DEVELOPMENT OF NEW, NON-INVASIVE TOOLS BASED ON FAECAL BACTERIAL SIGNATURES FOR THE EARLY DETECTION OF COLORECTAL CANCER

Marta Malagón Rodríguez

Universitat de Girona

DOCTORAL THESIS

Development of new, non-invasive tools based on faecal bacterial signatures for the early detection of colorectal cancer

Marta Malagón Rodríguez

2019

DOCTORAL THESIS

Development of new, non-invasive tools based on faecal bacterial signatures for the early detection of colorectal cancer

Marta Malagón Rodríguez

2019

Doctoral program in Molecular Biology, Biomedicine and Health

Directed by:
Dr. Xavier Aldeguer Manté

Directed and tutored by:
Prof. L. Jesús Garcia Gil

PhD thesis submitted in fulfillment of the requirements to obtain the doctoral degree from the University of Girona

Certificate of thesis direction

Hereby, Dr. Xavier Aldeguer Manté, and Dr. Jesús Garcia Gil, from the Universitat de Girona

CERTIFY:

That this doctoral thesis entitled "**Development of new, non-invasive tools based on faecal bacterial signatures for the early detection of colorectal cancer**", that Marta Malagón Rodríguez submitted to obtain the doctoral degree from Universitat de Girona, has been completed under their supervision.

In witness whereof and for such purposes as may arise, the following certification is signed:

Dr. Xavier Aldeguer Manté
MD, PhD. Chief of Gastroenterology
Associate Professor of Medicine

Prof. L. Jesús Garcia Gil
Chair of Microbiology
University of Girona

Girona, 2019

Agraïments

Ja ho diuen que tot el que comença acaba. I sembla que, al cap de gairebé tres anys, després d'haver pres la decisió d'endinsar-me en el món de la tesi doctoral aquesta arriba a la seva fi. És doncs, moment de fer balanç i donar les gràcies a totes les persones que m'heu acompanyat en aquest camí al llarg d'aquest temps.

Per començar, vull donar les gràcies als meus directors de tesi, el Dr. Xavier Aldeguer i el Prof. Jesús Garcia Gil. *Gràcies **Xevi** per haver-me donat l'oportunitat aquell abril de 2016 d'entrar a formar part del teu equip dins del grup de recerca de Malalties Digestives i Microbiota. Vas ser la primera persona després d'acabar els meus estudis de màster en donar-me l'oportunitat de treballar en el que realment m'agrada. I a tu, **Jesús,** vull agrair-te el temps que dediques a aportar-me coneixements del món de la microbiologia i la biologia molecular. Sempre que ho he necessitat m'has ajudat i m'has donat molt bons consells. Crec que sou els millors directors que hauria pogut tenir tant a nivell professional com personal.*

Crec però que una de les persones que m'ha aportat més, sense ella saber-ho, és la Mariona Serra. ***Mariona**, recordo estar parlant amb tu per telèfon des de l'escriptori del meu pis de Barcelona. Aleshores estava molt espantada ja que el meu futur era del tot incert. Em vas comentar que al grup de Digestiu de l'Hospital de Girona podrien necessitar a algú per fer el doctorat. Vas ser tu qui em vas posar en contacte amb en Xevi i l'Anna i, al cap i a la fi, ets qui ha fet possible que hagi arribat fins on soc ara. Mil gràcies per la confiança dipositada en mi, per posar-m'ho tot tant fàcil i per haver-me acollit amb els braços oberts a casa teva, GoodGut.*

I parlant de GoodGut... casa meva entre setmana... on hi tinc la meva segona família i on he forjat noves amistats. *Merci per fer el dia a dia més amè amb les nostres apostes sobre Game of Thrones, amb les tertúlies sobre els nostres amics els "Docus", amb les nostres frases top... Gràcies **Marta, Rosor**, **Sara** per cada moment compartit, cadascuna de vosaltres m'ha donat suport en tots els sentits sempre que ho he necessitat. Però sobretot gràcies a vosaltres, **Lia** i **Joan,** per tots els moments viscuts: Madrid (Gambas en Sol, el Mercadona, en gominas...), Washington (la simpàtica catalana, en Felix...), Tarragona (Bad liar, no he dormit res, en Joan de suplent per la presentació...), el CoDa course (viatjar de dimensió en dimensió, picar codi...). Us estimo molt a tots GoodGuteros!*

No em puc deixar a la meva primera companya de feina l'Anna Bahí. *Gràcies* ***Anna*** *per haver-me guiat pels passadissos del Trueta i dins del món dels estudis clínics. M'ha agradat molt poder compartir els primers mesos a Idibgi amb tu al despatxet de la planta 5 ala B!*

També vull reflectir el meu agraïment a les diferents persones que han participat d'alguna manera o altra en la meva tesi, com són el Dr. Joan Brunet, la Dr. Marta Pineda, la Dr. Virgina Piñol, el Dr. Gabi Capellà i el Dr. Antoni Castells. *Gràcies a tots per aportar de forma indirecta el vostre granet de sorra en aquesta tesi.*

I... tornant cap al meu estimat Alt Empordà... on hi tinc la meva gent. Potser no les veig tant com voldria, però sempre que ho he necessitat allà han estat: les meves amigues. *Mil gràcies* ***Irene***, ***Laura***, ***Marina***, ***Eli*** *i* ***Melissa*** *per totes les estones de rialles, bogeries, desconnexió, cerveses, vins, el bon menjar... Sou totes i cada una indispensable per mi!*

I finalment vull agrair a la meva família tot el que ha fet per mi. I són aquests els moments en què me n'adono que el temps va passant i que vaig deixant enrere persones que he estimat molt. Quan vaig presentar el Treball Final de Màster era el iaio Juan qui ens havia deixat... i ara sou vosaltres iaios els que ja no hi sou. ***Iaia Pepita*** *i* ***iaio Cecilio*** *allà on sigueu: us estimo i trobo a faltar molt... sempre animant-me, amb unes paraules dolces per mi... doncs sembla que al final acabo els estudis iaios! Semblava que mai acabarien oi? Us estimo.* ***Mama***, ***papa*** *i* ***Adrià****, tot i que ara ens veiem menys sou indispensables per mi. Els millors papes i germà que hauria pogut tenir! Gràcies pel suport incondicional i per estar sempre al meu costat tant en els bons moments com els no tant bons!!! Y a ti* ***Cris****, porque para mí eres Superman. Porque eres la persona que he escogido y lo haría las veces que hiciera falta. Porque siempre has tenido los brazos abiertos en los momentos de bajón y has estado a tope en los momentos de celebración, ¡que también han sido muchos! Te quiero.*

Acknowledgements

This work was partially founded by the Spanish Ministry of Economy, Industry and Competitivity (MINECO) though project RTC-2016-5017-1 and by the Centro para el Desarrollo Tecnológico Industrial (CDTI) through the NEOTEC project SNEO-20151529.

List of Publications

Some results of this Doctoral Thesis have been published or submitted in scientific journals included in the Journal Citation Report of the Institute of Scientific Information. Peer reviewed publication obtained from this thesis are listed below:

1. Malagón M, Ramió-Pujol S, Serrano M, Serra-Pagès M, Amoedo J, Oliver L, Bahí A, Mas-de-Xaxars T, Torrealba L, Gilabert P, Miquel-Cusachs JO, García-Nimo L, Saló J, Guardiola J, Piñol V, Cubiella J, Castells A, Aldeguer X, Garcia-Gil J. **Reduction of faecal immunochemical test false-positive results using a signature based on faecal bacterial markers**. *Alimentary Pharmacology & Therapeutics*. 2019; 49:1410-1420. https://doi.org/10.1111/apt.15251
   Gastroenterology Q1; IF_{2017}=7.357
2. Malagón M, Ramió-Pujol S, Serrano M, Amoedo A, Oliver L, Bahí A, Miquel-Cusachs JO, Ramirez M, Queralt X, Gilabert P, Saló J, Guardiola J, Piñol V, Serra-Pagès M, Castells A, Aldeguer X, Garcia-Gil J. **Faecal bacterial signature testing for colorectal cancer screening in a positive faecal immunochemical test population**. Submitted to *Gastroenterology* on May 20th.
3. Malagón M, Ramió-Pujol S, Serrano M, Serra-Pagès M, Amoedo A, Oliver L, Bahí A, Guardiola J, Navarro M, Rodríguez L, Uchima H, Pineda M, Capellà G, Castells A, Garcia-Gil J, Aldeguer X, Piñol V, Brunet J. **Use of new faecal bacterial signatures for the individualization of the Lynch syndrome surveillance**. Submitted to *International Journal of Cancer* on May 29th.

Patent

Part of the results of this Doctoral Thesis are included in a European patent application filed at the Spanish Patent and Trademarks Office (OEPM).

- Mariona Serra Pagès, Xavier Aldeguer Manté, Jesús Garcia Gil, Sara Ramió Pujol, Marta Malagón Rodríguez. **Improved method for the screening, diagnosis and/or monitoring of colorectal advanced neoplasia, advanced adenoma and/or colorectal cancer**. GoodGut S.L., Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta, Universitat de Girona (Application number: EP19382186.5; Filling date: 13th March 2019)

This patent application is currently under examination.

List of Abbreviations

AA	Advanced adenoma
AFAP	Attenuated familial adenomatous polyposis
APC	Adenomatous polyposis coli
AUC	Area under curve
BCTF	*Bacteroides fragilis*
BCTT	*Bacteroides thetaiotaomicron*
BFT	*Bacteroides fragilis* toxin
BMI	Body mass index
BRRS	Bannayan-Riley-Ruvalcaba syndrome
CFU	Colony-forming unit
CI	Confidence interval
CIMP	CpG island methylation
CIN	Chromosomal instability
CINT	*Collinsella intestinalis*
CRC	Colorectal cancer
CS	Cowden syndrome
DNA	Deoxyribonucleic acid
ECO	*Escherichia coli*
EGFR	Epidermal growth factor receptor
EUB	Eubacteria
FAP	Familial adenomatous polyposis
FIT	Faecal immunochemical test
FOBT	faecal occult blood test
FPRA	*Faecalibacterium prausnitzii*
FS	Flexible sigmoidoscopy
gFOBT	guaiac faecal occult blood test
GI	Gastrointestinal
GMLL	*Gemella morbillorum*
Hb	Haemoglobin
HNPCC	Hereditary nonpolyposis colorectal cancer
IBD	Inflammatory bowel disease
IEC	Intestinal epithelial cells
Ig	Immunoglobulin
IL	Interleukin
JPS	Juvenile polyposis syndrome

LR	log ratio
LS	Lynch Syndrome
MAP	MUTYH-associated polyposis
MAPK	MAP kinase
miRNA	micro ribonucleic acid
MMR	Mismatch repair
MSI	Microsatellite instability
MSI-H	High microsatellite instability
NA	Non-applicable
NAA	Non-advanced adenoma
NC	Normal colonoscopy
ND	No data
NPV	Negative predictive value
PCR	polymerase chain reaction
PHG	Phylogroup
PHTS	*PTEN* hamartoma tumour syndrome
PPV	Positive predictive value
PTST	*Peptostreptococcus stomatis*
qPCR	quantitative polymerase chain reaction
RAID	Risk Assessment Intestinal Disease
RNA	ribonucleic acid
ROC	Receiver operating curve
rRNA	ribosomal ribonucleic acid
RSBI	*Roseburia intestinalis*
SA	Serrated adenoma
SCFA	Short chain fatty acids
SPS	Serrated polyposis syndrome
TGF	Tumour growth factor
TNF	Tumour necrosis factor
UC	Ulcerative colitis

List of Figures

List of Tables

List of Boxes

Table of Contents

Summary

Colorectal cancer (CRC) is the third most common cancer and the second most common cancer cause of death worldwide, accounting for more than 800,000 deaths in 2018. Screening programs of CRC are recommended, as this disease constitutes an important health issue, its natural history has been widely studied, diagnostic methods for its early detection are available and its treatment is more effective when it is early diagnosed. The most commonly used non-invasive method is the faecal immunochemical test (FIT), which detects specifically human haemoglobin in faeces. Subjects who obtain a FIT-positive result are prompted to undergo a diagnosis colonoscopy. It is important to note that CRC-screening strategies show high performance in detecting not only CRC but also precancerous lesions. FIT sensitivity for advanced neoplasia (advanced adenomas and CRC) is around 30% and its positive predictive value (PPV) is not higher than 15%, which results in a high false-positive rate for advanced neoplasia. The considerable low PPV of FIT may be responsible for up to 30% unnecessary colonoscopies. Therefore, it is necessary to develop new, non-invasive analytical tools for CRC screening with higher specificity in order to reduce its false positive results.

The main purpose of this doctoral thesis was to search for faecal bacterial markers for CRC early detection, able to detect precancerous lesions before the appearance of clinical signs. To achieve this objective we analysed CRC and IBD-associated faecal bacterial markers through quantitative PCR in two clinically independent populations: (1) asymptomatic subjects aged between 50 and 69 years old (both included) who had obtained a positive result of FIT in an organised CRC-screening program (**Chapter 1**); and (2) subjects with symptoms compatible with CRC (i.e. presence of blood in stool, abdominal pain, changes in bowel habits, and/or unexplained weight loss) (**Chapter 2**). In Chapter 1, we report the development of a non-invasive tool (RAID-CRC) based on the combination of 4 bacterial markers (Eubacteria, *Peptostreptococcus stomatis*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*) with FIT (cut-off = 10 µg of haemoglobin/g of faeces) that was able to detect advanced neoplasia in symptomatic individuals (n=333) with 80% sensitivity and 90% specificity. In parallel, in Chapter 2 we describe the setup

of a second faecal bacterial signature in an asymptomatic population that participated in the national CRC-screening program (screening program FIT cut-off = 20 µg haemoglobin/g of faeces). In this case, the signature was firstly defined in a proof-of-concept study with 172 individuals and later validated on an external cohort of 327 subjects. This tool comprised 6 bacterial markers (Eubacteria, *Faecalibacterium prausnitzii*, *Gemella morbillorum*, *Bacteroides fragilis*, B46, and B48) and it was called RAID-CRC Screen. With this second system a sensitivity of 84% and a specificity of 16% were obtained for advanced neoplasia detection. In the validation cohort FIT produced 184 false-positive results. Using RAID-CRC Screen this value was reduced to 154, thus reducing the false-positive rate by more than 16%.

The second objective of the thesis was to define a faecal bacterial signature to be used for prediction of the absence of neoplastic lesions in Lynch syndrome carriers (LS), aiming at improving surveillance into that group of subjects genetically predisposed to develop CRC (**Chapter 3**). A new, non-invasive tool based on the combination of three faecal bacterial markers (Eubacteria, B46, and *Escherichia coli*) for the individualization of Lynch syndrome surveillance was developed (RAID-LS) obtaining a sensitivity of 100% and a specificity of 72%, which would suppose a reduction up to 70% of the unnecessary colonoscopies performed during LS surveillance.

The results obtained in this work support the use of microbiota as disease indicator in diseases such as inflammatory bowel disease, oral infections, or rheumatoid arthritis; whose microbiota has been reported to be altered. Hence, a broad field in human microbiota is yet to be explored, which surely will contribute to the development of new *in vitro* diagnosis methods.

Resum

El càncer colorectal (CCR) és el tercer càncer més freqüent i la segona causa de mort per càncer a nivell mundial, provocant més de 800.000 morts l'any 2018. Es recomana el cribratge del CCR ja que constitueix un important problema de salut, la seva història natural ha estat àmpliament estudiada, es disposa de mètodes diagnòstics per la seva detecció precoç i el seu tractament és més efectiu quan es detecta en un estadi inicial. El mètode no invasiu més utilitzat és el test de sang oculta en femta (FIT), amb el que es detecta l'hemoglobina humana en femta. Els subjectes que obtenen un resultat positiu del FIT són derivats a sotmetre's a una colonoscòpia. Cal destacar que els mètodes per al cribratge del CCR no només han de tenir una alta capacitat diagnòstica per al CCR pròpiament dit, sinó que també l'han de tenir per la detecció de lesions precanceroses. La sensibilitat del FIT per a la detecció de neoplàsia avançada (CCR i adenomes avançats) és del 30% i el seu valor predictiu positiu (VPP) és del 15%, fet que resulta en una elevada taxa de falsos positius per la neoplàsia avançada. De fet, el baix VPP pot portar a la realització de més d'un 30% de colonoscòpies de forma innecessària. Per tant, és necessari desenvolupar noves eines no invasives per al cribratge del CCR amb major especificitat i així poder reduir l'elevat nombre de falsos positius.

L'objectiu principal d'aquesta tesi doctoral és el desenvolupament d'una nova eina no invasiva basada en marcadors bacterians fecals per a la detecció precoç del CCR, que sigui capaç de detectar lesions precanceroses abans que apareguin signes clínics. Per tal d'assolir aquest objectiu vam analitzar marcadors bacterials fecals relacionats amb el CCR i la malaltia inflamatòria intestinal (MII) mitjançant la PCR quantitativa en dues poblacions clínicament independents: (1) subjectes asimptomàtics d'entre 50 i 69 anys (ambdós inclosos) que haguessin obtingut un resultat positiu del FIT en el marc d'un programa organitzat de cribratge del CCR (**Capítol 1**); i (2) subjectes amb símptomes compatibles amb el CCR (presència de sang en femta, dolor abdominal, canvi en el hàbits intestinals, i/o pèrdua de pes inexplicable) (**Capítol 2**). En el Capítol 1, vam desenvolupar una eina no invasiva (RAID-CRC) basada en la combinació de 4 marcadors bacterians (Eubacteria, *Peptostreptococcus stomatis*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*) amb el FIT (punt de tall = 10 µg d'hemoglobina/g de femta) capaç de

detectar en una població simptomàtica (n=333) la neoplàsia avançada amb una sensibilitat del 80% i una especificitat del 90%. En paral·lel, en el Capítol 2 vam definir una segona signatura bacteriana fecal en una població asimptomàtica participant en el programa nacional de cribratge (punt de tall del FIT en el programa de cribratge = 20 µg d'hemoglobina/g de femta). En aquest cas, la signatura es va definir en una prova de concepte amb 172 individus i seguidament es va validar en una cohort externa de 327 subjectes. La firma, que rep el nom de RAID-CRC Screen, està formada per 6 marcadors bacterians (Eubacteria, *Faecalibacterium prausnitzii*, *Gemella morbillorum*, *Bacteroides fragilis*, B46, and B48). Amb aquesta segona eina es va obtenir una sensibilitat del 84% i una especificitat del 16% per la detecció de la neoplàsia avançada. En aquesta cohort de validació hi va haver 184 resultats falsos positius. Utilitzant el RAID-CRC Screen aquest valor es va reduir a 154, el que suposa una reducció de la taxa de falsos positius de més del 16%.

L'objectiu secundari de la tesi va ser definir una signatura bacteriana fecal que permetés predir l'absència de lesions neoplàsiques en els portadors de la síndrome de Lync (LS), subjectes amb predisposició genètica de desenvolupar CCR, i així poder millorar el seu seguiment (**Capítol 3**). Es va desenvolupar una nova eina no invasiva (RAID-LS) basada en la combinació de 3 marcadors bacterians fecals (Eubacteria, B46, and *Escherichia coli*) per la individualització del seguiment de la síndrome de Lynch, obtenint una sensibilitat del 100% i una especificitat del 72%, el que suposaria una reducció de fins al 70% de les colonoscòpies innecessàries que es realitzen en el seguiment del LS.

Els resultats obtinguts en aquest treball postulen l'ús de la microbiota com a indicador de malaltia en malalties com la malaltia inflamatòria intestinal, infeccions orals, o l'artritis reumatoide; en les quals s'ha observat una alteració de la microbiota. Per tant, hi ha un ampli món per explorar pel que fa a la microbiota humana, que segurament contribuirà en el desenvolupament de nous mètodes per al diagnòstic *in vitro*.

Resumen

El cáncer colorrectal (CCR) es el tercer cáncer más frecuente y la segunda causa de muerte por cáncer a nivel mundial, provocando más de 800.000 muertes en 2018. Se recomienda el cribado del CCR ya que constituye un importante problema de salud, su historia natural ha estado ampliamente estudiada, se dispone de métodos diagnósticos para su detección precoz y su tratamiento es más efectivo cuando se detecta en un estadio inicial. El método invasivo más utilizado es el FIT, con el que se detecta la hemoglobina humana en heces. Los sujetos que obtienen un resultado positivo del FIT son derivados a someterse a una colonoscopia. Cabe destacar que los métodos para el cribado del CCR no solo tienen que tener una alta capacidad diagnóstica para el CCR, sino que también la deben tener para la detección de lesiones precancerosas. La sensibilidad del FIT para la detección de neoplasia avanzada (CCR y adenomas avanzados) es del 30% y su valor predictivo positivo (VPP) es del 15%, hecho que resulta en una elevada tasa de falsos positivos para la neoplasia avanzada. El bajo VPP puede llevar a la realización de más de un 30% de colonoscopias de forma innecesaria. Por lo tanto, es necesario desarrollar nuevas herramientas no invasivas para el cribado del CCR con una mayor especificidad y así poder reducir el elevado número de falsos positivos.

El principal objetivo de esta tesis doctoral es el desarrollo de una nueva herramienta no invasiva basada en marcadores bacterianos fecales para la detección precoz del CCR, que sea capaz de detectar lesiones precancerosas antes de que aparezcan signos clínicos. Para conseguir este objetivo analizamos marcadores bacterianos fecales relacionados con el CCR y la enfermedad inflamatoria intestinal (EII) mediante la PCR cuantitativa en dos poblaciones clínicamente independientes: (1) sujetos asintomáticos de entre 50 y 69 años (ambos incluidos) que hubiesen obtenido un resultado positivo del FIT en el marco de un programa organizado de cribado del CCR (**Capítulo 1**) y (2) sujetos con síntomas compatibles con el CCR (presencia de sangre en heces, dolor abdominal, cambio en los hábitos intestinales, y/o pérdida de peso inexplicable) (**Capítulo 2**). En el Capítulo 1, desarrollamos una herramienta no invasiva (RAID-CRC) basada en la combinación de 4 marcadores bacterianos (Eubacteria, *Peptostreptococcus stomatis*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron*) con el FIT (punto de corte = 10 µg de

hemoglobina/g de heces) capaz de detectar en una población sintomática (n=333) la neoplasia avanzada con una sensibilidad del 80% y una especificidad del 90%. En paralelo, en el Capítulo 2 definimos una segunda firma bacteriana fecal en una población asintomática participante en el programa nacional de cribado (punto de corte del FIT en el programa de cribado = 20 µg de hemoglobina/g de heces). En este caso, la firma bacteriana se desarrolló en una prueba de concepto con 172 individuos y seguidamente se validó en una cohorte externa de 327 sujetos. La firma, que recibe el nombre de RAID-CRC Screen, está formada por 6 marcadores bacterianos (Eubacteria, *Faecalibacterium prausnitzii*, *Gemella morbillorum*, *Bacteroides fragilis*, B46, and B48). Con esta segunda herramienta se obtuvo una sensibilidad del 84% y una especificidad del 16% para la detección de la neoplasia avanzada. En esta cohorte de validación hubo 184 resultados falsos positivos. Utilizando en RAID-CRC Screen este valor fue reducido a 154, lo que supone una reducción de la tasa de falsos positivos de más del 16%.

El objetivo secundario de la tesis fue definir una firma bacteriana fecal que permitiera predecir la ausencia de lesiones neoplásicas en los portadores del síndrome de Lynch (LS), sujetos con predisposición genética de desarrollar CCR, y así poder mejorar su seguimiento (**Capítulo 3**). Se desarrolló una nueva herramienta no invasiva basada en la combinación de 3 marcadores bacterianos fecales (Eubacteria, B46, and *Escherichia coli*) para la individualización del seguimiento del síndrome de Lynch, obteniendo una sensibilidad del 100% y una especificidad del 72%, lo que supondría una reducción de hasta el 70% de las colonoscopias innecesarias que se realizan en el seguimiento del LS. Estos resultados necesitan ser validados clínicamente en una cohorte externa más amplia.

Los resultados obtenidos en este trabajo postulan el uso de la microbiota como indicador de enfermedad en enfermedades como la enfermedad inflamatoria intestinal, infecciones orales, o la artritis reumatoide; en las cuales se ha observado una alteración de la microbiota. Por lo tanto, existe un amplio mundo para explorar en relación a la microbiota humana, que seguramente contribuirá en el desarrollo de nuevos métodos para el diagnóstico *in vitro*.

1

Scientific Background

1 Introduction to Colorectal Cancer

Colorectal cancer (CRC) is the second leading cause of cancer death worldwide, accounting for more than 880,000 deaths in 2018 [1]. Incidence varies globally and is directly related to elements of a so-called western life [2,3]. Although there are strong hereditary components, most colorectal cancers develop sporadically and over several years following the adenoma-carcinoma sequence [4,5]. The most commonly used therapies are surgery, radiotherapy for patients with rectal cancer and chemotherapy for stages III and IV colon cancer [6]. CRC screening has been shown to reduce incidence and mortality, however, only a few countries have implemented organised screening programs [7–9].

1.1 Epidemiology

Colorectal cancer incidence and mortality rates vary significantly around the world. Globally, CRC is the third most common cancer in men and the second in women [1]. The highest incidence rates are found in the most developed countries (i.e. Australia, New Zealand, Europe and North America), while the lowest rates are found in the poorest countries (Africa and South-Central Asia) (Figure 1). These geographic differences could be associated with the western diet and other environmental aspects. Low socioeconomic status is also related to a higher risk for the development of CRC [10].

CRC mortality rates have declined progressively in many western countries [2]. This reduction can be attributed to the application of organized CRC screening programs, as colonic polyps can be detected and removed, and CRC can be detected at an earlier stage; and primary and adjuvant treatments are more effective. On the contrary, death rates keep increasing in countries with more limited resources, specifically in Central and South America and Eastern Europe [2].

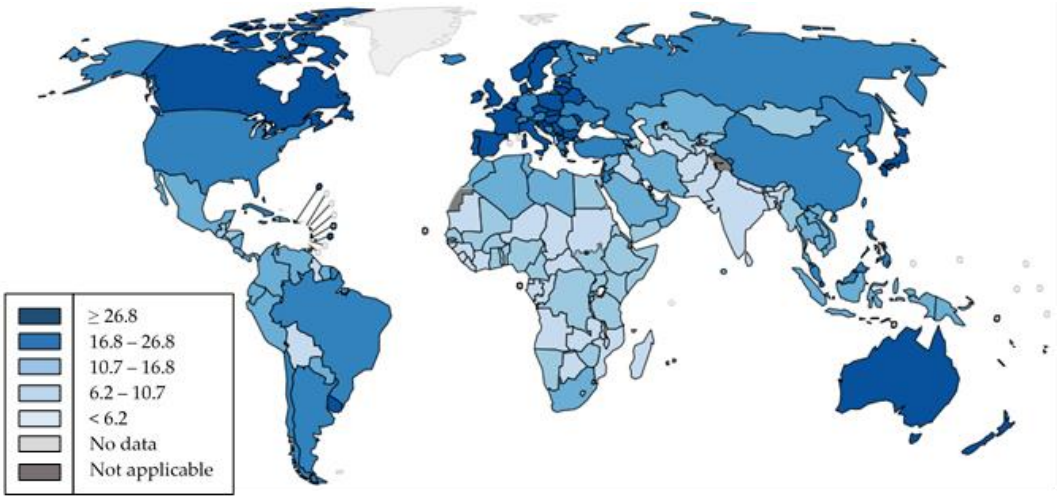

Figure 1. Worldwide colorectal cancer incidence in 2018 [1].

1.2 Risk factors

1.2.1 Non-modifiable risk factors

a) Age, sex and ethnicity

Colorectal cancer is uncommon before the age of 40, but the incidence increases significantly between 40 and 50 years and from then rates increase in each succeeding decade thereafter. Some studies suggest that CRC incidence is increasing in people under 50 years while it is decreasing in older [11,12]. Other studies report an increasing incidence rate of CRC among young people who are between 20 and 39 years old [13,14]. Regardless of these studies, the average age for colon cancer diagnosis for men is 68 and for women 72, and in terms of rectal cancer it is 63 in both sexes, hence CRC is a disease associated with oldness [15]. CRC incidence and mortality are higher in men than in women (Figure 2). Gender disparity might be due to different exposures to other risk factors such as smoking, sex hormones and complex interactions between these variables [16]. Both CRC and colonic adenomas appear to be distributed in the proximal colon in women, specifically in postmenopausal women [17]. For this reason, sigmoidoscopic examinations have been considered inadequate as screening test in women.

According to the American Cancer Society CRC incidence and mortality were highest in non-Hispanic blacks and lowest in Asians and Pacific Islanders (Figure 2) [15]. Differences among ethnicities may be explained by their socioeconomic status and

education. In 2015 in the USA 24% of non-Hispanic blacks lived in poverty, compared to 11% of Asians and Pacific Islanders and 9% of non-Hispanic whites [18]. Interestingly, Alaska Natives have the highest CRC incidence and mortality rates in the USA, being 80% higher than those in non-Hispanic blacks [19]. These rates might be caused due to specific CRC risk factors such as a diet high in animal fat and low in fruits and vegetables, vitamin D deficiency, smoking, obesity and diabetes [20,21].

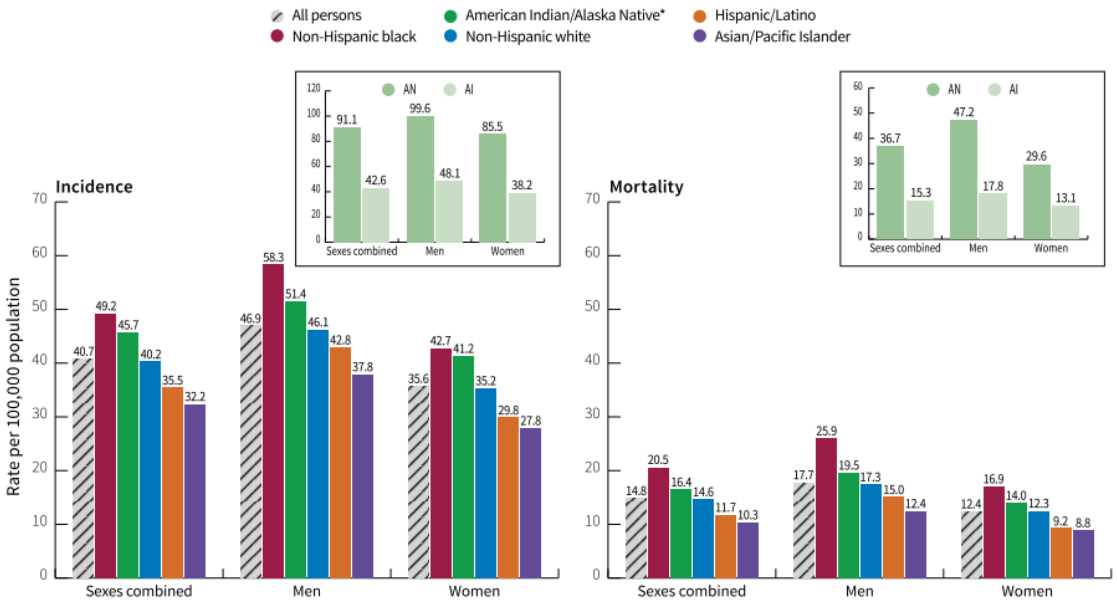

Figure 2. Colorectal cancer incidence (2009-2013) and mortality (2010-2014) rates by ethnicity and sex (USA data) [15]. AN: Alaska Native; AI: American Indian, excluding Alaska.

b) Hereditary CRC syndromes

The understanding of the molecular pathogenesis of colorectal cancer has led to the identification of several genetic disorders that are linked to a high risk of developing CRC. The most common familial colon cancer syndromes are familial adenomatous polyposis (FAP), MUTYH-associated polyposis (MAP) and Lynch syndrome, which is also known as hereditary nonpolyposis colorectal cancer (HNPCC). However, together these conditions account for only 5% of the CRC cases, being the majority Lynch syndrome. Despite its low proportion, these hereditary syndromes are important from a physiological, clinical and therapeutic point of view. Some of the genes involved in tumoral development appear mutated in the germline of hereditary syndromes carriers

[22]. The identification of these genes has made it possible to improve screening and surveillance strategies, and it has allowed the adoption of more radical therapeutic measures than those used in sporadic CRCs [23].

c) Personal or family history of sporadic CRCs

Subjects with a personal history of colonic adenomatous polyps or CRC have higher risk of developing future colorectal cancers. Between 0.5% and 9% of the patients who undergo colorectal resections because of CRC develop metachronous primary cancers in the first five years after surgery [24]. Those patients that have had a large (>1 cm) adenomatous polyp, polyps with villous component and polyps with high-grade dysplasia, show a higher risk of CRC [25].

Having at least one first-degree relative (parents, siblings or children) with CRC doubles the risk of developing a colorectal tumour when compared to general population [26]. This risk is further increased when two first-degree relatives or one first and one or more second-degree relatives have suffered from colon cancer [27,28]. If these relatives have been diagnosed of CRC below 50 years of age the risk is even more increased.

d) Inflammatory Bowel Diseases (IBD)

Patients with longstanding extensive colitis also show an increased risk of developing colorectal cancer (Figure 3) [29]. Population-based studies demonstrate that the risk for ulcerative colitis (UC) patients is 5.7 times that of the general population [30]. CRC is a leading cause of inflammatory bowel disease-associated death, accounting for 1 in 6 patients with UC [31].

In terms of Crohn's disease, when patients are classified into those with colonic involvement and those without, it is clear that the risk in Crohn's colitis is almost the same to that of ulcerative colitis, being 5.6 times that of general population [32].

The most important risk factor for developing colonic cancer is the anatomical extent of colonic inflammation as first reported in 1990 by Ekbom *et al* [30]. The severity of inflammation is also thought to be an independent risk factor for colonic cancer as demonstrated in a case controlled study from the UK [33]. Finally, the duration of colitis

is also an important risk factor. A metanalysis performed in 2001 by Eaden *et al* showed that when colitis persists around 10 years, the cumulative CRC risk to a patient with UC is 1.6%, increasing to 8.3% by 20 years and 18.4% by 30 years [34].

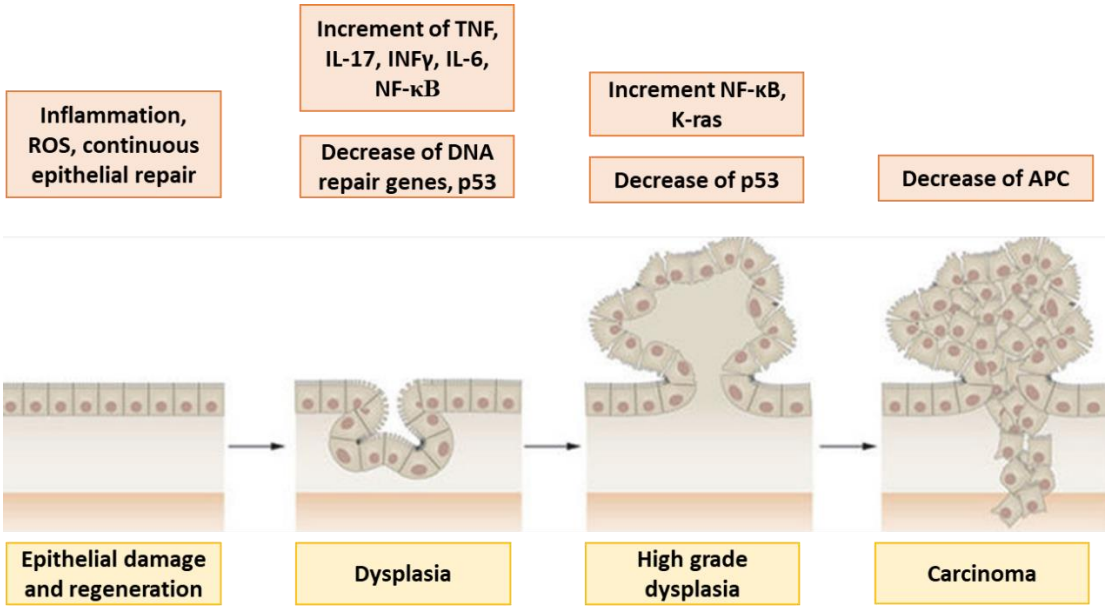

Figure 3. Mechanism of colitis-associated cancer development. Chronic inflammation is characterized by production of proinflammatory cytokines that can induce mutations in oncogenes and tumour suppressor genes (APC, p53) and subsequently the production of reactive oxygen species (ROS). Persistent inflammation prompts a favourable environment for tumour progression. TNF: Tumour Necrosis Factor; IL-17: Interleukin-17; INFγ: Interferon-γ; IL-6: Interleukin-6; NF-κB: Nuclear Factor-Kappa B; APC: Adenomatous Polyposis coli. Adapted from [35].

1.2.2 Modifiable risk factors

a) Diet:

Several epidemiological studies have reported an association between an equilibrated diet, high in fruits and vegetables, and protection from colorectal cancer [36–38]. Comparison between vegetarians and non-vegetarians showed that the first have a reduced risk of CRC, with an even more pronounced effect among those vegetarians who eat fish [39].

Fibre has been identified to be a protective agent against CRC. However, its role has generated opposed opinions in the literature. On the one hand, there are some studies that report that fibre acts as a diluent and absorbent of faecal carcinogens, modulates the colonic traffic, alters the biliary acids metabolism, reduces the intestinal pH and helps to increase the short chain fatty acids (SCFAs) production [40–43]. On the

other hand, other studies have not detected significant differences between fibre effect and CRC risk [44,45].

Folate, calcium and vitamin D are nutrients that have also been described as protective agents against colorectal cancer [46]. Lee *et al* observed that folate intake was inversely associated with risk of CRC only during early pre-adenoma stages [44]. In terms of calcium, it has been suggested that it has the potential to reduce CRC risk as it can reduce the proliferative effect of secondary bile acids in the colon [47]. These acids are produced during fats digestion and are highly irritating to the epithelial colonic cells. Calcium interacts with secondary bile acids forming insoluble soaps thus neutralizing their ability to irritate the epithelial surface, which consequently reduces CRC risk.

Red and processed meat long-term consumption is linked to an increased risk of CRC [48,49]. A factor that contributes to this risk is high cooking temperature as it may prompt polyaromatic hydrocarbons and other carcinogens production. However, modest consumption (once or twice per week) is acceptable for a healthy balanced diet [50].

b) Obesity and physical activity

Obesity is defined as an excessive fat accumulation which causes low-grade chronic inflammation and is closed linked to cancer [51]. It is probably because the adipose tissue is the largest endocrine organ in the body and it stimulates secretion of several cytokines such as IL-8, IL-6, IL-2, lactate dehydrogenase and TNFα (Figure 4). These cytokines are important in tumour initiation and progression [52,53] and in promoting epithelial mesenchymal transition and metastasis in obese patients [54].

Regular physical activity has been observed to be associated with CRC protection. When active individuals are compared to inactive ones, there is a reduction of 27% in the risk of suffering from proximal colon cancer [55]. The mechanism underlying this protective association is unknown and there have not been reported any intervention trials of physical activity for CRC protection [50]. However, the reduction of the accumulated fat as a result of constant physical activity could explain this association.

