.288	0.288
48.10	Ocimenol	$C_{10}H_{16}O$	5986-38-9	Exo	-	0.116
49.26	*α*-Terpinyl acetate	$C_{12}H_{20}O_2$	80-26-2	Exo	0.061	0.454
53.27	Cuminaldehyde	$C_{10}H_{12}O$	122-03-2	Exo	0.182	0.08
62.78	Nerolidol	$C_{15}H_{16}O$	7212-44-4	Exo	-	0.117

Table 5 – *Continuation*.

RT (min)	Metabolites	Formula	CAS nr	Possible Origin	Mean relative peak area	
					HC	CCR
Terpenoids						
21.10	Limetol	$C_{10}H_{18}O$	7392-19-0	Exo	-	0.181
46.37	β-Cyclocitral	$C_{10}H_{16}O$	432-25-7	Exo	-	0.063
51.30	Carvotanacetone	$C_{10}H_{16}0$	499-71-8	Exo	-	0.847
70.03	Cadalene	$C_{15}H_{18}$	483-78-3	Exo	0.480	0.375
Volatile phenols						
60.04	2-Bromophenol	C_6H_5BrO	95-56-7	Endo/Exo	-	0.008
61.28	Phenol	C_6H_5OH	108-95-2	Exo	0.263	0.398
64.05	*p*-Cresol	C_7H_8O	108-95-2	Endo/Exo	1.915	1.823
71.33	3-*tert*-Butylphenol	$C_{10}H_{14}O$	585-34-2	Exo	0.480	0.375
Others						
18.97	Hexanal	$C_6H_{12}O$	66-25-1	Endo/Exo	0.102	-
15.57	1,3-Dimethylcyclohexene	C_8H_{14}	2808-76-6	Exo	-	0.044
17.62	Toluene	C_7H_8	108-88-3	Exo	-	0.253
18.20	2-Ethyl-4-methylimidazole	$C_6H_{10}N_2$	931-36-2	Exo	-	0.251
20.54	2,6-Dimethyl-2,4-heptadiene	C_9H_{16}	4634-87-1	Exo	-	0.022
22.64	*m*-Xylene	C_8H_{10}	108-38-3	Exo	0.092	0.172
23.21	1,5,5-Trimethyl-6-methylene-cyclohexene	$C_{10}H_{16}$	514-95-4	Exo	-	0.02
32.81	Hemimellitene	C_9H_{12}	526-73-8	Endo/Exo	0.064	0.055
40.30	Trans-Edulan	$C_{13}H_{20}O$	41678-29-9	Exo	-	0.057
40.50	*p*-Xylene	$C_{10}H_{14}$	95-93-2	Exo	0.699	0.333
41.78	Benzaldehyde	C_7H_6O	100-52-7	Endo/Exo	-	0.093
48.32	Dihydro-5-methyl-2-(3H)-furanone	$C_5H_8O_2$	57129-69-8	Exo	-	0.128
50.18	α-Methylcinnamaldehyde	$C_{10}H_{10}O$	101-39-3	Exo	-	0.024

50.64	Indan-1,1,4,6-tetramethyl	$C_{13}H_{18}$	941-60-6	Endo/Exo	-	0.048
51.74	Naphthalene	$C_{10}H_8$	91-20-3	Exo	0.729	0.433
52.83	*o*-Acetyltoluene	$C_9H_{10}O$	577-16-2	Endo/Exo	-	0.029
62.06	1,6-Dimethylnaphthalene	$C_{12}H_{12}$	575-43-9	Exo	0.086	0.032
68.46	1,6,7-Trimehtlynaphthalene	$C_{13}H_{14}$	2245-38-7	Exo	0.154	0.136
68.69	Acetovanillone	$C_9H_{10}O_3$	498-02-2	Exo	-	0.142

Legend: Exo: exogenous; Endo: endogenous; HC: control group; CCR: colorectal cancer.; RT: retention time. TDN: 1,2-Dihydro-1,1,6-trimethyl-naphthalene

A detailed analysis of each sample group revealed significant differences in the relative abundance of several VOMs, as illustrated in Figure 12 and Table 5. The urinary volatilome is highly variable and subject to continuous changes, being conditioned by a series of physiological and environmental factors (59). Over time, urine composition can be modified due to enzymatic degradation, pH variations, bacterial activity, and the decomposition of urinary compounds (81). Furthermore, exogenous factors such as diet, psychological stress, physical activity, medications, and environmental exposures can contribute different compounds to the volatilomic profile, further increasing its complexity (82).

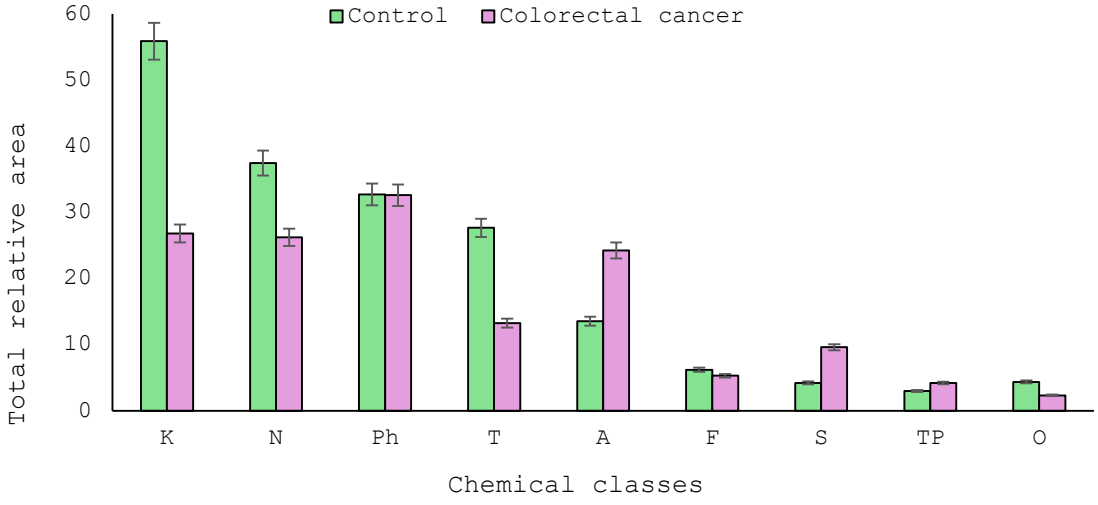


Figure 12 - Distribution of the chemical classes identified in the studied groups, control (HC, n= 19), and CCR (n= 19). Legend: K: ketones; N: norisoprenoids; Ph: phenolic compounds; T: terpenes; A: alcohols; F: furanic compounds; S: sulphur-containing compounds; TP: Terpenoids; O: others.

These intrinsic and extrinsic factors complicate the interpretation of volatilomic profiles in pathological circumstances such as cancer (6). The metabolic pathways involved in VOM biosynthesis are still poorly understood in many cases. Some metabolites may have solely endogenous origins, while others may originate from environmental sources or microbial interactions (83). In many cases, the same VOM may have a mixed origin (endogenous and exogenous), making its biological identification more difficult (84).

Understanding the metabolic origin of VOMs is essential for their potential use as a disease biomarker (61). Metabolic dysregulation associated with cancer leads to changes in biochemical pathways that subsequently manifest as changes in the urine volatilomic profile (85). Therefore, establishing links between specific VOMs and their endogenous sources can provide insight into the biochemical mechanisms underlying tumor development.

Ketones (56% and 26% for the total volatilomic profile of HC and CRC groups, respectively), norisoprenoids (37% and 26%), and phenolic compounds (33% in both) were the chemical classes that contributed most to the volatilomic profile of both study groups (Figure 12).

Ketones present in urine can be of endogenous and exogenous origin. Endogenous ketones are produced primarily as byproducts of metabolic pathways, such as carbohydrate metabolism and fatty acid β-oxidation (60,86). Acetone, one of the most prevalent and simple ketones, is formed by the decarboxylation of acetoacetate during lipid catabolism (85). In addition, the gut microbiota can contribute to the production of ketones through bacterial fermentation processes (58). Exogenous sources, such as dietary intake (e.g., flavorings, fermented foods, and alcoholic beverages) and environmental exposure, also play an important role in their composition (58). Among the detected VOMs, 3-hexanone, 4-heptanone, and 2-heptanone are ketones with mixed origins (endogenous and exogenous) (58). 3-Hexanone is an aliphatic ketone that belongs to the carbonyl class, found in several foods, such as bell peppers, cinnamon, and blackberries (87). It has been associated with several diseases, such as non-alcoholic fatty liver disease, autism spectrum disorders. Celiac disease and congenital metabolic disorders (88–90). 4-Heptanone, characterized by its sweet, cheese-like aroma, has been linked to kidney disorders, metabolic diseases, and celiac disease (88–90). 2-Heptanone, with a sweet aroma similar to coconut and cinnamon, is widely distributed throughout all organ systems and is found in foods such as corn, cow's milk, and peppermint (87).

Clinically, it has been associated with ulcerative colitis, Crohn´s disease, hepatic encephalopathy, and celiac disease (89–91). Other ketones identified, such as 2-pentanone, 3-methyl-2-heptanone, 3-octanone, carvotanacetone, are considered to be of exogenous origin, primarily derived from dietary sources (87). These VOMs have also been associated with inflammatory and metabolic diseases, including celiac disease, ulcerative colitis, and Crohn´s disease (89–91). In the present urinary volatilomic profile, notable differences in ketone abundance were observed between HC and CRC groups. 2-Pentanone and 3-hexanone showed increased levels in CRC patients, with relative areas 1.5 and 1.7 fold higher, respectively, compared to the HC group, which may reflect alterations in lipid metabolism or increased exposure to dietary and environmental sources (86). Conversely, the relative area of 4-heptanone and 2-heptanone were significantly reduced in CRC patients, being 2.6 and 3 fold higher in the HC group, potentially indicating disruptions in their metabolic synthesis or clearance. Specially, 3-methyl-2-heptanone and 3-octanone were detected in the CRC group and absent in the HC. Their specific presence in the CRC group may suggest cancer-associated metabolic dysregulation or distinct exposure patterns related to dietary, microbial, or therapeutic factors (92). Although 4-heptanone and 2-pentanone have been proposed in the literature as potential urinary biomarkers for CRC, the present data revealed different trends, while 2-pentanone levels were elevated in cancer patients, conforming to the literature, 4-heptanone showed a significant decrease, suggesting potential variability influenced by methodological differences, patient heterogeneity, or disease stage (62).

