UNIVERSITÀ DI SIENA 1240

Department of Medical Biotechnologies

Doctorate in Genetics, Oncology and Clinical Medicine

XXXVI° Cycle

Coordinator: Prof.ssa Ilaria Meloni

Association of endogenous and exogenous factors with early-onset colorectal cancer: germline mutations, epigenetic modifications, diet and lifestyle habits

PhD Candidate
Marta Puzzono

Signed by Marta Puzzono

Tutor
Andrea Galli
University of Florence, Florence, Italy

Co-Tutor
Giulia Martina Cavestro
Vita-Salute San Raffaele University, Milan, Italy

A.Y. 2022/2023

University of Siena
Doctorate in Genetics, Oncology and Clinical Medicine
XXXVI° Cycle

Final exam date:
20 March 2024

Commission:

Andrea Galli

Cavestro Giulia Martina

Oxana Bereshchenko

Jose Miguel Lizcano

Carlotta De Filippo

INDEX

ABSTRACT **5**
INTRODUCTION **8**
1.1 Early-onset Colorectal Cancer (eoCRC) 8
1.1.1 Definition and epidemiology 8
1.1.2 Pathophysiology of CRC 11
1.1.3 Clinical and histopathologic characteristics 13
1.1.4 Diagnosis and treatment 14
1.2 Non-modifiable endogenous risk factors 15
1.2.1 Family history of CRC 15
1.2.2 Hereditary cancer syndromes 16
1.3 Modifable exogenous risk factors 18
1.3.1 Diet 18
1.3.2 Physical activity 22
1.3.3 Obesity 24
1.3.4 Smoking 26
1.4 Epigenetics and CRC 28
1.4.1 DNA methylation 29
1.4.2 Histone modifications 29
1.4.3 Non-coding RNAs 30
1.4.3.1 microRNAs 30
1.4.3.2 long non-coding RNAs 34
AIMS **35**
MATERIALS AND METHODS **37**
3.1 Genetic risk assessment 37
3.2 DEMETRA project 37
3.2.1 Development of the online DEMETRA platform 38
3.2.2 The SQFFQ's validation study 39
3.2.2.1 Internal validation: SQFFQ's repeatability 39
3.2.2.2 External validation: SQFFQ's validity 40

3.3 Genome-wide discovery and identification of a miRNA signature 41
3.3.1 Patient cohorts 41
3.3.2 Nucleic acid isolation and miRNA expression analysis 42
3.3.3 Statistical analysis 43
3.3.3.1 Candidate miRNA selection 43
3.3.3.2 Candidate miRNA sequencing data analysis 43
3.3.3.3 Machine learning approach (XGBoost) 44
3.3.3.4 XGB model building, parameters hyper-tuning and application 46
3.3.3.5 Training and validation cohort results processing 47
3.3.3.6 Survival evaluation 48
3.3.3.7 Decision curve analysis and net benefit analysis 48
RESULTS **50**
4.1 Endogenous risk factors 50
4.2 Exogenous risk factors: DEMETRA project 51
4.2.1 Internal validation: SQFFQ's repeatability 51
4.3 miRNA signature to predict recurrence in stage I-III eoCRCs 55
4.3.1 Discovery phase 55
4.3.2 Training and validation phases 57
DISCUSSION **65**
5.1 Genetic risk assessment 65
5.2 SQFFQ repeatability 67
5.3 miRNA signature as prognostic biomarker 69
CONCLUSIONS AND FUTURE PERSPECTIVES **74**
6.1 Genetic risk assessment 74
6.2 The international case-control study 74
6.3 miRNA signature and personalized medicine 76
REFERENCES **78**

ABSTRACT

The incidence of early-onset colorectal cancer (eoCRC), defined as colorectal cancer occurring in young adults under the age of 50, is increasing globally. However, knowledge of the etiological factors and characteristics of CRC in young adults is far from complete.