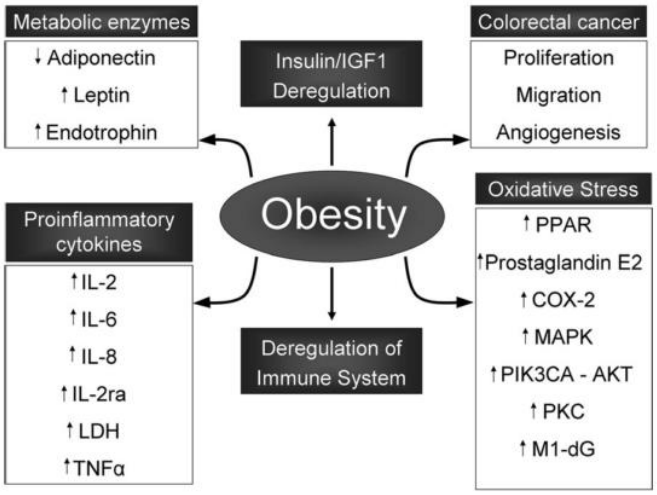

Figure 4. Schematic diagram of biochemical features of obesity. Up and down arrows denote up- or downregulation respectively [51].

c) Smoking

In the lung there is intimate exposure to inhaled tobacco smoke which results in about 80% of all primary lung cancers [56]. Smoking also increases the risk of cancers in organs for which exposure to tobacco degradation products is indirect (i.e. kidney, bladder, cervix, lower urinary tract and pancreas) [57]. In terms of digestive tract, oesophageal and gastric cancer have been strongly associated with tobacco. However, the smoking-colorectal cancer link remains controversial. Botteri *et al* performed a meta-analysis in 2008 in order to clarify the association of cigarette smoking and CRC [58]. It included 106 observational studies which estimated that the risk of developing CRC was higher among smokers compared to those who never smoked. For this reason, clinicians ask CRC survivors to avoid smoking.

d) Alcohol

Several studies have shown an association between alcohol consumption and an increased risk of CRC. A meta-analysis of 27 cohort and 34 case-control studies concluded that there is an association between alcohol drinking of more than 1 drink per day and colorectal cancer risk [59]. These results reproduce what other pooled analyses observed [60,61]. It is well-known that excessive alcohol consumption is a potentially modifiable risk factor for CRC and several other malignancies. Moreover, it complicates treatment and its outcomes by contributing to longer hospitalizations, longer recovery

periods, higher costs, and greater mortality. Consequently, several associations have made proposals for promoting awareness of the relationship between certain types of cancer and alcohol abuse [62].

1.3 Colonic epithelium biology

The function of the intestinal tract is to absorb and digest nutrients from ingested food and dispose of material that is not usable. The colon is the part where the terminal steps of the digestion take place. It functions to form and store faeces, which involves the further recovery of water and electrolytes and addition of bacteria and mucus [63].

Anatomically, the colon can be divided into ascending, transverse, descending and sigmoid colon (Figure 5A). At the clinic-pathological level it can be divided in two parts, the right or proximal colon, which includes the caecum, and the ascending and transverse colon; and the left or distal colon, which is formed by the descending and the sigmoid colon, and the rectum (Figure 5A). Histologically, the wall of the gastrointestinal tract is composed of four layers [64] (Figure 5B). The innermost layer is the mucosa, which can be divided into the epithelium, an underlying layer of loose connective tissue called the lamina propria, and a thin double layer of smooth muscle called the muscularis mucosa, which is often present. Underneath there is the submucosa, which is a loose connective tissue layer, with larger blood vessels, lymphatics, nerves, and can contain mucous secreting glands. This layer is followed by the muscularis propria, which is usually divided into two layers; the inner layer is circular and the outer layer is longitudinal. These layers are composed of smooth muscle used for peristalsis to move food down through the gut. Finally, the outermost layer is called the adventitia, also known as serosa, which is formed of loose connective tissue covered by visceral peritoneum and contains blood vessels, lymphatics and nerves.

The gut epithelium constitutes a barrier between our body and the surrounding environment. It is formed by several different cell types, including absorptive and secretory cells and cells that produce hormones [63]. Colonocytes are absorptive epithelial cells that constitute the majority of cells of this monolayer. They are responsible for absorbing nutrients and transporting them through the epithelium to allow uptake into capillaries located in the underlying layer [63]. Goblet cells are

secretory cells that are scattered between colonocytes. These cells synthesize and secrete bioactive molecules such as secretory and membrane-bound mucins, trefoil peptides, resistin-like molecule β, and Fc-γ binding protein, which are components of mucus [65]. Mucus is required for the movement and diffusion of intestinal contents, and it also provides protection against chemical damage [66]. Another cellular type are enteroendocrine cells, which are also scattered through the epithelial monolayer. They release hormones to regulate the secretion of molecules such as digestive enzymes or bile fluids from the pancreas and liver into the gut. Finally, Paneth cells, which are most commonly found in the small intestine [67], reside at the bottom of the crypts and secrete antimicrobial molecules when they are exposed to microbes or their antigens [68].

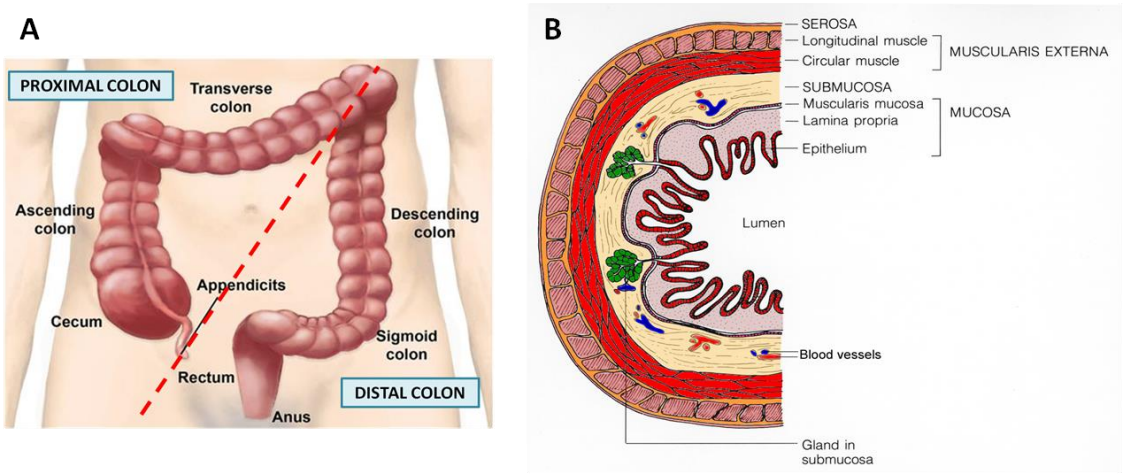

Figure 5. Large intestine anatomy: (A) Anatomic and clinic-pathological divisions. (B) Scheme of the histological layers of the digestive tube.

The origin of colorectal tumours take place in the intestinal epithelium [63]. Unlike the small intestine, the large intestine is a plain surface which gains surface thanks to epithelial cavities that are structured in crypts. These crypts, regulate the homeostasis of the gut epithelium because a dynamic equilibrium between cell production and death is created [63]. Therefore, this adult tissue constantly regenerates [69]. At the bottom of the crypts there are stem cells, which are constantly dividing and giving place to pluripotent cells and non-differentiated progenitors that migrate to take up their permanent positions at the base of the crypt [69]. Two main lineages of differentiated cells are discerned within the intestinal epithelium: the enterocyte or absorptive lineage and the secretory lineage [70] (Figure 6). Once differentiated cells reach the surface of the

intestinal epithelium, they are removed by apoptosis and released into gut lumen a few days later [70].

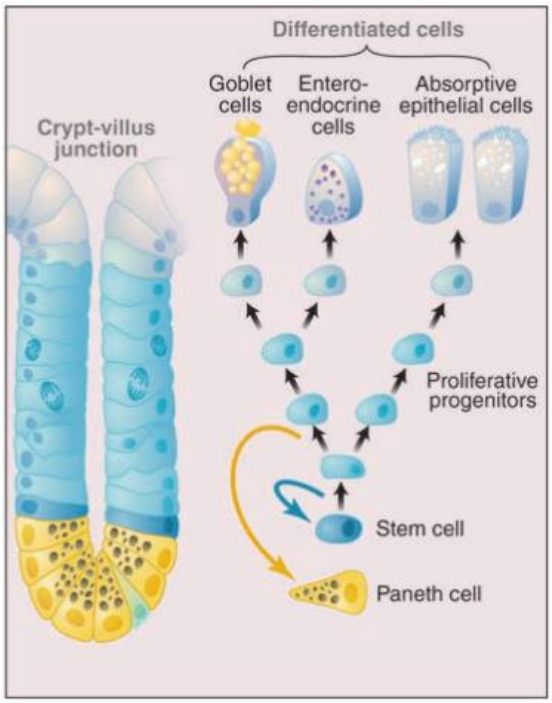

Figure 6. Large intestine epithelium. The large intestine epithelium is shaped into crypts. The lineage scheme depicts the stem cell, the transit-amplifying cells, and the two differentiated branches. The right branch constitutes the enterocyte lineage and the left is the secretory lineage [70].

1.4 Molecular pathogenesis in colorectal cancer

In 1990s, Fearon *et al* proposed a genetic model for colorectal carcinogenesis in which genomic alterations occur in a sequence that is parallel to the clinical progression of the tumour (Figure 7) [22]. Consequently, three different molecular carcinogenesis pathways have been described for CRC.

The most commonly identified in sporadic CRC is the chromosomal instability pathway (CIN), which is observed in 65%-70% of the patients [71]. It is characterized by an accelerated rate of gains or losses of whole or large portions of chromosomes that results in karyotypic variability among cells [72]. Consequently, there appears an imbalance in chromosome number (aneuploidy), subchromosomal genomic amplifications and a high frequency of loss of heterozygosity [71]. The second pathway is the microsatellite instability pathway (MSI), accounting for approximately 15% of the cases. Microsatellite instability is a hypermutable phenotype caused by the loss of DNA

mismatch repair activity [73]. Finally, the third pathway is the CpG island methylation (CIMP), which is more associated with serrated neoplasia. The molecular mechanism for CIMP remains unknown but it is present in nearly all tumours with aberrant methylation of *MHL1* [74], which is a mismatch repair gene involved in repairing base-to-base mismatches and insertion-deletion loops [75].

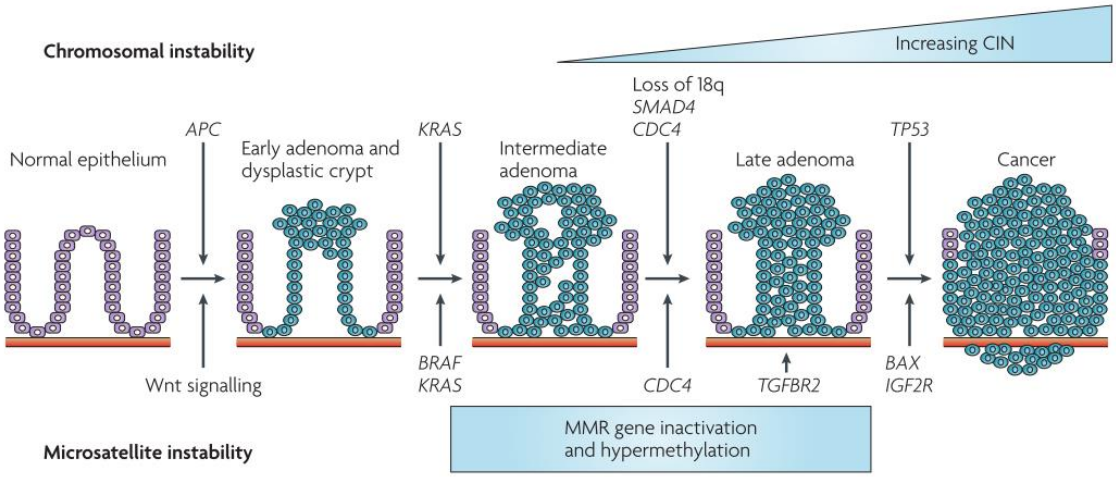

Figure 7. Genetic model in colorectal cancer [76].

1.4.1 Key Genes in CRC

a) Adenomatous Polyposis Coli

The adenomatous polyposis coli (*APC*) gene is a key tumour suppressor gene. Germ-line mutations in this gene result in familial adenomatous polyposis [77], in which hundreds to thousands of adenomatous colonic polyps develop [78]. Mutations in *APC* have been found in a 60% of sporadic carcinomas and adenomas. Several studies indicate that *APC* is essential for the development and homeostasis and that its inactivation facilitates tumorigenesis [79–82]. Mutated *APC* leads β-catenin to accumulate, which in turn activates the canonical Wnt pathway, which is sufficient for intestinal tumorigenesis [77], and increases the transcriptional activity of different downstream genes, including important oncogenes such as *MYC* and *CCND1* [83].

b) KRAS

KRAS is an oncogene member of the MAP kinase (MAPK) pathway, which regulates cell proliferation, differentiation, senescence, and apoptosis. There are

different RAS oncogenes, including *HRAS*, *NRAS* and *KRAS*, which is the most commonly mutated RAS family member in CRC and it is found mutated in 40% of sporadic CRCs [84]. *KRAS* encodes a small protein that transduces signals from the epidermal growth factor receptor (EGFR) family. *KRAS* mutation leads to constitutive activation of the MAPK cascade and promotes cell proliferation and survival independent of the EGFR [84]. The degree to which colorectal tumours depend on *KRAS* is still under investigation [85] but it is one of the most important drug targets for CRC because of its high rate of mutation in this cancer [86].

c) BRAF

BRAF is an oncogene member of the RAF family of the serine/threonine kinases, which as RAS pathway, regulates cell proliferation, differentiation, senescence, and apoptosis. *BRAF* mutations are found in about 10% of sporadic CRCs and are rare in Lynch syndrome carriers [84]. This is helpful to distinguish between subjects with familial CRC associated to Lynch syndrome and sporadic CRC. *BRAF* mutation has been seen to confer metastatic CRC a worse prognosis and resistance to therapies but at the same time a potential target for new drugs development [87].

d) PIK3CA

PIK3CA, the catalytic subunit of *PI3K*, is somatically mutated in over 25% of colorectal cancers [88]. Among patients with wild-type *KRAS* CRC, the presence of *PIK3CA* mutations has been correlated with a significant increase in CRC-specific mortality. Moreover, this mutation may be the long sought biomarker for successful adjuvant therapy with aspirin in patients with CRC [89]. Therefore, *PIK3CA* mutations appear to be a promising predictive biomarker.

e) SMAD4

SMAD proteins are essential mediators of the TGF-β signalling pathway [90]. Among this protein family, *SMAD4* has a central role as a common downstream regulator and tumour suppressor gene [91]. Sporadic mutations of *SMAD4* are present in 2.1-20.0% of colorectal cancers but data are limited [90]. Previous studies showed that the loss of *SMAD4* function was an independent prognostic factor for decreased

recurrence-free and overall survival in CRC patients, particularly those with advanced-stage disease [92,93].

f) p53

The *TP53* gene is involved in the control of the cell cycle and apoptosis and is commonly mutated in CRC [94,95]. The p53 protein induces G1 cell-cycle arrest and facilitates DNA repair prior to a cell committing to the process of DNA replication [84]. When DNA repair is unsuccessful, p53 induces apoptosis. It is believed that *TP53* mutation occurs at the transition from adenoma to carcinoma (Figure 7) and it has been found in 50-70% of CRC [22].

1.4.2 Epigenetics of CRC

Epigenetics are heritable changes in gene expression that do not arise as a consequence of alterations in the DNA sequence and that are mostly mediated by DNA methylation and histone modifications [96]. DNA methylation consists of the attachment of a methyl-group to the 5′ position of cytosine residues in cytosine-guanine dinucleotides [97]. Methylation can occur both in gene promoters and in non-promoter regions, the so-called CpG shores, which are located 2 kbp away from CpG islands and are also regulatory regions of gene activity [98]. It is important to recognize that DNA methylation is a normal mechanism in the mammalian genome by which cells regulate gene expression.

Although gene mutations are important in the pathogenesis of cancer and it is widely accepted, the role of epigenetic alterations has been controversial until recently [97]. Feinberg and Vogelstein were the first to report an epigenetic alteration in cancer. Specifically, they observed that there was an extensive global loss of 5′-methylcytosine content in colon cancers when compared to normal colon [99]. This global hypomethylation is usually found in an age-dependent fashion and in the early steps of the CRC process [100–102]. DNA methylation pattern of normally unmethylated CpG islands in gene promoters of many types of neoplasms have been shown to be hypermethylated [100]. In fact, genome-wide studies of cancer epigenomes have revealed that between 1-10% of CpG islands are aberrantly methylated, suggesting that

a high number of gene promoters may be hypermethylated in the average cancer [103,104].

Epigenetics instability in colorectal cancer is manifested through different pathways, including both hypermethylation of gene promoters that contain CpG islands and global DNA hypomethylation (Figure 8) [97]. Toyota and Issa observed that some CRCs show a high frequency of methylated genes, and they proposed the concept that these type of tumours present a unique molecular pathogenesis, calling them CpG Island Methylator Phenotype (CIMP) cancers [105]. Approximately the 20% of CRCs are CIMP tumours, and they are currently identified by having more than a 60% of methylated genes in a panel of "CIMP marker" genes [97]. This panel is still not standardized; however, the most commonly used panel is the one defined by Weisenberger *et al*, which includes the genes *NEUROG1*, *SOCS1*, *RUNX3*, *IGF2*, and *CACNA1G* [106].

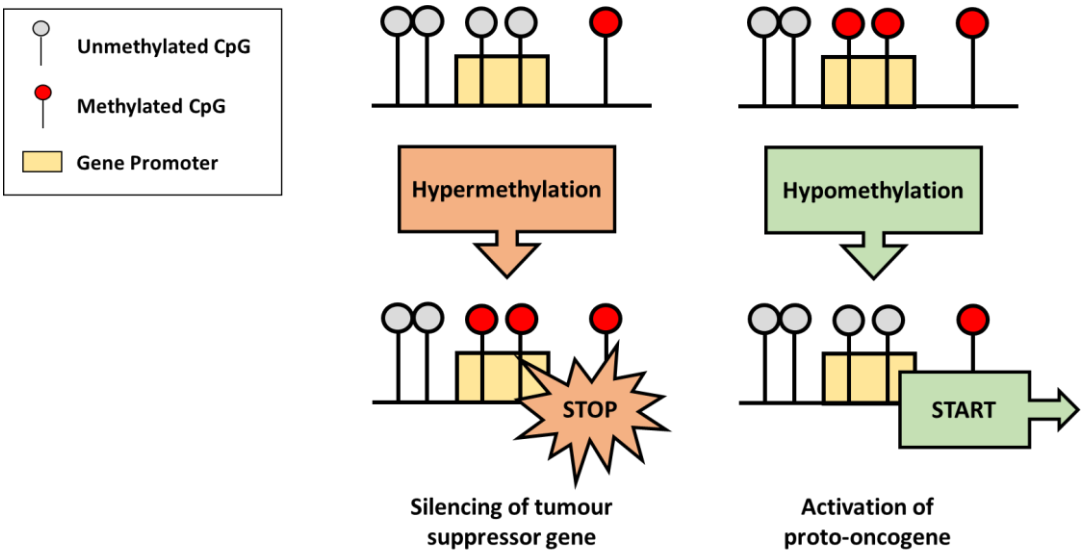

Figure 8. Epigenetic regulation of gene expression by methylation. Adapted from [107].

1.5 Normal epithelium – adenoma – carcinoma sequence

Colorectal cancer is one of the best-known tumours not only because of its localization and accessibility but also because different tumoral stages can coexist in the digestive tract of the same patient (Figure 9).

There are two different models that explain the first histological changes associated with the colorectal neoplasia transformation. In 2001, Vogelstein suggested a model in which genetically altered cells in the superficial portions of the mucosae spread

laterally and downward to form new crypts that first connect to pre-existing normal crypts and eventually replace them [108]. Another model is the one that postulates that adenomas start as unicryptal adenomas and grow initially by crypt fission, a bottom-up pattern [109].

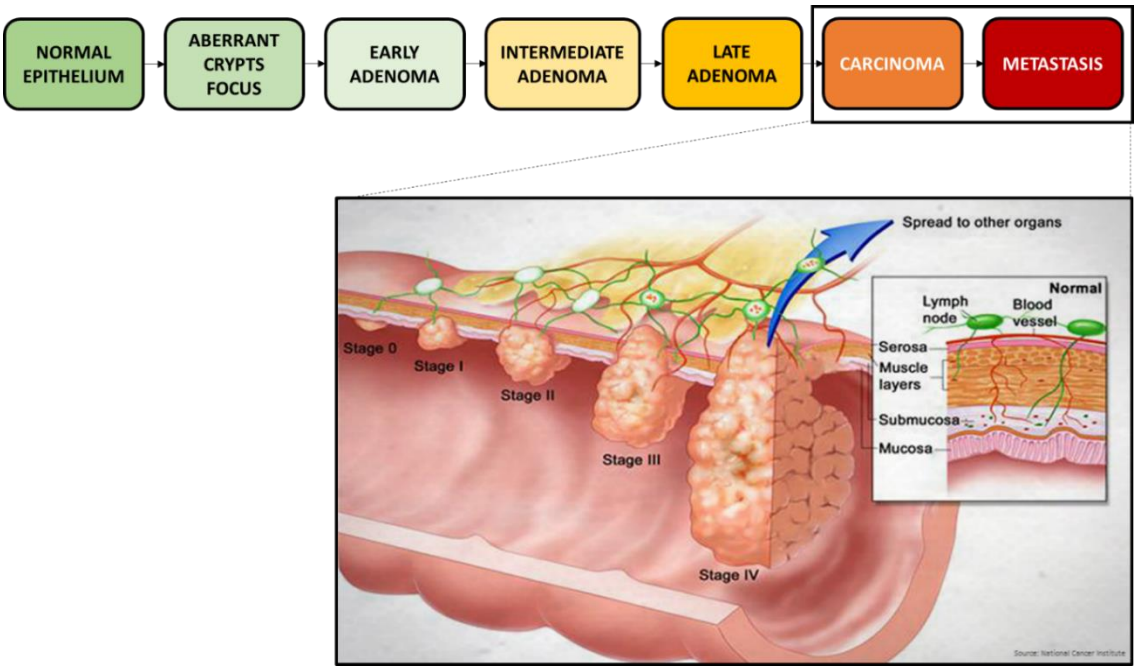

Figure 9. Normal epithelium – adenoma – carcinoma sequence. Adapted from [22] and [110].

Aberrant crypt foci, particularly those that are large and have dysplastic features, seem to be the precursors of adenoma and carcinoma [111]. Early adenomas are polyps smaller than 1 cm of diameter and do not show neither villous component nor high-grade dysplasia, while intermediate adenomas have a diameter equal or higher than 1 cm or present villous component or high-grade dysplasia. Late adenomas show a focus of intraepithelial carcinoma, which is commonly known as carcinoma *in situ*. Early adenomas are also known as non-advanced adenomas, and intermediate and late adenomas as advanced adenomas. Carcinoma is commonly referred to as cancer and it begins when tumoral invasive cells reach the epithelium basal membrane. Is in this stage when patients show the first clinical manifestations. There are 4 differentiated stages through carcinoma progression that are divided according to the extension of the primary tumour (category T), the presence of gangliolar metastasis (category N), and the presence of distal metastasis (category M) (Table 1). The last and most aggressive stage is metastasis, in which cells break away from the colon or the rectum, travel through the

blood or lymph system, and form new tumours in other parts of the body, such as liver, lung, peritoneum or bone [112].

Table 1. TNM categories in colorectal cancer.

Tumour stage (T)	**Definition**
TX	Cannot be assessed
T0	No evidence of cancer
Tis	Carcinoma *in situ*
T1	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades through muscularis propria into submucosa
T4	Tumour directly invades other organs or tissues
Nodal stage (N)	
NX	Regional lymph nodes cannot be assessed
N0	No lymph node metastasis
N1	Metastasis to one to three pericolic or perirectal lymph nodes
N2	Metastasis to four or more pericolic or perirectal lymph nodes
N3	Metastasis to any lymph node along a major named vascular trunk
Distant metastasis (M)	
MX	Presence of distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis present

1.6 Types of Colorectal Cancer

Colorectal cancer can be divided into three main groups according to the family background of the subjects: sporadic, familial and hereditary. Moreover, hereditary CRCs can be subdivided into different syndromes: Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC); familial adenomatous polyposis (FAP); and harmatomatous polyposis syndromes (Figure 10).

This classification is useful in order to determine which screening method is the best to follow for each individual at risk.

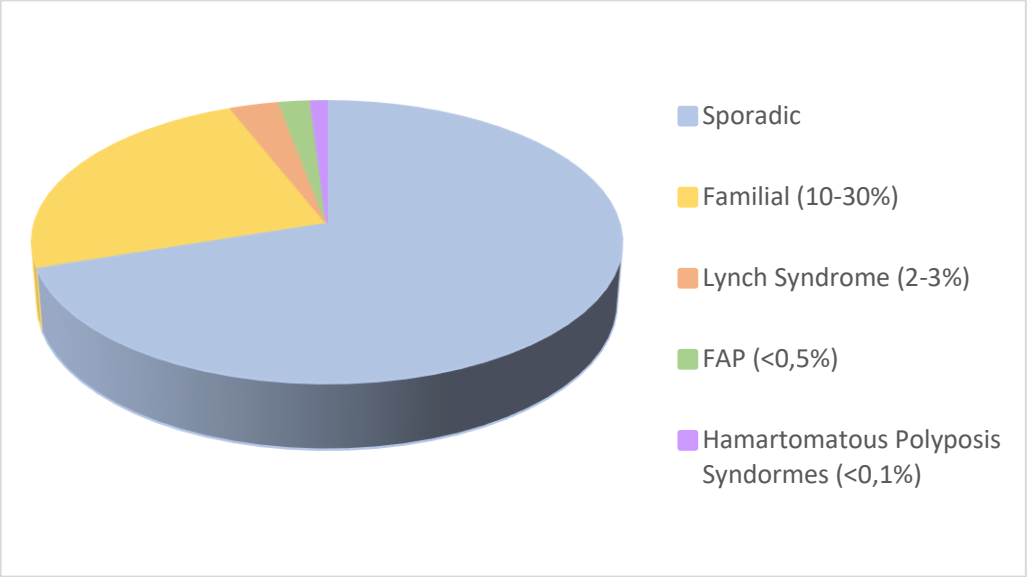

Figure 10. Colorectal cancer classification. In brackets there is the prevalence of each group. Adapted from [113].

1.6.1 Sporadic CRC

Sporadic CRC is developed in individuals with no familial or personal risk background and comprises between 65-85% of the total colorectal cancers [113,114]. It occurs in patients with a median age of 70-75 years and approximately 70% of CRCs develop in the distal colon [115]. Genetically, sporadic CRCs develop by accumulation of a series of abnormalities in tumour suppressor genes and oncogenes. The most common theory for the adenoma-carcinoma sequence is that in which *APC* mutation serves as an initiating event, followed by the accumulation of multiple mutations of genes, such as *KRAS*, *SMAD4* and *TP53* (Figure 11) [22,116]. According to this theory, at least seven different mutations are needed for CRC pathogenesis.

As stated before, the vast majority of sporadic colorectal cancers arise via the well-characterized chromosomal instability pathway. However, over the last two decades many of the molecular mechanisms of a "serrated neoplasia pathway" accounting for approximately 30% of CRCs have been determined [117]. This novel pathway is characterized because it requires the MAPK and the CIMP pathways activation [118]. MAPK pathway activation occurs primarily by either *BRAF* or *KRAS* mutation, and CIMP can be either low level or high level (Figure 11). Finally, although important, MSI is not a requirement of the serrated neoplasia pathway.

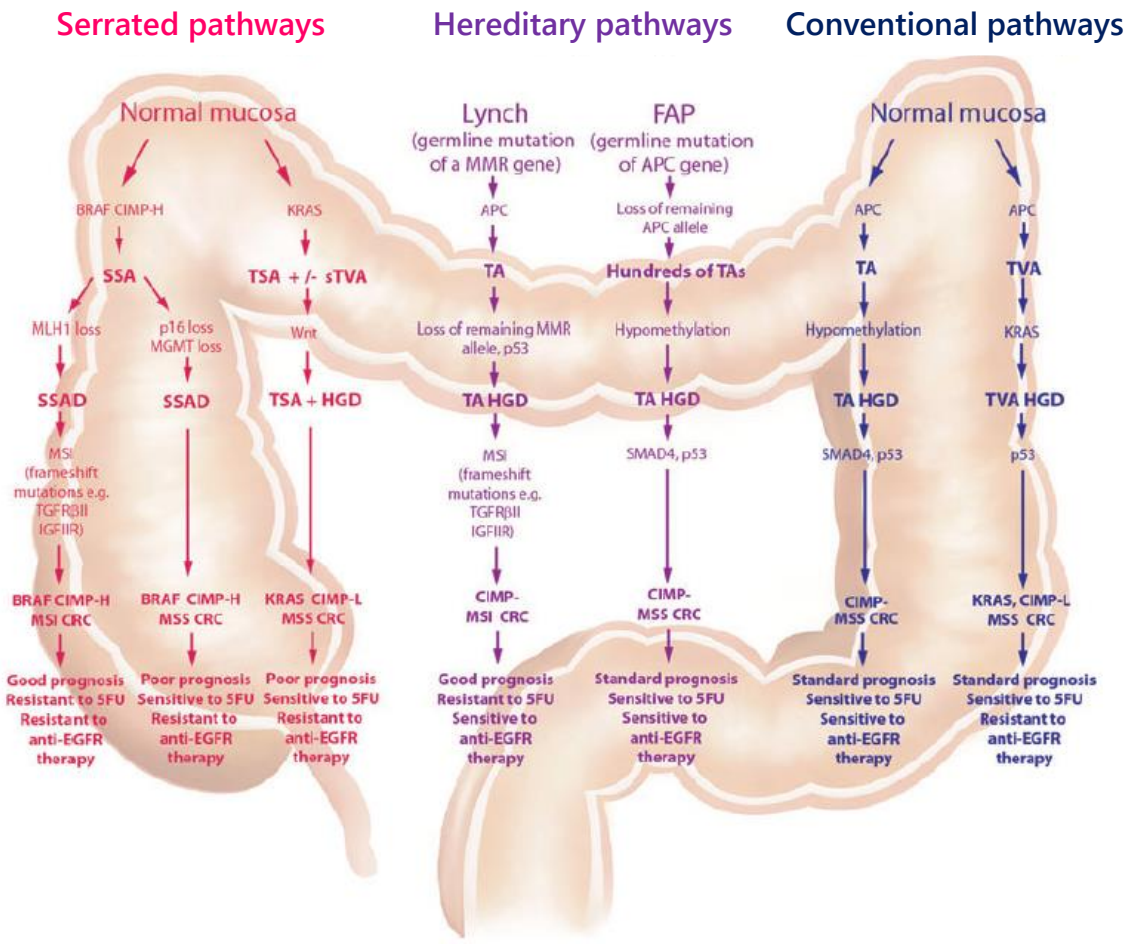

Figure 11. Putative molecular pathways to colorectal cancer [118]. CIMP-H, CpG island methylator high degree; CIMP-L, CpG methylator low degree; MMR, mismatch repair genes; MSI, microsatellite instability; MSS, microsatellite stability; FAP, familial adenomatous polyposis; SSA, sessile serrated adenoma; SSAD, sessile serrated adenoma with dysplasia; TSA, traditional serrated adenoma; TA, tubular adenoma; TVA, tubulovillous adenoma; sTVA, tubulovillous adenoma with serrated features; HGD, high grade dysplasia.

1.6.2 Familial CRC

Familial CRC is a heterogeneous condition that includes patients with no recognized hereditary syndromes and patients with seemingly sporadic forms that aggregate in families [119]. It represents between 25-30% of CRCs. The molecular mechanism that occur in these patients has not been established but a combination of environmental and inherited genetic factors may play a role in CRC development in these families. Colonoscopic surveillance is already offered to people with moderate risk due CRC family background [120–122]. However, there is lack of evidence supporting reduced mortality. Table 2 shows an estimation of CRC risk among first degree relatives according to a meta-analysis performed by Johns *et al* [123].

Table 2. Relative and absolute risk of developing colorectal cancer according to family history. Adapted from [119].

Family history	Relative risk of CRC	Absolute risk of CRC by age 79
No family history	1.00	5%
One first degree relative with CRC	2.25 (95% CI: 2.00-2.53)	11%
More than one first degree relative with CRC	4.25 (95% CI:3.01-6.08)	21%
One first degree relative diagnosed with CRC before age 45	3.87 (95% CI: 2.40-6.22)	19%

1.6.3 Hereditary CRC

Between 2-5% of all colorectal cancers occur within inherited syndromes [124]. CRC syndromes are commonly subclassified as Lynch syndrome (previously known as hereditary non-polyposis colorectal cancer syndrome, HNPCC) or as one of the polyposis syndromes, characterised by the presence of multiple colorectal polyps [125].

Lynch syndrome (LS) is the most common hereditary CRC predisposing syndrome. It is implicated in 2-4% of CRC cases [126]. LS is characterized by a high penetrance, early onset CRC, and an increased risk of extra-intestinal cancers (Figure 12A). Cancer risk in LS carriers vary among different germline mismatch repair gene mutations (Table 3) [127–129].

LS is inherited in an autosomal dominant fashion, developing from a mutation in one allele of one of the DNA mismatch repair genes (MMR) [125]. The most common are *MLH1* and *MSH2* (90%), followed by *MSH6* (10%) and *PMS2* (rare) [130,131]. Loss of functional MMR proteins leads to defects in DNA repair, and consequently, high DNA microsatellite instability. LS carriers have a 50% lifetime-risk of developing CRC, however, they do not show an increase in the number of adenomatous polyps, which are common precursors of CRC [132].

Table 3. Risk of all type of cancers by age 70 in Lynch syndrome (LS). NA: Non-available. Adapted from [125].

Cancer	Risk general population	Risk in LS	Average age of diagnosis (years)
Endometrium	2.7%		65
MLH1/MSH2		14-54%	48-62
MSH6		17-71%	54-57
PMS2		15%	49
Stomach	<1%	0.2-13%	49-55
Ovary	1.6%	4-20%	43-45
Breast	12.4%	5-18%	52
Prostate	16.2%	9-30%	59-60
Urinary tract	<1%	0.2-25%	52-60
Colorectal	4.5%		
MLH1/MSH2		22-74%	27-46
MSH6		10-22%	54-63
PMS2		15-20%	47-66
Small bowel	<1%	0.4-12%	46-49
Pancreas	1.5%	0.4-4.0%	63-65
Hepatobiliary tract	<1%	0.02-4%	54-57
Brain/central nervous system	<1%	1-4%	50
Sebaceous neoplasm	<1%	1-9%	NA

Gastrointestinal polyposis syndromes are subclassified according to the predominant type of polyps (i.e. adenomatous polyps, hamartomatous polyps or serrated polyps). The most predominant polyposis syndrome is the familial adenomatous polyposis (FAP), which accounts approximately 1% of CRCs [133]. The main characteristic of FAP is the emergence of hundreds to thousands of adenomatous polyps throughout the entire colon beginning from late childhood (Figure 12B). This syndrome is autosomal dominant and is caused by a germline heterozygous mutation in the *APC* [134]. There is a subgroup of FAP in which patients have milder manifestations compared to classic FAP named attenuated familial adenomatous polyposis (AFAP). AFAP patients have less than 100 polyps, later onset of CRC adenomas and a reduced lifetime risk of CRC [125]. Around 10% of AFAP patients carry an *APC* mutation and 7% have a mutation in the *MUTYH* gene [135]. *MUTYH*-associated polyposis (MAP) is a recessively inherited syndrome caused by biallelic mutations in

MUTYH gene [136]. Its main characteristic is the development of 10-100 adenomas in the colorectum (Figure 12C) [137]. MAP phenotype is similar to AFAP, however, there are reports of patients presenting with mainly hyperplastic or serrated polyps [138]. Moreover, biallelic *MUTYH* mutations have been found in Lynch-like patients who do not present mutation in the MMR genes [139]. Peutz-Jeghers syndrome (PJS) is an inherited, autosomal dominant disorder distinguished by hamartomatous polyps in the gastrointestinal tract and pigmented mucocutaneous lesions (Figure 12D) [140]. The prevalence of PJS differs between studies, being the widest estimated range from 1 in 8,300 to 1 in 280,000 individuals [141,142]. Currently, the only identifiable mutations causing PJS affect the *STK11* gene, which encodes for a multifunctional serine-threonine kinase, important in second messenger signal transduction [140]. Juvenile polyposis syndrome (JPS) is a rare autosomal dominant hereditary disorder characterized by the presence of multiple distinct juvenile polyps (Figure 12E) [143]. The estimated life-time risk of CRC in JPS patients varies from 9-50% [144], and CRC is found in 17-22% of JPS patients by age 35 [145]. Between 50-60% of patients diagnosed with this syndrome present a germline mutation in the genes *SMAD4* or *BMPR1A* [146,147]. Another rare polyposis syndrome is *PTEN* hamartoma tumour syndrome (PHTS). PHTS is a heterogeneous group of disorders with autosomal dominant inheritance caused by germline mutation of *PTEN* gene (phosphatase and tensin homolog) [125]. Most PHTS cases correspond to Cowden syndrome (CS), which shows an increased risk for developing tumours in various organs (i.e. breast and thyroid), and patients present with a wide range of clinical phenotypes including orocutaneous lesions, papillomatous papules, macrocephaly, dolichocepaly and polyps throughout the colon (Figure 12F) [148]. PHTS also includes Bannayan-Riley-Ruvalcaba syndrome (BRRS), which is a congenital syndrome characterised by lipomatosis and hamartomatous polyposis, and Proteus-like syndrome, a disorder involving hemihypertrophy and hamartomatous overgrowth of multiple tissue [149]. Finally, there is the serrated polyposis syndrome (SPS), which is a clinically defined syndrome characterised by the presence of multiple serrated polyps in the colorectum and an increased risk of CRC (Figure 12G). The average age of diagnosis is between 55-65 years, but ranges from 11 to 83 years [150]. Patients with SPS have an increased risk of CRC with might be higher than 50% during

the lifetime [151]. Accumulated somatic changes have been found in these serrated polyps including *BRAF* mutations and a high level of promoter methylation (CIMP). These different molecular aberrations suggest that serrated polyps develop via a serrated neoplasia pathway which differs from the traditional adenoma-carcinoma sequence (Figure 11) [152].