Norisoprenoids represent one of the most important chemical classes contributing to the urine volatilomic profile, often of exogenous origin (93). These VOMS are formed by the enzymatic degradation of dietary carotenoids, often found in fruits such as grapes, apples, and mangoes, and are identified by their aromatic properties (93). In this study, theaspirane, β-ionene, TDN, damascenone and dihydro-β-ionene showed notable reductions in CRC patients compared to HCs, with mean relative areas 2.9, 1.5, 1.4, 1.8, and 1 fold higher, respectively, in the HC group. Additionally, α-ionene was detected exclusively in CRC group, demonstrating a possible link with exogenous disease exposures or altered metabolic pathways linked to carcinogenesis. According to the literature, TDN has been suggested as a potential biomarker for CRC (60). However, study data reveal a reduction in its levels in CRC patients compared to HCs. This may be associated with changes related to the metabolism of the disease or the dietary processing of carotenoid-derived compounds (93).

Phenolic compounds were the third family that contributed the most to the urinary volatilomic profile of the CRC and HC groups, and represent endogenous and exogenous metabolic processes. p-Cresol is a well-characterized VOM, produced primarily by the intestinal microbiota through the fermentation of amino acids, particularly tyrosine and, to a lesser extent, phenylalanine (60,88). This VOM has been associated with the progression of renal dysfunction and has been indicated as a fecal biomarker for *Clostridium difficile* infection (87). Furthermore, it has been linked to several gastrointestinal disorders, including ulcerative colitis, irritable bowel syndrome, celiac disease, Crohn´s disease, and CRC (89–91). p-Cresol was present in high relative area in both groups analyzed (1.915 HCs vs. 1.823 CCRs), with a reduction only in CRC, which may reflect changes in the microbiota related to this disease (87). Phenol and 3-tert-butylphenol are VOMs of exogenous origin. Phenol is a toxic compound and has been shown to have slightly elevated levels in cancer patients (94). In contrast, 3-tert-butylphenol, a xenobiotic compound structurally similar to tyrosine known for its cytotoxicity, has been shown to have reduced levels in CRC patients (95). This decrease may reflect the excretion of phenolic compounds in the context of the disease. 2-Bromophenol, a primary metabolite, has been detected exclusively in the urine of CRC patients (60). Although found in low abundance, its presence may represent disease-specific metabolic alterations or interactions with environmental agents not observed in healthy individuals, warranting further investigation into its potential. According to the literature, p-cresol and phenol have been associated as potential biomarkers for CRC, given their link to gut microbial metabolism (59–61,63). In this study, phenol levels were elevated in CRC patients compared to HC, being 1.5 fold higher, supporting its potential as a disease-associated biomarker. However, p-cresol was detected at relatively high levels in both groups, with a slight decrease in patients with CRC, indicating that despite the frequent alteration of microbial activity in patients with CRC, there may be other limiting factors.

Terpenes comprise a structurally diverse and chemically complex class of natural products that significantly contribute to the volatilomic profile of human biological matrices, particularly urine (85). Terpenes are among the most abundant and varied VOMs, demonstrating primarily exogenous and, to a lesser extent, endogenous origin (85). The exogenous source of these VOMs comes primarily from the ingestion of terpene-rich foods, such as fruits, vegetables, herbs, spices, and beverages (e.g., tea, wine, and citrus juices) (93). In addition, exposure to cosmetics, perfumes, cleaning agents, and

pharmaceuticals containing essential oils and aromatic compounds is also common (87). Terpenes, after absorption through the gastrointestinal or respiratory tract, undergo hepatic metabolism and are consequently excreted in the urine (87). Biosynthetically, terpenes are derived from the mevalonate or methylerythritol phosphate pathways (93). Although monoterpenes can be synthesized endogenously via the mevalonate pathway, current evidence does not support significant endogenous production in humans; thus, the detection of these VOMs is interpreted as a consequence of environmental or dietary exposure (93). In this study, several terpenes were identified, including cosmene, limonene, γ-terpinene, o-cymene, carvomenthene, p-cymene, linalool oxide, nerol oxide, linalool, menthol, α-terpinyl acetate, cuminaldehyde, and nerolidol. p-Cymene emerged as the most abundant terpene, with levels 3 fold higher in HCs compared to CRC patients, as did o-cymene, which was 1.4 fold higher. In contrast, γ-terpinene, limonene, carvomenthene, cosmene, linalool oxide, nerol oxide, and linalool were detected only in CRC patients, which may represent changes associated with the disease, such as in nutrition and gastrointestinal absorption (87). α-Terpinyl acetate was higher in CRC patients (0.454) compared to the HC group (0.061), while menthol remained constant in both groups, and cuminaldehyde, a spice-derived terpene, was partially decreased in CRC patients (0.182 HCs vs. 0.08 CRCs) (87). These data suggest that terpenes may serve not only as dietary or environmental markers but also as potential metabolic biomarkers associated with CRC (85). According to the literature, p-cymene and γ-terpinene have been presented as potential biomarkers for CRC (60). In the present study, p-cymene showed a reduction in CRCs, whereas γ-terpinene was detected exclusively in the urine of CRC patients, suggesting a possible link with cancer-specific metabolic activity or exogenous exposure (60).

Alcohols constitute an important chemical class that can originate from various sources, including both endogenous and exogenous metabolic processes, and accounted for 14% and 24% of the volatilomic profile in HCs and CRC patients, respectively. In endogenous processes, alcohols can arise through the reduction of fatty acids in the gastrointestinal tract or through intermediate metabolic pathways, such as glycolysis, pyruvate metabolism, and citric acid cycle (24). In addition, the metabolism of some hydrocarbons can contribute to the formation of alcohols (86). Metabolism, especially of bacteria present in the intestine, also plays an important role in the production of alcohols, especially short-chain alcohols, reflecting the host-microbiota metabolic interaction (24). Exogenous sources include dietary intake of foods and beverages, since many alcohols

are naturally present or are generated during fermentation, cooking, or food processing (87). The identification of alcohols in urine may represent a combination of gut microbial activity and environmental or dietary exposure (24). This study demonstrated notable differences between HC and CR groups. Tetrahydromyrcerol was detected only in colorectal cancer patients, indicating a possible link with disease-related metabolic or microbial alterations. Dihydromyrcerol, tetrahydrolinalool, and 2-ethyl-1-hexanol were higher in CRC compared to HCs, being 1.4, 3.3, and 1.5 fold higher, respectively. 2-Ethyl-1-hexanol, a VOM of mixed origin, has been associated with ulcerative colitis, Crohn´s disease, and celiac disease, suggesting a possible link with intestinal inflammation (89–91). Although this compound has not been reported in the literature as a potential biomarker for CRC, it has been identified as a potential biomarker in lung cancer studies (55).

Furanic compounds constitute a significant class of VOMs that can arise from both endogenous and exogenous sources, particularly the diet, representing 6% and 5% of the total volatilomic profile in HCs and CRC patients, respectively (85). These VOMs are generally found in processed and thermally conventionalized foods, as they are formed by the thermal degradation and rearrangement of carbohydrates, such as glucose, fructose, and lactose, during industrial cooking (49). Although the human body can produce small amounts of these VOMs through oxidative stress and lipid peroxidation pathways, the main source is exogenous, through the consumption of heated or processed foods (85). Additionally, exposure to tobacco smoke can contribute to the presence of furans and derivatives in the body (96). These VOMs have been carefully analyzed due to their potential toxicity and carcinogenicity, and the International Agency for Research on Cancer (IARC) has classified them as possible human carcinogens (87). That said, they have been compounds of interest in volatilomic studies because they reflect dietary patterns, metabolic alteration, and exposure to harmful compounds (85). Furanic compounds, mainly of exogenous origin, show distinct patterns between HCs and CRC patients. 2-Methylfuran and 2,5-dimethylfuran were detected only in CRC patients. 2-Acetylfuran (0.323 HCs vs. 0.382 CRCs) showed a slight, non-significant increase in CRC patients, while 2-pentylfuran (0.064 HCs vs. 0.034 CRCs) and 4,7-dimethylbenzofuran (0.523 HCs vs. 0.296 CRCs) were reduced in CRC patients compared to HCs. In the literature, 4,7-dimethylfuran, 2-methylfuran, 2-pentylfuran, and 2-acetylfuran are associated with lung, breast, and colon cancer, and 2-methylfuran and 2,5-dimethylfuran in particular are associated with CRC patients (68). However, reduced

levels of 2-pentylfuran and 4,7-dimethylfuran in CRC patients may reflect changes in dietary intake and oxidative stress associated with cancer progression (68).

Sulphur-containing compounds are mostly formed by the incomplete metabolism of sulfur-containing amino acids, such as methionine and cysteine, through transamination pathways, and represented 4% and 10% of the total urinary volatilomic profile in HCs and CRC patients, respectively (97). Methanethiol is a key intermediate in the transamination process, which can undergo oxidation to form dimethyl disulfide and dimethyl trisulfide, two common sulfur volatiles (97). In addition, the gut microbiota, particularly Gram-negative bacteria, play an important role in the production of methanethiol and disulfide derivatives (85,87). These microbial processes can affect the sulfur volatile composition, particularly under conditions of dysbiosis or altered gastrointestinal function (85). In addition to the endogenous source, sulphur-containing volatiles are also present in various foods and beverages, such as garlic, onions, cabbage, and fermented products (85). Sulphur-containing compounds such as dimethyl disulfide and 2-methoxythipheno were found in higher amounts in the urine of CRC patients, being 2,3 and 1.7 fold higher than HCs, respectively. In addition,1,3-dithiane and 3-sulfolene, both of exogenous origin, were detected only in the cancer group, suggesting a relationship with disease-related metabolic or environmental alterations (87). According to the literature, dimethyl disulfide is a potential volatile biomarker of CRC due to its frequent association with altered sulfur metabolism and gut microbiota activity in cancer patients (62). In the present study, dimethyl disulfide levels were significantly higher in CRC patients compared to HCs, reinforcing previous findings and supporting its diagnostic relevance. This increase may result from enhanced microbial degradation of methionine or shifts in host-microbiota sulfur metabolism, both commonly observed in CRC (97). 2-Methoxythiophene has been linked to tobacco and industrial chemical exposure and may also result from the microbial transformation of sulfur-containing amino acids. (98). According to the literature, it has been described as a potential biomarker for CRC, supporting the results of the present study (62).