Estimating the impact of pathogenic variants (PVs) in eoCRC predisposition remains an active field of research. *Therefore the 1st aim was to evaluate the associations of germline PVs and family history of CRC with eoCRC*. A total of 105 eoCRCs were enrolled (mean age at diagnosis of 41.2 ± 6.7 years; 48.6% females, 51.4% males) for genetic testing through next-generation sequencing and multiplex ligation-dependent probe amplification. 20% of eoCRC carried a germline PV of genes known to be associated with CRC, of which 12.4% of mismatch repair genes, 2.9% of BRCA1-2, 3.8% of MUTYH, 0.9% of ATM, 0.9% of SDHAF2. One patient exhibited mosaicism of PVs in MSH2/MUTYH genes. 71.4% of eoCRCs didn't have a family history of CRC; 19% of eoCRCs reported having a first degree relative (FDR) with CRC and 12.4% had a second degree relative (SDR) with CRC; three patients had both a FDR and SDR with CRC and were all Lynch patients. When comparing mutated-eoCRCs with non-mutated eoCRCs, no statistically significant differences were found in terms of age at diagnosis or sex, location of CRC, presence of FDR with CRC. Mutated eoCRC differed significantly from non-mutated eoCRCs in terms of SDRs with CRC ($p < 0.001$).
In conclusion, 20% of eoCRCs are caused by germline pathogenic variants (PV) and a negative family history does not exclude hereditary cancer syndromes. Therefore, thorough family history should be routinely collected for all individuals with eoCRC and all patients with eoCRC should be offered multi-gene panel germline genetic testing. Identification of LS individuals as well as carriers of other relevant germline PVs (e.g. BRCA2, MUTYH) will allow at-risk individuals to take personalised preventive measures for both proband and relatives.

Questionable eoCRCs' exogenous risk factors are represented by diet and lifestyle, even though a still scant literature is available to date. Moreover, most studies are small, heterogeneous, focused exclusively on peculiar dietary and drinking habits of single countries without analyzing cooking, processing, and storage techniques. Therefore, *our 2nd aims were to (i) develop a unique and shared semi-quantitative food frequency questionnaire (SQFFQ) able to accurately describe dietary and drinking habits of eoCRCs and healthy controls of different countries at global level that will be involved in the future DEMETRA study (international case-control study evaluating the association of dietary, lifestyle and anthropometric factors with eoCRC of countries with different eoCRC incidence); (ii) validate the SQFFQ, making data obtained from different dietary questionnaires comparable.* We designed an ad-hoc, shared, online SQFFQ to investigate the usual consumption of 329 foods, grouped into 61 food groups, over the past year. In addition to frequency of consumption, the tool investigates the portions habitually consumed using validated photographs, household measures, and standard units, as well as types of

seasoning and methods of cooking. A special software was then developed to analyze responses and link them to food composition tables in order to provide a nutritional breakdown of individual and collective diets. The SQFFQ was then validated for repeatability by administering it twice, 3 weeks apart, to a sample of 30 young adults under 50 years (Internal Validation). To evaluate repeatability, the measurement error for each food group was estimated as the percentage change between the estimates of food consumption for the same individual. Afterwards, the agreement between the two measurements for each food group was measured with Cohen's kappa coefficient. Agreement levels, represented by the calculation of Cohen's kappa coefficient, were as follows: 12 food groups showed fair/sufficient agreement (Cohen's kappa 20-40), 23 foods exhibited good/moderate agreement (Cohen's kappa >40-60), and 22 food groups demonstrated high substantial agreement (Cohen's kappa >60). Therefore, the ad hoc designed SQFFQ provides a reasonably repeatable measure of dietary intake and can be used to assess the dietary and drinking habits of volunteers in this age group. Indeed, most staple foods in the Italian diet, including pasta, fruit, vegetables, legumes, eggs, meat, coffee, and tea, are well estimated by the SQFFQ. However, as yet described in literature, challenges persist in estimating the consumption of foods assumed sporadically (such as snacks) or in small quantities such as spices. Definitive conclusions will be drawn after the completion of the ongoing External validation, involving 100 volunteers from the same age group. In this phase, the SQFFQ will be validated against the gold standard, represented by a 4 days-food diary followed by a dietary recall. Once validation is complete, the international, multicenter, case-control study will start to evaluate the associations of diet, lifestyle and anthropometric factors with eoCRC comparing patients from countries with different incidence of eoCRC.