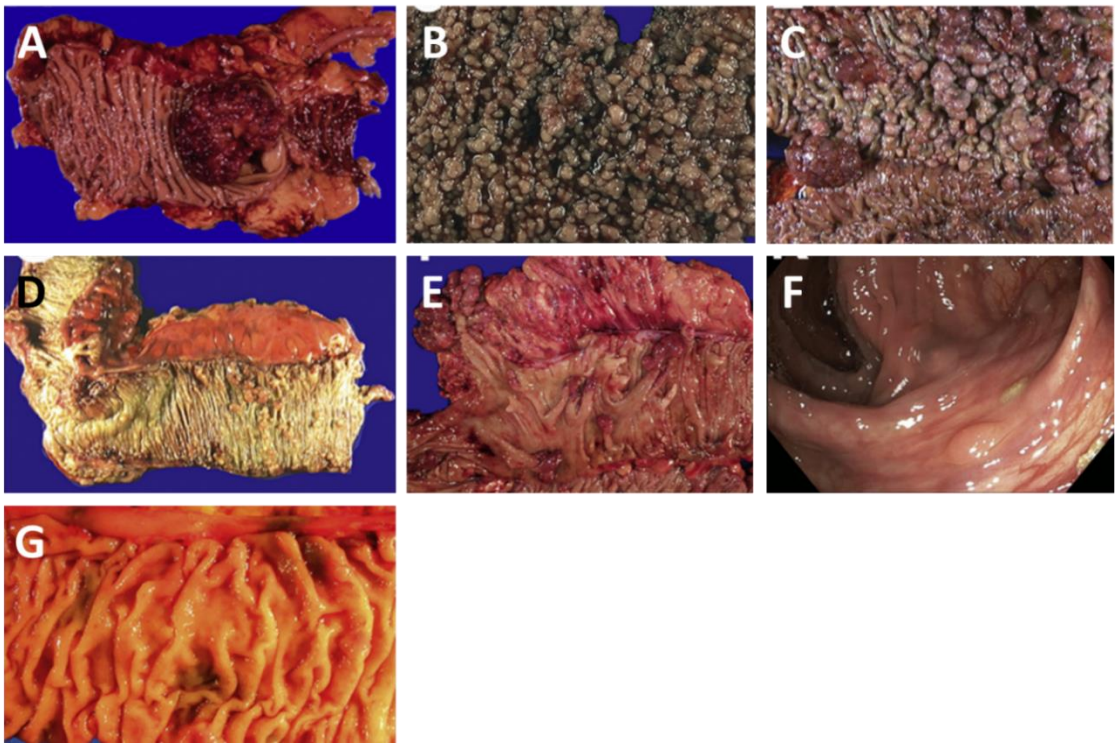

Figure 12. Macroscopic characteristics of each hereditary CRC syndromes. (A) Lynch syndrome, this resection specimen shows a malignant tumour in the caecum and an otherwise unremarkable surrounding mucosa without polyps; (B) Familial adenomatous polyposis, classic FAP is characterised by innumerable mainly small adenomatous polyps carpeting the entire colonic mucosa; (C) MUTYH-associated polyposis, MAP is characterised by development of 10-100 adenomas in the colorectum, and resembles attenuated FAP; (D) Peutz-Jeghers syndrome, lower number of polyps than in FAP and typically around a dozen, the hamartomas show a tree-like configuration with arborizing strands of smooth muscle crypts; (E) Juvenile polyposis syndrome, there are between a few to hundreds of polyps which vary in size between 5-50 mm and are typically spherical with a smooth surface due to erosion; (F) Cowden syndrome, CS patients present polyps throughout the entire gastrointestinal tract; (G) Serrated polyposis syndrome, SPS type 1 is depicted in G, which presents relatively few large, right-sided mostly sessile serrated adenomas. Adapted from [125].

2 Colorectal Cancer Screening

2.1 CRC prevention

Although colorectal cancer is the third leading cause of cancer death worldwide, it is one of the easiest tumours to prevent. Therefore, it is important that governments focus their efforts at different prevention levels: primary, secondary and tertiary level.

The main objective of primary prevention is decreasing the risk for developing colorectal neoplasms by following a healthy diet and lifestyle, and chemoprevention [153]. Chemopreventive measures consist of the use of drugs, vitamins, or other agents to try to reduce the risk of or delay the development or recurrence of cancer [154]. One of the most commonly evaluated approach for CRC chemoprevention is the use of nonsteroidal anti-inflammatory drugs such as aspirin [155].

Colorectal cancer screening is the secondary prevention strategy. The aim of CRC screening is to reduce its incidence, by detecting and removing precursor lesions such as colorectal adenomas; and mortality [156]. CRC is a disease susceptible for screening as it constitutes an important health issue, its natural history has been widely studied, diagnostic methods for its early detection are available and its treatment is more effective when it is early diagnosed (Box 1). Moreover, CRC screening programmes have been demonstrated to be more cost-effective than other screening programmes such as those for breast and cervix cancers [157].

The tertiary prevention strategy is CRC surveillance. The objective of this approach is to minimize the impact of previously established colorectal neoplasms on the prognosis of the patient. It is based on colonoscopic follow up after CRC or adenomas removal [158,159]. In 2015, Castells *et al* designed a proposal for risk stratification of patients with colorectal adenomas and serrated lesions detected at screening colonoscopies [160]. These recommendations are necessary to standardize the follow up of the large proportion of participants in organized CRC screening programs in whom neoplastic lesions are diagnosed.

Box 1. Screening programmes principles according to the World Health Organisation. Adapted from [161].

- The screening programme should respond to a recognized need.
- The objectives of screening should be defined at the outset.
- There should be a defined target population.
- There should be scientific evidence of screening programme effectiveness.
- The programme should integrate education, testing, clinical services and programme management.
- There should be quality assurance, with mechanisms to minimize potential risks of screening.
- The programme should promote equity and access to screening for the entire target population.
- Programme evaluation should be planned from the outset.
- The overall benefits of screening should outweigh the harm.

2.2 CRC screening methods

Continued improvement in screening quality and adherence remains key in order to further reduce the prevalence of colorectal cancer associated deaths. The most extensively used CRC screening methods are the following:

a) Colonoscopy

Colonoscopy allows direct visualization of the colonic mucosa, biopsy of lesions, and polyp removal over the entire colon [162]. Sensitivity and specificity for CRC and advanced adenoma approaches 100% in experienced hands [163]. Actually, colonoscopy is the confirmatory test used when a positive result of a non-invasive CRC-screening strategy is obtained [121,164]. Evidence of efficacy derives from observational studies, with a relevant impact on both incidence (reduced by 66-90%) and mortality (reduced by 31-65%) [158,165–167]. Screening colonoscopy has not only advantages but also limitations, i.e. higher inter-operator variability in adenoma detection rate has been reported [168], and some retrospective analyses have questioned the capability of reducing incidence and mortality from proximal CRC [169,170]. Moreover, colonoscopy is an invasive examination, which requires an exhaustive bowel cleansing, and it is time consuming, painful, and expensive [162]. Indeed, it is estimated that the cost to screen

the whole target average-risk population in Europe (i.e. approximately 146 M people) would exceed 3,650 M euros annually [153]. All this together with the fact that the prevalence of advanced neoplasia in average-risk population does not exceed 10% of these subjects [171], makes it rational to limit the use of screening colonoscopy to those individuals with the highest likelihood of presenting these lesions.

b) Flexible sigmoidoscopy

Flexible sigmoidoscopy (FS) allows visual inspection of colonic mucosa, biopsy taking and polyp removal in the distal tract of the colon. Randomized controlled trials have showed that screening with FS reduces CRC incidence by 18-23% and mortality by 22-31% [172]. Although demonstration of CRC-specific incidence and mortality reduction, FS has long been questioned because of its lower ability to detect proximal advanced neoplasms with respect to colonoscopy [173,174]. Actually, in the *post hoc* analysis of the ColonPrev study, simulated sigmoidoscopies detected between 35-43% fewer subjects with advanced neoplasms than colonoscopy [175].

c) Guaiac-based faecal occult blood test

The guaiac faecal occult blood test (gFOBT) relies on peroxidase-like activity between the group heme and alpha guaiaconic acid. The red blood cells released from colorectal polyps or other lesions contain haemoglobin as a major intracellular constituent. The haemoglobin heme groups can catalyse the oxidation reaction of alpha guaiaconic acid, which will develop a highly conjugated blue quinine compound (Figure 13) [176]. gFOBT exhibits a sensitivity of 12.9-79.4% and a specificity of 86.7-97.7% for CRC screening in many studies [177,178]. The main inconvenient of this test is the requirement of a prescribed diet in order to avoid false-positive results that can occur because of the consumption of specific foods which contain blood, alcohol or non-steroidal anti-inflammatory drugs.

d) Faecal immunochemical test

Unlike gFOBT, faecal immunochemical tests (FIT) are specific for human blood and insensitive to upper gastrointestinal tract bleeding (Figure 13) [164]. The quantitative nature of FIT allows the selection of an optimal cut-off concentration for the

desired target population [179]. This way, FIT could contribute to identifying those subjects with higher risk for developing advanced neoplasms in order to prioritize them in organized CRC-screening programs with a large colonoscopy demand [153,180]. Data from 19 studies, where FIT performance in CRC-screening in asymptomatic was assessed, showed that the overall sensitivity for CRC was 79% and the overall specificity was 94% [181]. It is important that CRC-screening strategies show high performance in detecting not only colorectal cancer but also precancerous lesions. Sensitivity for precancerous lesions is 28%, which results in a high false-positive rate [182]. Therefore, new CRC-screening methods with increased sensitivity for advanced adenomas must be developed.

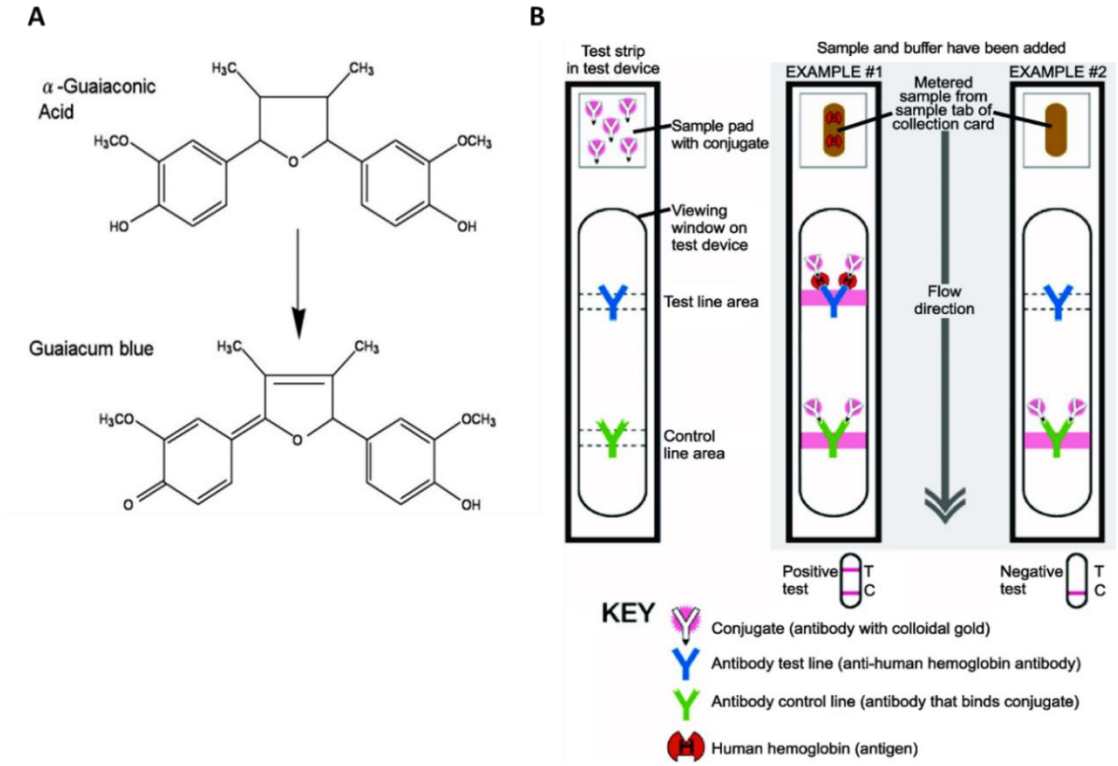

Figure 13. Faecal tests. A) Chemical oxidation reaction of alpha guaionic acid for guaiac. Step 1: apply faecal smear; Step 2: add developing reagent containing H_2O_2 to oxidize guaiaconic acid to guaiacum blue. B) Illustration of lateral flow immunochromatographic analysis principle of faecal immunochemical test for haemoglobin. Adapted from [176] and [183].

e) Faecal DNA biomarkers

Colorectal tumoral lesions contain cells with altered DNA which are continuously released into the large bowel lumen. Faeces stabilize DNA, thus making it possible to

extract it for analysis. As cancers are highly heterogeneous, panels combining different biomarkers have been developed in order to obtain higher detection rates for both CRC and advanced adenomas. Recently, Imperiale *et al* [184] compared a multitarget stool DNA test (Cologuard®), which includes quantitative molecular assays for *KRAS* mutations, aberrant *NDRG4* and *BMP3* methylation, and β-*actin*, plus a haemoglobin immunoassay, with a commercial FIT among a large series of individuals at average-risk for CRC. This new test showed a sensitivity for CRC and advanced adenomas of 92.3% and 42.4%, respectively, while FIT showed a sensitivity of 73.8% for CRC and 23.8% for advanced adenomas. Specificities with DNA testing and FIT for CRC detection were 89.8% and 96.4%, respectively. These results indicate that the combination of specific stool DNA biomarkers detects more advanced neoplastic lesions than FIT at expenses of a higher false-positive rate [184]. The main limitation for broad application of Cologuard® is its cost (599 $), which is too high for a routine screening assay [181].

f) Plasma molecular biomarkers

A CRC-screening tool based on blood biomarkers would favour screening uptake. Currently, there have been found two kind of molecules present in blood used for CRC early detection. On the one hand, aberrant DNA methylation patterns, which have been found in plasma or serum samples of CRC patients, are potential biomarkers for screening [185]. On the other hand, Giraldez *et al* showed that microRNAs (miRNAs), which are small non-coding RNAs that have been shown to play important roles in tumorigenesis, seem to have high potential as non-invasive biomarkers [186]. Nowadays, a plasma *SEPT9* gene methylation test (Epi proColon 2.0) is the only commercially available blood test for CRC screening. The PRESEPT study, which is the largest multicentric clinical trial aimed at assessing the ability of *SEPT9* to detect CRC in average-risk population, showed a sensitivity of 48.2% and a specificity of 91.5% for CRC detection [187]. Although this assay was designed for CRC detection, its performance for advanced adenoma detection was also assessed obtaining a sensitivity of 11% and a specificity of 92%. These results indicate that *SEPT9* can be detected in asymptomatic average-risk individuals' blood in a CRC-screening scenario [187]. However, this test needs to be improved as sensitivity for the detection of precancerous lesions is too low.

2.3 Organized CRC-screening programs

Colorectal cancer screening can be offered to the population as an organized program or as opportunistic screening. The first strategy is addressed to the whole population and the second to people who attends a doctor's office (Table 4). Evidence shows that organized CRC-screening programs allow broader coverture [188].

The European Union (EU) recommends population-based screening for breast, cervical and colorectal cancer using evidence-based methods with quality assurance of the entire screening process [189]. Currently, 20 out of 28 EU countries are initializing or implementing CRC-screening organized programs [190]. Several countries from other parts of the world such as USA, Australia, Japan, etc, have also implemented CRC-screening politics. Nowadays, CRC screening programs are focused on people between 50 and 74 years. The best CRC-screening strategy has not been defined, but the most widely used are faecal occult blood tests (both gFOBT and FIT), sigmoidoscopy, and colonoscopy (Figure 14).

Figure 14. Worldwide distribution of CRC-screening methods. Adapted from [190].

Table 4. Main differences between organized and opportunistic CRC-screening. gFOBT, guaiac faecal occult blood test; FIT, faecal immunochemical test; CRC, colorectal cancer. Adapted from [191].

Topic	Organized Screening	Opportunistic Screening
Screening method (i.e. gFOBT, FIT, colonoscopy, etc)	Fixed and chosen by the Government	Variable: chosen by the individual or by the provider
Target population	Population with a specific age (equity)	Variable: by medical recommendation or individual requirement
Invitation strategy	Active: anyone from target population	Passive: there is no consistent strategy
Objective	Mortality and incidence reduction (populational level)	Mortality and incidence reduction (individual level)
Screening method sensitivity	The highest sensitivity is not sought	The method with highest sensitivity is used
Screening method specificity	A high specificity is sought in order to reduce the false-positive rate and the associated costs	Lower importance
Screening interval	Stablished to maximize populational benefits with reasonable costs	Variable: usually more frequent in order to maximize the protection of the individual
Available health resources	Limited to the population. Follow health politics	Limited to individuals. It depends on the income and private insurances
Quality guarantee	The objectives have to be achieved and monitorization is needed	Objectives are established but they are not always monitored
Participation	Specified and monitored	Can be specified and monitored or not
Relationship with CRC risk	The age group with the highest screening benefits, not necessarily high-risk people	It can lead to an excessive screening of low-risk individuals
Benefits	Maximized benefits for the population within the existing resources	Maximized benefits for the individual
Risks	Minimized for the population within the existing resources	Not necessarily minimized

2.4 Risks and benefits of CRC-screening programs

Although CRC-screening programmes are aimed to reduce the incidence and mortality associated to colorectal tumoral lesions, they do not only offer benefits to the population but also adverse effects. The most common inconvenient derived from a

CRC-screening program are the false positive results. The false-positive rate will depend on the screening method cut-off. In Spain positive FIT results are those with a concentration equal or higher than 20 μg of haemoglobin per gram of faeces. False positive results lead to patients' concern [192], additional costs [193], and performance of unnecessary colonoscopies and their potential risks. FIT false positives have been associated to different factors such as patients' age and sex, smoking, nonsteroidal anti-inflammatory drugs and antiaggregant consumption, and haemorrhoids [194–196]. From 2000 to 2010, in the Catalan CRC screening programme the false-positive proportion was 55.2% [197]. It is necessary to develop new cheap non-invasive CRC screening tools with higher specificity than FIT in order to reduce those costs derived from its false positive results. Another adverse effect of CRC-screening programmes are FIT false negative results. These results cause false patients' tranquillity and generates a delay in the diagnosis of the neoplasia. Previous studies have shown that factors associated to FIT false negatives are the presence of the lesion in the proximal intestine, the size and the severity of the lesion, the non-polypoid morphology and the delay on the analysis of the test [182,198,199]. Nevertheless, it has been demonstrated that successive CRC-screening rounds (every 1 or 2 years) allow the detection of lesions previously undetected, thus reducing CRC risk [200].

3 Lynch Syndrome

Approximately 2-5% of all colorectal cancers arise from Lynch syndrome (LS), which is an autosomal dominant genetic disorder. The global incidence of heritable CRC is relatively low when compared to sporadic CRC; however, it is important to detect early such high-risk subjects and their families in order to set up periodic screening and surveillance programs [201].

3.1 Clinical and molecular features

Lynch syndrome is characterized by an early onset of CRC, usually before 50 and in the right colon. LS carriers have the tendency to develop synchronic or metachronous neoplastic lesions at the colon or other organs such as the endometrium, stomach, ureter, ovarium, biliary tract, and small intestine [202].

LS is caused by heterozygous germline mutations in the DNA MMR genes [203]: *MLH1*, located on chromosome 3p21.3; *MSH2* and *MSH6*, both located on 2p21; and *PMS2*, located on 7p22. Individuals with mutations in these genes have lower ability to repair base pair mismatches that occur during cell division. In normal conditions, MSH2 and MSH6 from a heterodimer that slides along DNA as a clamp to identify insertion-deletion loop errors (Figure 15). MLH1 partners with PMS2 and binds to the MSH2-MSH6 complex after ATP is bound, and together this group of 4 proteins targets *EXOI* to perform the repair. MMR genes germinal mutations have two major consequences: microsatellite instability (MSI) and loss of protein expression.

Microsatellites are short repetitive DNA fragments that are present throughout the whole genome. Microsatellite instability is a phenomenon found in the colorectal cancer DNA of subjects with MMR gene mutations, but not in the adjacent normal colorectal mucosa [204]. MSI constitutes a LS phenotypic marker and it has been found in more than 90% of LS carriers that fulfil the Amsterdam criteria and in 15% of sporadic CRC [205]. LS carriers also show loss of expression of the proteins encoded by MMR genes at the level of the tumour. This alteration can be detected through immunohistochemistry techniques, which is an effective strategy to detect subjects with mutations in DNA repairing genes [206].

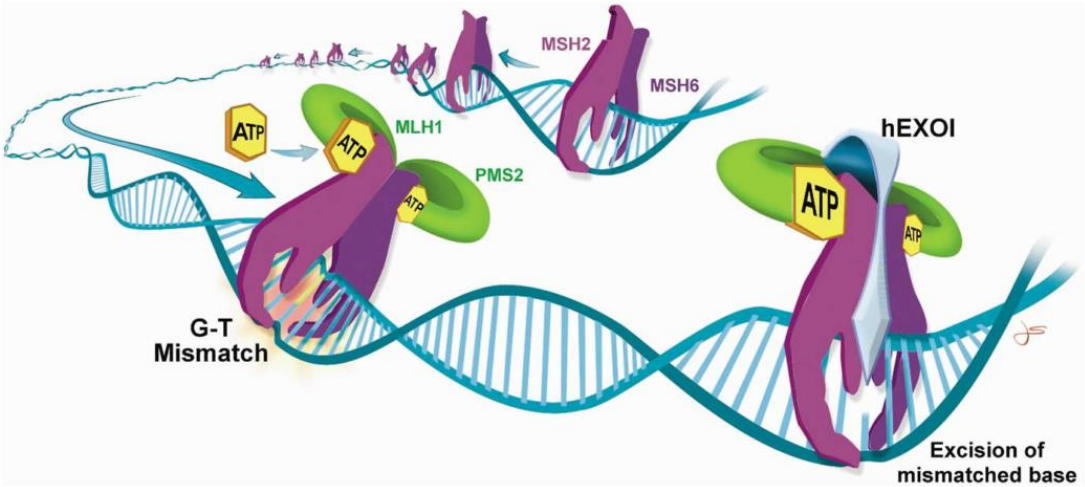

Figure 15. Model of human mismatch repair. Adapted from [207].

3.2 LS identification and screening strategies

Lynch syndrome identification is usually performed through the analysis of the presence of microsatellite instability in the tumour and/or with an immunohistochemical study of MMR proteins expression together with clinical criteria (Amsterdam criteria II and revised Bethesda guidelines) (Box 2) [208,209].

Currently, the most accepted strategy for Lynch syndrome identification follows the revised Bethesda guidelines [206]. MSI and/or immunohistochemistry analysis are performed using tumour biopsies of patients with CRC that meet revised Bethesda criteria. When there is microsatellite stability or normal expression of the MMR proteins no more studies are performed. Conversely, if there is high MSI or protein expression loss, patients are invited to undergo a germinal genetic study.

Box 2. Clinical criteria for Lynch syndrome identification. Adapted from [208,209].

Amsterdam Criteria II
There should be at least 3 relatives with a LS-associated cancer (CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis) • One should be a first-degree relative of the other 2 • At least 2 successive generations should be affected • At least 1 should be diagnosed before age 50 • Familial adenomatous polyposis should be excluded in the CRC case(s) if any • Tumours should be verified by pathological examination
Revised Bethesda Guidelines
Tumours from individuals should be tested for MSI in the following situations: • CRC diagnosed in a patient who is less than age 50 • Presence of synchronous, metachronous colorectal, or other LS-associated tumours, regardless of age • CRC diagnosed in two or more first- or second-degree relatives with LS-related tumours, regardless of age

3.3 LS surveillance

Genetic analysis should be offered to all LS first-degree relatives, so that only those members that carry mutations undergo endoscopic examinations. Colonoscopic surveillance aims to identify and remove adenomatous polyps and detect carcinomas at initial stages, thus improving LS carriers prognosis [210]. A major issue is to define an optimal procedure and interval for surveillance. Several observational studies on the

efficacy of regular colonoscopic surveillance have suggested intervals between 1 and 2 years [211–213]. Individuals with germline mutations are recommended to start colonoscopy at age 25 or 5 years younger than the youngest age at diagnosis in the family, whichever comes first. However, as colonoscopy is an invasive procedure that presents several inconveniences for patients (i.e. unpleasant bowel preparation, sedation, possibility of bowel perforation) there is low compliance [214]. Moreover, endoscopic examinations are associated to high costs. Therefore, it is necessary to develop novel tools which could lead to the reduction of unnecessary surveillance colonoscopies, hence improving the quality of life of LS carriers. In terms of extracolonic tumoral lesions, experts recommend to follow annual screening for endometrial cancer, beginning at age 25-35 [204].

4 Intestinal Microbiota

4.1 Gut microbiota composition

The human body is inhabited by a huge number of bacteria, archaea, viruses, and unicellular eukaryotes. It is estimated that the human microbiota contains around 10^{14} bacterial cells, which is 10 times greater than the number of human cells present in our bodies [215]. By far the most heavily colonized organ is the gastrointestinal tract (GI), with more than a 70% of all the human body microorganisms [216]. This can be explained by the big surface area (200 m^2) and by its richness in molecules that can be used as nutrients by microbes [217]. Most of the gut microbiota is composed of strict anaerobes. Over 50 bacterial phyla have been seen to colonize the gastrointestinal tract [218]; however, gut microbiota is dominated only by Bacteroidetes and Firmicutes. Other phyla such as Proteobacteria, Verrucomicrobia, Actinobacteria, Fusobacteria, and Cyanobacteria are present in minor proportions [219].

Nevertheless, gut microbiota is distributed heterogeneously along the gastrointestinal tract. Stomach is the organ with the lowest load of bacterial cells followed by the small intestine and the colon (10^1-10^3, 10^4-10^7, and 10^{11}-10^{12} cells per gram, respectively) [215]. Moreover, microbiota composition varies among these sites as many factors shape microbial diversity in the human gut (i.e. pH, redox potential, intestinal motility, availability of nutrients and water, salt contents, microbial competition, etc).

Wang *et al* reported that the microbial community in jejunum is different from those in distal ileum, colon and rectum [220]. While the upper GI tract is enriched in facultative anaerobes such as *Streptococcus* and Enterobacteriaceae, strict anaerobic microbes such as Firmicutes and Bacteroidetes are more abundant in the lower GI. Differences in microbial composition are also found between gut lumen and surfaces mainly due to the distribution of the tissue-associated mucus and the radial oxygen gradient (Figure 16). The epithelial surface and the inner mucin layer have low colonization in healthy individuals; the outer mucin layer contains specialist colonizers such as *Akkermansia muciniphila*; the gut lumen-liquid phase has the most variable microbial community; and the gut lumen-substrate particles are rich in specialized colonizers such as *Ruminococcus* spp [221].

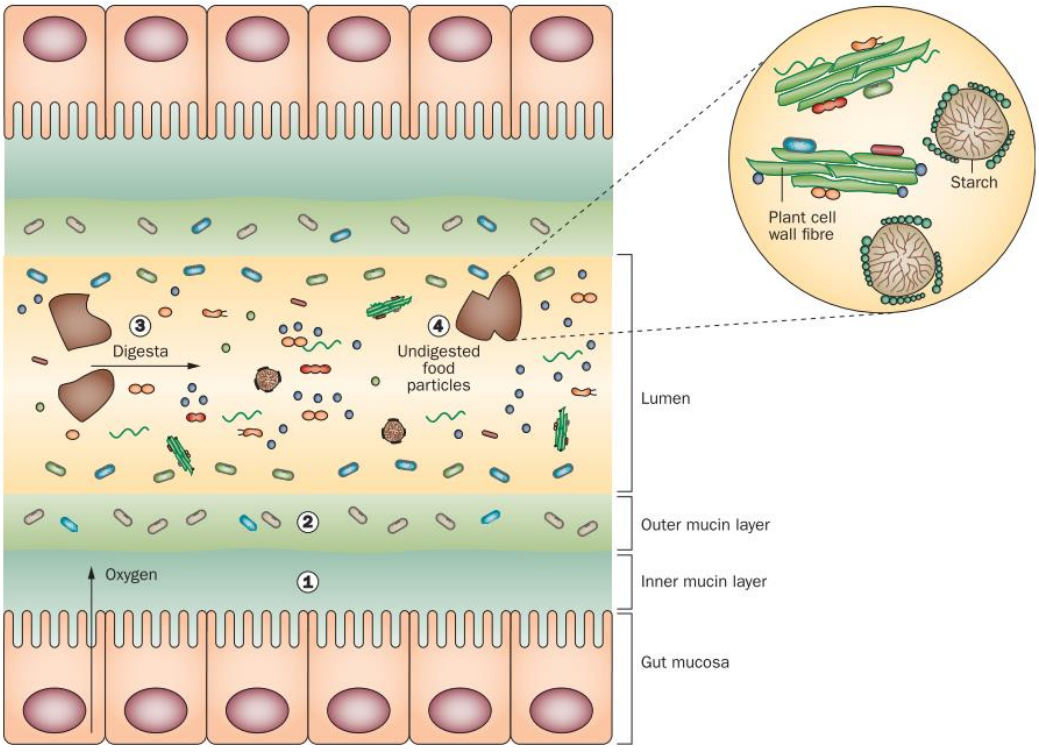

Figure 16. Microbial microenvironments within the large intestine. Extracted from [221].

4.2 Gut microbiota evolution though life stages

Human gut microbiota changes its structure and function from early infancy to old age. Infants have a peculiar gut microbiota composition, which exerts specific functions for the infant biology, training the immune system, and being involved in brain development and host nutrition [222]. When we get older the gut microbiota gains

diversity and develops new physiological functions, in order to fulfil adult requirements, such as food energy extraction.

Microbes colonize human for the first time immediately during the delivery [223]. The infants' intestinal tract is colonized just a few hours after birth, being facultative anaerobic bacteria (i.e. enterobacteria, streptococci, and staphylococci) the first colonizers (Figure 17). Over time, oxygen concentration decreases allowing strictly anaerobic bacteria to settle in the intestine [224]. The delivery mode is one of the factors that influences the early microbiota composition of infants. Dominguez-Bello *et al* observed that infants delivered through vagina acquire bacteria that resembles their mother's vaginal microbiota, whereas infants delivered by caesarean section acquire bacterial communities similar to those found on the mothers' skin surface [225]. Infants' gut microbiota fluctuates until solid food is introduced to their diet, when it shifts toward the adult-type microbiota. While infants have a microbiota enriched in Bifidobacteria, solid food introduction causes a shift into a bacterial community dominated by Firmicutes and Bacteroidetes, which resembles more the microbiota of adults [226]. This adult-type microbiota is capable of metabolizing complex polysaccharides and provides mutual benefits to the host [227]. Its high plasticity allows it to change and adapt in response to environmental and endogenous factors. Elderly people change their diet and lifestyle (i.e. taste and smell alterations, changes in gut motility and mastication), which can cause alterations that result in a nutritionally imbalanced diet [228]. These modifications together with immunosenescence affect both the phylogenetic and functional structure of gut microbiota, favouring the increase of pathobionts such as Enterobacteriaceae and the reduction of immunomodulatory groups such as *Clostridium* cluster IV and XIVa, and *Bifidobacterium* [229].

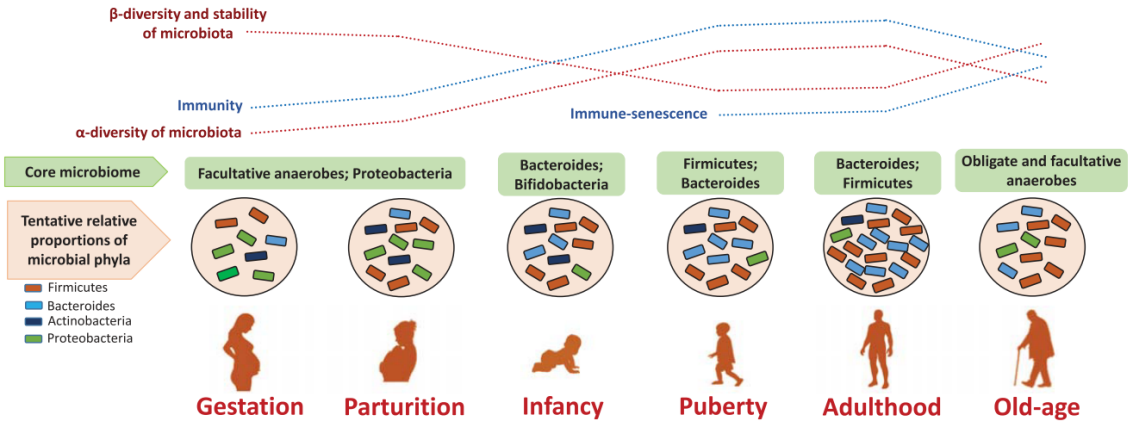

Figure 17. Age-related changes in human gut microbial ecosystem. α-diversity: species taxa richness within a single host/microbial environment, β-diversity: diversity in microbiota community (taxonomic abundance profiles) between different environments/samples. Extracted from [230].

4.3 Factors that determine gut microbiota composition

Microbiota composition is shaped by host and environmental selective pressures. Gut microbes must be adapted to a certain type of lifestyle due to the fewer number of biochemical niches available in the gut, compared with other microbial-rich environments [231].

Diet is the factor that exerts the largest effect on the gut microbiota. Extreme "animal-based" or "plant-based" diets result in wide-ranging alterations of the gut microbiota in humans [232]. Fibre-rich diets also influence gut microbe composition as it has been demonstrated that diets high in resistant starch or in non-starch polysaccharide fibre result in strong enrichment of different bacterial species [233]. New-born feeding methods can also affect the abundance of some bacterial species. *Bifidobacterium* spp. is highly abundant in breast-fed infant microbiota whereas it is reduced in formula-fed infants [234,235]. Moreover, formula-fed infant gut microbiota has altered levels of groups such as *Escherichia coli*, *Clostridium difficile*, *Bacteroides fragilis* and lactobacilli.

Intestinal mucus also provides a source of carbohydrates to the gut microbiota. The erosion of the colonic mucus barrier under dietary fibre deficiency is associated with a switch of the gut microbiota towards the utilization of secreted mucins as a nutrient source [236].

Co-operation between gut microbes also allows colonisation by a more diverse set of organisms, shaping the gut microbiota community [231]. A mechanism proposed to mediate this effect is microbial cross-feeding. For example, *Ruminococcus bromii* produces acetate by fermentation of resistant starch [237], which provides substrate for other microbiota members such as *Eubacterium hallii* and *Anaerostipes caccae*, which in turn convert it into butyrate [238].

Bile acids appears to be a major regulator of the gut microbiota. Kakiyama *et al* analysed the levels of faecal bile acids and the microbiome community of patients with early and advanced cirrhosis and those of control patients [239]. As cirrhosis progressed, they observed an increased bacterial dysbiosis, which was linked to low bile acid levels entering the intestine. This dysbiosis was characterized by reduction of gram-positive members such as *Blautia* and Rumminococcaceae and an increase in proinflammatory and potentially pathogenic taxa.

Another important factor that can shape microbiota is the host immune system. This effect is limited to stratification and compartmentalisation of bacteria to avoid opportunistic invasion of host tissue [216]. Species-specific effects are less probable as there is high amount of functional redundancy within the microbiota. Administered and host-derived antimicrobials also shape the gut microbiota. Many secreted antimicrobial proteins kill bacterial through direct interaction and disruption of the bacterial cell wall through enzymatic attack [240]. In patients with ileal Crohn's disease the reduction of mucosal α-defensin expression, an antimicrobial protein, has been seen, which highlights the importance of this kind of proteins [241].

Environmental factors such as geographic location, surgery, smoking, depression and living arrangements (urban or rural) are also implicated in shaping the gut microbiota. One of the environmental factors with higher impact on the microbial community are antibiotics, which dramatically disrupts both short- and long-term microbial balance, including decreases in the richness and diversity of the community [231]. The produced imbalance can result in an increased susceptibility to pathogen colonization, for example the elimination of some bacterial species that control the growth of pathogens such as *Clostridium difficile* can cause its overgrowth [242].

4.4 Functions of gut microbiota

There is still lack of complete knowledge about all functions of intestinal microbiota and its interactions with the host [243]. Nevertheless, it is well documented that gut microbiota is essential for intestinal development, homeostasis and protection against opportunistic pathogens (Table 5). Indeed, some researchers refer to intestinal microbiota as an extra organ of the host [244].

Table 5. Functions of the intestinal microbiota. Commensal bacteria exert a protective, structural and metabolic effects on the intestinal mucosal. IEC: Intestinal epithelial cells. Adapted from [244].

Protective functions	Structural functions	Metabolic functions
● Pathogen displacement	● Barrier fortification	● Control of IEC differentiation and proliferation
● Nutrient competition	● Induction of IgA	● Metabolize dietary carcinogens
● Production of antimicrobial factors (i.e. bacteriocins, lactic acids)	● Apical tightening of tight junctions	● Synthesize vitamins (i.e. biotin, folate)
	● Immune system development	● Ferment non-digestible dietary residues
		● Ion absorption

Several studies have demonstrated that gut microbiota plays a key role in extracting energy from the diet and producing vitamins [245]. Germfree rats are known to require vitamin K in their diets, while conventional rats do not [246]. Likewise, germfree rats and animals of certain other species require diets in their diets certain B vitamins in higher concentration than those required by their conventional counterparts. These vitamins may derive predominantly from microorganisms residing in the colon epithelium [247]. Gut microbiome, which is the genome of our microbiota, encodes many metabolic functions that the host cannot perform itself. For instance, intestinal microbiota can process non-digestible diet components such as vegetable-derived polysaccharides. Recently, gut microbiota has been proposed as a factor responsible for the weight gain and altered energy metabolism that accompanies the obese state [248]. In murine models it was shown that the majority of the intestinal microbiota is based on two bacterial phyla (Bacteroidetes and Firmicutes) [249]. Obese mice had a higher proportion of Firmicutes to Bacteroidetes (50% greater) than lean mice, results that were

observed also in humans. This was explained because Firmicutes produce more complete metabolism of a given energy source than do Bacteroidetes, hence promoting more efficient absorption of calories. Interestingly, when germ-free mice were colonized with the intestinal microbiota from obese mice, an increase in the total body fat of the recipient mice was observed despite maintaining the diet [249].

Beyond participating in the digestion and fermentation of food, the gut microbiota is also essential in the defence against opportunistic pathogens once they compete for nutrients and adhesions sites, some even actively eliminating competition by secreting antimicrobial peptides [250]. A stable microbiota together with mucus layer are essential to prevent pathogenic bacteria from causing host infections [251]. Absence of microbiota causes intestinal mucosal immunity underdevelopment and germ-free animals present smaller mesenteric lymph nodes and reduced numbers of immune cells such as IgA-producing plasma cells, CD4+ T-cells and intraepithelial $\alpha\beta$ T-cell receptor CD8+ cells, which results in a weakened capacity to fight off pathogenic bacteria [252]. Antibiotics disrupt the ecosystem in the gut of the young infant and could possibly augment the risk of autoinflammatory diseases later in life [253]. The indiscriminate depletion of commensal bacteria following antibiotic intake results in vacating niches, which can increase host vulnerability to excessive colonization by opportunistic pathogens and create dysbiosis [251].

5 Gut Microbiota in Colorectal Cancer

Colorectal cancer is essentially a genetic disease mainly caused by gradual accumulation of mutations in oncogenic genes that lead to non-controlled colonic epithelial cells proliferation [254]. It still remains unknown which specific events prompt the initiation and progression of the disease. In last decades growing attention has been given to the role of the microbiota in carcinogenesis. Indeed, microbes are thought to be involved in approximately 20% of cancers [255], specifically CRC [256]. Colonic mucosa is constantly exposed to the intestinal microbiota and its metabolites, which stimulate immune responses that have the potential to cause continuous inflammation. Interestingly, inflammation in the absence of the gut microbiota or microbial products is not sufficient to induce CRC [257].