Terpenoids are predominantly of exogenous origin, accounting for 3% and 4% of the total volatilomic profile in HCs and CRC patients, respectively (99). They can be absorbed through the skin, orally, or via inhalation, with uptake influenced by compound properties, skin characteristics, or respiratory mechanics (99). Once absorbed, terpenoids become systemically available, and their elimination involves enterohepatic circulation as well as renal and pulmonary excretion, appearing feces, expired air, and urine (99).

Limetol and β-cyclocitral are VOMs found in higher concentrations in certain foods and plants, such as limes (*Citrus aurantiifolia*) and Orange mints (*Mentha aquática*), respectively (87). Both have also been detected in various teas and other foods, indicating their potential as biomarkers of dietary intake (87). Cadalene is an aromatic hydrocarbon derived from sesquiterpenoids, presented in the essential oils of many plants (87). In this study, limetol, β-cyclocitral, and carvonacetone were found only in patients with CRC. However, cadalene was found in higher concentrations in HCs (o.480 HCs vs. 0.375 CRC). These VOMs are not described in the literature as potential biomarkers of this disease; however, β-cyclocitral has been linked to nonalcoholic fatty liver disease (89).

In addition to the main chemical classes, such as ketones, phenolic compounds, terpenes, and norisoprenoids, a wide range of VOMs of distinct chemical natures contribute to the urine volatilomic profile. Other VOMs include aromatic hydrocarbons, aldehydes, imidazoles, thiophenes, lactone and polycyclic derivatives, most of which are exogenous, such as environmental pollutants, food, and medication (98,100–102). Aromatic VOMs such as toluene, m-xylene, p-xylene, and naphthalene were found in greater abundance or present only in the urine of CRC patients. Specially, toluene and 2-ethyl-4-methylimidazole were detected only in the CRC group, indicating a metabolic relationship with the disease or environmental susceptibility (103). p-Xylene (0.699 HCs vs. 0.333 CRCs) and naphthalene (0.729 HCs vs. 0.433 CRCs), although present in HCs, showed decreased levels in CRC patients, possibly reflecting altered hepatic clearance or differences in exposure or microbiota composition (104). The aldehyde hexanal, typically associated with lipid peroxidation, was observed only in HCs, indicating possible disruption in oxidative stress responses or fatty acid metabolism in CRC (65). In contrast, benzaldehyde, which may originate from amino acid degradation and intestinal microbial metabolism, was only present in CRC patients, potentially serving as a potential biomarker (102). These VOMs may originate from oxidative stress, lipid peroxidation, or microbial activity (101). Other compounds such as, α-methylcinnamaldehyde, 1,3-dimethylcyclohexene, dihydro-5-methyl-2-(3H)furanone and 1,5,5-trimethyl-6-methylene-cyclohexene, typically derived from flavoring agents and natural products, were found exclusively in the CRC group, possibly reflecting altered dietary habits and metabolic alterations (100). Overall, this group of VOMs reveals important qualitative and quantitative differences between CRC patients and HCs. Although many of these VOMs have not been validated as potential biomarkers of CRC, the altered abundance between the groups highlights their potential and requires further investigation.

3.2 Statistical analysis results

Statistical analysis was performed using MetaboAnalyst 6.0 (80). Prior to analysis, variables were normalized to ensure data homogeneity and enable the construction of robust and interpretable models. The normalized data matrix was subjected to univariate analysis via one-way ANOVA followed by Turkey´s post-hoc test ($p < 0.05$). This analysis revealed that 30 of the 68 identified VOMs showed statistically significant differences between the two study groups: HCs and CRCs (Figure 13). Notably, several of these 30 VOMs have previously been associated with oncological conditions, as described in the Human Metabolome Database (87). These include 2-pentanone, 2-acetylfuran, 3-hexanone, dimethyl disulfide, p-cymene, 2-methoxythiphene, 2-methylfuran, and 4-heptanone, which have been previously reported in the literature as potential urinary biomarkers for CRC (59,60,62,86). These VOMs are frequently associated with pathophysiological processes such as microbial dysbiosis, lipid peroxidation, and chronic intestinal inflammation, with hallmarks of colorectal carcinogenesis (27,89,91). In addition, several VOMs emerged as candidate biomarkers, based on their significant differential expression and known or suspected involvement in disease-related metabolic pathways. These include α-terpinyl acetate, carvomenthene, nerol oxide, cosmene, damascenone and 3-sulfolene. Although their diagnostic relevance for CRC has not been fully validated, their consistent presence in cancer patients and statistical significance suggest a potential association with CRC-related metabolic alterations or exogenous exposure. Conversely, certain statistically significant metabolites, such as p-xylene and naphthalene, are primarily associated with environmental or industrial sources and are therefore less associated as disease-specific biomarkers, despite their altered levels in CRC patients (86). These data highlight a set of VOMs that exhibit statistical significance and biological origin as non-invasive biomarkers for colorectal cancer. Further validation is needed to assess their clinical utility for early detection and disease monitoring in colorectal cancer patients.

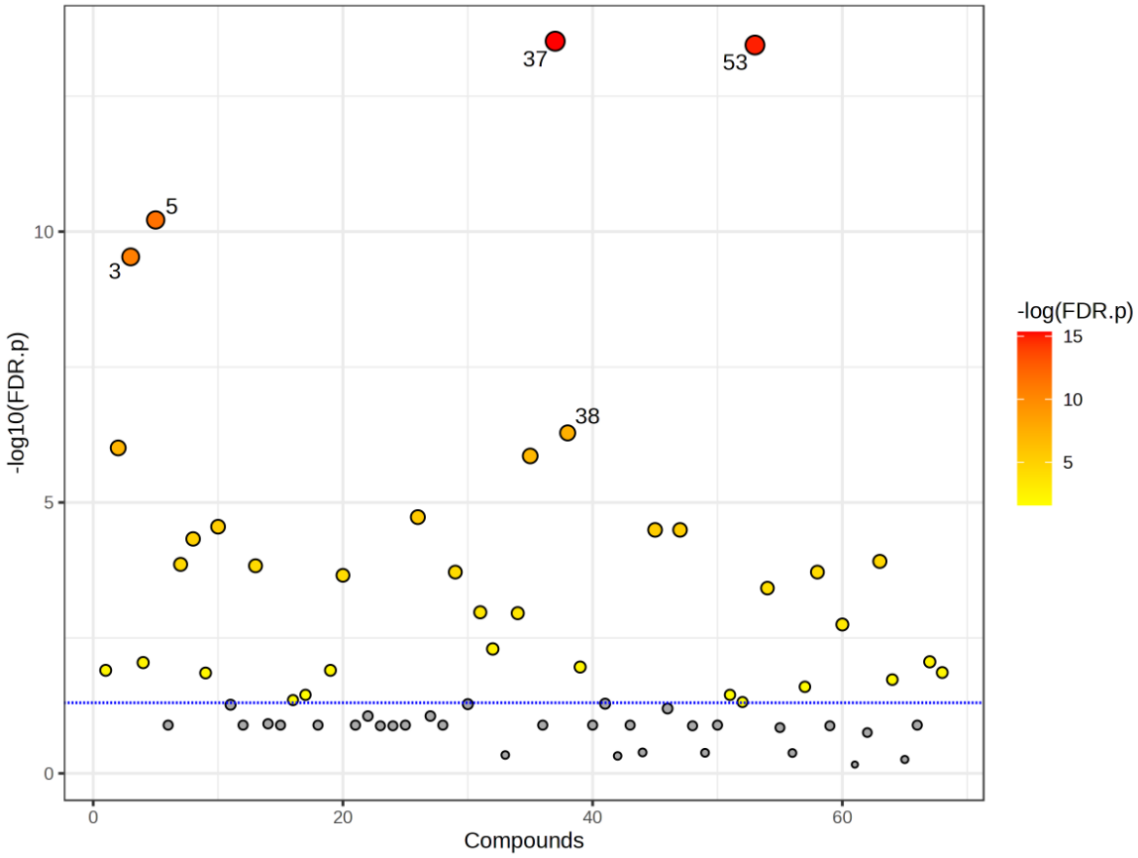


Figure 13 - Statistically significant diferences ($p < 0.05$) between the HC and CRC groups were identified using one-way ANOVA with Tukey´s post-hoc test. Legend: 3: 2-Pentanone; 5: 2-Acetylfuran; 37: p-xylene; 38: Theaspirane; 53: Naphthalene.

To investigate the overall metabolic differences between the HC and CRCgroups, an orthogonal partial least squares discriminant analysis (OPLS-DA) was performed using the normalized VOM data. The resulting scores plot (Figure 14) shows a clear separation between the two study groups, indicating distinct urinary volatilomic profiles. In the scores plot, the first component (T score\[1]) accounts for 11.1% of the total variance and is primarily responsible for group discrimination along the x-axis. The orthogonal component (Orthogonal T score\[1]), which explains 15.4% of the variance, comprises the within-group variability. Samples from HC cluster on the left side of the plot (in red), while samples from CRC patients are located on the right side (in green), demonstrating clear class separation. The confidence intervals do not show overlap, further confirming the robustness of the discrimination between groups and the reliability of the model. This distinct separation highlights the presence of significant biochemical differences in the urinary volatilome of CRC patients compared to HCs. These differences likely reflect alterations in host metabolism, microbiota composition, or environmental exposures associated with cancer pathophysiology. The model´s ability to separate the groups confirms the biological relevance of the VOMs identified in univariate and multivariate analysis, reinforcing their potential utility as non-invasive biomarkers for colorectal cancer.