Epigenetic changes are crucial in the pathogenesis of CRC, representing the missing link between CRC, specific gene expression patterns and the absence of genetic alterations. One of the most important epigenetic modifications involved in carcinogenesis is represented by an altered expression of microRNAs (miRNAs). Distinct miRNAs could be differentially expressed in patients with recurrent vs. non-recurrent eoCRC and used as simple predictive biomarkers to select the optimal post-treatment surveillance regimen in managing patients with stage I-III eoCRC. Therefore, the *3rd aim was to identify candidate miRNAs that were differentially expressed in eoCRC patients with and without recurrence and define which patients could benefit most from more aggressive surveillance.* We employed a five-layer approach for the development of a simple and clinically feasible test that may be translated into clinical practice. The first phase of the study (Discovery) consisted in the systematic interrogation and profiling of miRNA expression levels in 20 formalin-fixed paraffin-embedded (FFPE) samples of stage II-III eoCRCs that did (n 10) or did not (n 10) develop recurrence in five years following curative-intent surgery. In phase two (Assay development), we performed several bioinformatic analyses to identify the best candidate miRNAs that were differentially expressed in eoCRCs with and without recurrence and provided the highest discriminatory power between the two groups. At the end of phase two, we selected 10 best performing miRNAs and quantified their expression via qPCR. This third phase utilized 88 FFPE from a larger, independent training cohort of stage I-III eoCRC

who received curative-intent surgery (24 recurrent and 63 non-recurrent eoCRCs). Because there is no unique and universally accepted normalized miRNA, we employed several bioinformatic approaches to rigorously establish the ideal candidate based on intra- and inter-group expression stability. In the Assay Training phase, we optimized and trained an advanced machine learning algorithm (XGBoost) to predict the development of eoCRC recurrence based on RT-qPCR data and performed several interrogations to the model to understand its functioning. The optimized XGBoost-based 9-miRNAs risk-assessment model demonstrated a high accuracy in predicting recurrence in stage I-III eoCRCs in the training cohort with an AUC value of 0.90 (95% CI 83-95%) with a Youden index of 64.9% (CI 95%, 55%-82%), an accuracy of 81.8% (77-93%), sensitivity 84.0% (65-96%), specificity 81.0% (72-98%). We then evaluated the survival characteristics of the XGB model and observed that our 9-miRNA model can discriminate effectively between recurrent and non-recurrent cases up to 20 years after surgical resection: patients predicted to be at a high risk of recurrence by the XGB model had a statistically significant cumulative hazard of disease recurrence than those classified as low-risk ($p<0.001$). Finally, we performed an independent validation of our assay in a distinct and ethnically different validation cohort of 69 FFPE of stage I-III eoCRCs who received curative-intent surgery for eoCRC (9 recurrent and 60 non-recurrent). The XGBoost-based risk-assessment model, incorporating 9-miRNAs, exhibited good accuracy in predicting recurrence among stage I-III eoCRCs in the validation cohort, achieving an AUC value of 0.77 (95% CI 67.0-87.0%), with a sensitivity of 100% (88-100%), specificity of 62.9% (55-82%), accuracy 66.7% (59.0-83.0%), and Youden index of 62.9% (54-73%). To truly gauge the effects that this assay would have in a real-world scenario, we performed several decision-curve analyses. We observed a net benefit of the surveillance based on our 9-miRNA signature compared to the clinical-based surveillance especially in stage I and II high risk eoCRC, representing the subgroups of patients that could benefit more from more aggressive post-treatment follow-up strategies. Therefore, if confirmed in prospective trials, this 9-miRNA signature could establish the fundamentals of personalized medicine.

INTRODUCTION

1.1 Early-onset Colorectal Cancer (eoCRC)

1.1.1 Definition and epidemiology

Early-onset colorectal cancer (eoCRC) is defined as a colorectal cancer (CRC) diagnosed under the age of 50 years [1]. CRC screening typically commenced at age 50 for individuals at average-risk in the United States (US), leading to the recognition of CRC diagnosed in individuals younger than 50 years as early- or young-onset CRC in the literature.
Initially reported in the US [2], there is substantial evidence supporting a global increase in the incidence of eoCRC [3-9] (Fig. 1).