In 1975, the first observation associating gut microbiota with colorectal cancer was reported, when germ-free rats developed less chemically induced colorectal tumoral lesions than conventional rats [258]. While gastric cancers seem to result from a single pathogen, different hypotheses have emerged to explain the role of bacteria in colorectal cancer development. On the one hand, CRC can be caused when a dysbiotic gut microbiota with pro-carcinogenic features is capable of remodelling the microbiome as a whole and drives pro-inflammatory responses that promotes carcinogenesis [254]. On the other hand, there is a "driver-passenger" theory in which intestinal bacteria "bacteria drivers", initiate CRC by causing damage to epithelial DNA and tumorigenesis, thus promoting "passenger bacteria" proliferation that have growth advantage in the tumoral microenvironment (Figure 18) [259]. Nevertheless, the exact mechanisms that contribute to dysbiosis are not well understood, and it is still unknown if dysbiosis is a cause or a consequence of colorectal cancer.

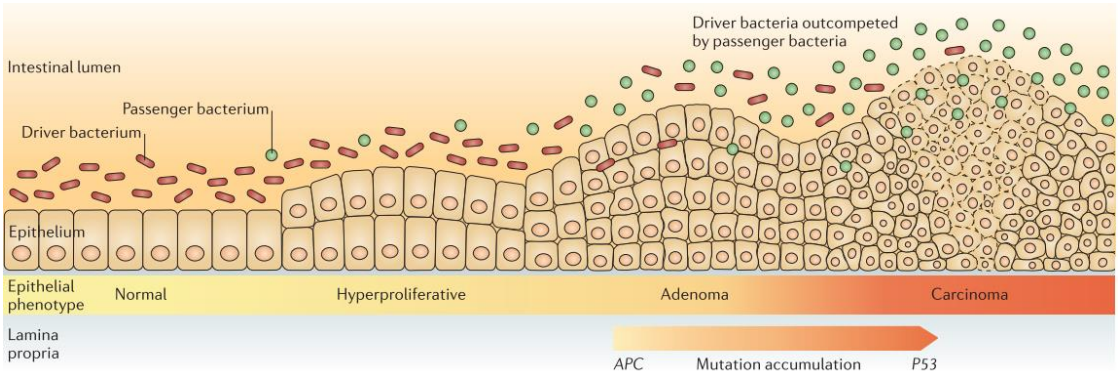

Figure 18. A bacterial driver-passenger model for colorectal cancer. Extracted from [259].

Several research groups have focused their studies on sequencing the 16S rRNA of bacteria from intestinal mucosa or faeces and have reported that patients with CRC show colorectal dysbiosis (Table 6) [260–267]. Some bacterial species have been identified and are suspected to play a role in colorectal carcinogenesis. These species mainly include *Helicobacter pylori*, *Streptococcus gallolyticus*, *Escherichia coli*, *Bacteroides fragilis*, *Enterococcus faecalis*, *Fusobacterium nucleatum*, *Peptostreptococcus stomatis*, and *Porphyromonas assacharolytica*.

Streptococcus gallolyticus

S. gallolyticus, formerly known as *S. bovis*, is a gram-positive bacterium and an opportunistic pathogen that can cause infections such as endocarditis or bacteremia in humans. In 1951 McCoy suggested an association between colorectal cancer and the presence of enterococcal endocarditis [268], which was later confirmed in 1974 by Keush *et al*, who reported that 64% of the *S. gallolyticus* endocarditis cases had gastrointestinal disease [269]. However, the extent and the basis of this relationship are not completely understood. More recently, Rusniok *et al* sequenced the *S. gallolyticus* genome and analysed the encoded proteins [270]. They found that this bacterial species synthetizes different biomolecules that assembly its capsule, collagen-binding proteins, and three types of pili, giving *S. gallolyticus* high efficiency to cause bacteremia, endocarditis, and colorectal cancer. The mechanisms through which this bacterium could cause carcinogenesis are not clear but three steps have been raised [271]: (1) specific adherence factors, such as collagen exposure in the tumour microenvironment may lead to increased *S. gallolyticus* colonization; (2) due to the altered nutritional status of the tumour microenvironment, this species may gain competitive growth advantage; and (3) upon colonization of premalignant tissue, *S. gallolyticus* may induce inflammatory or pro-carcinogenic pathways that increase the progression of CRC. These results suggest that patients infected with *S. gallolyticus* should undergo colonoscopic exploration as they might be at increased risk of CRC.

Helicobacter pylori

H. pylori is a gram-negative bacterium that colonizes the human gastric epithelium and causes infections to more than half of the adult population in the world throughout their lives [272]. It was classified as a human carcinogen by the International Agency for Research on Cancer in 1994 [273]. *Helicobacter* infection is usually diagnosed through histological detection in gastric biopsies, but stool antigen tests have been clinically accepted as a non-invasive alternative, which indicates that *H. pylori* also resides in the large intestine [274]. Intestinal pathologies induced by *Helicobacter* have not been established; however, there are some epidemiological studies conducted to examine if *H. pylori* infection can increase the risk of colorectal cancer development. A meta-

analysis conducted between 1991 and 2002 reported a 1.4-fold increased risk of CRC in patients with *H. pylori* infection [275]. Recently, in a meta-analysis of 7679 Asian patients a statistical association between this bacterium and colorectal adenomas was reported, suggesting a tumorigenic role of *H. pylori* at an early stage of carcinogenesis [276].

Table 6. Summary of 16 rRNA pyrosequencing studies involving colorectal cancer (CRC) and control specimens addressing microbial community structure.

Authors (year)	Study subjects (n)	Type of specimens	Genus/Species	Variation in CRC
Sobhani *et al* (2011)	Control (119), CRC (60)	Stool	*Bacteroides, Prevotella*	↑
Kostic *et al* (2012)	CRC (95)	Paired tissue (cancer, normal)	*Fusobacterium*	↑
			Bacteroides, Clostridia, Faecalibacterium	↓
Chen *et al* (2012)	Control (56), CRC (46)	Stool, rectal swab, cancer tissue, adjacent normal mucosa	*Lactobacillales* (tumour)	↑
			Erysipelotrichaceae, Prevotellaceae, Coriobacteriaceae (stool)	↑
			Faecalibacterium (tumour)	↓
Wang *et al* (2012)	Control (56), CRC (46)	Stool	*Porphyromonas, Escherichia, Enterococcus, Streptococcus, Peptostreptococcus*	↑
			Bacteroides, Roseburia, Alistipes, Eubacterium, Parasutterella	↓
Ahn *et al* (2013)	Control (94), CRC (47)	Stool	*Fusobacterium, Porphyromonas*	↑
			Clostridia	↓
Wu *et al* (2013)	Control (20), CRC (19)	Stool	*Bacteroides, Fusobacterium, Campylobacter*	↑
			Faecalibacterium, Roseburia	↓
Zeller *et al* (2014)	Control (61), adenoma (42), CRC (53)	Stool	*Fusobacterium nucleatum, Porphyromonas assaccharolytica, Peptostreptococcus stomatis*	↑

Escherichia coli

E. coli strains are aero-anaerobic gram-negative bacteria that reside in the healthy intestine [274]. The *E. coli* species can be classified into four phylotypes (A, B1, B2, and D). The majority of *E. coli* strains are commensal bacteria that coexist with their host and promote intestinal homeostasis. Nevertheless, phylotype B2 strains usually carry virulence genes that can often cause extraintestinal infections [277]. Several studies have observed that *E. coli* is frequently found attached to tumoral lesions and adjacent normal mucosa, usually accumulated in large amounts [278–280]. In 1998, Swidinski *et al* detected *E. coli* using PCR in 62% of adenoma and 77% of carcinoma biopsy specimen, compared to 12% from symptomatic patients, and 3% from asymptomatic subjects [278]. Another study observed that 66% of CRC biopsies were positive for *E. coli* strains, compared to 19% in the control group (patients with diverticulosis) [280]. *E. coli* of phylotype B2 has also shown to be cytotoxic in *in vitro* models, it also can induce double-strand DNA breaks, and the presence of the pks gene enables the synthesis of a genotoxin called Colibactin, which gives it tumour-inducing properties [281].

Bacteroides fragilis

B. fragilis is an anaerobic bacterium that although it is detected in 80% of children and adults, it represents less than 1% of gut microbiota [282]. This bacterial species can be classified in two molecular subtypes, nontoxigenic and enterotoxigenic [283]. The later produces a metalloprotease named *B. fragilis* toxin (BFT), which is the only *B. fragilis* virulence factor known until now [254]. This enterotoxin alters the colonic epithelial cells structure and function increasing colonic permeability, which is an early pathophysiologic change related with initial CRC tumorigenesis [284]. A study conducted by Toprak *et al* reported that 38% of stool samples from CRC patients were positive to *BFT* gene, whereas only 12% of those from control subjects were positive, indicating the association between *B. fragilis* and CRC [285].

Fusobacterium nucleatum

F. nucleatum is one of the most prevalent species in the oral cavity, and it has been associated with oral inflammation diseases [286]. Moreover, it has also been related to

pancreatic and oral cancer, liver abscess, appendicitis and other infections [287–289]. Metagenomic analyses, transcriptome sequencing and bacterial 16S rRNA gene DNA sequencing have shown enrichment of *Fusobacterium* species in colorectal cancer relative to surrounding tissues [290]. Recently, this bacterial species has been considered to be a potential initiator of CRC susceptibility. In 2013, Kostic *et al* confirmed that *F. nucleatum* is capable of promoting colorectal tumorigenesis in $APC^{min/+}$ mice [290]. Another study, reported that it also stimulates tumour cell growth in CRC through β-catenin signalling and it can activate oncogenes via the FadA adhesion virulence factor [291]. Viljoen *et al* identified significant associations between high colonization by *F. nucleatum* and CRC at stages II-IV and it was associated with tumoral microsatellite instability [292].

Salmonella enterica

S. enterica is a gram-negative and facultative anaerobe that acts as an intracellular pathogen in humans and animals [293]. *Salmonella* infection is mainly caused through the ingestion of contaminated food such as eggs and meat [294]. This infection can result in mild self-limiting gastroenteritis or even severe systemic infection that can be fatal. *Salmonella* species secrete effector proteins into host cells that activate AKT and ERK pathways, which are also activated in many cancers, and are essential for transforming pre-transformed cells [295]. AvrA is another *Salmonella* effector that activates host β-catenin signalling and has been observed to induce colorectal tumorigenesis in mice [296]. A recent study performed in two different populations in the USA and the Netherlands, reported that colorectal cancer and pre-cancer cases had increased amounts of antibody against *Salmonella* flagellin when compared to controls [297]. These results support a possible association between *Salmonella* and CRC.

Enterococcus faecalis

E. faecalis is a gram-positive facultative anaerobic bacterium that belongs to the lactic acid bacteria [298]. It is the most prevalent cultured strain among the enterococci found in human faeces (10^5-10^7 CFU). The role of *E. faecalis* in colorectal cancer is controversial since some studies suggest a protective role in CRC while others have observed pathogenic activity. For instance, Grootaert *et al* co-cultivated HTC-116 (an aggressive CRC lineage), with *E. faecalis* obtaining a downregulation in the expression

of the *FIAF* gene (angiopoietin-like protein 4), which is normally related to the development of some cancer types [299]. Another study conducted by Wang *et al* showed that exposure of murine primary colon epithelial cells to *E. faecalis*-infected macrophages activated their Wnt/β-catenin signalling, thus inducing pluripotent transcription factors associated with dedifferentiation [300]. In consequence, colon epithelial cells were reprogrammed, which suggested a role of these bacteria in inducing CRC. *E. faecalis* harmful role has also been suggested because it is able to produce ROS and extracellular superoxide that can cause genomic instability, damaging colonic DNA, which can predispose the host to mutations and thus cancer [301].

5.1 Potential use of gut microbiota as a non-invasive tool for CRC screening

Sporadic colorectal cancer shows a slow progression from detectable precancerous lesions, thus the prognosis for patients with early stages of CRC is encouraging. As stated before, the most robust methods for the detection of colorectal tumoral lesions are endoscopic tests; however, these methods are invasive and costly. The alternative to colonoscopies are tests that are patient-friendly, but they suffer from lack of sensitivity and specificity. These data highlight the necessity to develop CRC-screening methods with higher sensitivity and specificity that enhance the chances of a cure. Intestinal microbiota has not only beneficial functions for the host, but also there is evidence that suggest that it plays an important role in tumorigenesis. Hence, stool gut microbes could be used as a non-invasive tool for CRC screening.

Several research groups have reported an imbalance in the gut microbiota of CRC patients, which suggests that it could be used as a potential non-invasive tool for CRC screening [267,302–305]. In 2014, Zackular *et al* characterized the faecal gut microbiome from healthy controls (30), and adenoma (30) and carcinoma (30) patients, in order to establish a classification model [302]. They observed that patients with tumoral lesions had enrichment and depletion of several bacterial populations. The combination of known clinical CRC risk factors (i.e. age, ethnicity, BMI) with gut microbiota data improved the ability to differentiate between healthy, adenoma and carcinoma patients relative to risk factors alone, with an accuracy of 0.798 AUC for predicting CRC. The

same year, Zeller *et al* used metagenomic sequencing of faecal samples to identify taxonomic markers that distinguished healthy controls (61) from CRC patients (53) [267]. When the metagenomic classifier was compared to the standard faecal occult blood test (FOBT) accuracy values were similar and when both approaches were combined sensitivity improved greater than 45% relative to FOBT alone, while specificity values were maintained. Recently, Ai *et al* evaluated the effectiveness of different machine-learning models in predicting CRC analysing the intestinal microbiota in faeces [303]. Bayes Net was the model with the highest accuracy, showing the lowest false negative rate. As reported in the study conducted by Zeller *et al*, gut microbiota (0.93 AUC) was more accurate than FOBT alone, results that were improved when both approaches were combined. Although a clear association between CRC and gut microbiota was demonstrated in these studies, no common stool bacterial markers were identified.

Shae *et al* decided to re-analyse raw sequence and metadata from several studies uniformly, in order to identify a composite and generalisable microbial marker for CRC [304]. Results from 9 studies were processed with a total of 509 samples (235 healthy controls, 79 adenomas, and 195 CRC) analysed. They highlighted that *Parvimonas micra* ATCC 33270, *Streptococcus anginosus*, *Parabacteroides distasonis* and other uncultured members of Proteobacteria were increased in stools from patients with CRC compared with controls across studies and had high discriminatory capacity in diagnostic classification. Likewise, Amitay *et al* performed a systematic review of 19 studies and examined the differences in the faecal microbial population from healthy controls and from adenoma or CRC patients [305], aiming to create stool multi-bacterial models for early detection of adenomas and CRC. The authors concluded that further studies should focus on developing unified documented and reproducible protocols for studying the human faecal gut microbiota in order to develop new non-invasive approaches that can complement the current methods used for CRC early detection.

Another tool that has been explored to detect colorectal tumoral lesions is the faecal metabolome, as with it a unique metabolic fingerprint to diagnose CRC could be obtained. However, until now there are only a few studies that analyse the potential of faecal bacterial metabolome for CRC screening (Table 7).

The increasing interest in studying the differences of specific bacterial species between CRC patients and healthy subjects demonstrates the potential that researchers visualize in the use of the intestinal microbiota as a non-invasive tool for the early detection of colorectal carcinogenesis. Currently, one of the most commonly methods used for this purpose is the faecal immunochemical test (FIT). As stated in former sections, this test shows a sensitivity value around 80% and a specificity around 90% for the detection of CRC. However, in terms of precancerous lesions the sensitivity value is reduced until less than a 30%. Therefore, it is necessary to develop new methods capable of detecting not only CRC, but also advanced neoplasia (CRC + advanced adenomas), with greater accuracy values than those obtained with tools used in the current screening organized programs. Hence, intestinal microbiota is postulated as a potential alternative to the prevailing methods.

Table 7. Summary of studies in CRC faecal metabolic profiling. Adapted from [306].

Authors [ref]	**Study subjects (n)**	**Type of specimens**	**Study observations**
Bezabeh *et al* [307]	Controls (412) CRC (111)	Aqueous dispersion of stools	Potential to detect colorectal neoplasia
Monleón *et al* [308]	Controls (11) CRC (21)	Faecal water extract	Reproducible and effective method for detecting CRC markers. SCFA (↓) (acetate, butyrate) appears to be the most effective marker in CRC
Weir *et al* [309]	Controls (10) CRC (11)	Lyophilized human faeces	Butyric acid, linoleic acid, glycerol (↓) Secondary bile acid (↓) associated with (↓) *Ruminococcus* spp. Leucine, valine, acetic acid, valeric acid, isobutyric acid, isovaleric acid (↑) *A. muciniphila* (↑) associated with proline, serine, threonine (↑)
Goedert *et al* [310]	Controls (102) CRC (48)	Lyophilized human faeces	41 metabolites significantly associated with CRC Xenobiotics (↓) Heme, peptides/amino acids, vitamins, co-factors (↓) Other CRC associated molecules
Phua *et al* [311]	Controls (10) CRC (11)	Lyophilized human faeces	Fructose, linoleic acid, and nicotinic acid (↓) in CRC stools

Authors [ref]	Study subjects (n)	Type of specimens	Study observations
Batty *et al* [312]	Controls with positive FOBT (31) CRC (31)	Human faeces	Discrimination of CRC samples with better specificity and sensitivity than FOBT Ammonia, sulphides, acetaldehyde (↓)
Sinha *et al* [313]	Controls (89) CRC (42)	Lyophilized human faeces	Microbe-metabolite correlation in CRC: *Clostridia, Lachnospiraceae,* p-aminobenzoate and conjugated linoleate (↑)

5.2 Potential use of gut microbiota in Lynch Syndrome surveillance

The vast majority of gut microbiota studies conducted with colorectal cancer patients are focused on the study of the intestinal bacteria of sporadic CRC. In fact, there are no published research papers that analyse the gut microbiota of Lynch syndrome (LS) carriers. There is only an abstract written in 2017 by Lu *et al*, where possible interactions between gut microbiota and LS carriers were analysed [314]. Parallelly, the Memorial Sloan Kettering Cancer Centre is performing a clinical trial with the objective to perform a metagenomic analysis of the intestinal microbiota of LS carriers and other hereditary colonic syndromes, but it is still in the recruitment process [315].

LS carriers undergo periodic endoscopic surveillance, which is necessary for early detection of neoplastic lesions in these patients. However, colonoscopies require bowel preparation and sedation, can cause intestinal perforation, are time consuming and have high associated costs; aspects that cause low acceptance among LS carriers. If there was a non-invasive tool capable of detecting those LS carriers with no colorectal lesions, the intervals between endoscopic examinations could be expanded, thus improving their quality of life and reducing colonoscopy-associated costs. The study of the gut microbiota of LS carriers would allow to determine if there are differences with that of those subjects that develop CRC sporadically and healthy controls. Moreover, it would be useful to define a faecal bacterial signature that indicated the presence of colorectal tumoral lesions, which could be used for LS carriers' surveillance.

2

Objectives

The main objective of this thesis is to develop a novel non-invasive tool based on faecal bacterial markers for CRC screening, able to detect precancerous lesions before the appearance of clinical signs. The secondary objective of the thesis is to define a faecal bacterial signature which enables the prediction of the absence of neoplastic lesions in Lynch syndrome carriers in order to improve their surveillance. To address these challenges, faecal DNA extraction, quantitative polymerase chain reaction of specific bacterial species, and Machine Learning technology have been used. The results are organized into three chapters, which present the following specific aims:

Chapter 1. To develop a potential new non-invasive CRC-screening tool based on faecal bacterial markers capable of complementing FIT (RAID-CRC) in a symptomatic population, in order to assess the risk of intestinal disease, in particular advanced neoplasia (advanced adenomas + CRC).

Chapter 2. To clinically validate a new non-invasive tool for CRC-screening based on faecal bacterial signatures (RAID-CRC Screen) capable of decreasing FIT false-positive rate in FIT-positive population.

Chapter 3. To compare a specific faecal bacterial signature of sporadic CRC and IBD patients to that of Lynch syndrome (LS) carriers; and develop a non-invasive tool based on this signature that enables the prediction of the absence of neoplasia in LS carriers (RAID-LS).

3

Methodology

1 Patients and data sampling

Participants were recruited from four hospitals: Hospital Universitari de Girona Dr. Josep Trueta – Institut d'Assistència Sanitària Consortium (Girona and Salt, Spain), Hospital Universitari de Bellvitge (l'Hospitalet de Llobregat, Spain), Consorci Hospitalari de Vic (Vic, Spain), and Complexo Hospitalario de Ourense (Ourense, Spain). Enrolled subjects can be classified in three main groups: (1) participants of the Catalan CRC screening program (asymptomatic, FIT positive, 50-69 years); (2) individuals with symptomatology compatible with CRC (blood in stool, significant weight loss, abdominal pain, inexplicable tiredness, changes in bowel habits); and (3) Lynch syndrome carriers. The first were recruited in Girona, Bellvitge and Vic; the second in Ourense, Girona and Bellvitge; and the third in Girona and Bellvitge (Figure 19).

All subjects underwent colonoscopy in order to determine their colorectal status. According to the endoscopic examination and the pathology results, diagnosis was classified into four groups: normal colonoscopy (colonoscopy with no findings or with sigma and/or rectum hyperplastic polyps <10 mm), non-advanced adenomas (tubular adenomas <10 mm with low grade dysplasia, and serrated polyps <10 mm without dysplasia), advanced adenomas (adenomas >10 mm or with villous component or high grade dysplasia, serrated polyps >10 mm or with dysplasia, and pTis adenocarcinoma) and invasive CRC.

Participants were kindly asked to collect a stool sample from one bowel movement in a sterile faeces' container before colonoscopy and prior to bowel cleanse. Samples were immediately frozen after deposition. Then, subjects brought samples to the hospital, where they were kept frozen at -20 ºC for short-term storage and stored at -80 ºC upon arrival at the GoodGut S.L. facilities in Girona (Spain). The day of the colonoscopy subjects were also asked to answer a questionnaire on clinical data (i.e. age, gender, ethnicity, diseases, medication, familiar CRC background).

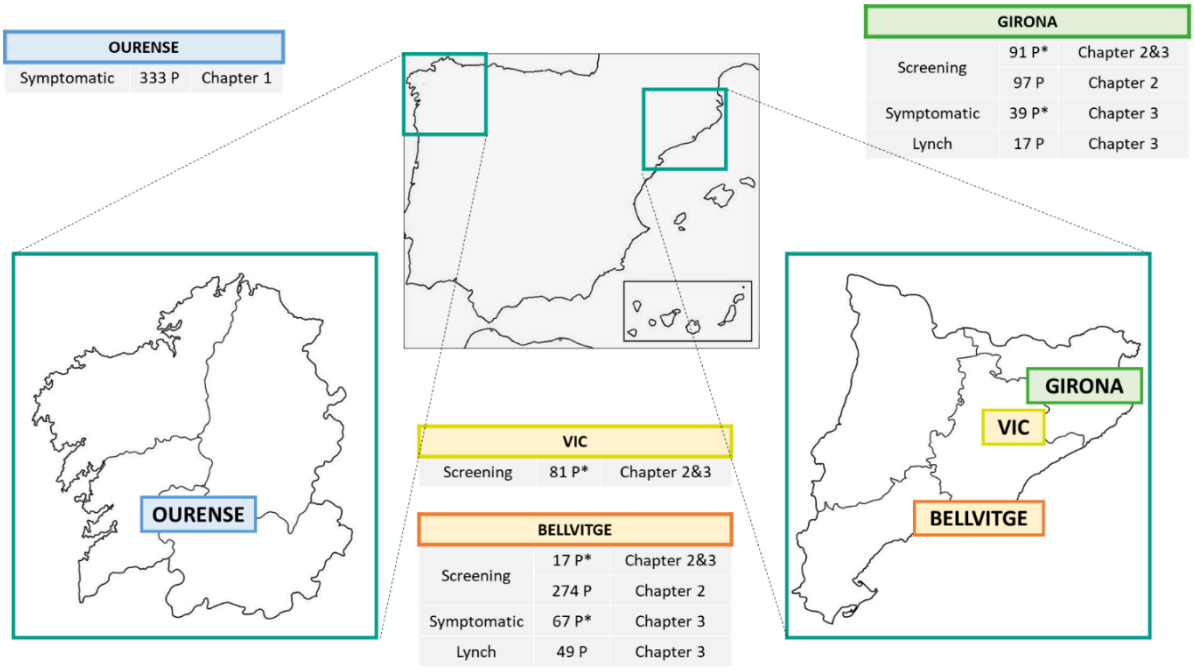

Figure 19. Distribution of the patients recruited in the studies of this thesis according to their enrolment group (screening, symptomatic, Lynch). Chapters in which each group was used are also exposed. P, participants; * subjects used as control group (no Lynch syndrome carriers) in Chapter 3 and in the proof-of-concept in Chapter 2 (only screening subjects).

2 Faecal DNA extraction and quantification

Stool samples were thawed for two hours. Genomic DNA was extracted from frozen faecal samples after homogenization using the NucleoSpin Soil Kit (Macherey-Nagel GMbH & Co., Duren, Germany). The instructions of manufacturer were followed, DNA was finally eluted in a 100 µl final volume of SE Elution Buffer and stored at -20 ºC until use. DNA concentration was determined with Qubit fluorometric quantification (ThermoFisher Scientific, Waltham, USA). All samples were adjusted to a final concentration of 8 ng/µL except for samples of the clinical validation in Chapter 2, which were directly analysed by qPCR after DNA extraction and the total bacterial load was used to normalize data.

3 Bacterial markers quantification through qPCR

Fourteen bacterial markers were targeted in the different studies included in this thesis. Eubacteria (EUB) was quantified as the total bacterial load; B10 (best match BLAST *Faecalibacterium prauznitzii*), B46 (best match BLAST *Subdoligranulum variabile*), B48 (best match BLAST *Ruminococcus*, *Roseburia*, *Coprococcus*), *Faecalibacterium prausnitzii*

(FPRA), *F. prausnitzii* phylogroup I (PHGI), *F. prausnitzii* phylogroup II (PHGII), and *Roseburia intestinalis* (RSBI) as butyrate-producing bacteria; *Escherichia coli* (ECO) as proinflammatory species; *Gemella morbillorum* (GMLL), *Peptostreptococcus stomatis* (PTST), and *Bacteroides fragilis* (BCTF) as opportunistic pathogens usually found in the oral cavity or in the bowel of mammals; *Collinsella intestinalis* (CINT) as a hydrogen and ethanol producer specific biomarker; and *Bacteroides thetaiotaomicron* (BCTT) as a saccharolytic species. B10, B46, B48, GMLL, PTST, BCTF, BCTT, and RSBI are associated with colorectal cancer; while FPRA, ECO, PHGI, and PHGII have been linked to inflammatory bowel disease.

Quantification of CRC-specific biomarkers was performed by preparing specific reactions for each biomarker using SYBR Green Master Mix (Promega, Madison, USA). Each reaction consisted of 20 µl containing 1× GoTaq qPCR Master Mix, between 150 nM and 300 nM of each primer, and up to 20 ng of genomic DNA template (Table 8). In terms of IBD-specific bacterial markers, quantification was performed by preparing single reaction for each biomarker using Probe Master Mix (Promega, Madison, USA). Each reaction consisted of 20 µl containing 1x GoTaq qPCR Master Mix, between 300 nM and 900 nM of each primer, between 150 nM and 300 nM of probe, and up to 20 ng of genomic DNA template. The species-specific primers were purchased at Macrogen (Macrogen, Seoul, South Korea). All quantitative PCR were run on an AriaMx Realtime PCR System (Agilent Technologies, Santa Clara, USA). Thermal profiles were different according to the biomarker analysed (Table 9). A melting curve step was added to the end of each qPCR to verify the presence of the expected amplicon size as well as to control primer dimmer formation. Data were collected and analysed with the Aria Software version 1.3 (Agilent Technologies, Santa Clara, USA). All samples were amplified in duplicates, which were considered valid when the difference between threshold cycles (Ct) was less than 0.6. A No-template control reaction and standards (10^7-10^3 genomic copies/µL) were included in each PCR run.

Table 8. Forward and reverse primers, and probes used in this work. EUB, Eubacteria; GMLL, *G. morbillorum*; PTST, *P. stomatis*; BCTF, *B. fragilis*; CINT, *C. intestinalis*; BCTT, *B. thetaiotaomicron*; RSBI, *R. intestinalis*; FPRA, *F. prausnitzii*, ECO, *E. coli*, PHGI, *F. prausnitzii* phylogroup I, PHGII, *F. prausnitzii* phylogroup II; F, Forward primer; R, Reverse primer; PR, probe. All probes were 5'-labelled with FAM (6-carboxyfluorescin) as the reporter dye. TAMRA was used as quencher dye at the 3'-end for FPRA and ECO probes, whereas BHQ1 was used for PHGI and PHGII.

Target	Primers / Probe	Sequence 5' → 3'	Primer/Probe concentration	Ref
EUB	EUB_F	ACT CCT ACG GGA GGC AGC AGT	200 nM	modified [316]
	EUB_R	GTA TTA CCG CGG CTG CTG GCA C		
B10	B10_F	CAA CAA GGT AAG TGA CGG C	300 nM	[317,318]
	B10_R	CGC CTA CCT GTG CAC TAC TC		
B46	B46_F	TCC ACG TAA GTC ACA AGC G	300 nM	[317,318]
	B46_R	CGC CTA CCT GTG CAC TAC TC		
B48	B48_F	GTA CGG GGA GCA GCA GTG	300 nM	[317,318]
	B48_R	GAC ACT CTA GAT GCA CAG TTT CC		
GMLL	GMLL_F	AAG AGT TCC AAG GCG TTC TC	150 nM	This study
	GMLL_R	CCA TTT CAA GAT CCG CTT TCT ATT T		
PTST	PTST_F	AGG TTG ATG CTC TGA GTA GTA G	150 nM	This study
	PTST_R	ATG AAT ACT AGC CTC TCC TCT TT		
BCTF	BCTF_F	TGA AAG CGT GCT CTT ACT ATT G	150 nM	This study
	BCFT_R	TAT TGG CTG TTG TGC TTT GT		
CINT	CINT_F	GAC CAT CAT GAA CTC TTC CTC	150 nM	This study
	CINT_R	CCG TTG CCT TCC AGT TC		
BCTT	BCTT_F	AGT GAC CTG AAA GAA TCC TAA T	150 nM	This study
	BCTT_R	GAC CGT CAA TAC CGA GAA AC		
RSBI	RSBI_F	GTG CCA GTA ACA GTC CAT ATT	150 nM	This study
	RSBI_R	TAG CAA AGC AGA GTG GAA AG		
FPRA	FPRA_F	TGT AAA CTC CTG TTG TTG AGG AAG ATA A	300 nM	[319]
	FPRA_R	GCG CTC CCT TTA CAC CCA	300 nM	
	FPRA_PR	**6FAM**-CAA GGA AGT GAC GGC TAA CTA CGT GCC AG-**TAMRA**	250 nM	
ECO	ECO_F	CAT GCC GCG TGT ATG AAG AA	300 nM	[320]
	ECO_R	CGG GTA ACG TCA ATG AGC AAA	300 nM	
	ECO_PR	**6FAM**-TAT TAA CTT TTA CTC CCT TCC TCC CCG CTG AA-**TAMRA**	100 nM	
PHGI & PHGII	PHG_F	CTC AAA GAG GGG GAC AAC AGT T	900 nM	[321]
	PHG_R	GCC ATC TCA AAG CGG ATT G	900 nM	
	PHGI_PR	**6FAM**-TAA GCC ACG ACC CGG CAT CG-**BHQ1**	300 nM	

Table 9. qPCR conditions. EUB, Eubacteria; GMLL, *G. morbillorum*; PTST, *P. stomatis*; BCTF, *B. fragilis*; CINT, *C. intestinalis*; BCTT, *B. thetaiotaomicron*; RSBI, *R. intestinalis*; FPRA, *F. prausnitzii*; ECO, *E. coli*; PHGI, *F. prausnitzii* phylogroup I; PHGII, *F. prausnitzii* phylogroup II; NA, non-applicable.

Bacterial markers	Total cycles	Denaturing		Annealing and Extension		Melting curve	
		Ta (ºC)	Time (min)	Ta (ºC)	Time (min)	Ta (ºC)	Time (min)
EUB	40	95	10:00	95 54	00:15 01:00	95 55 95	01:00 00:30 00:30
B10, B46, B48	40	95	10:00	95 62	00:15 00:45	95 55 95	01:00 00:30 00:30
GMLL, PTST, CINT, BCTT, RSBI	40	95	10:00	95 60	00:15 01:00	95 55 95	01:00 00:30 00:30
BCTF	40	95	10:00	95 55 72	00:15 00:30 01:00	95 55 95	01:00 00:30 00:30
FPRA, ECO	40	50 95	02:00 10:00	95 60	00:15 01:00	NA	NA
PHGI, PHGII	40	50 95	02:00 10:00	95 64	00:15 01:00	NA	NA

4 Statistical Analysis

In terms of qualitative analysis, absence of biomarker was considered if the obtained C_t value was not comprised within its dynamic range. All statistical analyses were performed using SPSS 23.0 statistical package (IBM, NYC, USA). Significance levels were established for P values ≤ 0.05.

Data normality was assessed through the Kolmogorov-Smirnov test. The non-parametric Kruskal-Wallis test was used to test differences in variables with more than two categories. Pairwise comparisons of subcategories of these variables were analysed using a Mann-Whitney test. The Bonferroni correction was used to correct for multiple comparisons. All comparisons using bacterial markers were performed between the relative abundances, which were normalized by the dynamic range of each bacterial marker.

The receiver operating characteristic (ROC) curve analysis was applied to determine the usefulness of each biomarker to distinguish among different colonic

neoplasia status. The accuracy of discrimination was measured by the area under the ROC curve (AUC).

Sensitivity, specificity, positive and negative predictive values of the designed algorithms were calculated using the software Epidat 3.1 (SERGAS, Xunta de Galicia, Spain).

4

Results

Chapter 1

Reduction of faecal immunochemical test false-positives results using a signature based on faecal bacterial markers in a symptomatic population

1 Background

Colorectal cancer is the third most common cancer in men and the second in women worldwide and a leading cause of cancer mortality. Around 75% of CRC are sporadic and they usually develop without symptomatology. Guidelines recommend routine screening for CRC in asymptomatic adults starting at age 50. Some countries are already implementing CRC screening politics in order to detect lesions at an early stage by using non-invasive tools. One of the most common tools is the faecal immunochemical test (FIT). Despite affordability, this test shows low sensitivity (29%) and a low positive predictive value (8%) for the detection of precancerous lesions. These values produce a high rate of false positive results among screening population. Experts from the United European Gastroenterology (UEG) are calling on European governments to focus their efforts on developing new, non-invasive, early diagnosis techniques able to complement the existing ones, thus saving thousands of lives from digestive cancers, and reducing the number of unnecessary colonoscopies due to the high false-positive rate.

Recently, it has been proved that bacterial communities in the intestinal mucosa of CRC patients are different from those of healthy individuals. In fact, evidences suggest that gut microbiota may play an important role in CRC pathogenesis. In 2012, Mas de Xaxars *et al* performed a preliminary and prospective study with 60 individuals (41 CRC patients and 19 patients with normal colonoscopy), in which a bacterial cluster highly correlated with CRC was defined using mucosal biopsies. Based on these results, a quantitative polymerase chain reaction system (qPCR) specifically targeted to those bacterial markers which make up the found bacterial signature was designed. Bacterial markers detection was later tested on stool samples (7 from healthy controls and 9 from CRC patients) looking for different abundances to check which one was suitable to be used as a non-invasive tool for CRC screening. A retrospective clinical study including 46 patients of the Hospital Universitari de Girona Dr. Josep Trueta (Girona, Spain) confirmed the suitability of some bacterial species as CRC markers (Table 10).

Table 10. Results obtained by Mas de Xaxars *et al.* A set of specific phylotypes determined in intestinal biopsy from CRC patients presented a high relatedness with CRC risk. Four of these bacterial markers (B3, B10, B46, B48) were quantified in faecal samples of healthy individuals (C=19) and patients with CRC (CRC=27). Results are expressed in Cts. Statistically significative p-values > 0.05. C, control; CRC, colorectal cancer.

		B3	B10	B46	B48
Mean ± SD	**CRC**	25.64 ± 3.97	16.13 ± 3.57	22.87 ± 3.45	20.16 ± 2.89
	C	22.47 ± 1.95	14.12 ± 1.96	20.58 ± 1.93	17.98 ± 1.52
p-value (t-test)	**C vs. CRC**	0.038	0.049	0.021	0.019
C vs. CRC	**AUC**	0.700	0.690	0.698	0.690
	Sensitivity	48.0%	57.0%	61.5%	36.0%
	Specificity	94.0%	89.5%	84.0%	100.0%

In this chapter a potential new non-invasive CRC-screening tool based on faecal bacterial markers capable of complementing FIT is developed in a symptomatic population, in order to assess the risk of intestinal disease, in particular advanced neoplasia (advanced adenomas + CRC).

2 Experimental Design

A cohort consisting of 333 consecutive patients with CRC-related symptoms referred for a diagnostic colonoscopy from primary and secondary health care to Complexo Hospitalario de Ourense (Ourense, Spain) was recruited (Table 11). Exclusion criteria were: (1) asymptomatic subjects undergoing colonoscopy for CRC screening, (2) patients with a previous history of colonic disease undergoing surveillance colonoscopy, (3) patients requiring hospital admission, (4) patients whose symptoms had ceased within 3 months before evaluation, and (5) patients who had received antibiotic treatment within the last month prior to inclusion. The study protocol was approved by the Biobanco del Complexo Hospitalario Universitario de Vigo (Vigo, Spain). Written informed consent was obtained from all study patients.

Table 11. Patients characteristics classified according to colonoscopy diagnostic. Hb, haemoglobin; FIT20 (20 µg Hb/f of faeces); CRC, colorectal cancer; AA, advanced adenoma; NAA, non-advanced adenoma; NC, normal colonoscopy.

Characteristics	CRC	AA	NAA	NC
n (%)	48 (14.4)	30 (9.0)	88 (26.4)	167 (50.2)
Age (mean, range)	73 (53-91)	65 (44-83)	67 (37-89)	61 (20-87)
Sex, female (%)	17 (10.0)	15 (8.8)	32 (18.8)	106 (62.4)
FIT100 (%)	47 (97.9)	18 (60.0)	30 (34.1)	21 (12.6)

All subjects underwent colonoscopy in order to determine their colorectal status. According to the endoscopic examination and the pathology results, subjects were classified into four groups as indicated in section 3.1 (Patients data and sampling). Patients diagnosed with CRC were also classified according to the stage of the tumour (Table 12). Individuals were also asked to answer a questionnaire in order to record clinical and epidemiologic data.

Table 12. Patients with colorectal cancer according to tumour TNM stage. CRC, colorectal cancer.

CRC stage	n (%)
0	3
I	6
II	10
III	21
IV	8

Participants provided a stool sample from one bowel movement before colonoscopy and prior to bowel cleanse. Genomic DNA was extracted from frozen samples and its concentration was determined by fluorimetry. After having adjusted DNA concentration to 8 ng/µl qPCR assays were performed. In this chapter, ten specific bacterial sequences associated with CRC were targeted: Eubacteria (EUB) as the total bacterial load; B10 (best match BLAST *Faecalibacterium prausnitzii*), B46 (best match BLAST *Subdoligranulum variabile*), B48 (best match BLAST *Ruminococcus, Roseburia, Coprococcus*), and *Roseburia intestinalis* (RSBI) as butyrate-producing bacteria biomarkers; *Gemella morbillorum* (GMLL), *Peptostreptococcus stomatis* (PTST), and *Bacteroides fragilis*

(BCTF) as opportunistic pathogens usually found in the oral cavity or in the bowel of mammals; *Collinsella intestinalis* (CINT) as a hydrogen and ethanol producer specific biomarker; and *Bacteroides thetaiotaomicron* (BCTT) as a saccharolytic species. The species-specific primers used in this study are shown in Table 8 and the thermal profiles were different according to the analysed biomarker (Table 9).