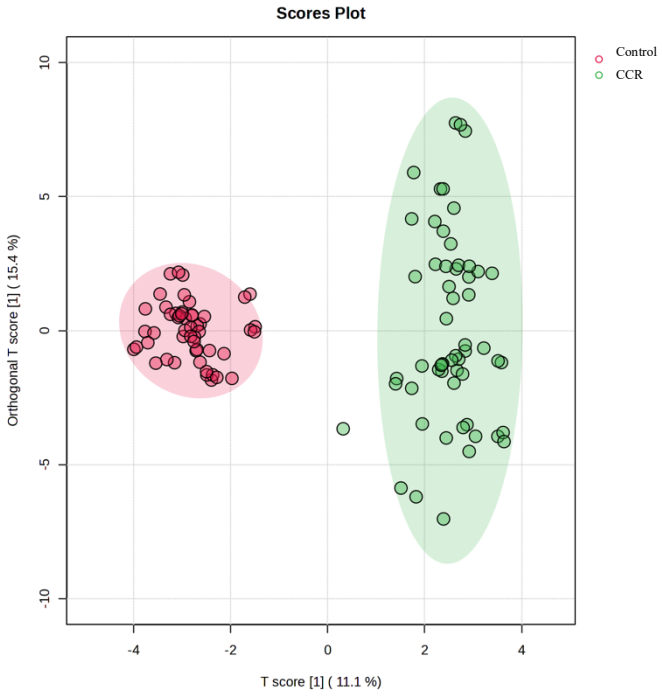
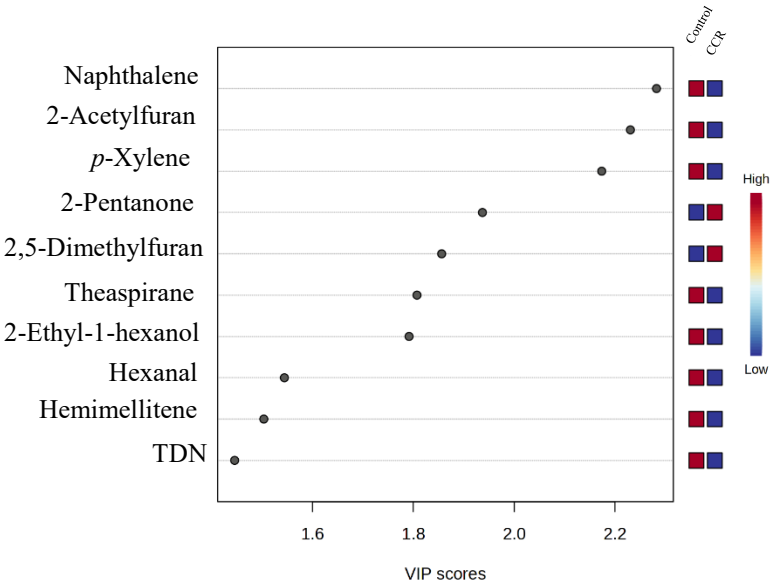


Figure 14- Scores plot from the OPLS-DA model based on urinary profiles from healthy individuals (HC group, n=17) and colorectal cancer patients (CRC group, n=19).

To explore the metabolic distinctions between HCs and CRC patients, an OPLS-DA was performed on the urinary volatilomic data. The robustness and statistical reliability of the model were assessed using a permutation test, while the Variable Importance in Projection (VIP) scores were used to identify the most discriminative VOMS contributing to group separation (Figure 15). The permutation test confirmed the statistical validity of the model. An R^2Y value of 0.951 was obtained, indicating excellent explanatory power, and a Q^2 value of 0.916, reflecting strong predictive ability. Both metrics were statistically significant ($p < 0.05$), and none of the 20 permutations exceeded the observed values. These results demonstrate that the separation observed between the CRC and HC groups is not random but based on systematic differences in the urinary volatile profiles, validating the use of the OPLS-DA model. Additionally, the VIP scores plot revealed the metabolites that contributed most significantly to the discrimination between groups. Metabolites with VIP values greater than 1.5 were considered highly relevant. Among these, the most important VOMs included naphthalene, 2-acetylfuran, p-xylene, 2-pentanone, 2,5-dimethylfuran, theaspirane, 2-ethyl-1-hexanol, hexanal, hemimellitene, and TDN. 2-Pentanone and 2,5-dimethylfuran were most closely associated with CRC patients, and others were most closely associated with HCs. These VOMs represent a combination of endogenous and exogenous origin, and several have been previously associated with colorectal cancer, intestinal inflammation, Crohn´s disease, microbial metabolism, and lipid peroxidation. Their high VIP values support their importance in the multivariate classification and highlight their potential as non-invasive biomarkers for CRC. Furthermore, the color gradient in the VIP plot visually explains the relative concentration of each metabolite in both groups, showing clear trends in metabolite upregulation or downregulation between CRC patients and healthy controls. Together, the results from the permutation test and VIP analysis confirm the discriminative power of the selected metabolomic features and reinforce their biological relevance in the context of CRC.

A.



B.

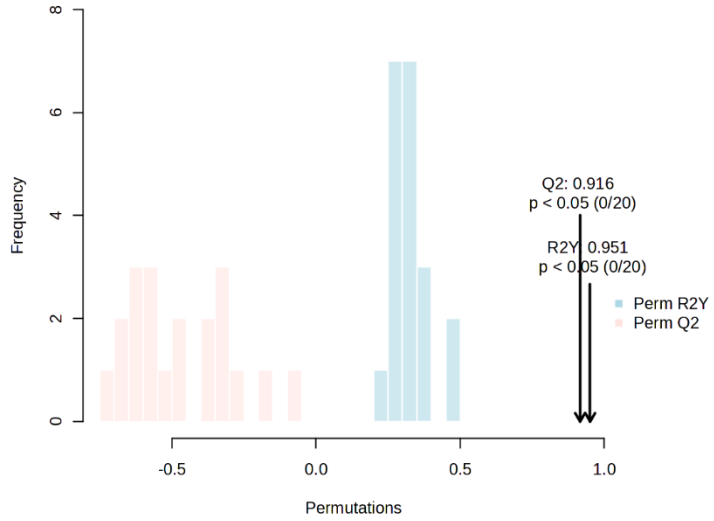


Figura 15 - **A**: VIP score plot highlighting the top 10 volatile organic metabolites contributing to the discrimination between CRC patients and healthy controls. **B**: The histogram shows the distribution of R2Y (blue) and Q2 (pink) values obtained from 20 random permutations.

The hierarchical clustering heatmap provides a clear visual representation of the discriminatory power of urinary VOMs in differentiating between CRC patients and HCs (Figure 16). The data, normalized for consistent comparison, revealed two distinct clusters corresponding to the two study groups. This observation indicates that the urine volatilomic profile is sufficiently divergent to allow classification based on disease state. The heatmap shows that a set of VOMs are consistently elevated in the CRC group while reduced in HCs, and vice versa. These differential patterns of expression highlight metabolic alterations likely associated with the presence of cancer. The clustering structure not only confirms the ability of selected VOMs to segregate the two groups but also indicates the consistency and reliability of the dataset. Furthermore, this visual clustering is consistent with previous multivariate (OPLS-DA) and univariate (ANOVA) analyses, which can also identify significant differences between the two groups.

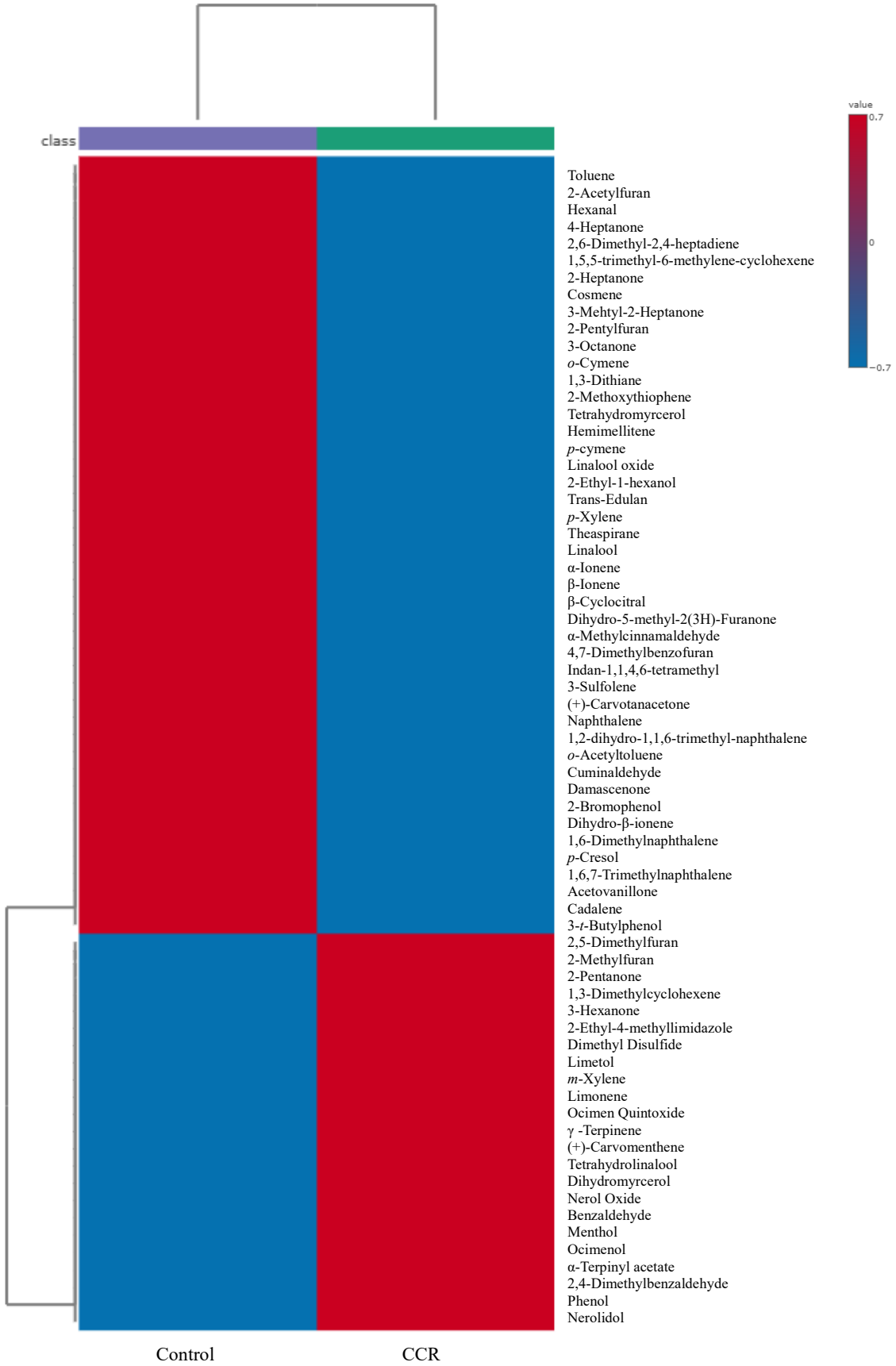


Figure 16 – Heatmap of all identified urinary VOMs in control and colorectal cancer subjects. The heatmap displays the normalized intensity values of all VOMs detected in urine samples from HCs and CRC patients.