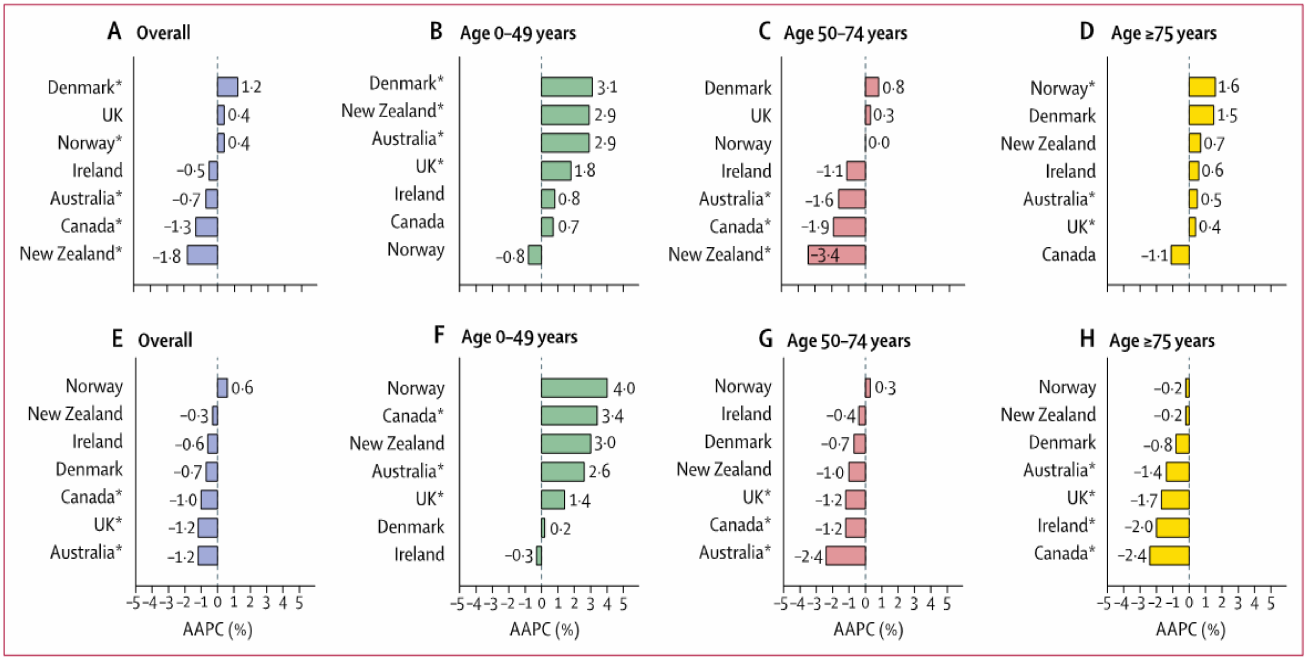


Figure 1: Average annual percentage change (AAPC) in the incidence of colon and rectal cancer by age group. A-D) AAPC of colon cancer; E-H) AAPC of rectal cancer. *p<0.05 [4].

Across Europe, there has been a significant increase in the incidence of early-onset colorectal cancer (eoCRC) among individuals aged 20–39 years in 12 out of 20 countries over the past 25 years. These countries include Belgium, Germany, the Netherlands, the UK, Norway, Sweden, Finland, Ireland, France, Denmark, the Czech Republic, and Poland. Similarly, in the 40–49 age group, a comparable rise was observed in 8 out of 20 countries, namely the UK, Greenland, Sweden, Slovenia, Germany, Finland, Denmark, and the Netherlands (Fig. 2). Conversely, Italy has experienced a decreasing trend, while no significant changes were reported in the remaining European countries [7].