FIT analysis was performed at Complexo Universitario de Ourense employing the same sample used in the CRC-specific biomarkers analysis. Stool samples for faecal haemoglobin determination were analysed using the OC-Sensor tube collector and the assay was performed using the automated OC-Sensor, which detects gastrointestinal bleeding associated with disorders such as CRC, polyps and diverticulitis (Eiken Chemical Co., Tokyo, Japan). Positive tests were those with a concentration of faecal haemoglobin equal or higher than 20 μg Hb/g of faeces (FIT20).

Those bacterial markers which in combination were capable of distinguishing subjects with advanced neoplasia lesions from those with normal colonoscopy or non-advanced adenomas were used to design a decision tree algorithm, which was tested by Machine Learning. The specific methodology consisted of an initial training iteration on 100 aleatory partitions of the dataset and a further validation of the predictive models generated using 4 different machine learning algorithms (neural network, logistic regression, gradient boosting tree, random forest). RAID-CRC algorithm was eventually designed using four of the bacterial markers analysed together with FIT results.

3 Results

3.1 Faeces biomarkers in neoplasia progression

The relative abundance of each bacterial marker was determined for each diagnostic (normal colonoscopy, non-advanced adenoma, advanced adenoma, CRC) (Figure 20). Regardless of the colonoscopy diagnosis, three different butyrate producing species (B10, B46, and B48) were the most prevalent biomarkers with relative abundance values of 20.4%, 19.0%, and 20.0%, respectively. GMLL and PTST were significantly more abundant in CRC population than in normal colonoscopy individuals ($p=0.006$ and $p<0.001$, respectively) or non-advanced adenoma subjects ($p=0.047$ and $p<0.001$,

respectively). Although with no significant differences, it could be observed a tendency of B46, being more abundant in CRC patients rather than in subjects with advanced adenomas (p=0.087). Interestingly, EUB abundance was maintained constant regardless of neoplasia status. Comparison among the different CRC stages (0, I, II, III, and IV) did not show significant differences in the abundance of any bacterial marker.

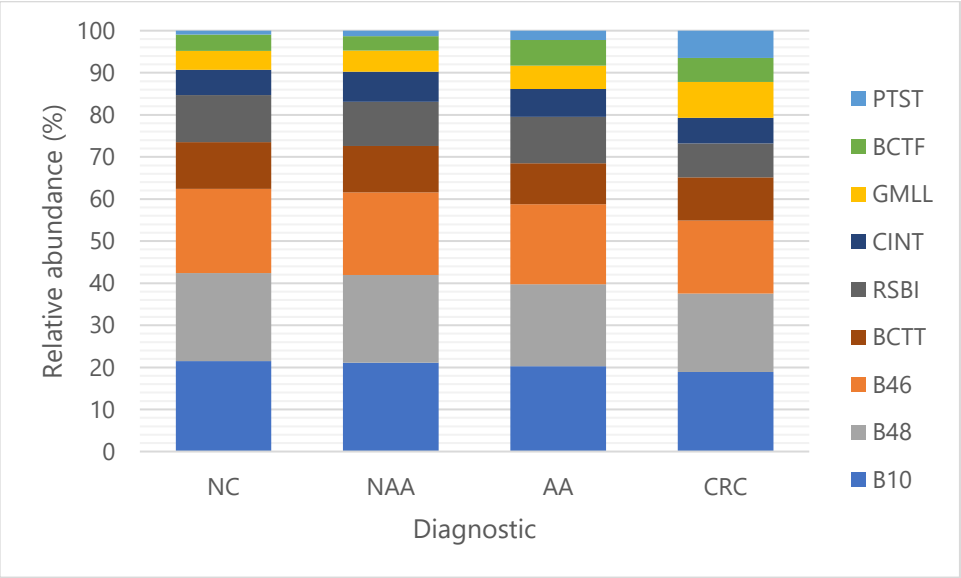

Figure 20. Relative abundance in percentage of the analysed bacterial markers (B10, B46, B48, *G. morbillorum* (GMLL), *P. stomatis* (PTST), *B. fragilis* (BCTF), *C. intestinalis* (CINT), *B. thetaiotaomicron* (BCTT) and *R. intestinalis* (RSBI)); for subjects with normal colonoscopy (NC), non-advanced adenoma (NAA), advanced adenoma (AA), and colorectal cancer (CRC).

3.2 CRC specific biomarkers can detect advanced neoplasia lesions

The relative abundance of bacterial markers was compared after grouping subjects as follows: (1) normal colonoscopy, (2) neoplasia (non-advanced adenoma + advanced adenoma + CRC), (3) advanced neoplasia (advanced adenoma + CRC), and (4) CRC (Figure 21). PTST was found to be highly correlated with neoplasia lesions (p<0.001). Regarding the detection of advanced neoplasia lesions, GMLL, PTST, and BCTF were potential biomarkers for their detection (p=0.006, p<0.001, and p=0.030, respectively). In terms of prevalence, these three opportunistic pathogens were found more often in patients with advanced neoplasia (GMLL, 64.9%; PTST, 58.4%; and BCTF, 44.7%) than in healthy subjects (GMLL, 53.5%; PTST, 26.1%; and BCTF, 29.8%).

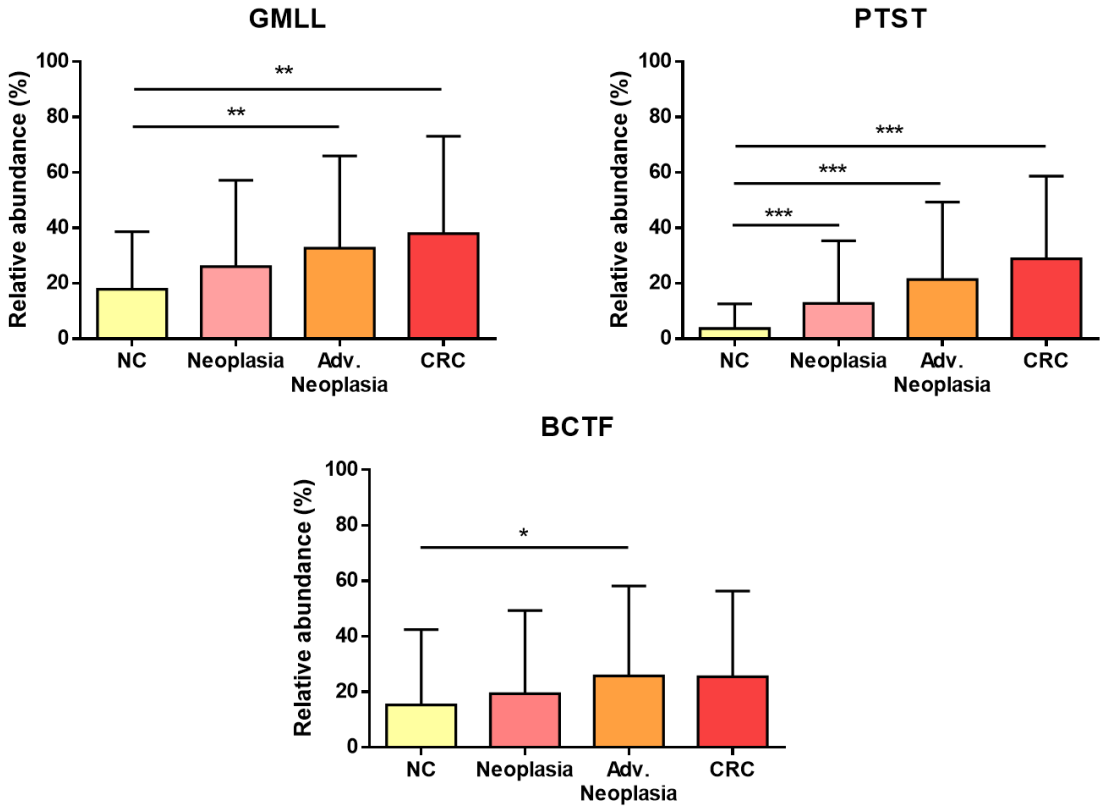

Figure 21. Comparison of the biomarkers' abundances among different diagnostics. NC, normal colonoscopy; neoplasia, non-advanced adenoma + advanced adenoma + colorectal cancer; advanced neoplasia, advanced adenoma + colorectal cancer; CRC, colorectal cancer. Level of significance: * when p-value < 0.05, ** when p-value < 0.01, and *** when p-value < 0.001.

3.3 Combination of CRC bacterial markers and FIT allows a substantial reduction of false positive results

One the one hand, when FIT20 was used, significant differences were observed between subjects with normal colonoscopies or non-advanced adenomas and advanced neoplasia (p<0.001). A 17.1% (19) of the subjects who showed a normal colonoscopy and a 24.3% (27) of those who had non-advanced adenomas showed FIT positive values. These results led to obtain a sensitivity and specificity of 84% and 81%, respectively, and positive and negative predictive values of 59% and 94%, respectively (AUC=0.828, 95% CI (0.773-0.883)) for the detection of advanced neoplasia. On the other hand, when FIT10 (10 μg Hb/g of faeces) was used, a 21.1% (27) of the subjects who showed normal colonoscopy and a 24.2% (31) of those who had non-advanced adenomas showed a false-positive result. FIT10 let to obtain a sensitivity and specificity of 91% and 76%, respectively, and positive and negative predictive values of 55% and 96%, respectively (AUC=0.836, 95% CI (0.787-0.886)) for the detection of advanced neoplasia. Sensitivity

values for bacterial markers alone were much lower being 39% for GMLL (AUC=0.622, 95% CI (0.541-0.694)), 53% for PTST (AUC=0.710, 95% CI (0.628-0.776) and 33% for BCTF (AUC=0.571, 95% CI (0.499-0.656)), while specificity values were comparable.

FIT results, both FIT20 and FIT10, were combined with the faecal bacterial markers in order to know which offered higher performance in terms of sensitivity and specificity values for the detection of advanced neoplasia lesions. The combination of the bacterial markers and FIT20 led to obtain a sensitivity of 76% and a specificity of 91% (Table 13). Nevertheless, these results were slightly improved by FIT10 as it showed a 4% higher sensitivity for the detection of advanced neoplasia, which therefore was the cut-off value of choice. Thus, RAID-CRC test is based on the combination of EUB, PTST, BCTF, and BCTT with a faecal haemoglobin concentration equal or higher than 10 μg Hb/g of faeces (FIT10). Although BCTT did not show significant differences between subjects with normal colonoscopies or non-advanced adenomas and advanced neoplasia subjects, once in combination with EUB, PTST, and BCTF, it was able to increase specificity by detecting healthy subjects.

The final algorithm consists of the combination of FIT10 and three ratios between bacterial markers (PTST/EUB, BCTF/EUB, BCTT/EUB). The application of the algorithm to the detection of advanced neoplasia resulted in a decrease in the number of false positive results, with a 9.7% of the subjects showing a normal colonoscopy and an 11.7% of subjects having non-advanced adenomas. Altogether, we obtained a sensitivity and a specificity of 80% and 90% (AUC=0.837, 95% CI (0.730-0.944)), respectively, and positive and negative predictive values of 70% and 94%, respectively. More importantly, the false-positive rate was reduced by 50%, being 46 subjects the false-positive results for FIT20 and 23 subjects for RAID-CRC.

Table 13. Diagnostic performance of RAID-CRC (using FIT20 and FIT10), FIT20 and FIT10 of the studied symptomatic population compared to FIT20 results of general screening population. FIT20 (20 µg haemoglobin/g of faeces); FIT10 (10 µg haemoglobin/g of faeces); PPV, positive predictive value; NPV, negative predictive value.

		Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
RAID-CRC (using FIT20)	Precancerous lesion	50	91	40	94
	Colorectal cancer	93	87	55	99
	Advanced neoplasia	76	91	72	92
RAID-CRC (using FIT10)	Precancerous lesion	59	90	43	95
	Colorectal cancer	94	85	51	99
	Advanced neoplasia	80	90	70	94
FIT20 *(this study)*	Precancerous lesion	62	81	28	95
	Colorectal cancer	98	75	42	99
	Advanced neoplasia	84	81	59	94
FIT10 *(this study)*	Precancerous lesion	76	76	28	96
	Colorectal cancer	100	71	38	100
	Advanced neoplasia	91	76	55	96
FIT20 [181,182,322]	Precancerous lesion	28	93	13	97
	Colorectal cancer	78	92	2	99
	Advanced neoplasia	30	93	15	97

4 Discussion

Early detection of advanced colorectal neoplasia through population-based screening and surveillance strategies is a critical step to reduce CRC mortality [121,267,323,324]. The ideal technique should be non-invasive, cost-effective, reproducible and capable to detect premalignant lesions with high risk of tumour development and high sensitivity and specificity values. In this study, we have defined a faecal bacterial signature that complements FIT and is able to reduce FIT-associated false positive results by increasing its specificity, in a symptomatic population.

Analysis of CRC-specific bacterial markers revealed that subjects with different colonoscopy diagnosis (i.e. normal colonoscopy, non-advanced adenoma, advanced adenoma, and CRC) showed different microbiological patterns. The total bacterial load does not seem to be affected when neoplasia appears, according to Sobhani *et al* [260]. Therefore, tumour lesions affect gut microbiota diversity but not its total amount. Using

a qPCR-based approach, our results clearly indicate the existence of a bacterial dysbiosis in patients with CRC. The studied bacterial markers were classified according to gut health related phenotypes: butyrate producers (B10, B46, B48, RSBI), opportunistic pathogens (GMLL, PTST, BCTF), hydrogen and oxygen producers (CINT), and saccharolytic species (BCTT) (Figure 22). Relative abundance of these phenotypes was found to change progressively as progression of the disease status. In particular, between subjects with normal colonoscopies and those with CRC we found a decrease in relative abundance of butyrate producers which were replaced by pathogenic bacteria group, being more abundant in CRC and advanced adenoma individuals than in subjects with normal colonoscopies. It was already reported that patients with CRC show a reduction of butyrate producers and an increase of opportunistic pathogens, which constitutes a major structural imbalance of their gut microbiota [264]. Bacterial dysbiosis can alter the balance of host cell proliferation and death, guide the immune system function and influence the metabolism of host-produced factors, ingested foodstuffs and pharmaceuticals [325]. Changes in bacterial composition, represented by a decrease in the amount of butyrate producing species and an increase in the opportunistic pathogens load, are likely to be a consequence of neoplastic lesion progression [4,254,331,260,267,290,326–330]. However, it has been reported that the increase in the abundance of opportunistic pathogens can lead to the release of bacterial toxins that can directly damage host DNA [325]. Other factors like reactive oxygen and nitrogen species, chemokines and cytokines released by these microorganisms can also contribute to tumour growth [254,325,329]. In other studies, there has already been observed how some other opportunistic pathogenic species are more abundant in subjects with CRC. In the same way, the abundances of the three bacterial markers representing opportunistic pathogens increased with the tumour lesion progression. Among all the bacterial markers hereby analysed, opportunistic pathogens were the best candidates to distinguish between advanced neoplasia and subjects with normal colonoscopies or non-advanced adenomas. However, despite the capability of detecting advanced neoplasia, GMLL, PTST and BCTF are not able to improve the sensitivity obtained with the current screening strategies. Therefore, we proposed to combine bacterial markers with FIT in a new tool, called RAID-CRC, in order to increment the

specificity values and consequently reduce the number of false positive results translated to unnecessary colonoscopies.

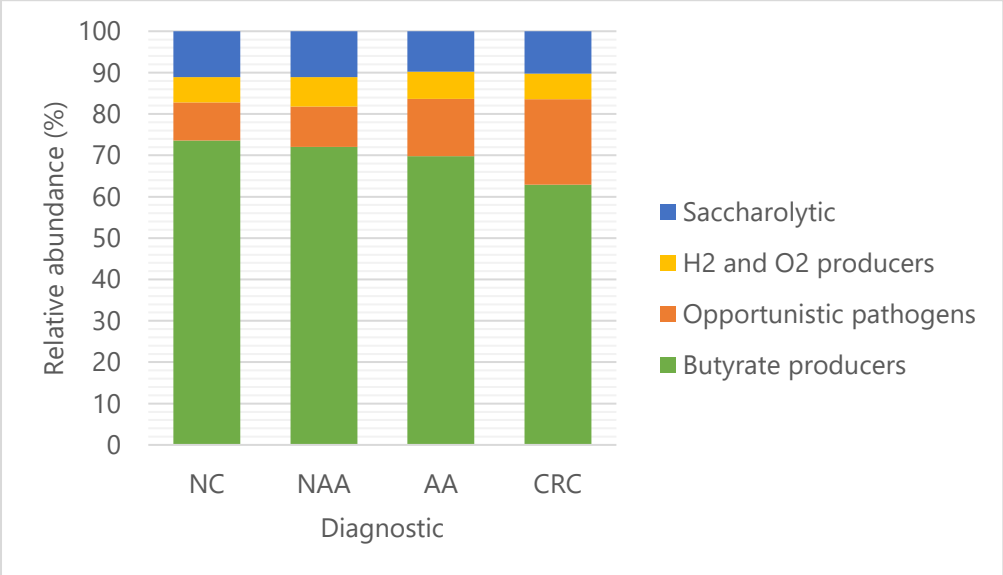

Figure 22. Relative abundance in percentage of the analysed bacterial markers (butyrate producing species: B10, B46, B48, and *R. intestinalis* (RSBI); opportunistic pathogens: *G. morbillorum* (GMLL), *P. stomatis* (PTST), and *B. fragilis* (BCTF); H_2 and O_2 producers: *C. intestinalis* (CINT); and saccharolytic bacteria: *B. thetaiotaomicron* (BCTT)) for subjects with normal colonoscopy (NC), non-advanced adenoma (NAA), advanced adenoma (AA), and colorectal cancer (CRC).

The RAID-CRC algorithm combines three bacterial markers abundance ratios (PTST/EUB, BCTF/EUB, BCTT/EUB) with FIT10. Our results show that high abundances of PTST and BCTF correlate with advanced neoplasia, whereas BCTT abundance is correlated with healthy individuals. BCTT is a commensal bacterium commonly found in the gut microbiota of healthy individuals. Commensal bacteria have been observed to attenuate gut inflammation and to contribute to colonization resistance [332,333]. Hence, high abundances of BCTT correlate with a good intestinal health. Using ratios allowed data normalization, which is critical to control qPCR-associated variables in order to differentiate true biological changes from experimentally induced variation [334]. Reduction in the faecal haemoglobin concentration from 20 µg Hb/g of faeces to 10 µg haemoglobin/g of faeces allows capturing positive subjects that otherwise would be considered false negative with a cut-off value of 20 µg/g, at expenses of increasing the false-positive rate. However, the RAID-CRC algorithm led to an important reduction of false negative results due to an increase of the sensitivity for detection of precancerous lesions with respect to FIT (Table 13). It has been reported that the sensitivity for

precancerous lesions obtained by FIT20 in a screening population is substantially lower than the one which could be obtained by RAID-CRC (increase of 30%) [181,182,322], as well as in terms of advanced neoplasia (increase of 50% by RAID-CRC). In terms of advanced neoplasia, sensitivity might be also much higher for RAID-CRC (80%) than for FIT20 (30%). It is important to highlight that comparisons have been made using different populations: patients with clinical symptoms and average-risk population. Therefore, RAID-CRC results point out the use of this new tool as a potential alternative to FIT20 in a CRC-screening population, nevertheless it must be validated in a screening scenario.

Although gut microbiota and its effects on the human body are becoming highly relevant in Medicine, bacteria had never been used before as indicators of any relevant change in the bowel physiology such as the development of a neoplasm. In the present work we developed a new methodology suitable to be used in national CRC-screening programs using stool samples. Bacterial signatures used in this work were originally retrieved from mucosa samples. Therefore, their presence in faeces is not heavily subjected to the variability caused by diet and some external factors [335,336] but is a measure of the real abundance in the colonic mucosa. This helps to overcome the enormous background noise present in stools and provides physiological meaningfulness to the biomarkers.

No metadata on body mass index (BMI), smoking or feeding habits were available on our dataset. Although these parameters have been reported to influence the microbiota composition of faecal samples [337–341], our biomarkers arise from mucosa samples which are not so dependent on external factors [342,343]. Biedermann *et al* reported that smoking withdrawal increases microbial diversity [344]. Other studies observed that chronic alcohol consumption leads to an increase in Proteobacteria and a decrease in Bacteroidetes [239,345]. Regarding BMI, its effect on microbiota is controversial [346–348]. Since RAID-CRC is not addressed to a specific population, which includes a variety of conditions and habits, a non-stratification strategy on the basis of these variables is a good way to reproduce with utmost reliability the CRC screening scenario. Another limitation of the study is the method used by patients to

collect and conserve stool samples, as they had to collect the sample in a sterile faeces' container, froze it using home freezers and transport it to the hospital under cooling conditions. Although acceptability among study participants was high, the procedure followed for sample collection, conservation and transport may be too complex to be implemented in massive CRC screening programs. Moreover, it would take longer to obtain the final results as not only the FIT value would have to be determined but also the relative abundance of the bacterial markers. To overcome this limitation, we have considered to set up the detection of RAID-CRC (FIT and bacterial signature analysis) in the FIT tube collector.

Cost-effectiveness is also a critical issue in population-based screening [349–351]. Wong and co-workers made a comparison of FIT and colonoscopy in this scenario, showing that FIT was cost-effective manner in average-risk screening, whereas colonoscopy was cost-effective among higher-risk subject [352]. Therefore, combination of FIT with faecal bacterial markers may be superior in terms of cost-effectiveness, since the use of RAID-CRC would permit to save up to 30% of total colonoscopies. More specifically, the implementation of RAID-CRC in a CRC-screening program would result in a reduction of 33,000 colonoscopies due to false positive results when compared to a screening program based on the FOBT (55,000 vs. 22,000 false positives, respectively) [353]. Considering the cost of RAID-CRC comparable to that of FOBT, the estimated savings in follow-up colonoscopies after positive screening results would be 77 million € per 100,000 participants in the screening program (Table 14). In addition, using the CRC biomarkers presented in this work may achieve both in developed and in resource-deprived regions, where colonoscopy facilities are limited, since RAID-CRC represents a potentially viable, cost-effective tool in a CRC-screening scenario.

In conclusion, RAID-CRC is a promising tool for CRC screening because it may allow to achieve a similar sensitivity as the current methodology used in most of the CRC-screening programs (FIT20), with a higher specificity and PPV. We will next seek validation in a screening setting of the obtained proof of concept in symptomatic individuals.

Table 14. Comparison of costs associated to follow-up colonoscopies among different CRC-screening programs.

Factor	**CRC-screening FIT-based** [353]	**CRC-screening with combined gFOBT and microbiota profiling** [267]	**CRC-screening with combined FIT and faecal bacterial signature (RAID-CRC)**
Nº of colonoscopies due to false positive results (per 100,000 screening participants)	47,600	24,750*	22,000 (this study)
Costs associated to follow-up colonoscopies after a positive screening result	111 M €**	70 M €**	51 M €***

*The savings have been calculated assuming that combining gFOBT with microbiota profiling can increase screening sensitivity more than a 45% relative to gFOBT alone.

**Associated costs have been calculated considering that the cost of the test is 25 €.

***Associated costs have been calculated considering that the cost of the test is 10 €.

Chapter 2

Faecal bacterial signature testing for colorectal cancer screening in a positive faecal immunochemical test population

1 Background

Guidelines recommend routine screening for CRC in asymptomatic adults starting at age 50. The most extensively used non-invasive test for CRC screening is the faecal immunochemical test (FIT). Although its overall sensitivity for CRC is around 61-91% and for advanced adenomas between 27-67%, these figures still contain a high false-positive rate and a low positive predictive value.

In this chapter a new non-invasive tool for CRC-screening (i.e. RAID-CRC Screen) based on a faecal bacterial signature that complements FIT increasing its specificity and positive predictive value for advanced neoplasia detection among FIT-positive participants has been developed. Proof-of-concept and clinical validation of the designed tool in two independent cohorts of an organized, population-based CRC screening program are reported.

2 Experimental Design

2.1 Proof of concept study

A cohort consisting of 189 consecutive FIT-positive participants in the national CRC screening program (asymptomatic, 50-69 years, FIT cut-off at 20 μg of haemoglobin/g of faeces) was recruited (Table 15). The recruiting centres were the Hospital Universitari Dr. Josep Trueta-IAS (Girona, Spain), the Hospital Universitari de Bellvitge (L'Hospitalet de Llobregat, Spain), and the Consorci Hospitalari de Vic (Vic, Spain). The study protocol (clinical investigation code: RAID-CRC 20202015) was approved by the Clinical Research Ethics Committee of the three participating centres. From 189 recruited subjects, 17 had to be excluded because of poor sample condition. Finally, 172 samples from asymptomatic subjects were used in the proof-of-concept study.

Exclusion criteria were: (1) subjects who had received antibiotic treatment within the last month prior to inclusion; (2) subjects who had received chemotherapy and/or radiotherapy within the last 6 months prior to inclusion; (3) subjects with severe comorbidity which, in opinion of the investigator, should preclude participation in the study; (4) subjects who had gastrointestinal adverse effects of chemotherapy and/or

radiotherapy (received 6 months prior to inclusion) that may compromise function of the digestive system; and (5) pregnancy at the time of the inclusion. Written informed consent was obtained from all participants.

Table 15. Characteristics of patients included in the proof-of-concept (RAID-CRC 20202015) and clinical validation (GG-RAIDCRC-1002) studies, classified according to their diagnosis. CRC, colorectal cancer; AA, advanced adenoma; NAA, non-advanced adenoma; NC, normal colonoscopy.

	Characteristics	CRC	AA	NAA	NC
Proof of concept	**n (%)**	11 (6.3)	67 (39.0)	38 (22.1)	56 (32.6)
	Age (mean, range)	61 (50-69)	61 (50-69)	60 (50-69)	59 (49-69)
	Sex, female (%)	6 (54.5)	19 (28.3)	15 (39.5)	37 (66.1)
Clinical validation	**n (%)**	19 (5.8)	124 (37.9)	85 (26.0)	99 (30.3)
	Age (mean, range)	61 (54-69)	61 (50-73)	61 (50-70)	58 (49-69)
	Sex, female (%)	6 (31.6)	52 (41.9)	42 (49.4)	53 (53.5)

All subjects underwent colonoscopy in order to determine their colorectal status. According to the endoscopic examination and the pathology results, subjects were classified into four groups as indicated in section 3.1 (Patients data and sampling). Individuals were also asked to answer a questionnaire in order to record clinical and epidemiologic data.

2.2 Clinical validation study

The algorithm developed in the proof-of-concept study (*RAID-CRC Screen*) was validated in an independent cohort consisting of 359 consecutive FIT-positive participants in the national CRC screening program (Table 15). Inclusion and exclusion criteria were the same as in the proof-of-concept study. From the 359 recruited subjects, 32 had to be excluded from the study for different reasons: subject did not collect the stool sample, incorrect sample conservation, absence of colonoscopic diagnosis, and poor sample condition. Finally, 327 samples were used for the clinical validation of *RAID-CRC Screen*.

Clinical validation was designed as a cross-sectional, multicentre study. The study protocol (clinical investigation code: GG-RAIDCRC-1002) was approved by the Clinical

Research Ethics Committee of Hospital Universitari Dr. Josep Trueta-Institut d'Assistència Sanitària (IAS) Consortium and Hospital Universitari de Bellvitge. According to the endoscopic findings, subjects were classified into the four groups previously mentioned in section 3.1 (Patients data and sampling). Written informed consent was obtained from all study participants.

All participants, both from the proof-of-concept and the clinical validation, provided a stool sample from one bowel movement before colonoscopy and prior to bowel cleanse. Genomic DNA was extracted from frozen samples and its concentration was determined by fluorimetry. In the proof-of-concept, after having adjusted DNA concentration to 8 ng/µl qPCR assays were performed; however, in the clinical validation DNA concentration was not assessed since the total bacterial load (Eubacteria) was used to normalize data. In this chapter, the specific bacterial sequences targeted were classified in four different groups according its characteristics: Eubacteria (EUB) as the total bacterial load; B10 (best BLAST match *Faecalibacterium prausnitzii*), B46 (best BLAST match *Subdoligranulum variabile*), B48 (best BLAST match *Ruminococcus, Roseburia, Coprococcus*), *F. prausnitzii* (FPRA), and *Roseburia intestinalis* (RSBI) as butyrate-producing bacteria biomarkers; *Gemella morbillorum* (GMLL), *Peptostreptococcus stomatis* (PTST), and *Bacteroides fragilis* (BCTF) as opportunistic pathogens usually found in the oral cavity or in the bowel of mammals; and *Bacteroides thetaiotaomicron* (BCTT) as a saccharolytic species. The species-specific primers used in this study are shown in Table 8 and the thermal profiles were different according to the analysed biomarker (Table 9).

FIT analysis was performed at Hospital Universitari de Girona Dr. Josep Trueta – IAS, Hospital Universitari de Bellvitge, or Consorci Hospitalari de Vic, employing a different sample than the one used in the CRC-specific biomarkers analysis which was collected in the prior month to colonoscopy. Stool samples for faecal haemoglobin determination were analysed using the OC-Sensor tube collector and the assay was performed using the automated OC-Sensor, which detects gastrointestinal bleeding associated with disorders such as CRC, polyps and diverticulitis (Eiken Chemical Co., Tokyo, Japan). Positive tests were those with a concentration of faecal haemoglobin equal or higher than 20 µg Hb/g of faeces (FIT20).

In the proof-of-concept study, machine learning was used to determine which combination of the bacterial markers were capable of distinguishing subjects with AN from those with normal colonoscopy or NAA. The specific methodology consisted of an initial training iteration with the 70% of 100 random partitions of the dataset and a further validation with the 30% left of the predictive models generated using 4 different machine learning algorithms (neural network, logistic regression, gradient boosting tree, and random forest). *RAID-CRC Screen* was eventually designed using the combination of six of the bacterial markers.

Sample size of the clinical validation was calculated in 359 subjects using the online platform GRANMO v7.12 estimating the population by proportions, with 95% confidence interval, an accuracy of +/- 5%, estimating a population percentage that is expected to be around 30%, with a 10% repositions foreseen.

3 Results

3.1 Faecal bacterial markers in neoplasia progression

In the proof-of-concept study, stool samples were used to evaluate the relative abundance of each bacterial marker according to each diagnosis (Figure 23). Regardless of the colonoscopy result, the most prevalent bacterial species were butyrate producers (B10, B46, B48, FPRA, and RSBI), with relative abundance values of 18.53%, 15.76%, 17.94%, 15.68%, and 12.02%, respectively. Subjects with either NAA or AA showed higher relative abundances of B10, B46 and FPRA when compared to healthy individuals ($p=0.018$, $p=0.004$, and $p=0.023$, respectively). PTST and BCTF were significantly more abundant in CRC population than in individuals with normal colonoscopy ($p=0.002$ and $p=0.017$, respectively). Although with no significant differences, it could be observed a tendency of GMLL, being more abundant in CRC patients ($p=0.073$); similarly, PTST and RSBI were more abundant in the presence of either NAA or AA ($p=0.056$ and 0.060, respectively).

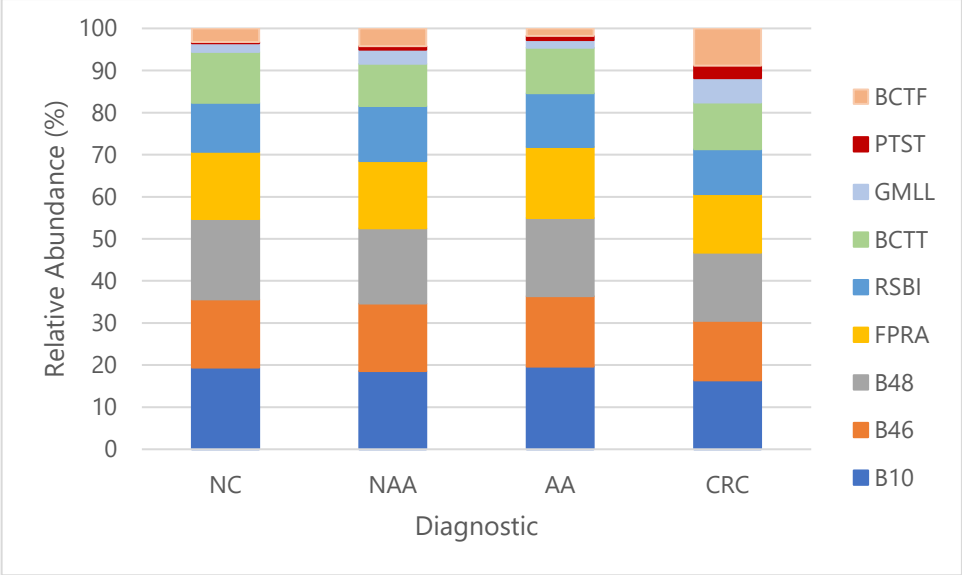

Figure 23. Relative abundance of the analysed bacterial markers (B10, B46, B48, *F. prausnitzii* (FPRA), *G. morbillorum* (GMLL), *P. stomatis* (PTST), *B. fragilis* (BCTF), *B. thetaiotaomicron* (BCTT) and *R. intestinalis* (RSBI)); in subjects with normal colonoscopy (NC), non-advanced adenoma (NAA), advanced adenoma (AA), and colorectal cancer (CRC).

On the one hand, a higher abundance of opportunistic pathogens (GMLL, PTST, and BCTF) was observed in CRC patients with respect to healthy subjects (p=0.073, p=0.002, and p=0.017, respectively). In terms of prevalence, these bacterial markers were found more often in patients with CRC (GMLL, 75.0%; PTST, 50.0%; and BCTF, 66.6%) than in healthy subjects (GMLL, 40.8%; PTST, 11.1%; and BCTF, 31.3%). On the other hand, when subjects with normal colonoscopy were compared to patients with AN, three butyrate-producing bacteria (B10, B46, and FPRA) showed significant differences in their relative abundance (p=0.035, p=0.030, and p=0.042, respectively).

3.2 RAID-CRC Screen algorithm development

The development of *RAID-CRC Screen* algorithm was focused on the reduction of false positive results for AN among the FIT-positive subjects, while maintaining 100% sensitivity for CRC. Although FPRA, B46, and B10 were the only bacterial markers that showed significant differences between subjects with normal colonoscopy or NAA and those with AN within the FIT-positive population, they did not provide the desired sensitivity values. The combination of the two first markers (FPRA and B46) with three other bacterial species (B48, GMLL, and BCTF) and the total bacterial load led to the achievement of an algorithm with a sensitivity of 95%, a specificity of 26%, a PPV of 50%, and a negative predictive value of 86%, for AN (Table 16). More specifically, *RAID-CRC*

Screen consists of 5 bacterial ratios: FPRA/EUB, B46/EUB, B48/EUB, GMLL/EUB, and BCTF/EUB applied in the FIT-positive population. While FIT 20 μg/g lead to 94 false-positive results for AN detection, *RAID-CRC Screen* reduced this value to 62 implying a reduction of the false positive rate of 34%. Notably, developers focused on sensitivity for AN which approached 100%.

Table 16. Diagnostic performance of RAID-CRC Screen in the proof-of-concept study (n=172). NC, normal colonoscopy; NAA, non-advanced adenoma; PPV, positive predictive value; NPV, negative predictive value.

Most advanced finding	Groups used for calculating specificity	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Colorectal cancer (n=11)	NAA+NC	100	28	13	100
Advanced adenomas (n=67)	NAA+NC	95	26	49	88
Advanced neoplasia (n=78)	NAA+NC	95	26	50	86

3.3 Clinical validation of RAID-CRC Screen

When *RAID-CRC Screen* was scaled up to the 327 FIT-positive individuals of the validation cohort, a sensitivity of 95%, 82% and 84% were obtained for the detection of CRC, AA and AN, respectively (Table 17). Specificities were 16% among participants with NAA or normal colonoscopy, and 18% among those with negative results on colonoscopy (Table 17). In this validation study, the algorithm detected 30 true negative subjects but generated 23 false negatives from the 327 FIT-positive individuals, 22 of them with AA and 1 with CRC. More importantly, while using FIT 20 μg/g there were 184 false-positive results for AN, *RAID-CRC Screen* reduced this figure to 154, which implies a reduction of the false positive rate of 16.3%.

Table 17. Diagnostic performance of RAID-CRC Screen in the validation study (n=327). NC, normal colonoscopy; NAA, non-advanced adenoma; PPV, positive predictive value; NPV, negative predictive value.

Most advanced finding	Groups used for calculating specificity	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Colorectal cancer (n=19)	NC+NAA	95	16	11	97
	NC	95	18	18	95
Advanced adenoma (n=124)	NC+NAA	82	16	40	58
	NC	82	18	56	45
Advanced neoplasia (n=143)	NC+NAA	84	16	44	57
	NC	84	18	60	44

4 Discussion

Early detection of CRC is crucial to reduce its incidence and mortality. The European Union recommends population-based, organized, screening for CRC using evidence-based methods with quality assurance of the entire screening process [189]. The best CRC-screening strategy has not been defined yet, but the most widely used in Western countries FIT, which shows 79% sensitivity and 99% negative predictive value for CRC detection [182]. The most common limitation of FIT-based CRC-screening programs are false-positive results [192], as they lead to patient's concerns, additional costs [193] and adverse events associated with unnecessary colonoscopies. In this study, we have defined and clinically validated a faecal bacterial signature that complements FIT by reducing its associated false-positive rate among FIT-positive participants.

Different bacterial markers associated with CRC have been analysed in subjects with colorectal neoplastic lesions at different stages: normal colonoscopy, NAA, AA and CRC. Using a qPCR-based approach, our results clearly indicate the existence of a microbial dysbiosis in patients with CRC. The analysed bacterial markers were classified according to their gut health related phenotypes: butyrate producing bacteria (B10, B48, B48, FPRA, RSBI), saccharolytic bacteria (BCTT), and opportunistic pathogens (GMLL, PTST, BCTF) (Figure 24). Relative abundance of these microbiological groups differed significantly when healthy subjects were compared to CRC patients. Specifically, we found a decrease in the abundance of butyrate producing bacteria in CRC when compared to subjects with normal colonoscopy, being replaced by pathogenic species.

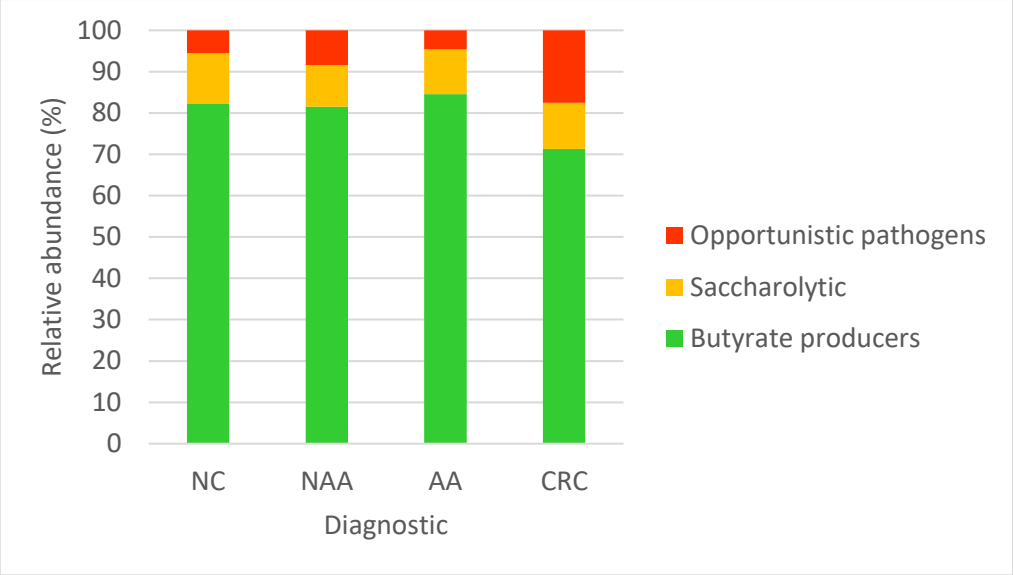

Figure 24. Relative abundance of the analysed bacterial markers (butyrate producing species: B10, B46, B48, FPRA, and RSBI; opportunistic pathogens: GMLL, PTST, BCTF; saccharolytic bacteria: BCTT); in subjects with normal colonoscopy (NC), non-advanced adenoma (NAA), advanced adenoma (AA), and colorectal cancer (CRC).