During the course of this study, which aimed at establishing the urinary volatilomic profile of CRC patients to identify potential biomarkers, an unexpected and clinically relevant observation emerged. A hierarchical clustering analysis (dendogram) based on the urinary volatilomic profile revealed a distinct grouping pattern between the HC and CRC groups (Figure 17). However, one individual initially classified as a healthy control (CTRL4) clustered among the CRC patient group. After the study was concluded, it was confirmed that this individual, previously assumed to be cancer-free, had in fact received a late-stage diagnosis of metastatic CRC. This impressive result highlights the strong discriminative power of the urinary volatilomic profile developed in this study and supports the hypothesis that metabolic changes related to CRC can be detected in urine even before clinical diagnosis. This observation reinforces the potential of urinary volatilomic as a non-invasive, sensitive approach for the early detection of CRC. The fact that the data-driven analysis was able to identify a metabolomic signature indicative of CRC in a misclassified control highlights the clinical relevance and translational value of this approach. The study, therefore, not only contributes to the growing field of cancer metabolomics but also provides compelling evidence for the practical utility of VOM-based biomarkers in early cancer detection.

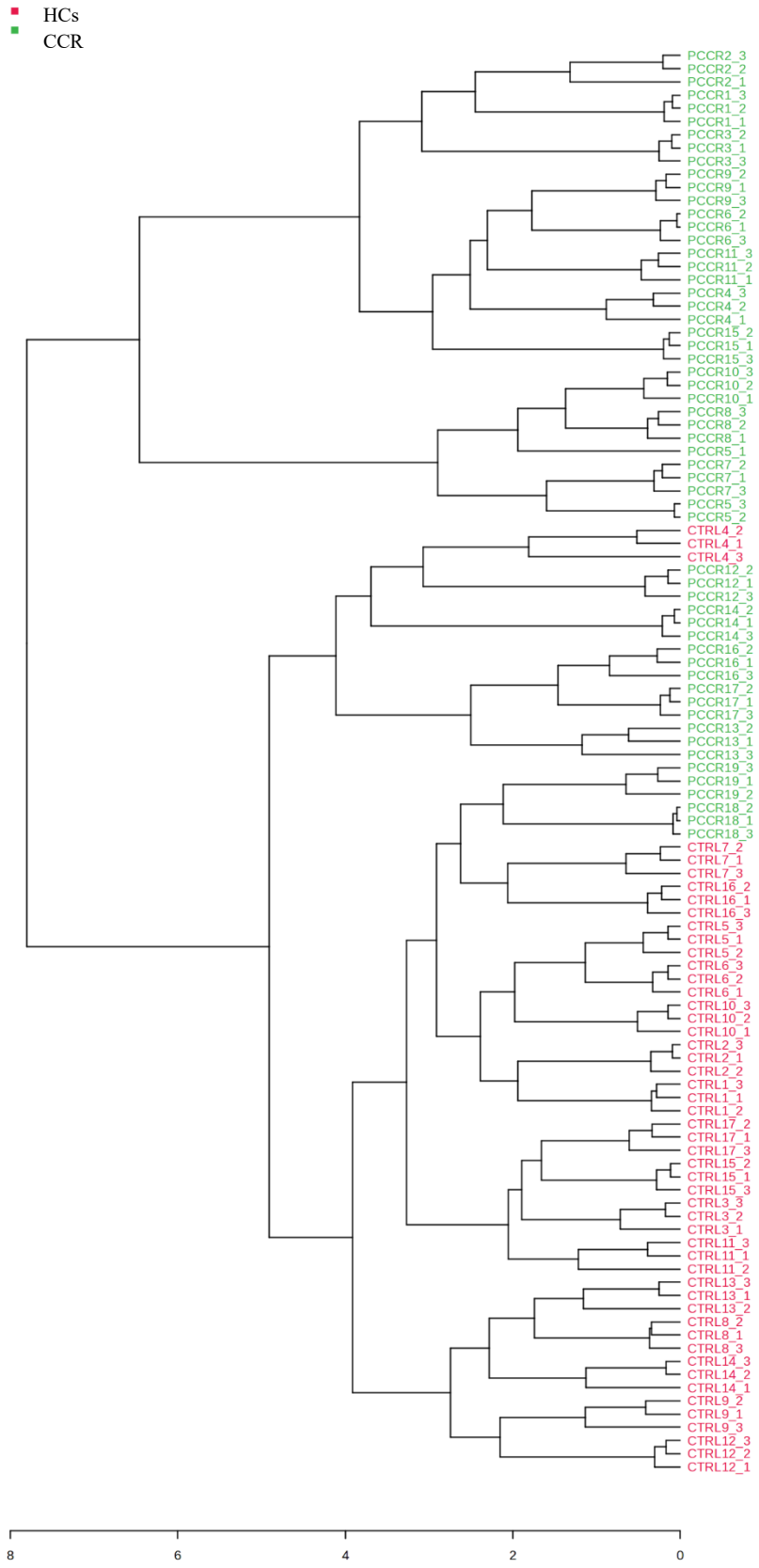


Figura 17 - Hierarchical clustering dendrogram of urinary volatilomic profiles from control and colorectal cancer subjects.

4 Chapter IV. Conclusions and future perspectives

Through this study, 68 urinary VOMs were identified in the volatilomic profile of CRC patients and HCs. The chemical classes that the most contributed were ketones, phenolic compounds, and norisoprenoids. By standardizing the data, multivariate ANOVA followed by Turkey´s test ($p < 0.05$) identified 30 of the 68 VOMs with significant differences between the groups, several of which are associated with oncological conditions and processes such as microbial dysbiosis, lipid peroxidation, and chronic intestinal inflammation such as p-cresol, 2-ethyl-1-hexanol, 2-pentanone, and p-cymene. Subsequently, multivariate OPLS-DA analysis revealed a clear separation between the groups, confirmed by the permutation test and VIP scores, which highlighted important VOMs. 2-Pentanone and 2,5-dimethylfuran were most closely associated with CRC patients, whereas naphthalene, 2-acetylfuran, p-xylene, theaspirane, 2-ethyl-1-hexanol, hexanal, hemimellitene, and TDN were most closely associated with HCs. An important finding was the early detection of an individual initially classified as a control but later diagnosed with metastatic cancer. This result highlights the strong discriminatory power of the urinary volatilomic profile developed in this study and confirms the hypothesis that metabolic alterations related to CRC can be detected in urine even before clinical diagnosis. However, to increase the robustness of this study, it would be necessary to increase the number of participants and obtain more relevant information, such as diet and disease stage, and investigate other profiles, such as proteomics and genomics. These factors should be explored in the future with the aim of developing and optimizing a potential noninvasive screening tool for the population.

References

1. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. Science. 2020;368:eaaw5473.
2. Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. Annual Review of Pathology: Mechanisms of Disease. 2013;8:277.
3. GLOBOCAN. Global Cancer Observatory [Internet]. 2024 [cited 2024 Nov 4]. Available from: https://gco.iarc.fr/today/en/dataviz/maps-heatmap?mode=population&cancers=41
4. Mármol I, Sánchez-de-Diego C, Dieste AP, Cerrada E, Yoldi MJR. Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. International Journal of Molecular Sciences. 2017;18:197.
5. Ionescu VA, Gheorghe G, Bacalbasa N, Chiotoroiu AL, Diaconu C. Colorectal Cancer: From Risk Factors to Oncogenesis. Medicina. 2023;59:1646.
6. Sawicki T, Ruszkowska M, Danielewicz A, Niedźwiedzka E, Arłukowicz T, Przybyłowicz KE. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers. 2021;13:2025.
7. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi DJ, John A, et al. Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. Cancers. 2022;14:1732.
8. Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database of Systematic Reviews. 2019;9:CD002200.
9. Shaukat A, Levin TR. Current and future colorectal cancer screening strategies. Nature Reviews Gastroenterology and Hepatology. 2022;19:521.
10. Iwasaki H, Shimura T, Kataoka H. Current status of urinary diagnostic biomarkers for colorectal cancer. Clinica Chimica Acta. 2019;498:76.
11. Salimoglu S, Kilinc G, Calik B. Anatomy of the colon, rectum, and anus. In: Colon Polyps and Colorectal Cancer: Second Edition. Springer International Publishing; 2020. p. 1.
12. American Cancer Society. Colorectal Cancer Facts & Figures 2023-2025. American Cancer Society. 2023.

13. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. Gastroenterology Review. 2019;14:89.
14. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: Incidence and mortality estimates from GLOBOCAN. Gut. 2023;72:338.
15. Bento MJ, Leite P, Rita S, Revisão C, Gonçalves AF, Sousa A, et al. Registo Oncológico Nacional de Todos os Tumores na População Residente em Portugal, em 2020. 2023.
16. Kanth P, Inadomi JM. Screening and prevention of colorectal cancer. BMJ. 2021;374:1855.
17. Berenguer P, Müller S, Fraga C, Serrão P, Cordeiro M, Camacho C. Colorectal cancer - Epidemiological Approach Of The Autonomous Region of Madeira, A Portugal Insular Region. Vol. 68, CA: A Cancer Journal for Clinicians. Wiley; 2023.
18. American Cancer Society. Colorectal Cancer Risk Factors Colorectal Cancer Causes, Risk Factors, and Prevention [Internet]. [cited 2025 Jan 21]. Available from: https://www.cancer.org/content/dam/CRC/PDF/Public/8605.00.pdf
19. Ye P, Xi Y, Huang Z, Xu P. Linking obesity with colorectal cancer: Epidemiology and mechanistic insights. Cancers. 2020;12:1408.
20. An S, Park S. Association of Physical Activity and Sedentary Behavior With the Risk of Colorectal Cancer. Journal of Korean Medical Science. 2022;37:e158.
21. Song M, Chan AT, Sun J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. Gastroenterology. 2020;158:322.
22. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, et al. Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. Journal of the National Cancer Institute. 2010;102:1012.
23. National Cancer Institute. Alcohol and Cancer Risk [Internet]. 2021. p. 1. Available from: https://www.cancer.gov/about-cancer/causes-prevention/risk/alcohol/alcohol-fact-sheet
24. Johnson CH, Golla JP, Dioletis E, Singh S, Ishii M, Charkoftaki G, et al. Molecular mechanisms of alcohol-induced colorectal carcinogenesis. Cancers. 2021;13:4404.
25. Hultcrantz R. Aspects of colorectal cancer screening, methods, age and gender. Journal of Internal Medicine. 2021;289:493.