In the proof-of-concept, we prioritized the correct classification of subjects with CRC, thus a 100% sensitivity was sought. Therefore, *RAID-CRC Screen* was developed to maximize the reduction of false-positive results, while maintaining a high sensitivity for AN. The defined microbiological signature combines five bacterial markers abundance ratios (FPRA/EUB, GMLL/EUB, B46/EUB, BCTF/EUB, B48/EUB). Our results showed that high FPRA/EUB, GMLL/EUB, and BCTF/EUB bacterial ratios correlated with the presence of AN. On the contrary, low B46/EUB and B48/EUB bacterial ratios correlated with healthy individuals and advanced neoplasia, respectively. In this setting, it is important to note that use of the total bacterial load for data normalization is critical to control variables associated to qPCR, thus differentiating true biological changes from experimentally induced variation [334].

Clinical validation of *RAID-CRC Screen* confirmed a sensitivity and negative predictive values for CRC detection approaching 100% among FIT-positive individuals. With respect to the detection of AA, promising results were also obtained with a sensitivity of 82%. Altogether, *RAID-CRC Screen* reached a sensitivity and specificity of 84% and 18%, respectively, for the detection of AN. More importantly, the use of this new non-invasive tool achieved a reduction of up to 20% unnecessary colonoscopies among subjects who tested positive to FIT. In addition, cancer screening programs seek

methods with high negative predictive values, since all affected individuals should be detected. On this line, our method obtained a negative predictive value for CRC detection of almost 100%.

A two-step approach, in which individuals with a FIT-positive result undergo colonoscopy, is a cost-effective CRC screening strategy in average-risk population [171,352]. In this setting, use of the defined bacterial signature after testing positive in FIT may increase cost-effectiveness, as the number of unnecessary colonoscopies would be further reduced. More specifically, the implementation of *RAID-CRC Screen* in a CRC screening program would result in a 20% reduction of colonoscopies indicated due to false-positive results of FIT, thus representing a saving of at least 138 M € per year in UK. In a similar manner, the use of faecal bacterial signature may allow the introduction of CRC screening programs in resource-deprived regions, where colonoscopy availability is limited.

Based on our results, we propose a three-step CRC screening strategy, in which the bacterial signature (*RAID-CRC Screen*) is applied to FIT-positive individuals for a better selection of those who should undergo colonoscopy. This approach was associated with a significant reduction of false-positive results of FIT among participants in a population-based, organized CRC screening program. Therefore, *RAID-CRC Screen* is being postulated as a new non-invasive tool for CRC screening, adding specificity and PPV to FIT while maintaining high sensitivity for AN.

Chapter 3

Use of new faecal bacterial signatures for the individualization of the Lynch syndrome surveillance

1 Background

While most colorectal tumours are sporadic (70-80%), a small part of them have an hereditary component, i.e. Lynch syndrome (2-5%), different forms of colorectal polyposis (1%) or CRC associated to *MUTYH* gene (1%) [354]. Lynch syndrome is an autosomal dominant inherited condition caused by germline mutation in DNA mismatch repair (MMR) genes (i.e. *MLH1, MSH2, MSH6 and PMS2*) [355]. MMR genes repair errors that happen during DNA replication before cell division. When this mechanism is deficient, microsatellite instability (MSI) in tumours is triggered [356]. It is characterized by the development of early-onset CRC, endometrial, gastric, urinary tract, small bowel and other cancers [131]. It is also typified by accelerated carcinogenesis, being less than 2 years the period of time from a normal colonoscopy diagnosis to a colorectal carcinoma [131]. This interval of time is increased up to 10 years when talking about the development of a sporadic CRC [357]. There is an ongoing discussion about the optimal interval between colonoscopic examinations. Guidelines recommend intensive colonoscopy surveillance starting at the age of 25 years, with intervals of no more than 2 years [358,359]. Currently, there is special interest in the development of new non-invasive tools that lead to the reduction of unnecessary surveillance colonoscopies performed in Lynch syndrome (LS) carriers.

The vast majority of intestinal microbiota studies are focused on the analysis of patients who develop colorectal cancer sporadically. Nowadays, there are very few studies focused on the specific analysis of LS carriers' gut bacterial community [314,315]. Thus, a Chinese research group has analysed possible interactions between gut microbiota and LS and observed that, together with genetic factors, some gut microbes such as *Bacteroides fragilis*, *Parabacteroides distasonis* and Pseudomonadaceae family are also implicated in the development of LS [314]. On the other hand, a clinical trial lead by the Memorial Sloan Kettering Cancer Centre, that is still recruiting patients, will perform a metagenomic evaluation of the gut microbiome of patients with LS and other hereditary colonic polyposis syndromes [315].

In this chapter a specific faecal bacterial signature of sporadic CRC and IBD patients is compared to that of LS carriers. Moreover, a non-invasive tool based on this

signature that enables the prediction of the absence of neoplasia in LS carriers is developed. This new method, which is called RAID-LS "Risk Assessment Intestinal Disease for Lynch Syndrome", could lead to significant savings in the number of surveillance colonoscopies performed in LS carriers, expanding the intervals between endoscopic examinations, thus improving patients' quality of life and compliance.

2 Experimental Design

A cohort consisting of 66 Lynch syndrome carriers who underwent a surveillance colonoscopy was recruited (Table 18, Table 19). A second cohort consisting of 301 control individuals who joined the regional CRC screening program (asymptomatic, FIT-positive, 50-69 years) or presented CRC compatible symptomatology (blood in stool, abdominal pain, changes in bowel habits, unexplained tiredness, weight loss) was also recruited (Table 17). The recruiting centres were the Institut Català d'Oncologia (ICO) (Girona and l'Hospitalet de Llobregat, Spain), the Hospital Universitari Dr. Josep Trueta (Girona, Spain), the Hospital Universitari de Bellvitge (l'Hospitalet de Llobregat, Spain), and the Consorci Hospitalari de Vic (Vic, Spain). Exclusion criteria were: (1) subjects who had received antibiotic treatment within the last month prior to inclusion; (2) subjects who had received chemotherapy and/or radiotherapy within the last 6 months prior to inclusion; (3) subjects with severe comorbidity which, in opinion of the investigator, should preclude participation in the study; (4) subjects who had gastrointestinal adverse effects of chemotherapy and/or radiotherapy (received 6 months prior to inclusion) that may compromise function of the digestive system; and (5) pregnancy at the time of the inclusion.

The study protocol was approved by the Clinical Research Ethics Committee of Hospital Universitari de Girona Dr. Josep Trueta and Hospital Universitari de Bellvitge (clinical investigation code: 20202015). Written informed consent was obtained from all study patients.

All subjects underwent colonoscopy in order to determine their colorectal status. According to the endoscopic examination and the pathology results, subjects were classified into four groups as indicated in section 3.1 (Patients data and sampling).

Participants were also asked to answer a questionnaire in order to record clinical and epidemiological data.

Table 18. Main characteristics of the study participants. NC, normal colonoscopy; NAA, non-advanced adenoma; AA, advanced adenoma; SA, serrated adenoma; CRC, colorectal cancer; NA, not of application.

		Lynch	Control
Age	mean (range)	50 (25-83)	60 (33-85)
Sex (female)	n (%)	44 (66.7)	152 (50.5)
Diagnosis	NC (%)	54 (81.8) †	136 (45.2)
	NAA (%)	7 (10.6) ‡	63 (20.9)
	AA (%)	3 (4.5)	79 (26.2)
	SA (%)	1 (1.5)	9 (3.0)
	CRC (%)	1 (1.5)	14 (4.6)
Mutated gene	*MLH1* (%)	28 (43.0)	NA
	MSH2 (%)	14 (21.0)	NA
	MSH6 (%)	16 (24.0)	NA
	PMS2 (%)	8 (12.0)	NA
Previous CRC	n (%)	15 (23.0)	0
Total	n	66	301

† Endometrium cancer (4), ovarian cancer (1); ‡ Endometrium cancer (1)

Table 19. Detailed description of the LS population of the study. ID, identification; M, male; F, female; NA, non-applicable; ND, no data.

Subject ID	Gender	Current age	Mutated Gene	CRC age	CRC TNM	Type of CRC surgery	Other cancers (age)	Nº surveillance colonoscopies	Follow-up years
BG101	M	66	*MSH6*	NA	NA	NA	0	0	3
BG102	M	78	*MLH1*	66	pT3N0M0	Right hemicolectomy	prostate (72)	7	10
BG103	M	42	*MLH1*	NA	NA	NA	0	0	4
BG105	F	37	*MLH1*	NA	NA	NA	0	2	7
BG107	F	50	*MSH6*	NA	NA	NA	0	0	3
BG108	M	74	*MSH6*	44 / 63	pT3N0M0 / pT4N0M0	Subtotal colectomy	0	7	16
BG110	F	48	*MSH2*	NA	NA	NA	0	6	16
BG111	M	75	*MSH6*	NA	NA	NA	0	0	3
BG112	F	54	*MLH1*	NA	NA	NA	0	0	4
BG113	F	61	*MLH1*	NA	NA	NA	0	0	4
BG114	F	47	*MLH1*	NA	NA	NA	0	0	5
BG116	M	46	*MSH6*	NA	NA	NA	0	0	3
BG117	F	50	*MLH1*	37	pT4N0M0	Subtotal colectomy	0	5	13
BG118	F	54	*MLH1*	40 / 41	pT2N0M0 / ND	Right hemicolectomy	0	10	13
BG119	M	44	*MSH2*	NA	NA	NA	0	0	14
BG120	F	61	*MLH1*	NA	NA	NA	0	6	10
BG121	F	61	*MLH1*	37	pT4N1M0	Right hemicolectomy	0	7	10
BG122	F	36	*MLH1*	NA	NA	NA	0	4	10
BG123	M	51	*MSH2*	NA	NA	NA	4 sebaceous adenomas (36, 37, 39, 45), skin (40)	2	9
BG124	F	43	*MLH1*	28	pT2N0M0	Anterior rectal resection	0	5	15
BG126	F	34	*MLH1*	NA	NA	NA	0	4	9

Subject ID	Gender	Current age	Mutated Gene	CRC age	CRC TNM	Type of CRC surgery	Other cancers (age)	Nº surveillance colonoscopies	Follow-up years
BG127	F	61	*MLH1*	42	pT3N0M0	Left hemicolectomy	0	8	18
BG129	M	55	*MSH6*	NA	NA	NA	0	3	ND
BG130	M	27	*MSH6*	NA	NA	NA	0	0	4
BG131	F	74	*MSH6*	49	pT3N0M0	Right hemicolectomy	uterus (47)	4	9
BG132	F	56	*MSH6*	NA	NA	NA	uterus (53)	0	4
BG133	F	69	*MSH6*	NA	NA	NA	0	ND	19
BG134	F	63	*MSH2*	NA	NA	NA	uterus (43), 3 sebaceous adenomas (43, 47, 48)	6	20
BG136	M	39	*MLH1*	NA	NA	NA	0	1	7
BG137	F	41	*MLH1*	NA	NA	NA	ovary (34)	2	6
BG138	M	63	*MLH1*	55	pT3N0M0	Right hemicolectomy	appendix (55)	3	6
BG139	M	58	*MSH2*	37 / 48 / 48	pTxNxM0 / pT3N1M0 / pT1N1M0	Subtotal colectomy	sebaceous adenoma (ND), skin (54)	6	9
BG140	F	48	*MLH1*	NA	NA	NA	0	2	11
BG141	M	67	*MSH2*	48 / 48 / 54 / 66	pT3N1M0 / pT3N0M0 / pT1NxM0 / pT1NxM0	Subtotal colectomy	ureter (53), bladder (53), prostate (60)	ND	ND
BG142	F	60	*MSH2*	NA	NA	NA	uterus (56)	0	3
BG143	M	74	*MSH6*	63	pT1N0M0	Left hemicolectomy	0	8	13
BG144	F	63	*PMS2*	NA	NA	NA	0	5	13
BG145	F	55	*MLH1*	NA	NA	NA	uterus (52)	6	13
BG146	F	47	*PMS2*	NA	NA	NA	0	3	7
BG147	F	85	*PMS2*	NA	NA	NA	0	2	7
BG148	F	33	*MLH1*	NA	NA	NA	0	2	13

Subject ID	Gender	Current age	Mutated Gene	CRC age	CRC TNM	Type of CRC surgery	Other cancers (age)	Nº surveillance colonoscopies	Follow-up years
BG149	M	69	*MLH1*	55	pT1N0M0	Right hemicolectomy	0	8	13
BG150	M	51	*MSH2*	41	pT3N2Mx	Anterior rectal resection	0	5	10
BG251	F	55	*MLH1*	NA	NA	NA	0	8	17
BG252	F	32	*MSH2*	NA	NA	NA	0	0	6
BG253	F	54	*MLH1*	NA	NA	NA	uterus (48)	0	5
BG255	F	50	*PMS2*	NA	NA	NA	0	4	13
GR201	F	61	*MSH6*	NA	NA	NA	0	7	6
GR202	M	39	*MLH1*	NA	NA	NA	0	3	8
GR203	F	43	*MSH2*	NA	NA	NA	0	3	7
GR205	F	43	*MSH6*	NA	NA	NA	0	4	7
GR206	F	44	*MLH1*	39	pT3N1M0	Right hemicolectomy	0	4	4
GR207	F	39	*MSH2*	NA	NA	NA	bladder (33), uterus (35)	1	3
GR208	F	37	*MSH2*	NA	NA	NA	liver (32)	2	7
GR209	F	49	*MLH1*	NA	NA	NA	0	6	9
GR210	F	42	*PMS2*	NA	NA	NA	0	4	6
GR211	F	66	*MSH2*	NA	NA	NA	uterus (42), ovary (43)	1	3
GR212	M	36	*PMS2*	NA	NA	NA	0	4	6
GR213	F	52	*MLH1*	NA	NA	NA	0	8	9
GR214	M	44	*MSH2*	NA	NA	NA	sebaceous adenoma (38), ureter (40)	3	5
GR215	F	55	*MSH6*	NA	NA	NA	0	2	3
GR216	M	43	*PMS2*	NA	NA	NA	0	4	6
GR217	F	60	*MSH6*	NA	NA	NA	0	5	7
GR218	F	49	*MSH6*	NA	NA	NA	0	7	7

Participants provided a stool sample from one bowel movement before colonoscopy and prior to bowel cleanse. Genomic DNA was extracted from frozen samples and its concentration was determined by fluorimetry. After having adjusted DNA concentration to 8 ng/µl qPCR assays were performed. In this chapter, the specific bacterial sequences targeted were 14: Eubacteria (EUB) as the total bacterial load; B10 (best match BLAST *Faecalibacterium prausnitzii*), B46 (best match BLAST *Subdoligranulum variabile*), B48 (best match BLAST *Ruminococcus, Roseburia, Coprococcus*), *Faecalibacterium prausnitzii* (FPRA), *F. prausitzii* phylogroup I (PHGI), *F. prausnitzii* phylogroup II (PHGII) and *Roseburia intestinalis* (RSBI) as butyrate-producing bacteria biomarkers; *Escherichia coli* (ECO) as proinflammatory species; *Gemella morbillorum* (GMLL), *Peptostreptococcus stomatis* (PTST), and *Bacteroides fragilis* (BCTF) as opportunistic pathogens; *Collinsella intestinalis* (CINT) as a hydrogen and ethanol producer specific biomarker; and *Bacteroides thetaiotaomicron* (BCTT) as a saccharolytic species. B10, B46, B48, GMLL, PTST, BCTF, BCTT, and RSBI are associated with colorectal cancer; while FPRA, ECO, PHGI, and PHGII have been linked to inflammatory bowel disease. The species-specific primers used in this study are shown in Table 8 and the thermal profiles were different according to the analysed biomarker (Table 9).

The RAID-LS algorithm was designed using a decision tree. This method consists of the division of a data set into smaller data sets using different cut-offs. With this method we identified those bacterial markers that could distinguish healthy LS carriers from those with neoplasia lesions. The receiver operating characteristic (ROC) curve analysis was applied to determine the usefulness of each biomarker to distinguish among different colonic neoplasia status. The accuracy of discrimination was measured by the area under the ROC curve (AUC). Likelihood ratios (LR) were calculated to analyse the performance of the designed algorithm.

3 Results

3.1 Comparison of the abundance of specific faecal bacterial markers between healthy LS carriers and patients with sporadic CRC

The relative abundance of bacterial markers was compared after grouping LS carriers with a normal colonoscopy in two groups: (1) those who had already suffered

from CRC (n=13) and (2) those without previous colorectal lesions (n=41). EUB, which corresponds to the total bacterial load, was equivalent in both groups. In spite of the lack of differences those with no previous CRC (41 subjects) were used in this analysis, thus avoiding the potential bias linked to prior surgery.

Comparison between healthy control subjects (n=136) and NC LS population (n=41), showed that GMLL, PTST and BCTF were significantly more abundant in LS carriers (Table 20). Regarding the comparison between NC LS (41) and control subjects with adenomas, both NAA (63) and AA (79), significant differences were found in the abundance of GMLL, PTST, PHGI and PHGII for NAA, and B10, GMLL, PTST, BCTF, PHGI and PHGII for AA. Noteworthy, no bacterial markers showed significant differences when NC LS (41) were compared to control subjects with CRC (14).

Table 20. Comparison of the mean relative abundance (mean ± standard deviation) of the analysed bacterial markers between healthy Lynch syndrome carriers and control subjects with different colonoscopy diagnostics. Level of significance: when p-value < 0.05 (highlighted). NC, normal colonoscopy; NAA, non-advanced adenoma; AA, advanced adenoma; CRC, colorectal cancer; EUB, Eubacteria; GMLL, *G. morbillorum*; PTST, *P. stomatis*; BCTF, *B. fragilis*; BCTT, *B. thetaiotaomicron*; RSBI, *R. intestinalis*; FPRA, *F. prausnitzii*; ECO, *E. coli*; PHGI, *F. prausnitzii* phylogroup I; PHGII, *F. prausnitzii* phylogroup II.

	Lynch	**Control**				**p-value (Mann-Whitney test)**			
Diagnostic	**NC**	**NC**	**NAA**	**AA**	**CRC**	**NC Lynch vs. NC Control**	**NC Lynch vs. NAA Control**	**NC Lynch vs. AA Control**	**NC Lynch vs. CRC Control**
n	41	136	63	79	14	NA	NA	NA	NA
EUB	9.49 ± 2.41	9.41 ± 0.43	9.43 ± 0.66	9.60 ± 2.02	9.35 ± 0.34	0.055	0.350	0.424	0.700
B10	14.42 ± 3.14	13.64 ± 2.34	12.92 ± 1.05	12.94 ± 1.80	13.22 ± 1.46	0.985	0.161	**0.039**	0.802
B46	21.69 ± 2.23	20.90 ± 1.92	20.16 ± 1.10	20.03 ± 1.27	20.25 ± 1.53	0.226	**<0.001**	0.438	0.129
B48	16.42 ± 1.32	16.43 ± 1.44	16.12 ± 1.42	16.20 ± 1.66	16.22 ± 1.11	0.998	0.335	0.051	0.975
GMLL	28.02 ± 1.88	30.47 ± 0.86	30.09 ± 1.39	30.20 ± 1.57	29.13 ± 2.42	**<0.001**	**<0.001**	**<0.001**	0.242
PTST	28.13 ± 1.51	29.11 ± 0.98	29.05 ± 0.77	29.11 ± 1.18	27.77 ± 3.01	**<0.001**	**<0.001**	**<0.001**	0.226
BCTF	29.49 ± 3.53	30.00 ± 3.58	28.92 ± 4.32	29.89 ± 3.91	27.38 ± 5.70	**0.012**	0.199	**<0.001**	1.000
BCTT	22.32 ± 3.10	22.30 ± 3.41	22.84 ± 4.25	23.09 ± 4.20	21.91 ± 2.59	0.999	1.000	0.990	0.999
RSBI	21.67 ± 3.62	20.01 ± 3.65	21.20 ± 3.35	21.56 ± 3.75	21.46 ± 3.52	0.982	0.988	0.999	0.999
FPRA	16.21 ± 1.32	15.54 ± 1.89	14.98 ± 1.12	14.79 ± 1.10	15.38 ± 1.55	0.998	0.335	0.051	0.993
ECO	21.83 ± 4.08	22.87 ± 4.95	24.03 ± 4.48	23.30 ± 3.65	23.83 ± 3.15	0.785	0.051	0.172	0.411
PHGI	22.26 ± 6.74	20.08 ± 5.53	19.39 ± 5.42	19.06 ± 5.11	18.27 ± 1.94	0.274	**0.047**	**0.012**	0.271
PHGII	19.84 ± 3.97	18.69 ± 3.89	17.74 ± 1.93	18.40 ± 4.28	18.60 ± 1.94	0.223	**0.012**	**0.02**	0.714

3.2 Study of the abundance of specific faeces bacterial markers in LS population

In terms of the LS gene mutation (*MLH1, MSH2, MSH6, PMS2)* no significant differences were observed in the abundance of the analysed bacterial markers between this group and neither CRC nor IBD-specific biomarkers, except for PTST, which showed higher relative abundance in *MLH1* mutation carriers (p=0.034) (Table 21).

Table 21. Comparison of the mean relative abundance (mean ± standard deviation) of the analysed bacterial markers in each of the mutation variants (MLH1, n=28; MSH2, n=14; MSH6, n=16; PMS2, n=8). Significant differences in the relative abundance of the bacterial markers obtained among the different MMR genes were obtained using the non-parametric test Kruskal-Wallis. Level of significance: * when p-value < 0.05. EUB, Eubacteria; GMLL, *G. morbillorum*; PTST, *P. stomatis*; BCTF, *B. fragilis*; BCTT, *B. thetaiotaomicron*; RSBI, *R. intestinalis*; FPRA, *F. prausnitzii*; ECO, *E. coli*; PHGI, *F. prausnitzii* phylogroup I; PHGII, *F. prausnitzii* phylogroup II.

	MLH1	***MSH2***	***MSH6***	***PMS2***	**p-value (Kruskal-Wallis test)**
EUB	9.87 ± 3.26	9.12 ± 0.88	9.14 ± 0.50	9.20 ± 0.39	0.915
B10	14.42 ± 3.36	14.27 ± 2.87	14.77 ± 3.84	13.85 ± 1.80	0.999
B46	21.71 ± 2.47	21.21 ± 1.80	21.89 ± 2.45	21.66 ± 2.41	0.860
B48	16.32 ± 1.22	16.52 ± 1.19	15.95 ± 1.48	16.88 ± 1.38	0.579
GMLL	28.22 ± 1.56	28.70 ± 1.82	27.96 ± 2.51	28.63 ± 1.48	0.835
PTST	27.67 ± 1.58	28.98 ± 0.65	28.37 ± 1.66	28.50 ± 0.88	**0.034***
BCTF	29.57 ± 3.08	29.30 ±3.96	29.71 ± 3.17	29.47 ± 4.15	0.709
BCTT	22.39 ± 2.69	22.46 ± 3.41	21.56 ± 2.83	21.99 ± 3.08	0.880
RSBI	21.42 ± 3.45	21.97 ± 4.30	21.56 ± 3.83	21.23 ± 2.56	0.995
FPRA	16.32 ± 3.35	15.84 ± 2.09	16.51 ± 3.53	15.70 ± 1.66	0.998
ECO	21.46 ± 3.79	22.47 ± 4.71	22.24 ± 5.24	19.77 ± 3.88	0.624
PHGI	20.80 ± 5.38	24.62 ± 8.44	24.85 ± 8.69	18.38 ± 2.89	0.156
PHGII	19.69 ± 4.58	19.98 ± 2.57	20.20 ± 5.36	19.89 ± 2.18	0.419

3.3 Definition of a bacterial signature for neoplasia detection in LS carriers

After grouping LS carriers, the relative abundance of the bacterial markers was compared among the different segments as follows: (1) NC and (2) patients with neoplasia (NAA+AA+CRC). We only used those subjects who did not have any personal CRC background (51 subjects) since it is the population in which we wanted to focus the

development of the non-invasive tool for LS carriers' surveillance. Four subjects were excluded because of poor sample condition (i.e. low DNA concentration or qPCR inhibition). In this preliminary analysis GMLL was the only bacterial marker that showed significant differences in its abundance when NC LS individuals were compared to LS with neoplasia (p=0.010). While GMLL sensitivity for neoplasia was high (90%), its specificity value was low (50%), (AUC=0.760, 95% CI (0.600-0.910)). Noteworthy, a sensitivity of 100% is sought since all LS with neoplastic lesions need to be detected. Although the sensitivity of GMLL for the detection of neoplastic lesions was high we designed an algorithm (RAID-LS) in order to improve the sensitivity. The best bacterial markers combination comprised EUB, B46 and ECO. More specifically, two ratios of bacterial markers (B46/EUB, ECO/EUB) were used. RAID-LS application led to high sensitivity and specificity values, 100% and 72%, respectively (LR(+)=3.55, 95% CI (2.15-5.85)). The positive predictive value and the negative predictive value were 42% and 100%, respectively, implying an accuracy of 77%.

4 Discussion

In this study, we have compared the abundance of specific bacterial markers associated to sporadic CRC and IBD patients with that of LS carriers. LS carriers show a CRC-like bacterial signature irrespective of the presence of lesions. We also present preliminary evidence suggesting that some specific bacterial marker may pinpoint those carriers that do not harbour lesions.

Comparison between healthy LS carriers and control subjects brought new findings. Significant differences in the abundance of three of the analysed bacterial markers were evidenced when healthy LS were compared to healthy controls, which suggests a distinct basal gut environment in LS carriers. Unexpectedly, LS carriers with a normal colonoscopy showed similar relative abundances of the analysed bacterial markers than that of control subjects with sporadic CRC.

MMR genes dysfunction in LS carriers leads to the loss of functional MMR proteins, which in turn leads to defects in DNA repair and, subsequently, high DNA microsatellite instability (MSI) [360]. Moreover, MMR gene mutations affect growth-regulatory genes that play an important role in CRC carcinogenesis such as *APC* or *TGFβ*

[76]. A proinflammatory microenvironment is found around colorectal tumour lesions as high levels of several cytokines and chemokines, such as CCL2 (MCP-I), CXCL1 (GROα) and CXCL5 (ENA-78), are secreted by tumours [361]. The appearance of a neoplastic lesion in LS carriers, both NAA and AA, prompts an increase in the number of bacterial markers that show significant differences on their relative abundance. These results were expected as the appearance of an adenoma in the gut of sporadic CRC patients has been seen to alter intestinal microbiota [362,363]. Some studies have proposed an association between pathogenic bacteria and sporadic colon tumorigenesis, and the depletion of protective bacteria have been also related to sporadic CRC pathology [364]. These results suggest that the inflammatory basal gut environment of LS carriers may mimic the gut environment of control subjects with tumour lesions, which results in gut microbial dysbiosis.

Among other observations we want to highlight that analysis of specific gut bacterial markers according to the affected gene revealed that only PTST seemed to be more abundant in subjects with the *MLH1* mutation which is the most frequent among LS carriers [78,354]. It will be of interest to validate whether this or other changes are validated when expanding the study. Also, our study shows that the long-term impact of surgery may not be that relevant in these patients. Several studies have reported that colorectal surgery leads to changes in the amount and diversity of gut bacteria [365–367]. In our study no differences in the abundance of the studied bacterial markers were observed between LS carriers with and without previous CRC likely linked to the fact that at least 5 years have elapsed since surgery at the time of sample collection allowing gut microbiota recover [368–370].

To date, a number of non-invasive methods directed to sporadic CRC screening have been developed. The most commonly used are the faecal immunochemical test (FIT), the guaiac faecal occult blood test (gFOBT), and Cologuard, among others [322,371,372]. Because of CRC development predisposition of LS, this population is exposed to undergo periodic endoscopic examinations [127,373,374]. Colonoscopic surveillance has demonstrated a significant reduction in CRC incidence and a considerable decrease in overall mortality in this at-risk population [375]. However, this

invasive procedure presents several inconveniences for patients such as the impact of bowel preparation, the use of sedation and the risk of intestinal perforation impacting compliance [214]. Moreover, the performance of colonoscopies is associated to high costs. In this work, a bacterial signature has been shown to be of potential use for predicting absence of neoplastic lesions in LS carriers and potential monitoring. Our observations point to the combination of three bacterial markers (EUB, B46, ECO) the so-called RAID-LS algorithm, as an informative tool. RAID-LS combines two bacterial markers abundance ratios: (B46/EUB, ECO/EUB). Low abundances of ECO and B46correlate with healthy LS carriers. *E. coli* is a proinflammatory bacterium, which is commonly found in patients with active ulcerative colitis and Crohn disease [376–378]. The appearance of a tumour lesion promotes a greater inflammatory environment in LS carriers, which can be associated to a higher abundance of ECO. As for B46, a butyrate producing bacterium usually found in high abundances in healthy guts [264,379], was less abundant in healthy LS carriers. Our results suggest that healthy LS gut microbiota is similar to that of sporadic CRC patients, thus B46 basal relative abundance in healthy LS carriers could be altered because of the proinflammatory environment of the gut. In addition, it seems that the appearance of a lesion could cause a modification in the abundance of this bacterial marker.

The main weakness of this study has to do with the number of enrolled subjects, as corresponds to a pilot study. However, all participants were well characterized in terms of their clinical features. Nevertheless, our results are encouraging enough to perform further validation on a higher number of participants. RAID-LS is able to identify LS carriers and may discriminate between healthy LS carriers and those with neoplastic lesions based on the relative abundance of the selected markers. However, the low prevalence of pre-malignant lesions in LS carriers precludes a fast recruitment.

In conclusion, the development of RAID-LS opens the door to the eventual non-invasive surveillance of LS carriers based on microbiota profiles. If confirmed intervals of colonoscopic surveillance could be tailored based on it. Under a health management optics, the use of RAID-LS could also result in a reduction of colonoscopy associated

costs, since up to a 60% of endoscopic examinations could be skipped. This must be corroborated after validation on a much larger and geographical diverse population.

5

General Discussion

Almost a century after Lockhart-Mummery and Dukes at St. Marks Hospital (London) found that colorectal cancers were associated with residual adenomatous tissue, which lead to the concept of early detection of curable cancers and cancer prevention through polypectomy [380]. The polyp-carcinoma sequence was challenged by many researchers for more than 70 years until it was finally proven in 1993 [165]. In 1948, Gilbertsen initiated the first CRC screening study enrolling 21,500 people who underwent rigid sigmoidoscopy [381]. Colorectal cancer screening changed substantially when in 1967 Gregor reported early-stage CRC by means of a new guaiac card test that could be prepared at home (gFOBT) [382]. However, it was not until 1973 when positive gFOBT patients could have an accurate diagnostic workup with the use of colonoscopes, which could also remove polyps adding a new and potentially huge preventive factor to CRC screening [383]. In early 90s a consensus appeared throughout the world literature that CRC screening is effective and should be offered to all people aged 50 and older who are at average CRC risk. The European Union Guidelines have recently advocated a 2-step approach, most commonly with FIT as the first step primarily for resource reasons [384].

Screening by means of FIT has shown advantages over gFOBT. FIT requires only one faecal sample instead of sampling from three bowel movements. Moreover, FIT tube collectors design facilitates sample collection and test handling, which increases adherence rates with higher detection rates of colorectal tumoral lesions [385]. FIT also has some disadvantages since its performance shows variability among different subgroups as it has been shown to have higher sensitivity for left-sided adenomas and its value has been seen to be higher for aspirin users compared to nonusers [386,387]. Although FIT is a quantitative technique, it has been mainly used using a given cut-off, thus limiting the use of the test to a qualitative result (positive or negative). High cut-off values lead to high positive predictive values and are useful when resources are limited. However, sensitivity for advanced adenomas decreases as the cut-off value increases. Conversely, lower cut-off values increase sensitivity for the detection of patients with advanced neoplasms, which results in higher needs for endoscopic resources.

Current research is focused on the development of new non-invasive tools for CRC screening capable of reducing FIT false-positive results. Newly developed tests include DNA, RNA and protein biomarker stool and blood analysis. Recent studies based on DNA sequences have shown an abnormal gut microbiome structure in CRC patients when compared to healthy patients [260,264,266,388]. Therefore, faecal microbiota may be considered as a feasible complement to FIT. In this work, we hypothesized that using novel bacterial biomarkers of CRC in combination with FIT could improve the ability to fine-tuning the eligibility of candidates for colonoscopy. Most research done to date fail to obtain high accuracy values for the detection of precancerous lesions [182,184]. Faecal microbial composition differs from that observed in intestinal mucosal samples, due to environmental factors such as diet, bile acids and host's immune system. Bacterial markers analysed in this thesis were originally studied on colorectal mucosal samples and their detection was later optimized in faeces. Thus, their presence in stool is not strongly subjected to diet and other external factors variability but in some extent reflects the real abundance in the colonic mucosa. This step-wise methodology allows to by-pass the huge background noise present in faecal samples, while giving physiological meaningfulness to bacterial markers. Another factor that has been a matter of controversy and that could impact the reproducibility of faecal bacterial signatures is the geographical or cultural origin of sample donors. This issue has been recently assessed by Thomas *et al*, who performed a meta-analysis with seven cohorts from different origins (USA, Germany, France, Italy, China, Canada, and Austria) and were able to identify reproducible microbiome biomarkers in patients with CRC.

Based on this strategy, we have developed three different faecal bacterial signatures capable of discriminating between healthy subjects and patients with advanced neoplasia. The first signature was aimed at individuals with symptomatology compatible with colorectal cancer (RAID-CRC); the second was directed to those asymptomatic subjects older than 50 years that obtained a FIT-positive result in a CRC-screening program (RAID-CRC Screen); and the third was aimed at Lynch syndrome carriers who underwent a surveillance colonoscopy (RAID-LS). These populations were treated separately since CRC development is different in each of them. Thus, we hypothesized that gut microbiota changes should be different when a tumoral lesion

appeared in each group of subjects. Indeed, all the studied groups showed different microbiological patterns according to colonoscopy diagnosis (i.e. normal colonoscopy, non-advanced adenoma, advanced adenoma, and CRC), which indicates that the appearance of the tumoral lesion causes shifts in gut microbiota composition. When symptomatic and asymptomatic subjects with CRC were compared, both displayed higher abundances of opportunistic pathogenic species (i.e. *Gemella morbillorum*, *Peptostreptococcus stomatis*, *Bacteroides fragilis*), but the microbiota profile of advanced adenomas between these populations was clearly different. Since our aim was to distinguish between healthy subjects and those with advanced neoplasia (CRC and/or advanced adenomas) we decided to develop an algorithm for each group of subjects, which best fitted each gut microbiota profile. The development of the microbiological signature for discriminating between healthy LS carriers and those with neoplastic lesions was performed separately, since the latter show genetic predisposition to develop CRC and they present an inflammatory colonic environment.

Despite significant progress has been made in the last years in CRC prevention and diagnostic, through the implementation of organised screening programs in developed countries, there is still much to do. The faecal bacterial signatures defined in this thesis are tools with a potential successful use in the public healthcare system. Nevertheless, two of them (RAID-CRC and RAID-LS) still seek clinical validation in larger external cohorts. Moreover, technical work is yet to be done to apply these tools to the entire population, in order to cope with massive samples processing.

Other non-invasive tools for CRC screening have been or are being developed by other research groups or by private companies. Recently, cross-cohort microbial diagnostic signatures have been identified in a metagenomic analysis of CRC datasets [388]. In this study, the authors identified in the combined analysis of heterogeneous CRC cohorts reproducible faecal microbiome biomarkers and accurate disease-predictive models that could constitute the basis for clinical prognostic tests. However, they state that microbiome signatures for adenoma detection are only partially predictive. Most private companies are developing tools based on blood biomarkers. The German company Epigenomics AG developed Epi proColon® 2.0, test that comprises a

qualitative assay for the PCR detection of methylated *Septin9* DNA, the presence of which is associated with CRC [389]. Despite it is already in the market, its clinical utility has not yet been examined in prospective studies and the effect of *Septin9*-based screening on the incidence of CRC and CRC-associated mortality is unknown. VolitionRx, a Belgian *in vitro* diagnosis company, has also developed a blood-based approach for CRC early detection called Nu-Q [390]. This test is based on the combination of different cell-free nucleosomes found in serum to identify patients at risk of colorectal cancer [391]. Although Nu-Q shows high sensitivity and specificity values, 75% and 90%, respectively; no data is provided regarding advanced adenoma detection, either. In addition, sample size used in their study is small, with only 58 symptomatic patients enrolled. Other *in vitro* diagnostic companies that are currently undergoing validation of their CRC detection tests are the Spanish UniversalDx (UDX test) and Amadix (ColoFast test) [392,393]. UDX test is based on the detection of specific blood metabolites associated with CRC, while ColoFast is focused on the detection of miRNA. When compared to the above-mentioned tests, the main advantage of RAID-CRC is that it offers high diagnostic accuracy not only for CRC, but also for advanced adenoma detection.

Several reasons have led us to use faecal bacterial signatures for *in vitro* diagnostic purposes. In recent years, gut microbiota has been extensively characterized in patients from different clinical groups: healthy, adenoma, and carcinoma. Analysis of intestinal bacteria from stool samples has revealed both an enrichment and depletion of several microbiological populations associated with adenomas and CRC. Moreover, some bacterial species have been identified to be significantly more abundant in the presence of a tumoral lesion [260,264,267,388,394]. These studies have opened a range of possibilities to develop non-invasive tools for CRC diagnosis. Whether CRC-associated bacterial species abundance changes as a consequence of the intestinal mucosa inflammation, or if these species are responsible for prompting tumour generation, is yet to be elucidated. Nonetheless, gut microbiota is postulated as CRC indicator since bacterial communities are extremely sensitive to physicochemical changes that occur in their habitat. Considering that the disease most often causes these changes, the rationale

for the use of bacteria as disease indicator is more than justified by the Microbial Ecology foundations and laws.

The use of stool samples for the determination of different bacterial signatures facilitates the participation of patients, since samples can be obtained on a daily basis and their collection is easy, it can be performed at home using a sterile faeces tube collector. Furthermore, there exists a great variety of kits for DNA extraction from faecal samples. Since different types of DNA are extracted with these extraction kits, qPCR systems that target specific regions of 16S rRNA genes or genomic DNA from the desired bacterial species must be designed. Approaches based on qPCR combine the high selectivity of PCR with quantification by recording the amplification of a PCR product. Quantification of the target gene is determined during the early exponential phase of the PCR, which avoids problems associated with the so-called "end-point" PCR, in which PCR products are only analysed once the final PCR cycle has been completed [395]. Therefore, qPCR represents a fast, sensitive, selective, robust, reproducible and reliable method enabling the quantification of bacterial markers, particularly useful in complex samples such as faeces.