26. Song M, Emilsson L, Roelstraete B, Ludvigsson JF. Risk of colorectal cancer in first degree relatives of patients with colorectal polyps: Nationwide case-control study in Sweden. BMJ. 2021;373:877.
27. Shah SC, Itzkowitz SH. Colorectal Cancer in Inflammatory Bowel Disease: Mechanisms and Management. Gastroenterology. 2022;162:715.
28. Kastrinos F, Samadder NJ, Burt RW. Use of Family History and Genetic Testing to Determine Risk of Colorectal Cancer. Gastroenterology. 2020;158:389.
29. Leslie A, Carey FA, Pratt NR, Steele RJC. The colorectal adenoma - carcinoma sequence. British Journal of Surgery [Internet]. 2002;89:845. Available from: www.bjs.co.uk
30. Snyder C, Hampel H. Hereditary Colorectal Cancer Syndromes. Seminars in Oncology Nursing. 2019;35:58.
31. Cavalcanti Carneiro da Silva F, Dominguez Valentin M, de Oliveira Ferreira F, Carraro DM, Rossi BM. Mismatch repair genes in Lynch syndrome: a review. Sao Paulo Medical Journal. 2009;127:46.
32. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clinical Genetics. 2009;76:1.
33. Lawler T, Walts ZL, Steinwandel M, Lipworth L, Murff HJ, Zheng W, et al. Type 2 Diabetes and Colorectal Cancer Risk. JAMA Network Open. 2023;6:e2343333.
34. Tárraga López PJ, Solera Albero J, Rodríguez-Montes JA. Primary and secondary prevention of colorectal cancer. Clinical Medicine Insights: Gastroenterology. 2014;7:33.
35. US Department of Health and Human Services. Reducing Tobacco Use: A report of the Surgeon General. National Library of Medicine [Internet]. 2000 [cited 2024 Nov 4];49:1. Available from: https://pubmed.ncbi.nlm.nih.gov/11190114/
36. Katona BW, Weiss JM. Chemoprevention of Colorectal Cancer. Gastroenterology. 2020;158:368.
37. Terry P, Jain M, Miller AB, Howe GR, Rohan TE. Dietary intake of folic acid and colorectal cancer risk in a cohort of women. International Union Against Cancer. 2002;97:864.
38. Bonithon-Kopp C, Kronborg O, Giacosa A, Räth U, Faivre J. Calcium and fibre supplementation in prevention of colorectal adenoma recurrence: a randomised intervention trial. The Lancet. 2000;356:1286.

39. Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. Gastroenterology. 2020;158:291.
40. Golas MM, Gunawan B, Cakir M, Cameron S, Enders C, Liersch T, et al. Evolutionary patterns of chromosomal instability and mismatch repair deficiency in proximal and distal colorectal cancer. Colorectal Disease. 2022;24:157.
41. Cefalì M, Epistolio S, Palmarocchi MC, Frattini M, De Dosso S. Research progress on KRAS mutations in colorectal cancer. Journal of Cancer Metastasis and Treatment. 2021;7:26.
42. Sheng J, Sun H, Yu FB, Li B, Zhang Y, Zhu YT. The role of cyclooxygenase-2 in colorectal cancer. International Journal of Medical Sciences. 2020;17:1095.
43. Liebl MC, Hofmann TG. The role of p53 signaling in colorectal cancer. Cancers. 2021;13:2125.
44. Sun BL. Current Microsatellite Instability Testing in Management of Colorectal Cancer. Clinical Colorectal Cancer. 2021;20:e12.
45. De Palma FDE, D'argenio V, Pol J, Kroemer G, Maiuri MC, Salvatore F. The molecular hallmarks of the serrated pathway in colorectal cancer. Cancers. 2019;11:1017.
46. American Cancer Society. American Cancer Society. 2024. Colorectal Cancer Early Detection, Diagnosis, and Staging. Available from: https://seer.cancer.gov/csr/1975_2016/,
47. Monahan KJ, Davies MM, Abulafi M, Banerjea A, Nicholson BD, Arasaradnam R, et al. Faecal immunochemical testing (FIT) in patients with signs or symptoms of suspected colorectal cancer (CRC): a joint guideline from the Association of Coloproctology of Great Britain and Ireland (ACPGBI) and the British Society of Gastroenterology (BSG). Gut. 2022;71:1939.
48. Hosoe N, Limpias Kamiya KJL, Hayashi Y, Sujino T, Ogata H, Kanai T. Current status of colon capsule endoscopy. Digestive Endoscopy. 2021;33:529.
49. Newton KF, Newman W, Hill J. Review of biomarkers in colorectal cancer. Colorectal Disease. 2012;14:3.
50. Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: Current research and future prospects. International Journal of Molecular Sciences. 2020;21:5311.
51. J. YIU A, Y. YIU C. Biomarkers in Colorectal Cancer. Anticancer Research. 2016;36:1093.

52. Liu G, Li B. Role of miRNA in transformation from normal tissue to colorectal adenoma and cancer. Journal of Cancer Reseasrch and Therapeutics. 2019;15:278.
53. Rapado-González Ó, Álvarez-Castro A, López-López R, Iglesias-Canle J, Suárez-Cunqueiro MM, Muinelo-Romay L. Circulating microRNAs as promising biomarkers in colorectal cancer. Cancers. 2019;11:898.
54. Schmidt K, Podmore I. Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer. Journal of Biomarkers. 2015;2015:1.
55. da Costa BRB, De Martinis BS. Analysis of urinary VOCs using mass spectrometric methods to diagnose cancer: A review. Clinical Mass Spectrometry. 2020;18:27.
56. Becker R. Non-invasive cancer detection using volatile biomarkers: Is urine superior to breath? Medical Hypotheses. 2020;143:1.
57. Altobelli E, Angeletti PM, Latella G. Role of urinary biomarkers in the diagnosis of adenoma and colorectal cancer: A systematic review and meta-analysis. Journal of Cancer. 2016;7:1984.
58. Mills GA, Walker V. Headspace solid-phase microextraction profiling of volatile compounds in urine: application to metabolic investigations. Journal of Chromatography B [Internet]. 2001;753:259. Available from: www.elsevier.com/locate/chromb
59. Qiu Y, Cai G, Su M, Chen T, Liu Y, Xu Y, et al. Urinary metabonomic study on colorectal cancer. Journal of Proteome Research. 2010;9:1627.
60. Silva CL, Passos M, Cmara JS. Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry. British Journal of Cancer. 2011;105:1894.
61. Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. Journal of Proteome Research. 2012;11:1354.
62. Arasaradnam RP, Mcfarlane MJ, Ryan-Fisher C, Westenbrink E, Hodges P, Thomas MG, et al. Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis. PLoS One. 2014;9:e108750.
63. Liesenfeld DB, Habermann N, Toth R, Owen RW, Frei E, Böhm J, et al. Changes in urinary metabolic profiles of colorectal cancer patients enrolled in a prospective cohort study (ColoCare). Metabolomics. 2015;11:998.

64. Mozdiak E, Wicaksono AN, Covington JA, Arasaradnam RP. Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP). Techniques of Coloproctology. 2019;23:343.
65. Monedeiro F, Monedeiro-milanowski M, Ligor T, Buszewski B. A review of gc-based analysis of non-invasive biomarkers of colorectal cancer and related pathways. Journal of Clinical Medicine. 2020;9:3191.
66. Delphan M, Lin T, Liesenfeld DB, Nattenmüller J, Böhm JT, Gigic B, et al. Associations of branched-chain amino acids with parameters of energy balance and survival in colorectal cancer patients: results from the ColoCare study. Metabolomics. 2018;14:22.
67. Wang X, Wang J, Rao B, Deng LI. Gut flora profiling and fecal metabolite composition of colorectal cancer patients and healthy individuals. Experimental and Therapeutic Medicine. 2017;13:2848–54.
68. Porto-Figueira P, Pereira JAM, Câmara JS. Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatomic biosignature. Analytica Chimica Acta. 2018;1023:53.
69. Brieley James GMK. TNM Classification of Malignant Tumors. 8th ed. Brierley JD, Gospodarowicz MK, Wittekind C, editors. Union for International Cancer Control; 2017. 1 p.
70. Mahmoud NN. Colorectal Cancer: Preoperative Evaluation and Staging. Surgical Oncology Clinics of North America. 2022;31:127.
71. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. The Lancet. 2019;394:1467.
72. Kosmider S, Lipton L. Adjuvant therapies for colorectal cancer. World Journal of Gastroenterology [Internet]. 2007;13:3799. Available from: http://www.wjgnet.com/1007-9327/13/3799.asp
73. Shinji S, Yamada T, Matsuda A, Sonoda H, Ohta R, Iwai T, et al. Recent Advances in the Treatment of Colorectal Cancer: A Review. Journal of Nippon Medical School. 2022;89:246.
74. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nature Reviews Disease Primers. 2015;1:1.

75. Mody K, Baldeo C, Bekaii-Saab T. Antiangiogenic Therapy in Colorectal Cancer. The Cancer Journal [Internet]. 2018;24:165. Available from: www.journalppo.com
76. Cancer Care Ontario. Cancer Care Ontario. 2021 [cited 2025 Jan 22]. p. 1 Colorectal Cancer Follow-up Care Patient Pathway. Available from: https://www.rvh.on.ca/wp-content/uploads/2021/05/Colorectal-Cancer-Follow-up-Care-Patient-Pathway.pdf
77. Costa Ferreira Bandarra MM, Escoval A, Marques Da Costa L. Is There a Role for Primary Care in the Follow-Up of Colorectal Cancer Patients? the Case of a Portuguese University Hospital. Portuguese Journal of Public Health. 2018;36:72.
78. Viveiros Berenguer C. Urinary Metabolomic Fingerprints as a Powerful Tool for the Next Generation of Prostate Cancer Diagnosis [Mater's thesis]. Universidade da Madeira; 2022.
79. Riccio G, Berenguer C V., Perestrelo R, Pereira F, Berenguer P, Ornelas CP, et al. Differences in the Volatilomic Urinary Biosignature of Prostate Cancer Patients as a Feasibility Study for the Detection of Potential Biomarkers. Current Oncology. 2023;30:4904.
80. Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C, et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. Nucleic Acids Research. 2024;52:W398.
81. Amann A, Costello BDL, Miekisch W, Schubert J, Buszewski B, Pleil J, et al. The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. Journal of Breath Research. 2014;8:1.
82. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The Human Urine Metabolome. PLoS One. 2013;8:e73076.
83. Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P, Wai J. Volatile biomarkers in the breath of women with breast cancer. Journal of Breath Research. 2010;4:1.
84. Tait E, Perry JD, Stanforth SP, Dean JR. Identification of volatile organic compounds produced by bacteria using HS-SPME-GC-MS. Journal of Chromatographic Science. 2014;52:363.
85. De Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. Journal of Breath Research. 2014;8:1.
86. Janfaza S, Khorsand B, Nikkhah M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. Biology Methods and Protocols. 2019;4:1.

87. Wishart DS, Guo AC, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: The Human Metabolome Database for 2022. Nucleic Acids Research. 2022 Jan 7;50:D622.
88. Silva CL, Passos M, Câmara JS. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers - A powerful strategy for breast cancer diagnosis. Talanta. 2012;89:360.
89. Raman M, Ahmed I, Gillevet PM, Probert CS, Ratcliffe NM, Smith S, et al. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. Clinical Gastroenterology and Hepatology. 2013;11:868.
90. Di Cagno R, De Angelis M, De Pasquale I, Ndagijimana M, Vernocchi P, Ricciuti P, et al. Duodenal and faecal microbiota of celiac children: Molecular, phenotype and metabolome characterization. BMC Microbiology. 2011;11:219.
91. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Alimentary Pharmacology and Therapeutics. 2016;43:596.
92. van Liere ELSA, van Dijk LJ, Bosch S, Vermeulen L, Heymans MW, Burchell GL, et al. Urinary volatile organic compounds for colorectal cancer screening: A systematic review and meta-analysis. European Journal of Cancer. 2023;186:69.
93. Riccio G, Neto J, Berenguer C V., Ornelas CP, Greco V, Pereira J, et al. Evaluation of the Potential of Urinary Volatilomic Patterns of COVID-19 Patients Infected With SARS-CoV-2 for Disease Diagnosis. An Exploratory Study. Pre print [Internet]. 2024;1. Available from: https://www.preprints.org/manuscript/202404.0355/v1
94. Duranton F, Cohen G, De Smet R, Rodriguez M, Jankowski J, Vanholder R, et al. Normal and pathologic concentrations of uremic toxins. Journal of the American Society of Nephrology. 2012;23:1258.
95. Hoang NMH, Park K. Applications of Tert-Butyl-Phenolic Antioxidants in Consumer Products and Their Potential Toxicities in Humans. Toxics. 2024;12:869.
96. Vevang KR, Zhang L, Grill AE, Hatsukami DK, Meier E, Nomura SO, et al. Furan Metabolites Are Elevated in Users of Various Tobacco Products and Cannabis. Chemical Research in Toxicology. 2023;36:157.

97. Wang Y, Nguyen LH, Mehta RS, Song M, Huttenhower C, Chan AT. Association between the Sulfur Microbial Diet and Risk of Colorectal Cancer. JAMA Network Open. 2021;4:e2134308.
98. Al-Jumaili MHA, Bakr EA, Huessien MA, Hamed AS, Muhaidi MJ. Development of heterocyclic-based anticancer agents: A comprehensive review. Heterocyclic Communications. 2025;31:1.
99. Masyita A, Mustika Sari R, Dwi Astuti A, Yasir B, Rahma Rumata N, Emran T Bin, et al. Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. Food Chem X. 2022;13:1.
100. Yannai S. Dictionary of food compounds with CD-ROM. CRC Press (Taylor & Francis Group); 2012.
101. Gurjar VK. Recent Advances in Oxygen-Containing Heterocycles as Anticancer Agents. International Journal of Medical, Pharmaceutical and Health Sciences. 2024;1:80.
102. De Vietro N, Aresta A, Rotelli MT, Zambonin C, Lippolis C, Picciariello A, et al. Relationship between cancer tissue derived and exhaled volatile organic compound from colorectal cancer patients. Preliminary results. Journal of Pharmaceutical and Biomedical Analysis. 2020;180:1.
103. Lv J, Chen Q, Yang L, Guan J, Wang S, Gui G, et al. Comparative serum imidazole propionate profiling in colorectal adenoma and cancer by UPLC-MS/MS. BMC Gastroenterology. 2025;25:426.
104. Bayat M, Asban P, Koshki Nasab F, Taherian M, Sepahvand A, Farhadi A, et al. A Review of the Effect of Outdoor Polycyclic Aromatic Hydrocarbons on Bladder Cancer. Health Scope. 2023;12:1.
105. Instituto Nacional de Estatística I. Estatísticas da Saúde - 2022. 2024th ed. Instituto Nacional de Estatística, editor. 2022.
106. Dinis J, Portugal C, Sousa N, Fernandes I, Netto E. Avaliação e monitorização dos rastreios oncológicos organizados de base populacional. 2021. 1 p.
107. Fontana CM. BioRender [Internet]. [cited 2025 Jan 21]. Available from: https://app.biorender.com/biorender-templates/t-642593000236b76192a5d523
108. Cancro Online. Cancro Online [Internet]. 2021. Available from: https://www.cancro-online.pt/cancro-colorretal/informacao-basica/fatores-de-risco/

109. Ullah F, Pillai AB, Omar N, Dima D, Harichand S. Early-Onset Colorectal Cancer: Current Insights. Cancers. 2023;15:3202.

APPENDICES

SUPPLEMENTARY MATERIAL

Additional figures and tables

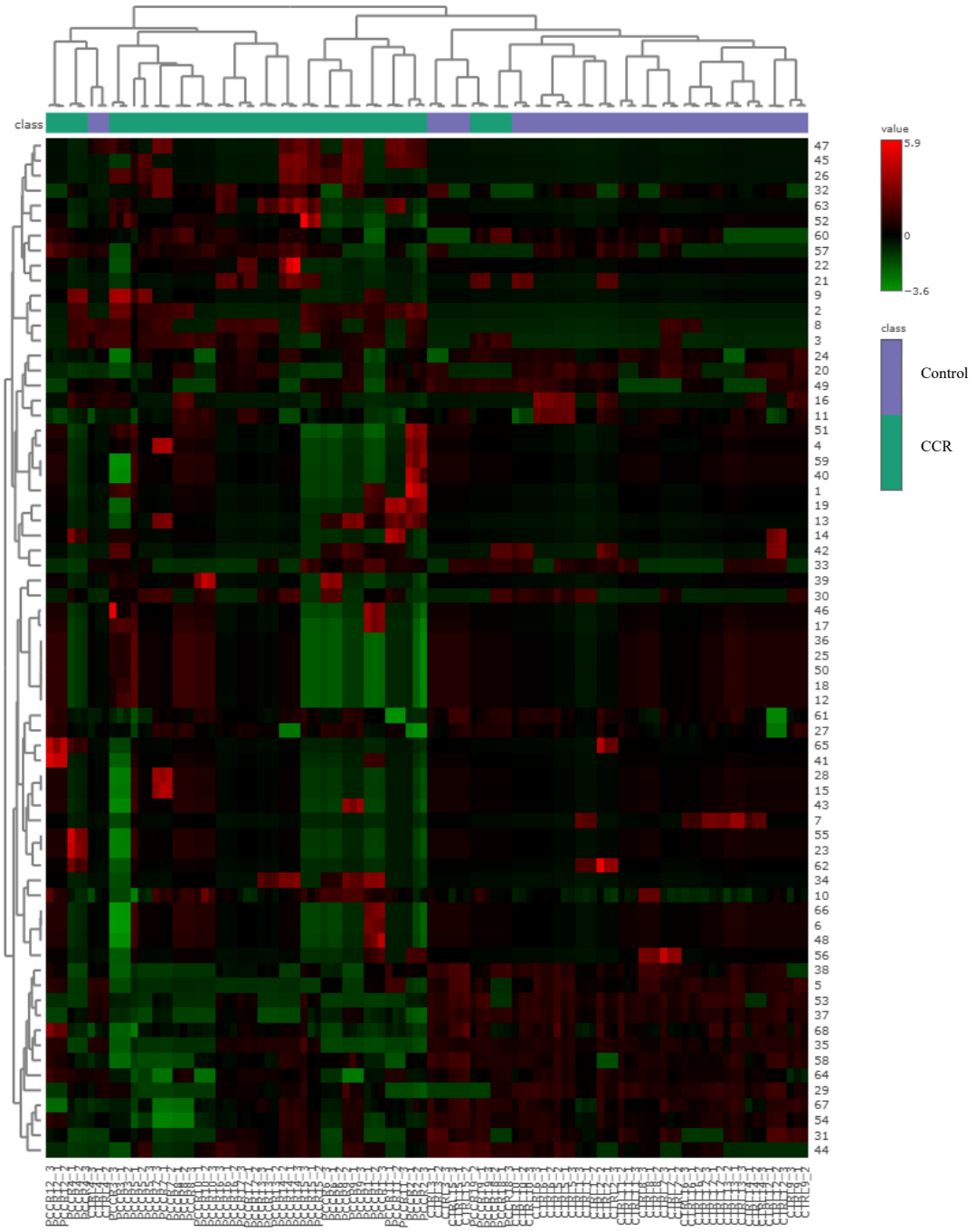


Figure S.1- Heatmap representing the relative abundance of all urinary VOMs identified in the control (purple) and CRC (green) groups. Columns represent individual samples and rows represent VOMs. Color intensity determines abundance (red= highest, green= lowest).