Finally, findings reported in this doctoral thesis should encourage to find similar signatures in other diseases, since changes in gut microbiota have been reported in many disorders such as inflammatory bowel disease [377], gingivitis and other periodontal affectations [396], or rheumatoid arthritis [397]. All things considered, the use of bacterial markers as disease indicators is more than promising, and a bright future in the contribution of microbiota to *in vitro* clinical diagnosis is foreseen.

6

Conclusions

1. Subjects with symptomatology compatible with colorectal cancer, FIT-positive asymptomatic individuals and Lynch syndrome carriers show distinctive intestinal microbiota changes during colorectal neoplasia progress.

2. The opportunistic pathogenic species *Gemella morbillorum*, *Peptostreptococcus stomatis*, and *Bacteroides fragilis*, show a higher relative abundance in faecal samples of patients with colorectal cancer.

3. Faecal bacterial markers are good complements of the faecal immunochemical test (FIT) for early detection of cancer, since they add specificity to this test. Consequently, the number of FIT false positive results is reduced, which implies a decrease in the unnecessary colonoscopies performed in a CRC-screening scenario.

4. Faecal bacterial markers increase sensitivity for advanced adenomas, which is one of the main objectives of organised CRC-screening programs.

5. Main gut microbiota indicators of healthy Lynch syndrome carriers resemble that of sporadic CRC, which indicates that this population has an altered basal intestinal environment. Faecal bacterial markers are postulated as good indicators for discriminating between healthy Lynch syndrome carriers and those with neoplastic lesions waiting for clinical validation.

6. RAID-CRC, RAID-CRC Screen and RAID-LS are promising, new, non-invasive tools for early detection of CRC with the potential to improve CRC-related organised programs management.

7

References

1 World Health Organization. GLOBOCAN 2018. 2018. http://gco.iarc.fr/today (accessed 27 Feb 2019).

2 Center MM, Jemal A, Smith RA, *et al.* Worldwide variations in colorectal cancer. *Cancer J Clin* 2009;**59**:366–78. doi:10.1007/DCR.0b013e3181d60a51

3 Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers Prev* 2009;**18**:1688–94. doi:10.1158/1055-9965.EPI-09-0090

4 Feng Q, Liang S, Jia H, *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun* 2015;**6**:6528. doi:10.1038/ncomms7528

5 Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 2007;**50**:113–30. doi:10.1111/j.1365-2559.2006.02549.x

6 Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2014;**383**:1490–502. doi:10.1016/S0140-6736(13)61649-9

7 Elmunzer BJ, Hayward RA, Schoenfeld PS, *et al.* Effect of Flexible Sigmoidoscopy-Based Screening on Incidence and Mortality of Colorectal Cancer: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PLOS Med* 2012;**9**:e1001352. doi:10.1371/journal.pmed.1001352

8 Hewitson P, Glasziou P, Watson E, *et al.* Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (Hemoccult): An update. *Am J Gastroenterol* 2008;**103**:1541–9. doi:10.1111/j.1572-0241.2008.01875.x

9 Zauber AG. The Impact of Screening on Colorectal Cancer Mortality and Incidence – Has It Really Made a Difference? *Dig Dis Sci* 2015;**60**:681–91. doi:10.1007/s10620-015-3600-5

10 Doubeni CA, Laiyemo AO, Major JM, *et al.* Socioeconomic status and the risk of colorectal cancer: An analysis of over one-half million adults in the NIH-AARP Diet and Health Study. *Cancer* 2012;**118**:3636–44. doi:10.1002/cncr.26677

11 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;**68**:7–30. doi:10.3322/caac.21442

12 Davis DM, Marcet JE, Frattini JC, *et al.* Is it time to lower the recommended screening age for colorectal cancer? *J Am Coll Surg* 2011;**213**:352–61. doi:10.1016/j.jamcollsurg.2011.04.033

13 Singh KE, Taylor TH, Pan C-JG, *et al.* Colorectal Cancer Incidence Among Young Adults in California. *J Adolesc Young Adult Oncol* 2014;**3**:176–84. doi:10.1089/jayao.2014.0006

14 Tawadros PS, Paquette IM, Hanly AM, *et al.* Adenocarcinoma of the rectum in patients under age 40 is increasing: impact of signet-ring cell histology. *Dis Colon Rectum* 2015;**58**:474–8. doi:10.1097/DCR.0000000000000318

15 American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019. 2017.

16 Murphy G, Devesa SS, Cross AJ, *et al.* Sex Disparities in Colorectal Cancer Incidence by Anatomic Subsite, Race and Age. *Int J Cancer Int J Cancer April* 2011;**1**:1668–75. doi:10.1002/ijc.25481

17 Schoenfeld P, Cash B, Flood A, *et al.* Colonoscopic Screening of Average-Risk Women for Colorectal Neoplasia. *N Engl J Med* 2005;**352**:2061–8. doi:10.1056/NEJMoa042990

18 Semega JL, Fontenot KR, Kollar MA. Income and Poverty in the United States: 2016. 2017.

19 National Cancer Institute. SEER*Stat Databases: November 2015 Submission. https://seer.cancer.gov/data-software/documentation/seerstat/nov2015/ (accessed 9 May 2018).

20 Perdue DG, Haverkamp D, Perkins C, *et al.* Geographic variation in colorectal cancer incidence and mortality, age of onset, and stage at diagnosis among American Indian and Alaska native people, 1990-2009. *Am J Public Health* 2014;**104**:S404–14. doi:10.2105/AJPH.2013.301654

21 Kelly JJ, Alberts SR, Sacco F, *et al.* Colorectal Cancer in Alaska Native People, 2005-

2009. *Gastrointest Cancer Res* 2012;**5**:149–55.

22 Ft Fearon E, Vogelstein B. A Genetic Model for Colorectal Tumorigenesis. *Cell* 1990;**61**:759–67.

23 Castellsagué Torrents E. *Càncer colorectal hereditari: aplicacions diagnòstiques de l'estudi de la dosi dels gens APC, MLH1 i MSH2 (Doctoral Thesis)*. 2009.

24 Nava HR, Pagana TJ. Postoperative surveillance of colorectal carcinoma. *Cancer* 1982;**49**:1043–7. doi:10.1002/1097-0142(19820301)49:5<1043::AID-CNCR2820490533>3.0.CO;2-Y

25 Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med* 1992;**326**:658–62.

26 Tuohy TMF, Rowe KG, Mineau GP, *et al.* Risk of colorectal cancer and adenomas in the families of patients with adenomas: a population-based study in Utah. *Cancer* 2014;**120**:35–42. doi:10.1002/cncr.28227

27 Taylor DP, Stoddard GJ, Burt RW, *et al.* How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet Med* 2011;**13**:385–91. doi:10.1097/GIM.0b013e3182064384

28 Taylor DP, Burt RW, Williams MS, *et al.* Population-based family-history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;**138**:877–85. doi:10.1053/j.gastro.2009.11.044

29 Rutter MD. Surveillance programmes for neoplasia in colitis. *J Gastroenterol* 2011;**46**:1–5. doi:10.1007/s00535-010-0309-2

30 Ekbom A, Helmick C, Zack M, *et al.* Ulcerative Colitis and Colorectal Cancer. A Population-Based Study. *N Engl J Med* 1990;**323**:1228–33.

31 Gyde S, Prior P, Dew MJ, *et al.* Mortality in ulcerative colitis. *Gastroenterology* 1982;**83**:36–43.

32 Ekbom A, Helmick C, Zack M, *et al.* Increased risk of large-bowel cancer in Crohn's disease with colonic involvement. *Lancet* 1990;**336**:357–9.

doi:10.1016/0140-6736(90)91889-I

33 Rutter M, Saunders B, Wilkinson K, *et al.* Severity of Inflammation Is a Risk Factor for Colorectal Neoplasia in Ulcerative Colitis. *Gastroenterology* 2004;**126**:451–9. doi:10.1053/j.gastro.2003.11.010

34 Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001;**48**:526–35.

35 Viennois E, Chen F, Merlin D. NF-κB pathway in colitis-associated cancers. *Transl Gastrointest Cancer* 2013;**2**:21–9. doi:10.3978/j.issn.2224-4778.2012.11.01

36 Slattery ML, Boucher KM, Caan BJ, *et al.* Eating patterns and risk of colon cancer. *Am J Epidemiol* 1998;**148**:4–16.

37 Terry P, Giovannucci E, Michels KB, *et al.* Fruit, vegetables, dietary fiber, and risk of colorectal cancer. *J Natl Cancer Inst* 2001;**93**:525–33.

38 Kim YI, Mason JB. Nutrition chemoprevention of gastrointestinal cancers: a critical review. *Nutr Rev* 1996;**54**:259–79.

39 Orlich MJ, Singh PN, Sabaté J, *et al.* Vegetarian Dietary Patterns and the Risk of Colorectal Cancers. *JAMA* 2015;**175**:767–76. doi:10.1001/jamainternmed.2015.59

40 Larsson S, Giovannucci E, Bergkvist L, *et al.* Whole grain consumption and risk of colorectal cancer: a population-based cohort of 60.000 women. *Br J Cancer* 2005;**92**:1803–7. doi:10.1038/sj.bjc.6602543

41 Kunzmann AT, Coleman HG, Huang W-Y, *et al.* Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial 1,2. *Am J Clin Nutr* 2015;**102**:881–90. doi:10.3945/ajcn.115.113282

42 Bingham S, Day NE, Luben R, *et al.* Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* 2003;**361**:1496–501. doi:10.1016/S0140-6736(03)13174-1

43 Peters U, Sinha R, Chatterjee N, *et al.* Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. *Lancet* 2003;**361**:1491–5. doi:10.1016/S0140-6736(03)13173-X

44 Lee JE, Willett WC, Fuchs CS, *et al.* Folate intake and risk of colorectal cancer and adenoma: modification by time 1–4. *Am J Clin Nutr* 2011;**93**:817–25. doi:10.3945/ajcn.110.007781

45 Michels KB, Edward Giovannucci, Joshipura KJ, *et al.* Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. *J Natl Cancer Inst* 2000;**92**:1740–52.

46 Lamprecht SA, Lipkin M. Chemoprevention of colon cancer by calcium, vitamin D and folate: Molecular mechanisms. *Nat Rev Cancer* 2003;**3**:601–14. doi:10.1038/nrc1144

47 Flood A, Peters U, Chatterjee N, *et al.* Calcium from Diet and Supplements is Associated With Reduced Risk of Colorectal Cancer in a Prospective Cohort of Women. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:126–32.

48 Chao A, Connell CJ, Mccullough ML, *et al.* Meat Consumption and Risk of Colorectal Cancer. *JAMA* 2016;**293**:172–82.

49 Norat T, Bingham S, Ferrari P, *et al.* Meat, fish, and colorectal cancer risk: The European prospective investigation into cancer and nutrition. *J Natl Cancer Inst* 2005;**97**:906–16. doi:10.1093/jnci/dji164

50 Uptodate. Colorectal cancer: Epidemiology, risk factors, and protective factors. 2018.https://www.uptodate.com/contents/colorectal-cancer-epidemiology-risk-factors-and-protective-factors#H90630250 (accessed 18 May 2018).

51 Martinez-Useros J, Garcia-Foncillas J. Obesity and colorectal cancer: molecular features of adipose tissue. *J Transl Med* 2016;**14**. doi:10.1186/s12967-016-0772-5

52 Hursting SD, Dunlap SM. Obesity, metabolic dysregulation, and cancer: a growing concern and an inflammatory (and microenvironmental) issue. *Ann N Y Acad Sci* 2012;**1271**:82–7. doi:10.1111/j.1749-6632.2012.06737.x

53 Yehuda-Shnaidman E, Schwartz B. Mechanisms linking obesity, inflammation and altered metabolism to colon carcinogenesis. *Obes Rev* 2012;**13**:1083–95. doi:10.1111/j.1467-789X.2012.01024.x

54 Hwang W, Yang M, Tsai M, *et al.* SNAIL regulates interleukin-8 expression, stem celllike activity, and tumorigenicity of human colorectal carcinoma cells. *Gastroenterology* 2011;**141**:279–91. doi:10.1053/j.gastro.2011.04.008

55 Boyle T, Keegel T, Bull F, *et al.* Physical Activity and Risks of Proximal and Distal Colon Cancers: A Systematic Review and Meta-analysis. *J Natl Cancer Inst* 2012;**104**:1548–61. doi:10.1093/jnci/djs354

56 Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet* 2003;**362**:847–52.

57 Gandini S, Botteri E, Iodice S, *et al.* Tobacco smoking and cancer: A meta-analysis. *Int J Cancer* 2008;**122**:155–64. doi:10.1002/ijc.23033

58 Botteri E, Iodice S, Bagnardi V, *et al.* Smoking and Colorectal Cancer. *JAMA* 2008;**300**:2765–78. doi:10.1001/jama.2008.839

59 Fedirko V, Tramacere I, Bagnardi V, *et al.* Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol* 2011;**22**:1958–72. doi:10.1093/annonc/mdq653

60 Cho E, Smith-Warner SA, Ritz J, *et al.* Alcohol Intake and Colorectal Cancer: A Pooled Analysis of 8 Cohort Studies. *Ann Intern Med* 2004;**140**:603. doi:10.7326/0003-4819-140-8-200404200-00007

61 Mizoue T, Inoue M, Wakai K, *et al.* Alcohol Drinking and Colorectal Cancer in Japanese: A Pooled Analysis of Results from Five Cohort Studies. *Am J Epidemiol* 2008;**167**:1397–406. doi:10.1093/aje/kwn073

62 LoConte NK, Brewster AM, Kaur JS, *et al.* Alcohol and cancer: A statement of the American society of clinical oncology. *J Clin Oncol* 2018;**36**:83–93. doi:10.1200/JCO.2017.76.1155

63 Näthke IS. The adenomatous polyposis coli protein: the achilles heel of the gut

epithelium. *Annu Rev Cell Dev Biol* 2004;**20**:337–66. doi:10.1146/annurev.cellbio.20.012103.094541

64 Paxton S, Peckham M, Adele K, *et al.* The Leeds Histology Guide. 2003.https://www.histology.leeds.ac.uk/oral/GI_layers.php (accessed 28 Jul 2018).

65 Kim YS, Ho SB. Intestinal Goblet Cells and Mucins in Health and Disease: Recent Insights and Progress. *Curr Gastroenterol Rep* 2010;**12**:319–30. doi:10.1007/s11894-010-0131-2

66 Paulus U, Loeffler M, Zeidler J, *et al.* The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies. *J Cell Sci* 1993;**106**:473–84.

67 Brittan M, A. Wright N. Gastrointestinal stem cells. *J Pathol* 2002;**197**:492–509. doi:10.1002/path.1155

68 Ayabe T, Satchell DP, Wilson CL, *et al.* Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol* 2000;**1**:113–8. doi:10.1038/77783

69 Heath JP. Epithelial Cell Migration in the Intestine. *Cell Biol Int* 1996;**20**:139–46. doi:10.1006/cbir.1996.0018

70 Radtke F, Clevers H. Self-renewal and cancer of the gut: Two sides of a coin. *Science (80-)* 2005;**307**:1904–9. doi:10.1126/science.1104815

71 Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. *Gastroenterology* 2010;**138**:2059–72. doi:10.1053/j.gastro.2009.12.065

72 Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998;**396**:643–9. doi:10.1038/25292

73 Boland CR, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 2010;**138**:2073–87. doi:10.1053/j.gastro.2009.12.064

74 Markowitz SD, Bertagnolli MM. Molecular basis of colorectal cancer. *N Engl J Med*

2010;**362**:1245–7. doi:362/13/1245 [pii] 10.1056/NEJMc1000949

75 Chang K, Taggart MW, Reyes-Uribe L, *et al.* Immune Profiling of Premalignant Lesions in Patients With Lynch Syndrome. *JAMA Oncol* 2018;**4**:1085–92. doi:10.1001/jamaoncol.2018.1482

76 Walther A, Johnstone E, Swanton C, *et al.* Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 2009;**9**:489–99. doi:10.1038/nrc2645

77 Aoki K, M. Taketo M. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci* 2007;**120**:3327–35. doi:10.1242/jcs.03485

78 Ewing I, Bogaert J, Prenen H. The molecular genetics of colorectal cancer. *Frontline Gastroenterol* 2014;**5**:26–30. doi:10.1136/flgastro-2013-100329

79 Fodde R, Edelmann W, Yang K, *et al.* A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Genetics* 1994;**91**:8969–73.

80 Oshima M, Oshima H, Kitagawa K, *et al.* Loss of APC heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated APC gene. *Med Sci* 1995;**92**:4482–6.

81 Ishikawa T-O, Tamai Y, Li Q, *et al.* Requirement for tumor suppressor APC in the morphogenesis of anterior and ventral mouse embryo. *Dev Biol* 2003;**253**:230–46.

82 Andreu P. Crypt-restricted proliferation and commitment to the Paneth cell lineage following APC loss in the mouse intestine. *Dev Dis* 2005;**132**:1443–51. doi:10.1242/dev.01700

83 Molenaar M, Van De Wetering M. XTcf-3 Transcription Factor Mediates-Catenin-Induced Axis Formation in Xenopus Embryos. *Cell* 1996;**86**:391–9.

84 Armaghany T, Wilson JD, Chu Q, *et al.* Genetic Alterations in Colorectal Cancer. *Gastrointest Cancer Res Cancer Res* 2012;**5**:19–27.

85 Mccormick F. KRAS as a Therapeutic Target. *Clin Cancer Res* 2015;**21**:1797–801. doi:10.1158/1078-0432.CCR-14-2662

86 Porru M, Pompili L, Caruso C, *et al.* Targeting KRAS in metastatic colorectal cancer: current strategies and emerging opportunities. *J Exp Clin Cancer Res* 2018;**37**:57. doi:10.1186/s13046-018-0719-1

87 Sanz-Garcia E, Argiles G, Elez E, *et al.* BRAF mutant colorectal cancer: Prognosis, treatment, and new perspectives. *Ann Oncol* 2017;**28**:2648–57. doi:10.1093/annonc/mdx401

88 Samuels Y, Wang Z, Bardelli A, *et al.* High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science (80-)* 2004;**304**:554. doi:10.1126/science.1096502

89 Liao X, Lochhead P, Nishihara R, *et al.* Aspirin Use, Tumor PIK3CA Mutation, and Colorectal-Cancer Survival. *N Engl J Med* 2012;**17**:1596–606. doi:10.1056/NEJMoa1207756

90 Mehrvarz A, Advani S, Overman MJ, *et al.* Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. *PLoS One* 2017;**12**:e0173345. doi:10.1371/journal.pone.0173345

91 Derynck R, Zhang Y, Feng X-H. Smads: Transcriptional Minireview Activators of TGF-beta Responses. *Cell* 1998;**95**:737–40.

92 Alazzouzi H, Alhopuro P, Salovaara R, *et al.* SMAD4 As a Prognostic Marker in Colorectal Cancer. *Clin Cancer Res* 2005;**11**:2606–11.

93 Miyaki M, Iijima T, Konishi M, *et al.* Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* 1999;**18**:3098–103.

94 Lane DP, Benchimol S. p53: Oncogene or anti-oncogene? *Genes Dev* 1990;**4**:1–8. doi:10.1101/gad.4.1.1

95 Vogelstein B, R. Fearon E, R. Hamilton S, *et al.* Genetic Alterations during Colorectal - Tumor Development. *N Engl J Med* 1988;**319**:525–32. doi:10.1056/NEJM198809013190901

96 Holliday R. The Inheritance of Epigenetic Defects. *Science (80-)* 1987;**238**:163–70.

97 Lao VV, Grady WM. Epigenetics and Colorectal Cancer. *Nat Rev Gastroenterol*

Hepatol 2012;**8**:686–700. doi:10.1038/nrgastro.2011.173

98 Irizarry RA, Ladd-Acosta C, Wen B, *et al.* Genome-wide methylation analysis of human colon cancer reveals similar hypo-and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009;**41**:178–86. doi:10.1038/ng.298

99 Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;**301**:89–92. doi:10.1038/301089a0

100 Feinberg AP. The epigenetics of cancer etiology. *Semin Cancer Biol* 2004;**14**:427–32. doi:10.1016/j.semcancer.2004.06.005

101 Ehrlich M. DNA methylation in cancer: too much, but also too little. *Oncogene* 2002;**21**:5400–13. doi:10.1038/sj.onc

102 Suzuki K, Suzuki I, Leodolter A, *et al.* Global DNA demethylation in gastrointestinal cancer is age dependent and precedes genomic damage. *Cancer Cell* 2006;**9**:199–207. doi:10.1016/j.ccr.2006.02.016

103 Pfeifer GP, Rauch TA. DNA methylation patterns in lung carcinomas. *Semin Cancer Biol* 2009;**19**:181–7. doi:10.1016/j.semcancer.2009.02.008

104 Costello JF, Frühwald MC, Smiraglia DJ, *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 2000;**24**:132–8. doi:10.1038/72785

105 Toyota M, Issa J-PJ. CpG island methylator phenotypes in aging and cancer. *Semin Cancer Biol* 1999;**9**:349–57. doi:10.1006/scbi.1999.0135

106 Weisenberger DJ, Trinh BN, Campan M, *et al.* DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res* 2008;**36**:4689–98. doi:10.1093/nar/gkn455

107 Verma M, Srivastava S. Review Epigenetics in cancer : implications for early Epigenetics and cancer. *Lancet* 2002;**3**:755–63. doi:10.1016/S1470-2045(02)00932-4

108 Shih I-M, Wang T-L, Traverso G, *et al.* Top-down morphogenesis of colorectal

tumors. *PNAS* 2001;**98**:2640–5.

109 L. Preston S, Wong W-M, On-On A, *et al.* Bottom-up Histogenesis of Colorectal Adenomas: Origin in the Monocryptal Adenoma and Initial Expansion by Crypt Fission. *Cancer Res* 2003;**63**:3819–25.

110 National Cancer Institute. Colorectal Cancer. 2018.https://www.cancer.gov/types/colorectal (accessed 7 Aug 2018).

111 Takayama T, Katsuki S, Takahashi Y, *et al.* Aberrant crypt foci of the colon as precursors of adenoma and cancer. *N Engl J Med* 1998;**339**:1277–84.

112 Riihimäki M, Hemminki A, Sundquist J, *et al.* Patterns of metastasis in colon and rectal cancer. *Sci Rep* 2016;**6**:29765. doi:10.1038/srep29765

113 Burt RW. Colon Cancer Screening. *Gastroenterology* 2000;**119**:837–53. doi:10.1053/gast.2000.16508

114 Winawer SJ, Fletcher RH, Miller L, *et al.* Colorectal Cancer Screening: Clinical Guidelines and Rationale The Adenoma-Carcinoma Sequence. *Gastroenterology* 1997;**112**:594–642.

115 Yamagishi H, Kuroda H, Imai Y, *et al.* Molecular pathogenesis of sporadic colorectal cancers. *Chin J Cancer* 2016;**35**:4. doi:10.1186/s40880-015-0066-y

116 White RL. Tumor suppressing pathways. *Cell* 1998;**92**:591–2. doi:10.1016/S0092-8674(00)81124-1

117 Snover DC. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol* 2011;**42**:1–10. doi:10.1016/j.humpath.2010.06.002

118 Bettington M, Walker N, Clouston A, *et al.* The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* 2013;**62**:367–86. doi:10.1111/his.12055

119 Armelao F, de Pretis G. Familial colorectal cancer: A review. *World J Gastroenterol* 2014;**20**:9292–8.

120 Davila RE, Rajan E, Baron TH. ASGE guideline: colorectal cancer screening and

surveillance. *Gastrointest Endosc* 2006;**63**:546–57. doi:10.1016/j.gie.2006.02.002

121 Levin B, Lieberman DA, McFarland B, *et al.* Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;**58**:130–60. doi:10.3322/CA.2007.0018

122 Rex DK, Johnson DA, Anderson JC, *et al.* American College of Gastroenterology Guidelines for Colorectal Cancer Screening 2008. *Am J Gastroenterol* 2009;**104**:739–50. doi:10.1038/ajg.2009.104

123 Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;**96**:2992–3003. doi:10.1111/j.1572-0241.2001.04677.x

124 Jasperson KW, Tuohy TM, Neklason DW, *et al.* Hereditary and Familial Colon Cancer. *Gastroenterology* 2010;**138**:2044–58. doi:10.1053/j.gastro.2010.01.054

125 Ma H, Brosens LA, Johan G, *et al.* Pathology and genetics of hereditary colorectal cancer. *Pathology* 2018;**50**:49–59. doi:10.1016/j.pathol.2017.09.004

126 Hampel H, Frankel WL, Martin E, *et al.* Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer. *J Clin Oncol* 2008;**26**:5783–8. doi:10.1200/JCO.2008.17.5950

127 Giardiello FM, Allen JI, Axilbund JE, *et al.* Guidelines on Genetic Evaluation and Management of Lynch Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2014;**147**:502–26. doi:10.1053/j.gastro.2014.04.001

128 Brosens LA, Offerhaus JA, Giardiello FM, *et al.* Hereditary Colorectal Cancer: Genetics and Screening. *Surg Clin North Am* 2015;**95**:1067–80. doi:10.1016/j.suc.2015.05.004

129 Møller P, Seppälä TT, Bernstein I, *et al.* Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective

Lynch Syndrome Database. *Gut* 2018;**67**:1306–16. doi:10.1136/gutjnl-2017-314057

130 Rustgi AK. The genetics of hereditary colon cancer. *Genes Dev* 2007;**21**:2525–38. doi:10.1101/gad.1593107

131 Lynch HT, Lynch PM, Lanspa SJ, *et al.* Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 2009;**76**:1–18. doi:10.1111/j.1399-0004.2009.01230.x

132 Engel C, Rahner N, Schulmann K, *et al.* Efficacy of Annual Colonoscopic Surveillance in Individuals With Hereditary Nonpolyposis Colorectal Cancer. *Clin Gastroenterol Hepatol* 2010;**8**:174–82. doi:10.1016/j.cgh.2009.10.003

133 Mulvihill JJ. The frequency of hereditary large bowel cancer. *Prog Clin Biol Res* 1983;**115**:61–75.

134 Kinzler KW, Vogelstein B. Lessons from Hereditary Review Colorectal Cancer. *Cell* 1996;**87**:159–70.

135 Grover S, Kastrinos F, Steyerberg EW, *et al.* Prevalence and Phenotypes of APC and MUTYH Mutations in Patients with Multiple Colorectal Adenomas. *JAMA* 2012;**308**:485–92. doi:10.1001/jama.2012.8780

136 Al-Tassan N, Chmiel NH, Maynard J, *et al.* Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 2002;**30**:227–32. doi:10.1038/ng828

137 Landon M, Ceulemans S, Saraiya DS, *et al.* Analysis of current testing practices for biallelic MUTYH mutations in MUTYH-associated polyposis. *Clin Genet* 2015;**87**:368–72. doi:10.1111/cge.12375

138 Guarinos C, Juárez M, Egoavil C, *et al.* Prevalence and characteristics of MUTYH-associated polyposis in patients with multiple adenomatous and serrated polyps. *Clin Cancer Res* 2014;**20**:1158–68. doi:10.1158/1078-0432.CCR-13-1490

139 Castillejo A, Vargas G, Castillejo MI, *et al.* Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. *Eur J Cancer* 2014;**50**:2241–50. doi:10.1016/j.ejca.2014.05.022

140 Kopacova M, Tacheci I, Rejchrt S, *et al.* Peutz-Jeghers syndrome: Diagnostic and therapeutic approach. *World J Gastroenterol* 2009;**15**:5397–408. doi:10.3748/wjg.15.5397

141 Chen HM, Fang JY. Genetics of the hamartomatous polyposis syndromes: A molecular review. *Int J Colorectal Dis* 2009;**24**:865–74. doi:10.1007/s00384-009-0714-2

142 Gammon A, Jasperson K, Kohlmann W, *et al.* Hamartomatous polyposis syndromes. *Best Pract Res Clin Gastroenterol* 20109;**23**:219–31. doi:10.1016/j.bpg.2009.02.007.HAMARTOMATOUS

143 Brosens LA, Langeveld D, Arnout van Hattem W, *et al.* Juvenile polyposis syndrome. *World J Gastroenterol* 2011;**17**:4839–44. doi:10.3748/wjg.v17.i44.4839

144 Howe JR, Mitros FA, Summers RW. The risk of gastrointestinal carcinoma in familial juvenile polyposis. *Ann Surg Oncol* 1998;**5**:751–6.

145 Schreibman IR, Baker M, Amos C, *et al.* The hamartomatous polyposis syndromes: A clinical and molecular review. *Am J Gastroenterol* 2005;**100**:476–90. doi:10.1111/j.1572-0241.2005.40237.x

146 Van Hattem WA, Brosens LA, De Leng WWJ, *et al.* Large genomic deletions of SMAD4, BMPR1A and PTEN in juvenile polyposis. *Gut* 2008;**57**:623–7. doi:10.1136/gut.2007.142927

147 Calva-Cerqueira D, Chinnathambi S, Pechman B, *et al.* The rate of germline mutations and large deletions of SMAD4 and BMPR1A in juvenile polyposis. *Clin Genet* 2009;**75**:79–85. doi:10.1111/j.1399-0004.2008.01091.x

148 Zhou XP, Hampel H, Thiele H, *et al.* Association of germline mutation in the PTEN tumour suppressor gene and Proteus and Proteus-like syndromes. *Lancet* 2001;**358**:210–1. doi:10.1016/S0140-6736(01)05412-5

149 Eng C, Ji H. Molecular Classification of the Inherited Hamartoma Polyposis Syndromes: Clearing the Muddied Waters. *Am J Hum Genet* 1998;**62**:1020–2.

150 Crowder CD, Sweet K, Lehman A, *et al.* Serrated polyposis is an underdiagnosed

and unclear syndrome: The surgical pathologist has a role in improving detection. *Am J Surg Pathol* 2012;**36**:1178–85. doi:10.1097/PAS.0b013e3182597f41

151 Edelstein DL, Cruz-Correa M, Soto-Salgado M, *et al.* Risk of Colorectal and Other Cancers in Patients With Serrated Polyposis. *Clin Gastroenterol Hepatol* 2015;**13**:1697–9. doi:10.1016/j.cgh.2015.02.003

152 Kambara T, Simms LA, Whitehall LJ, *et al.* BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* 2004;**53**:1137–44. doi:10.1136/gut.2003.037671

153 Castells A. Choosing the optimal method in programmatic colorectal cancer screening: current evidence and controversies. *Therap Adv Gastroenterol* 2015;**8**:221–33.

154 National Cancer Institute. Definition of chemoprevention - NCI Dictionary of Cancer Terms - National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/chemoprevention (accessed 14 Aug 2018).

155 Ferrández Á, Piazuelo E, Castells A. Aspirin and the prevention of colorectal cancer. *Best Pract Res Clin Gastroenterol* 2012;**26**:185–95. doi:10.1016/j.bpg.2012.01.009

156 Tárraga P, Solera J, Rodríguez-Montes JA. Primary and Secondary Prevention of Colorectal Cancer. *Clin Med Insights Gastroenterol* 2014;**7**:33–46.

157 López Bastida J, Sassi F, Bellas Beceiro B, *et al. Análisis coste-efectividad del cribado del cáncer colorrectal en la población general*. 2010.

158 Zauber AG, Winawer SJ, O MJ, *et al.* Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *N Engl J Med* 2012;**366**:687–96.

159 Rodríguez-Moranta F, Saló J, Arcusa À, *et al.* Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: A prospective, multicenter, randomized, controlled trial. *J Clin Oncol* 2006;**24**:386–93. doi:10.1200/JCO.2005.02.0826

160 Castells A, Binefa G, Fité A, *et al.* Postpolypectomy surveillance i patients with adenomas and serrated lesions: a proposal for risk stratification in the context of organized colorectal cancer-screening programs. *Endoscopy* 2015;**47**:86–7. doi:10.1093/aje/kwr426

161 Andermann A, Blancquaert I, Beauchamp S, *et al.* Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ* 2008;**86**:317–9. doi:10.2471/BLT.07.050112

162 Stracci F, Zorzi M, Grazzini G, *et al.* Colorectal cancer screening: tests, strategies, and perspectives. *Front Public Heal* 2014;**2**:1–9. doi:10.3389/fpubh.2014.00210

163 Groenen PJTA, Van Krieken JH, Murray ; S, *et al.* Current status of screening for colorectal cancer. *Ann Oncol* 2013;**24**:1963–72. doi:10.1093/annonc/mdt157

164 European Colorectal Cancer Screening Guidelines Working Group. European guidelines for quality assurance in colorectal cancer screening and diagnosis: Overview and introduction to the full Supplement publication. *Endoscopy* 2013;**45**:51–9. doi:10.1055/s-0032-1325997

165 Winawer SJ, Zauber AG, Ho MN, *et al.* Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;**329**:1977–81. doi:10.1056/NEJM199312303292701

166 Brenner H, Chang-Claude J, Seiler CM, *et al.* Protection from colorectal cancer after colonoscopy: a population-based, case-control study. *Ann Intern Med* 2011;**154**:22–30. doi:10.7326/0003-4819-154-1-201101040-00004

167 Citarda F, Tomaselli G, Capocaccia R, *et al.* Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut* 2001;**48**:812–5.

168 Kaminski MF, Regula J, Kraszewska E, *et al.* Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;**362**:1795–803.

169 Brenner H, Hoffmeister M, Arndt V, *et al.* Protection from right-and left-sided colorectal neoplasms after colonoscopy: Population-based study. *J Natl Cancer Inst*

2010;**102**:89–95. doi:10.1093/jnci/djp436

170 Singh H, Nugent Z, Mahmud SM, *et al.* Predictors of colorectal cancer after negative colonoscopy: A population-based study. *Am J Gastroenterol* 2010;**105**:663–73. doi:10.1038/ajg.2009.650

171 Quintero E, Castells A, Bujanda L, *et al.* Colonoscopy versus Fecal Immunochemical Testing in Colorectal-Cancer Screening. *N Engl J Med* 2012;**366**:697–706.

172 Schoen RE, Pinsky PF, Weissfeld JL, *et al.* Colorectal-Cancer Incidence and Mortality with Screening Flexible Sigmoidoscopy. *N Engl J Med* 2012;**366**:2345–57. doi:10.1056/NEJMoa1114635

173 Imperiale TF, Wagner DR, Lin CY, *et al.* Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *N Engl J Med* 2000;**343**:169–74.

174 Lieberman DA, Weiss DG, Bond JH, *et al.* Use of colonoscopy to screen asymptomatic adults for colorectal cancer. *N Engl J Med* 2000;**343**:162–8.

175 Castells A, Bessa X, Quintero E, *et al.* Risk of Advanced Proximal Neoplasms According to Distal colorectal Findings: Comparison of Sigmoidoscopy-Based Strategies. *J Natl Cancer Inst* 2013;**105**:878–86. doi:10.1093/jnci/djt117

176 Sanford KW, McPherson RA. Fecal Occult Blood Testing. *Clin Lab Med* 2009;**29**:523–41. doi:10.1016/j.cll.2009.06.008

177 Imperiale TF, Ransohoff DF, Itzkowitz SH, *et al.* Fecal DNA versus Fecal Occult Blood for Colorectal-Cancer Screening in an Average-Risk Population. *N Engl J Med* 2004;**23**:2704–14.

178 Lieberman DA. Screening for colorectal cancer. *N Engl J Med* 2009;**361**:1179–87.

179 Van Rossum LG, Van Rijn AF, Laheij RJ, *et al.* Random Comparison of Guaiac and Immunochemical Fecal Occult Blood Tests for Colorectal Cancer in a Screening Population. *Gastroenterology* 2008;**135**:82–90. doi:10.1053/j.gastro.2008.03.040

180 Auge JM, Pellise M, Escudero JM, *et al.* Risk stratification for advanced colorectal neoplasia according to fecal hemoglobin concentration in a colorectal cancer screening program. *Gastroenterology* 2014;**147**. doi:10.1053/j.gastro.2014.06.008

181 Song L-L, Li Y-M. Current noninvasive tests for colorectal cancer screening: An overview of colorectal cancer screening tests. *World J Gastrointest Oncol* 2016;**8**:793–800. doi:10.4251/wjgo.v8.i11.793

182 Chiu HM, Lee YC, Tu CH, *et al.* Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. *Clin Gastroenterol Hepatol* 2013;**11**:832–8. doi:10.1016/j.cgh.2013.01.013

183 Allison JE, Fraser CG, Halloran SP, *et al.* Population screening for colorectal cancer means getting FIT: The past, present, and future of colorectal cancer screening using the fecal immunochemical test for hemoglobin (FIT). *Gut Liver* 2014;**8**:117–30. doi:10.5009/gnl.2014.8.2.117

184 Imperiale TF, Ransohoff DF, Itzkowitz SH, *et al.* Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *N Engl J Med* 2014;**370**:1287–97. doi:10.1056/NEJMoa1311194

185 Mitchell PS, Parkin RK, Kroh EM, *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *PNAS* 2008;**105**:10513–8.

186 Giráldez MD, Lozano JJ, Ramírez G, *et al.* Circulating MicroRNAs as Biomarkers of Colorectal Cancer: Results From a Genome-Wide Profiling and Validation Study. *Clin Gastroenterol Hepatol* 2013;**11**:681–8. doi:10.1016/j.cgh.2012.12.009

187 Church TR, Wandell M, Lofton-Day C, *et al.* Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 2014;**63**:317–25. doi:10.1136/gutjnl-2012-304149

188 Senore C, Inadomi J, Segnan N, *et al.* Optimising colorectal cancer screening acceptance: A review. *Gut* 2015;**64**:1158–77. doi:10.1136/gutjnl-2014-308081

189 Council of the European Union. Council Recommendation of 2 December 2003 on cancer screening (2003/878/EC). 2003.

190 Ponti A, Anttila A, Ronco G, *et al.* Cancer Screening in the European Union (2017). Report on the implementation of the Council Recommendation on cancer screening. 2017.

191 Ibañez-Sanz G. *Personalització del cribratge del càncer colorectal (Doctoral Thesis)*. 2017.

192 Bobridge A, Bampton P, Cole S, *et al.* The psychological impact of participating in colorectal cancer screening by faecal immuno-chemical testing – the Australian experience. *Brithish J Cancer* 2014;**111**:970–5. doi:10.1038/bjc.2014.371

193 Lafata JE, Simpkins J, Lamerato L, *et al.* The Economic Impact of False-Positive Cancer Screens. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:2126–58.

194 Wong MCS, Ching JYL, Chan VCW, *et al.* Factors associated with false-positive and false-negative fecal immunochemical test results for colorectal cancer screening. *Gastrointest Endosc* 2015;**81**:596–607. doi:10.1016/j.gie.2014.08.006

195 Stegeman I, De Wijkerslooth TR, Stoop EM, *et al.* Risk factors for false positive and for false negative test results in screening with fecal occult blood testing. *Int J Cancer* 2013;**133**:2408–14. doi:10.1002/ijc.28242

196 Van Turenhout ST, Oort FA, Terhaar Sive Droste JS, *et al.* Hemorrhoids detected at colonoscopy: An infrequent cause of false-positive fecal immunochemical test results. *Gastrointest Endosc* 2012;**76**:136–43. doi:10.1016/j.gie.2012.03.169

197 Garcia M, Binefa G, Milà N. False-positive results in colorectal cancer screening: what do we know? *Color Cancer* 2012;**1**:447–50.