Legend:

1- 2-Methylfuran; 2- 2.5-Dimethylfuran; 3- 2-Pentanone; 4- 1.3-Dimethylcyclohexene; 5- 2-Acetylfuran; 6-Toluene; 7- Hexanal; 8- 3-Hexanone; 9- 2-Ethyl-4-methylimidazole; 10- Dimethyl Disulfide; 11- 4-Heptanone; 12- 2.6-Dimetil-2.4-heptadieno; 13- Limetol; 14- m-xylene; 15- 1.5.5-Trimethyl-6-methylene-cyclohexene; 16- 2-Heptanone; 17- Cosmene; 18- 3-Metil-2-heptanona; 19- Limonene; 20- 2-Pentylfuran; 21- Ocimen Quintoxide; 22- γ-Terpinene; 23- 3-Octanone; 24- o-Cymene; 25- m-Ditiano; 26- (+)-Carvomenthene; 27- 2-Methoxythiophene; 28- Tetrahydromyrcenol; 29- Hemimellitene; 30- Tetrahydrolinalool; 31- p-Cymene; 32- Dihydromyrcenol; 33- Linalool oxide; 34- Nerol Oxide; 35- 2-Ethyl-1-hexanol; 36- Trans-Edulan; 37- p-Xylene; 38- Theaspirane; 39- Benzaldehyde, 40- Linalool; 41- α-Ionene; 42- β-Ionone; 43- β-Cyclocitral; 44- Menthol; 45- Ocimenol; 46- 2(3H) Furanone. dihydro-5-methyl; 47- α-Terpinyl acetate; 48- α-Methylcinnamaldehyde; 49- 4.7-Dimethylbenzofuran; 50-Indan. 1.1.4.6-tetramethyl; 51- 3-Sulfolene; 52- (+)-Carvotanacetone; 53- Naphthalene; 54- 1.2-Dihydro-1.1.6-trimethyl-naphthalene; 55- o-acetiltolueno; 56- Cuminaldehyde; 57- 2.4-Dimethylbenzaldehyde; 58- Damascenone; 59- 2-Bromophenol; 60- Phenol; 61- Dihydro-β-ionone; 62- 1.6-Dimethylnaphthalene; 63- Nerolidol; 64-p-Cresol; 65- 1.6.7-Trimethylnaphthalene; 66- Acetovanillone; 67- Cadalene; 68- 3-t-Butilfenol

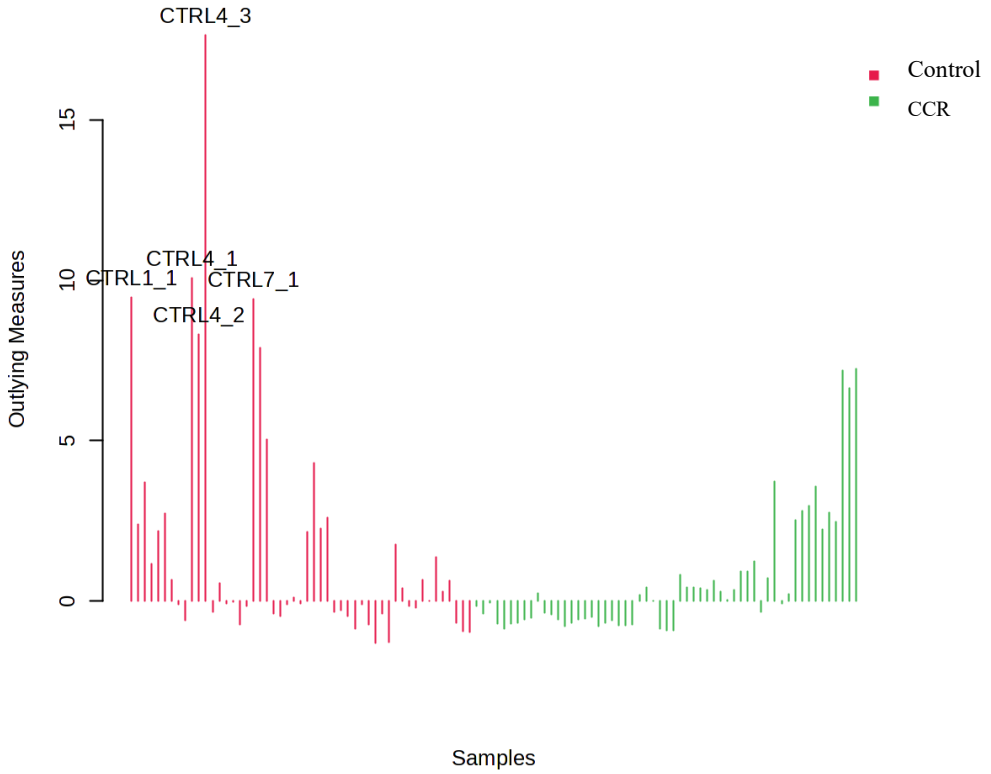


Figure S.2- Graph identifying discordant values for the control (red) and CRC (green) groups. The marked peaks correspond to the samples with the greatest deviation from the group distribution.

Supplementary information

Termo de Consentimento Informado para Estudo de Investigação Clínica, no SESARAM, EPERAM

(modelo 1- aplica-se aos estudos de investigação clínica/translacional em que os dados de saúde/produtos biológicos não saem do SESARAM, EPERAM)

FOLHETO DE INFORMAÇÃO AO PARTICIPANTE

Convidámos o Sr. (a) a participar em um estudo de investigação clínica que será realizado no SESARAM, EPERAM em colaboração com os Laboratórios do Centro de Química da Madeira, Universidade da Madeira.

Nome do Estudo de Investigação:

Estabelecimento do perfil volatilómico da urina de pacientes com cancro colorretal para a identificação de potenciais biomarcadores

1. **Introdução do tema do Estudo de Investigação:**

As lesões malignas colorretais podem representar um grande desafio, quer no seu diagnóstico, quer no seu tratamento. Estas neoplasias são as terceiras com maior incidência a nível mundial, contando para 9,6% do total de tumores (GLOBOCAN 2022).

2. **Objetivos deste Estudo de Investigação:**

Este estudo vai permitir desenvolver métodos laboratoriais, apoiado em plataformas integradas baseadas em OMICS, para diagnostico de tumores colorretais. Contribuindo assim, para o desenvolvimento de um meio de diagnóstico rápido, barato e não invasivo, para deteção destes tumores.

3. **Metodologia do Estudo de Investigação:**

Para participar no estudo, os participantes têm de responder a um questionário onde constam dados epidemiológicos, antecedentes patológicos, averiguando a história pessoal e familiar de cancro. A todos colhem sangue e urina para exames analíticos.

O estudo analítico do sangue é realizado no Laboratório do SESARAM, EPERAM e o da urina nos Laboratórios do Centro de Química da Madeira, na Universidade da Madeira.

4. **Riscos da participação neste Estudo de Investigação:**

Neste estudo o participante tem de fazer somente colheita de sangue para análises. A colheita de urina é feita de forma não invasiva

5. **Benefícios da participação neste Estudo de Investigação:**

Contribuir para desenvolver um método de diagnostico precoce de tumores colorretais.

6. **Confidencialidade e anonimato:**

Toda a confidencialidade e anonimato dos dados será garantida.

A identificação far-se-á através da atribuição de um número, não existindo nenhum material de referência aos dados de identificação. Os dados de saúde são integrados numa "base de dados" de forma anónima e não identificável, tendo como fim o estudo de investigação clínica acima referido que pode ser consultada por todos os investigadores do estudo.

Só o investigador clínico principal do estudo - Drª Isabel Jardim, Médica Gastroenterologista do SESARAM, EPERAM é que tem acesso à identificação do participante e ao código que lhe foi atribuído.

7. **Tratamento, armazenamento e destruição dos dados:**

Todos os dados investigados no estudo são armazenados numa "base de dados" que será protegida por uma palavra-passe. O tratamento desses dados, será feito no SESARAM, EPERAM em programas e equipamentos informáticos com acesso restrito aos membros da Equipa de Investigação do SESARAM, EPERAM.

Os dados de saúde/produto biológico colhido(s) serão para uso exclusivo do presente estudo de investigação, não serão cedidos para nenhum outro estudo.

A "base de dados" deste estudo será destruída no fim do prazo legal de conservação dos dados (de acordo com as políticas editoriais das revistas onde vierem a ser publicadas os estudos).

8. **Divulgação dos resultados deste Estudo de Investigação:**

Somente será divulgado/publicado o produto final da investigação, sob a forma de artigos, comunicações em reuniões científicas e mestrado ou introdução em repositórios científicos ou comunicações para a Comunidade.

9. **Participação neste Estudo de Investigação:**

A sua participação é voluntária, não existe nenhuma contrapartida financeira ou de outra natureza, à sua participação.

Informamos também, que poderá livremente recusar ou interromper a sua participação no estudo, sem qualquer tipo de penalização e que os seus dados de saúde/ produto biológico dados serão de imediato extintos da "base de dados" do estudo.

Nº de entrada no estudo ……

Perturbação psíquica/compreensão comprometida ☐

Nome do Estudo de Investigação:

Estabelecimento do perfil volatilómico da urina de pacientes com cancro colorretal para a identificação de potenciais biomarcadores

Nome……

(nome do participante no estudo)

Cartão de cidadão nº ……

Confirmo ter lido e compreendido toda a informação deste Consentimento Informado. Todas as minhas questões foram esclarecidas.

Concordo voluntariamente participar no estudo de investigação acima referido.

……	……
(data)	(Assinatura do participante ou responsável legal)

Concordo que os meus dados de saúde/produto biológico, sejam tratados de forma anonimizada pelos investigadores do SESARAM, EPERAM e do Centro de Química da Madeira e sejam publicados para a comunidade científica somente como produto final da investigação.

Sei que posso retirar o meu consentimento e que os meus dados serão extintos da "base de dados" do estudo de investigação e o produto biológico será eliminado.

Sei que os meus dados de saúde não sairão do SESARAM, EPERAM. Sei que a minha urina será analisada no laboratório do Centro de Química da Madeira

Sei que receberei uma fotocópia deste formulário de consentimento informado caso pretenda.

……	……
(data)	(Assinatura do participante ou responsável legal)
……	……
(data)	(Assinatura do Investigador)
……	……
(data)	(Assinatura da pessoa que explicou o consentimento)