198 Haug U, Knudsen AB, Brenner H, *et al.* Is fecal occult blood testing more sensitive for left- versus right-sided colorectal neoplasia? A systematic literature review. *Expert Rev Mol Diagn* 2011;**11**:605–16. doi:10.1586/erm.11.41

199 Brown LF, Fraser CG. Effect of delay in sampling on haemoglobin determined by faecal immunochemical tests. *Ann Clin Biochem* 2008;**45**:604–5. doi:10.1258/acb.2008.008024

200 Nishida H, Urano S. Effectiveness of repeated screening using the fecal occult

blood test and its impact on reducing false-negative cancer cases. *Eur J Cancer Prev* 2011;**20**:184–9. doi:10.1097/CEJ.0b013e3283447453

201 Sehgal R, Sheahan K, O'Connell PR, *et al.* Lynch Syndrome: An updated review. *Genes (Basel)* 2014;**5**:497–507. doi:10.3390/genes5030497

202 Lynch HT, Lanspa S, Shaw T, *et al.* Phenotypic and genotypic heterogeneity of Lynch syndrome: a complex diagnostic challenge. *Fam Cancer* 2017;**17**:403–14. doi:10.1007/s10689-017-0053-3

203 Peltomäki P. Lynch syndrome genes. *Fam Cancer* 2005;**4**:227–32. doi:10.1007/s10689-004-7993-0

204 Giardiello FM, Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* 2001;**121**:198–213. doi:10.1053/gast.2001.25581

205 Aaltonen LA, Salovaara R, Kristo P, *et al.* Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998;**338**:1481–7.

206 Piñol V, Castells A, Andreu M, *et al.* Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *J Am Med Assoc* 2005;**293**:1986–94. doi:10.1001/jama.293.16.1986

207 Gruber SB. New developments in Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer) and mismatch repair gene testing. *Gastroenterology* 2006;**130**:577–87. doi:10.1053/j.gastro.2006.01.031

208 Vasen H, Watson P, Mecklin J, *et al.* New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. *Gastroenterology* 1999;**116**:1453–6. doi:10.1016/S0016-5085(99)70510-X

209 Umar A, Boland CR, Terdiman JP, *et al.* Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and

Microsatellite Instability. *J Natl Cancer Inst* 2004;**96**:261–8.

210 Winawer S, Fletcher R, Rex D, *et al.* Colorectal Cancer Screening and Surveillance: Clinical Guidelines and Rationale-Update Based on New Evidence. *Gastroenterology* 2003;**124**:544–60. doi:10.1053/gast.2003.50044

211 De Jong AE, Hendriks YMC, Kleibeuker JH, *et al.* Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology* 2006;**130**:665–71. doi:10.1053/j.gastro.2005.11.032

212 Mecklin JP, Aarnio M, Läärä E, *et al.* Development of Colorectal Tumors in Colonoscopic Surveillance in Lynch Syndrome. *Gastroenterology* 2007;**133**:1093–8. doi:10.1053/j.gastro.2007.08.019

213 Dove-Edwin I, De Jong AE, Adams J, *et al.* Prospective Results of Surveillance Colonoscopy in Dominant Familial Colorectal Cancer With and Without Lynch Syndrome. *Gastroenterology* 2006;**130**:1995–2000. doi:10.1053/j.gastro.2006.03.018

214 Pylvänäinen K, Kairaluoma M, Mecklin JP. Compliance and satisfaction with long-term surveillance in Finnish HNPCC families. *Fam Cancer* 2006;**5**:175–8. doi:10.1007/s10689-005-5442-3

215 Sekirov I, Russell S, Antunes L. Gut microbiota in health and disease. *Physiol Rev* 2010;**90**:859–904. doi:10.1152/physrev.00045.2009.

216 Ley RE, Peterson DA, Gordon JI. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell* 2006;**124**:837–48. doi:10.1016/j.cell.2006.02.017

217 Gebbers JO, Laissue JA. Immunologic structures and functions of the gut. *Schweiz Arch Tierheilkd* 1989;**131**:221–38.

218 Schloss PD, Handelsman J. Status of the Microbial Census. *Microbiol Mol Biol Rev* 2004;**68**:686–91. doi:10.1128/MMBR.68.4.686-691.2004

219 Eckburg PB, Bik EM, Bernstein CN, *et al.* Diversity of the Human Intestinal Microbial Flora. *Science (80-)* 2005;**308**:1635–8.

220 Wang M, Ahrné S, Jeppsson B, *et al.* Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* 2005;**54**:219–31. doi:10.1016/j.femsec.2005.03.012

221 Flint HJ, Scott KP, Louis P, *et al.* The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 2012;**9**:577–89. doi:10.1038/nrgastro.2012.156

222 Candela M, Biagi E, Turroni S, *et al.* Dynamic efficiency of the human intestinal microbiota. *Crit Rev Microbiol* 2013;**41**:1–7. doi:10.3109/1040841X.2013.813900

223 Jost T, Lacroix C, Braegger CP, *et al.* New Insights in Gut Microbiota Establishment in Healthy Breast Fed Neonates. *PLoS One* 2012;**7**. doi:10.1371/journal.pone.0044595

224 Vael C, Desager K. The importance of the development of the intestinal microbiota in infancy. *Curr Opin Pediatr* 2009;**21**:794–800. doi:10.1097/MOP.0b013e328332351b

225 Dominguez-Bello MG, Costello EK, Contreras M, *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS* 2010;**107**:11971–5. doi:10.1073/pnas.1002601107

226 Koenig JE, Spor A, Scalfone N, *et al.* Succession of microbial consortia in the developing infant gut microbiome. *PNAS* 2010;**108**:4578–85. doi:10.1073/pnas.1000081107

227 Vanhoutte T, Huys G, De Brandt E, *et al.* Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. *FEMS Microbiol Ecol* 2004;**48**:437–46. doi:10.1016/j.femsec.2004.03.001

228 Biagi E, Nylund L, Candela M, *et al.* Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians. *PLoS One* 2010;**5**:e10667. doi:10.1371/journal.pone.0010667

229 Quercia S, Candela M, Giuliani C, *et al.* From lifetime to evolution: timescales of human gut microbiota adaptation. *Front Microbiol* 2014;**5**:1–9.

doi:10.3389/fmicb.2014.00587

230 Nagpal R, Mainali R, Ahmadi S, *et al.* Gut microbiome and aging: Physiological and mechanistic insights. *Nutr Heal Aging* 2018;**4**:267–85. doi:10.3233/NHA-170030

231 Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem J* 2017;**474**:1823–36. doi:10.1042/BCJ20160510

232 David LA, Maurice CF, Carmody RN, *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014;**505**:559–63. doi:10.1038/nature12820

233 Walker AW, Ince J, Duncan SH, *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* 2011;**5**:220–30. doi:10.1038/ISMEJ.2010.118

234 Yu Z-T, Chen C, Kling DE, *et al.* The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology* 2013;**23**:169–77. doi:10.1093/glycob/cws138

235 Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe* 2011;**17**:478–82. doi:10.1016/j.anaerobe.2011.03.009

236 Desai MS, Seekatz AM, Koropatkin NM, *et al.* A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* 2016;**167**:1339–53. doi:10.1016/j.cell.2016.10.043

237 Ze X, Duncan SH, Louis P, *et al.* Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J* 2012;**6**:1535–43. doi:10.1038/ismej.2012.4

238 Louis P, Scott KP, Duncan SH, *et al.* Understanding the effects of diet on bacterial metabolism in the large intestine. *J Appl Microbiol* 2007;**102**:1197–208. doi:10.1111/j.1365-2672.2007.03322.x

239 Kakiyama G, Pandak WM, Gillevet PM, *et al.* Modulation of the Fecal Bile Acid Profile by Gut Microbiota in Cirrhosis. *J Hepatol* 2013;**58**:949–55.

doi:10.1016/j.jhep.2013.01.003

240 Hooper L V., MacPherson AJ. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* 2010;**10**:159–69. doi:10.1038/nri2710

241 Wehkamp J, Salzman NH, Porter E, *et al.* Reduced Paneth cell-defensins in ileal Crohn's disease. *PNAS* 2005;**102**:18129–34.

242 Rafii F, Sutherland JB, Cerniglia CE, *et al.* Effects of treatment with antimicrobial agents on the human colonic microflora. *Ther Clin Risk Manag* 2008;**4**:1343–57.

243 Backhed F, Ley RE, Sonnenburg JL, *et al.* Hostal-bacterial mutualism in the human intestine. *Science (80-)* 2005;**307**:1915–20.

244 O'hara AM, Shanahan F. The gut flora as a forgotten organ. *EMBO Rep* 2006;**7**:688–93. doi:10.1038/sj.embor.7400731

245 Tsai F, Coyle WJ. The Microbiome and Obesity: Is Obesity Linked to Our Gut Flora? *Curr Gastroenterol Rep* 2009;**11**:307–13.

246 Wostmann BS. The Germfree Animal in Nutritional Studies. *Annu Rev Nutr* 1981;**1**:257–79. doi:10.1146/annurev.nu.01.070181.001353

247 Savage DC. Gastrointestinal Microflora In Mammalian Nutrition. *Annu Rev Nutr* 1986;**6**:155–78.

248 Harris K, Kassis A, Eve Major G, *et al.* Is the Gut Microbiota a New Factor Contributing to Obesity and Its Metabolic Disorders? *J Obes* 2012;**2012**:1–14. doi:10.1155/2012/879151

249 Kallus SJ, Brandt LJ. The intestinal microbiota and obesity. *J Clin Gastroenterol* 2012;**46**:16–24. doi:10.1097/MCG.0b013e31823711fd

250 Moens E, Veldhoen M. Epithelial barrier biology: good fences make good neighbours. *Immunology* 2011;**135**:1–8. doi:10.1111/j.1365-2567.2011.03506.x

251 Takiishi T, Fenero CIM, Câmara NOS. Intestinal barrier and gut microbiota: Shaping our immune responses throughout life. *Tissue Barriers* 2017;**5**:e1373208.

doi:10.1080/21688370.2017.1373208

252 Sommer F, Bäckhed F. The gut microbiota-masters of host development and physiology. *Nat Rev Microbiol* 2013;**11**:227–38. doi:10.1038/nrmicro2974

253 Carstens LE, Westerbeek EAM, van Zwol A, *et al.* Neonatal antibiotics in preterm infants and allergic disorders later in life. *Pediatr Allergy Immunol* 2016;**27**:759–64. doi:10.1111/pai.12614

254 L. Sears C, S. Garrett W. Microbes, Microbiota and Colon Cancer. *Cell Host Microbe* 2015;**15**:317–28. doi:10.1016/j.chom.2014.02.007.Microbes

255 Zur Hausen H. The search for infectious causes of human cancers: Where and why. *Virology* 2009;**392**:1–10. doi:10.1016/j.virol.2009.06.001

256 Collins D, Hogan AM, Winter DC. Microbial and viral pathogens in colorectal cancer. *Lancet Oncol* 2011;**12**:504–12. doi:10.1016/S1470-2045(10)70186-8

257 Arthur JC, Perez-Chanona E, Mühlbauer M, *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science (80-)* 2012;**338**:120–3. doi:10.1126/science.1224820

258 Weisburger JH, Reddy BS, Narisawa T, *et al.* Germ-free status and colon tumor induction by N-methyl-N'-nitro-N-nitrosoguanidine. *Proc Soc Exp Biol Med* 1975;**148**:1119–21.

259 Tjalsma H, Boleij A, Marchesi JR, *et al.* A bacterial driver-passenger model for colorectal cancer: Beyond the usual suspects. *Nat Rev Microbiol* 2012;**10**:575–82. doi:10.1038/nrmicro2819

260 Sobhani I, Tap J, Roudot-Thoraval F, *et al.* Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. *PLoS One* 2011;**6**:e16393. doi:10.1371/journal.pone.0016393

261 Kostic AD, Gevers D, Pedamallu CS, *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* 2012;**22**:292–8. doi:10.1101/gr.126573.111

262 Chen W, Liu F, Ling Z, *et al.* Human Intestinal Lumen and Mucosa-Associated

Microbiota in Patients with Colorectal Cancer. *PLoS One* 2012;**7**:e39743. doi:10.1371/journal.pone.0039743

263 Sanapareddy N, Legge RM, Jovov B, *et al.* Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. *ISME J* 2012;**6**:1858–68. doi:10.1038/ismej.2012.43

264 Wang T, Cai G, Qiu Y, *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* 2012;**6**:320–9. doi:10.1038/ismej.2011.109

265 Wu N, Yang X, Zhang R, *et al.* Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol* 2013;**66**:462–70. doi:10.1007/s00248-013-0245-9

266 Ahn J, Sinha R, Pei Z, *et al.* Human Gut Microbiome and Risk for Colorectal Cancer. *J Natl Cancer Inst* 2013;**105**:1907–11. doi:10.1093/jnci/djt300

267 Zeller G, Tap J, Voigt AY, *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;**10**:766–766. doi:10.15252/msb.20145645

268 McCoy WC, Mason JM. Enterococcal endocarditis associated with carcinoma of the sigmoid; report of a case. *J Med Assoc State Ala* 1951;**21**:162–6.

269 Keusch GT. Opportunistic infections in colon carcinoma. *Am J Clin Nutr* 1974;**27**:1481–5.

270 Rusniok C, Couvé E, Da Cunha V, *et al.* Genome sequence of streptococcus gallolyticus: Insights into its adaptation to the bovine rumen and its ability to cause endocarditis. *J Bacteriol* 2010;**192**:2266–76. doi:10.1128/JB.01659-09

271 Boleij A, Tjalsma H. Gut bacteria in health and disease: A survey on the interface between intestinal microbiology and colorectal cancer. *Biol Rev* 2012;**87**:701–30. doi:10.1111/j.1469-185X.2012.00218.x

272 Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006;**118**:3030–44. doi:10.1002/ijc.21731

273 International Agengy for Research on Cancer. Helicobacter pylori. 2018.

274 Sun J, Kato I. Gut microbiota, inflammation and colorectal cancer. *Gene&Diseases* 2016;**3**:130–43. doi:10.1016/j.gendis.2016.03.004

275 Zumkeller N, Brenner H, Zwahlen M, *et al.* Helicobacter pylori infection and colorectal cancer risk: A meta-analysis. *Helicobacter pylori Infect Color cancer risk a meta-analysis* 2006;**11**:75–80. doi:10.1111/j.1523-5378.2006.00381.x

276 Guo Y, Li H-Y. Association between Helicobacter pylori infection and colorectal neoplasm risk: A meta-analysis Based on East Asian population. *J Cancer Res Ther* 2014;**10**:263–6. doi:10.4103/0973-1482.151482

277 Johnson JR, Johnston B, Kuskowski MA, *et al.* Molecular Epidemiology and Phylogenetic Distribution of the Escherichia coli pks Genomic Island. *J Clin Microbiol* 2008;**46**:3906–11. doi:10.1128/JCM.00949-08

278 Swidsinski A, Khilkin M, Kerjaschki D, *et al.* Association between intraepithelial Escherichia coli and colorectal cancer. *Gastroenterology* 1998;**115**:281–6. doi:10.1016/S0016-5085(98)70194-5

279 Buc E, Dubois D, Sauvanet P, *et al.* High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One* 2013;**8**:e56964. doi:10.1371/journal.pone.0056964

280 Raisch J, Buc E, Bonnet M, *et al.* Colon cancer-associated B2 Escherichia coli colonize gut mucosa and promote cell proliferation. *World J Gastroenterol Gastroenterol* 2014;**20**:6560–72. doi:10.3748/wjg.v20.i21.6560

281 Wassenaar TM. E. coli and colorectal cancer: a complex relationship that deserves a critical mindset. *Crit Rev Microbiol* 2018;**44**:619–32. doi:10.1080/1040841X.2018.1481013

282 Gagnière J, Raisch J, Veziant J, *et al.* Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol* 2016;**22**:501–18. doi:10.3748/wjg.v22.i2.501

283 Sears CL. Enterotoxigenic Bacteroides fragilis: a Rogue among Symbiotes. *Clin Microbiol Rev* 2009;**22**:349–69. doi:10.1128/CMR.00053-08

284 Soler AP, Miller RD, Laughlin K V, *et al.* Increased tight junctional permeability is associated with the development of colon cancer. *Carcinogenesis* 1999;**20**:1425–31.

285 Toprak NU, Yagci A, Gulluoglu BM, *et al.* A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect* 2006;**12**:782–6. doi:10.1111/j.1469-0691.2006.01494.x

286 Kistler JO, Booth V, Bradshaw DJ, *et al.* Bacterial Community Development in Experimental Gingivitis. *PLoS One* 2013;**8**:e71227. doi:10.1371/journal.pone.0071227

287 Mitsuhashi K, Nosho K, Sukawa Y, *et al.* Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis. *Oncotarget* 2015;**6**:7209–20.

288 Ahmed Z, Bansal SK, Dhillon S. Pyogenic liver abscess caused by Fusobacterium in a 21-year-old immunocompetent male. *World J Gastroenterol* 2015;**21**:3731–5. doi:10.3748/wjg.v21.i12.3731

289 Salö M, Marungruang N, Roth B, *et al.* Evaluation of the microbiome in children's appendicitis. *Int J Colorectal Dis* 2017;**32**:19–28. doi:10.1007/s00384-016-2639-x

290 Kostic AD, Chun E, Robertson L, *et al.* Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* 2013;**14**:207–15. doi:10.1016/j.chom.2013.07.007

291 Rubinstein MR, Wang X, Liu W, *et al.* Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013;**14**:195–206. doi:10.1016/j.chom.2013.07.012

292 Viljoen KS, Dakshinamurthy A, Goldberg P, *et al.* Quantitative Profiling of Colorectal Cancer-Associated Bacteria Reveals Associations between Fusobacterium spp., Enterotoxigenic Bacteroides fragilis (ETBF) and Clinicopathological Features of Colorectal Cancer. *PLoS One* 2015;**10**:e0119462. doi:10.1371/journal.pone.0119462

293 Andino A, Hanning I. Salmonella enterica: Survival, Colonization, and Virulence

Differences among Serovars. *Sci World J* 2015;**2015**:1–16. doi:10.1155/2015/520179

294 Crum Cianflone NF. Salmonellosis and the GI Tract: More than Just Peanut Butter. *Curr Gastroenterol Rep* 2008;**10**:424–31.

295 Scanu T, Spaapen RM, Bakker JM, *et al.* Salmonella Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma. *Cell Host Microbe* 2015;**17**:763–74. doi:10.1016/j.chom.2015.05.002

296 Lu R, Bosland M, Xia Y, *et al.* Presence of Salmonella AvrA in colorectal tumor and its precursor lesions in mouse intestine and human specimens. *Oncotarget* 2017;**8**:55104–15.

297 Kato I, Boleij A, Kortman GAM, *et al.* Partial Associations of Dietary Iron, Smoking and Intestinal Bacteria with Colorectal Cancer Risk. *Nutr Cancer* 2013;**65**:169–77. doi:10.1080/01635581.2013.748922

298 de Almeida CV, Taddei A, Amedei A. The controversial role of Enterococcus faecalis in colorectal cancer. *Therap Adv Gastroenterol* 2018;**11**:1–11. doi:10.1177/1756284818783606

299 Grootaert C, Van de Wiele T, Van Roosbroeck I, *et al.* Bacterial monocultures, propionate, butyrate and H2O2 modulate the expression, secretion and structure of the fasting-induced adipose factor in gut epithelial cell lines. *Environ Microbiol* 2011;**13**:1778–89. doi:10.1111/j.1462-2920.2011.02482.x

300 Wang X, Yang Y, Huycke MM. Commensal-infected macrophages induce dedifferentiation and reprogramming of epithelial cells during colorectal carcinogenesis. *Oncotarget* 2017;**8**:102176–90.

301 Fearon ER. Molecular Genetics of Colorectal Cancer. *Annu Rev Pathol Mech Dis* 2011;**6**:479–507. doi:10.1146/annurev-pathol-011110-130235

302 Zackular JP, Rogers MAM, Ruffin Iv MT, *et al.* The Human Gut Microbiome as a Screening Tool for Colorectal Cancer. *Cancer Prev Res* 2014;**7**:1112–21. doi:10.1158/1940-6207.CAPR-14-0129

303 Ai L, Tian H, Chen Z, *et al.* Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. *Oncotarget* 2017;**8**:9546–56.

304 Shah MS, DeSantis TZ, Weinmaier T, *et al.* Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. *Gut* 2018;**67**:882–91. doi:10.1136/gutjnl-2016-313189

305 Amitay EL, Krilaviciute A, Brenner H. Systematic Review: Gut Microbiota in Fecal Samples and Detection of Colorectal Neoplasms. *Gut Microbes* 2018;**9**:293–307. doi:10.1080/19490976.2018.1445957

306 Villéger R, Lopès A, Veziant J, *et al.* Microbial markers in colorectal cancer detection and/or prognosis. *World J Gastroenterol* 2018;**24**:2327–47. doi:10.3748/wjg.v24.i22.2327

307 Bezabeh T, Somorjai R, Dolenko B, *et al.* Detecting colorectal cancer by 1H magnetic resonance spectroscopy of fecal extracts. *NMR Biomed* 2009;**22**:593–600. doi:10.1002/nbm.1372

308 Monleón D, Morales JM, Barrasa A, *et al.* Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR Biomed* 2009;**22**:342–8. doi:10.1002/nbm.1345

309 Weir TL, Manter DK, Sheflin AM, *et al.* Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. *PLoS One* 2013;**8**:e70803. doi:10.1371/journal.pone.0070803

310 Goedert JJ, Sampson JN, Moore SC, *et al.* Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* 2014;**35**:2089–96. doi:10.1093/carcin/bgu131

311 Phua LC, Chue XP, Koh PK, *et al.* Non-invasive fecal metabonomic detection of colorectal cancer. *Cancer Biol Ther* 2014;**15**:389–97. doi:10.4161/cbt.27625

312 Batty C, Cauchi M, Lourenço C, *et al.* Use of the Analysis of the Volatile Faecal Metabolome in Screening for Colorectal Cancer. *PLoS One* 2015;**10**:e0130301. doi:10.1371/journal.pone.0130301

313 Sinha R, Ahn J, Sampson JN, *et al.* Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. *PLoS One* 2016;**11**:e0152126.

314 Lu X, Kang Q, Jin P, *et al.* The Interactions Between Gut Microbiota and Lynch Syndrome. *Clin Gastroenterol Hepatol* 2017;**15**:e48. doi:10.1016/j.cgh.2016.09.116

315 Memorial Sloan Kettering Cancer Center. Metagenomic Evaluation of the Gut Microbiome in Patients With Lynch Syndrome and Other Hereditary Colonic Polyposis Syndromes. ClinicalTrials.gov. 2015.https://clinicaltrials.gov/ct2/show/NCT02371135 (accessed 26 Feb 2018).

316 Matsuda K, Iwaki KK, Garcia-Gomez J, *et al.* Bacterial Identification by 16S rRNA Gene PCR-Hybridization as a Supplement to Negative Culture Results. *J Clin Microbiol* 2011;**49**:2031–4. doi:10.1128/JCM.00615-10

317 Mas de Xaxars Rivero T. *Descripció i quantificació de la microbiota intestinal associada al càncer colorectal (Doctoral Thesis)*. 2012.

318 Serra-Pagès M, García-Gil J, Mas de Xarxars T, *et al.* PCT/EP2015/054451. Biomarkers for early detection, risk screening and monitoring of colorectal cancer and adenomatous polyps. 2015;:20.

319 Lopez-Siles M, Martinez-Medina M, Busquets D, *et al.* Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 2014;**304**:464–75. doi:10.1016/j.ijmm.2014.02.009

320 Huijsdens XW, Linskens RK, Mak M, *et al.* Quantification of Bacteria Adherent to Gastrointestinal Mucosa by Real-Time PCR. *J Clin Microbiol* 2002;**40**:4423–7. doi:10.1128/JCM.40.12.4423-4427.2002

321 Lopez-Siles M, Martinez-Medina M, Surís-Valls R, *et al.* Changes in the abundance of Faecalibacterium prausnitzii phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and colorectal cancer patients. *Inflamatory Bowel Dis* 2016;**22**:28–41.

322 Bailey JR, Aggarwal A, Imperiale TF. Colorectal Cancer Screening: Stool DNA and

Other Noninvasive Modalities. *Gut Liver* 2016;**10**:204–11. doi:10.5009/gnl15420

323 Brenner H, Altenhofen L, Stock C, *et al.* Prevention, Early Detection, and Overdiagnosis of Colorectal Cancer Within 10 Years of Screening Colonoscopy in Germany. *Clin Gastroenterol Hepatol* 2015;**13**:717–23. doi:10.1016/j.cgh.2014.08.036

324 Shah R, Jones E, Vidart V, *et al.* Biomarkers for Early Detection of Colorectal Cancer and Polyps: Systematic Review. *Cancer Epidemiol Biomarkers Prev* 2014;**23**:1712–28. doi:10.1158/1055-9965.EPI-14-0412

325 Garrett WS. Cancer and the microbiota. *Science (80-)* 2015;**348**:80–6. doi:10.1126/science.aaa4972

326 Borges-Canha M, Portela-Cidade JP, Dinis-Ribeiro M, *et al.* Role of colonic microbiota in colorectal carcinogenesis: A systematic review. *Rev española enfermedades Dig* 2015;**107**:659–71.

327 Kelly D, Mulder IE. Microbiome and immunological interactions. *Nutr Rev* 2012;**70**:S18–30. doi:10.1111/j.1753-4887.2012.00498.x

328 Gao Z, Guo B, Gao R, *et al.* Microbiota disbiosis is associated with colorectal cancer. *Front Microbiol* 2015;**6**:1–9. doi:10.3389/fmicb.2015.00020

329 Zhao L, Zhang X, Zuo T, *et al.* The Composition of Colonic Commensal Bacteria According to Anatomical Localization in Colorectal Cancer. *Engineering* 2017;**3**:90–7. doi:10.1016/J.ENG.2017.01.012

330 Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014;**12**:661–72. doi:10.1038/nrmicro3344

331 Narayanan V, Peppelenbosch MP, Konstantinov SR. Human fecal microbiome-based biomarkers for colorectal cancer. *Cancer Prev Res* 2014;**7**:1108–11. doi:10.1158/1940-6207.CAPR-14-0273

332 Macfarlane GT, Macfarlane S. Bacteria, Colonic Fermentation, and Gastrointestinal Health. *J AOAC Int* 2012;**95**:50–60. doi:10.5740/jaoacint.SGE_Macfarlane

333 Baümler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 2016;**535**:85–93. doi:10.1038/nature18849

334 Vandesompele J, De Preter K, Pattyn I, *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:34–41.

335 Conlon MA, Bird AR. The Impact of Diet and Lifestyle on Gut Microbiota and Human Health. *Nutrients* 2015;**7**:17–44. doi:10.3390/nu7010017

336 Vandeputte D, Falony G, Vieira-Silva S, *et al.* Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 2015;**65**:57–62. doi:10.1136/gutjnl-2015-309618

337 Davis SC, Yadav JS, Barrow SD, *et al.* Gut microbiome diversity influenced more by the Westernized dietary regime than the body mass index as assessed using effect size statistic. *Microbiologyopen* 2017;**6**:1–17. doi:10.1002/mbo3.476

338 Yun Y, Kim H-N, Kim SE, *et al.* Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. *BMC Microbiol* 2017;**17**:151–60. doi:10.1186/s12866-017-1052-0

339 Rogers MAM, Todd Greene M, Saint S, *et al.* Higher Rates of Clostridium difficile Infection among Smokers. *PLoS One* 2012;**7**:e42091. doi:10.1371/journal.pone.0042091

340 Capurso G, Lahner E. The interaction between smoking, alcohol and the gut microbiome. *Best Pract Res Clin Gastroenterol* 2017;**31**:579–88. doi:10.1016/j.bpg.2017.10.006

341 Baothman OA, Zamzami MA, Taher I, *et al.* The role of Gut Microbiota in the development of obesity and Diabetes. *Lipids Health Dis* 2016;**15**:108. doi:10.1186/s12944-016-0278-4

342 Watt E, Gemmell MR, Berry S, *et al.* Extending colonic mucosal microbiome analysis-assessment of colonic lavage as a proxy for endoscopic colonic biopsies. *Microbiome* 2016;**4**:1–15. doi:10.1186/s40168-016-0207-9

343 Durbán A, Abellán JJ, Jiménez-Hernández N, *et al.* Assessing Gut Microbial Diversity from Feces and Rectal Mucosa. *Microb Ecol* 2011;**61**:123–33. doi:10.1007/s00248-010-9738-y

344 Biedermann L, Zeitz J, Mwinyi J, *et al.* Smoking Cessation Induces Profound Changes in the Composition of the Intestinal Microbiota in Humans. *PLoS One* 2013;**8**:e59260. doi:10.1371/journal.pone.0059260

345 Rao RK, Seth A, Sheth P. Recent Advances in Alcoholic Liver Disease I. Role of intestinal permeability and endotoxemia in alcoholic liver disease. *Am J Physiol* 2004;**286**:881–4.

346 Turnbaugh PJ, Ley RE, Mahowald MA, *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;**444**:1027–31. doi:10.1038/nature05414

347 Schwiertz A, Taras D, Schäfer K, *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* 2010;**18**:190–5. doi:10.1038/oby.2009.167

348 Santacruz A, Marcos A, Wärnberg J, *et al.* Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity* 2009;**17**:1906–15. doi:10.1038/oby.2009.112

349 Sonnenberg A, Delcò F, Inadomi JM. Cost-Effectiveness of Colonoscopy in Screening for Colorectal Cancer. *Ann Intern Med* 2000;**133**:573–84. doi:10.7326/0003-4819-133-8-200010170-00007

350 McGrath JS, Ponich TP, Gregor JC. Screening for colorectal cancer: the cost to find an advanced adenoma. *Am J Gastroenterol* 2002;**97**:2902–7. doi:10.1111/j.1572-0241.2002.07059.x

351 Telford JJ, Levy AR, Sambrook JC, *et al.* The cost-effectiveness of screening for colorectal cancer. *Can Med Assoc J* 2010;**182**:1307–13.

352 Wong MC, Ching JYL, Chan VCW, *et al.* The comparative cost-effectiveness of colorectal cancer screening using faecal immunochemical test vs. colonoscopy. *Nat Sci Reports* 2015;**5**:1–9. doi:10.1038/srep13568

353 Garcia M, Mila N, Binefa G, *et al.* False-positive results from colorectal cancer screening in Catalonia (Spain), 2000-2010. *J Med Screen* 2012;**19**:77–82. doi:10.1258/jms.2012.012013

354 Vasen HFA, Tomlinson I, Castells A. Clinical management of hereditary colorectal cancer syndromes. *Nat Rev Gastroenterol Hepatol* 2015;**12**:88–97. doi:10.1038/nrgastro.2014.229

355 Moreira L, Balaguer F, Lindor N, *et al.* Identification of Lynch Syndrome Among Patients With Colorectal Cancer. *JAMA* 2012;**308**:1555–65. doi:10.1001/jama.2012.13088

356 Martín-López J V, Fishel R. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. *Fam Cancer* 2013;**1**:159–68. doi:10.1007/s10689-013-9635-x

357 Vasen H, Nagengast F, Meera Khan P. Interval cancers in hereditary non-polyposis colorectal cancer (Lynch syndrome). *Lancet* 1995;**345**:1183–4.

358 Vasen H, Blanco I, Aktán-Collán K, *et al.* Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* 2013;**62**:812–23. doi:10.1136/gutjnl-2012-304356

359 Seppälä T, Pylvänäinen K, Evans DG, *et al.* Colorectal cancer incidence in path_MLH1 carriers subjected to different follow-up protocols: a Prospective Lynch Syndrome Database report. *Hered Cancer Clin Pract* 2017;**15**. doi:10.1186/s13053-017-0078-5

360 Simona Pino M, Mino-Kenudson M, Mandes Wildemore B, *et al.* Deficient DNA Mismatch Repair Is Common in Lynch Syndrome-Associated Colorectal Adenomas. *J Mol Diagnostics* 2009;**11**:238–47. doi:10.2353/jmoldx.2009.080142

361 Michielsen AJ, Hogan AE, Marry J, *et al.* Tumour Tissue Microenvironment Can Inhibit Dendritic Cell Maturation in Colorectal Cancer. *PLoS One* 2011;**6**:e27944. doi:10.1371/journal.pone.0027944

362 Zou S, Fang L, Lee M-H. Dysbiosis of gut microbiota in promoting the

development of colorectal cancer. *Gastroenterol Rep* 2018;**6**:1–12. doi:10.1093/gastro/gox031

363 Flemer B, Lynch DB, Brown JMR, *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* 2017;**66**:633–43. doi:10.1136/gutjnl-2015-309595

364 Zackular JP, Baxter NT, Chen GY, *et al.* Manipulation of the Gut Microbiota Reveals Role in Colon Tumorigenesis. *Am Soc Microbiol* 2015;**1**:e00001-15.

365 Ohigashi S, Sudo K, Kobayashi D, *et al.* Significant Changes in the Intestinal Environment After Surgery in Patients with Colorectal Cancer. *J Gastrointest Surg* 2013;**17**:1657–64. doi:10.1007/s11605-013-2270-x

366 Zhang JW, Du P, Gao J, *et al.* Preoperative probiotics decrease postoperative infectious complications of colorectal cancer. *Am J Med Sci* 2012;**343**:199–205. doi:10.1097/MAJ.0b013e31823aace6

367 Usami M, Miyoshi M, Kanbara Y, *et al.* Effects of perioperative synbiotic treatment on infectious complications, intestinal integrity, and fecal flora and organic acids in hepatic surgery with or without cirrhosis. *J Parenter Enter Nutr* 2011;**35**:317–28. doi:10.1177/0148607110379813

368 Howard BM, Kornblith LZ, Christie SA, *et al.* Characterizing the gut microbiome in trauma: significant changes in microbial diversity occur early after severe injury. *Trauma Surg Acute Care Open* 2017;**2**:e000108. doi:10.1136/tsaco-2017-000108

369 Stavrou G, Kotzampassi K. Gut microbiome, surgical complications and probiotics. *Ann Gastroenterol* 2017;**30**:45–53. doi:10.20524/aog.2016.0086

370 Pullen LC. Gut Microbiome Rebounds After Colonoscopy Bowel Preparation. Medscape Med. news. 2015.https://www.medscape.com/viewarticle/837467 (accessed 23 Mar 2018).

371 Janz T, Lu K, Povlow MR, *et al.* A Review of Colorectal Cancer Detection Modalities, Stool DNA, and Fecal Immunochemistry Testing in Adults Over the

Age of 50. *Cureus* 2016;**8**:e931. doi:10.7759/cureus.931

372 Quintero E, Parra-Blanco A. Noninvasive diagnostic tools in colorectal cancer mass screening. *Curr Colorectal Cancer Rep* 2007;**3**:29–34. doi:10.1007/s11888-007-0013-7

373 Dove-Edwin I, Sasieni P, Adams J, *et al.* Prevention of colorectal cancer by colonoscopic surveillance in individuals with a family history of colorectal cancer: 16 year, prospective, follow-up study. *BMJ* 2005;**331**:e1047. doi:10.1136/bmj.38606.794560.EB

374 Sjöström O, Lindholm L, Melin B, *et al.* Colonoscopic surveillance – a cost-effective method to prevent hereditary and familial colorectal cancer. *Scand J Gastroenterol* 2017;**52**:1002–7. doi:10.1080/00365521.2017.1327615

375 Järvinen HJ, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 1995;**108**:1405–11. doi:10.1016/0016-5085(95)90688-6

376 Ni J, Wu GD, Albenberg L, *et al.* Gut microbiota and IBD: Causation or correlation? *Nat Rev Gastroenterol Hepatol* 2017;**14**:573–84. doi:10.1038/nrgastro.2017.88

377 Nishida A, Inoue R, Inatomi · Osamu, *et al.* Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clin J Gastroenterol* 2018;**11**:1–10. doi:10.1007/s12328-017-0813-5

378 Reinoso Webb C, Koboziev I, Furr KL, *et al.* Protective and Pro-Inflammatory Roles of Intestinal Bacteria. *Pathophysiology* 2016;**23**:67–80. doi:10.1016/j.pathophys.2016.02.002

379 Rivière A, Selak M, Lantin D, *et al.* Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Front Microbiol* 2016;**7**:e979. doi:10.3389/fmicb.2016.00979

380 Lokhart-Mummery J, Dukes C. The precancerous changes in the rectum and colon. *Surgery, Gynecol Obstet* 1927;**36**:591–6.

381 Gilbertsen V, Nelms J. The prevention of invasive cancer of the rectum. *Cancer*

1978;**41**:1137–9.

382 Greegor DH. Diagnosis of Large-Bowel Cancer in the Asymptomatic Patient. *JAMA J Am Med Assoc* 1967;**201**:943–5. doi:10.1001/jama.1967.03130120051012

383 Wolff WI, Shinya H. Polypectomy via the fiberoptic colonoscope. *N Engl J Med* 1973;**3288**:329–32. doi:10.1056/nejm198811243192103

384 Commission E. European guidelines for quality assurance in colorectal cancer screening and diagnosis. 2010.

385 Hol L, Wilschut JA, Van Ballegooijen M, *et al.* Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. *Br J Cancer* 2009;**100**:1103–10. doi:10.1038/sj.bjc.6604961

386 Haug U, Kuntz KM, Knudsen AB, *et al.* Sensitivity of immunochemical faecal occult blood testing for detecting left- vs right-sided colorectal neoplasia. *Br J Cancer* 2011;**104**:1779–85. doi:10.1038/bjc.2011.160

387 Brenner H, Tao S, Haug U. Low-dose aspirin use and performance of immunochemical fecal occult blood tests. *JAMA* 2010;**304**:2513–20.

388 Thomas AM, Manghi P, Asnicar F, *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* 2019;**In press**. doi:10.1038/s41591-019-0405-7

389 Lamb YN, Dhillon S. Epi proColon® 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer. *Mol Diagnosis Ther* 2017;**21**:225–32. doi:10.1007/s40291-017-0259-y

390 Volition Rx. Volition RX | Nu-Q. https://volitionrx.com/ (accessed 23 Apr 2019).

391 Rahier J-F, Druez A, Faugeras L, *et al.* Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer. *Clin Genet* 2017;**9**:1–7. doi:10.1186/s13148-017-0351-5

392 Universal Diagnostics DX. Universal DX. https://www.universaldx.com/ (accessed 18 Apr 2019).

393 Amadix. Amadix. Right Treatment, Right Patient. http://www.amadix.com/ (accessed 18 Apr 2019).

394 Goedert JJ, Gong Y, Hua X, *et al.* Fecal Microbiota Characteristics of Patients with Colorectal Adenoma Detected by Screening: A Population-based Study. *EBioMedicine* 2015;**2**:597–603. doi:10.1016/j.ebiom.2015.04.010

395 Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol* 2009;**67**:6–20. doi:10.1111/j.1574-6941.2008.00629.x

396 Liu P, Liu Y, Wang J, *et al.* Detection of fusobacterium nucleatum and fada adhesin gene in patients with orthodontic gingivitis and non-orthodontic periodontal inflammation. *PLoS One* 2014;**9**:3–7. doi:10.1371/journal.pone.0085280

397 Picchianti-Diamanti A, Panebianco C, Salemi S, *et al.* Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int J Mol Sci* 2018;**19**:2938. doi:10.3390/ijms19102938