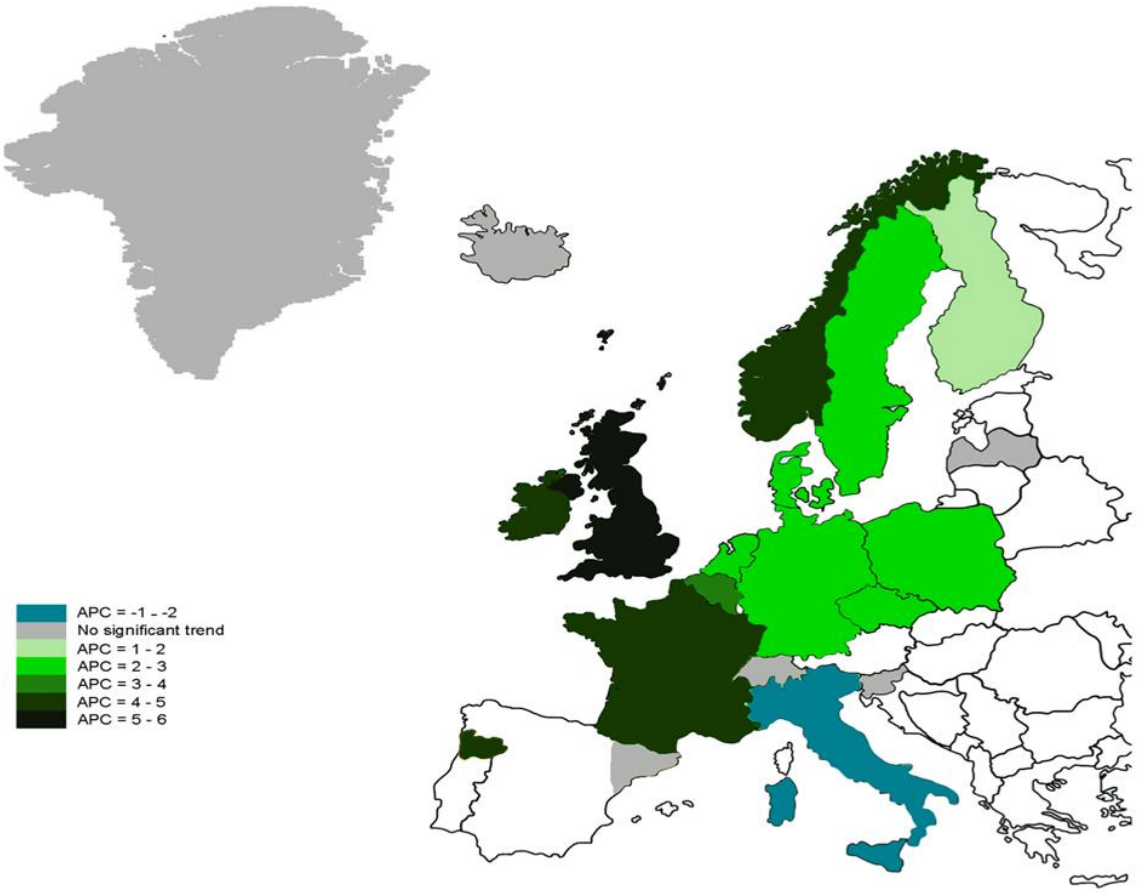


Figure 2: Annual percentage change (APC) in CRC incidence in Europe in young adults aged 20-39 y, 1990-2016 [7]

This decreasing trend of eoCRCs in Italy, together with a decline of late-onset CRCs (loCRCs), was corroborated by data gathered from 48 Italian cancer registries spanning the period from 2003 to 2014, encompassing 60% of the population and nearly 15 million individuals aged 20–49 years [8,9] (Fig. 3).

CRC rates have experienced an increase among subjects born in the US since the 1960s [10]. Thereafter, a birth cohort effect is almost evident on a global scale, despite variation in population characteristics and screening approaches worldwide. Generation X, encompassing those born between 1965 and 1980, witnessed an initial eoCRC rise [11,12], with rates continuing to escalate after reaching the age of 50 [13–15]. In particular, individuals born in 1965–1969 exhibit a 1.22-fold (95% CI 1.15-1.29) increase in rates, while those born in 1975–1979 show a 1.58-fold (95% CI 1.43-1.75) increase, compared to those born between 1950 and 1954 [16]. Similar trends were experienced by millennials, born between 1981 and 1996, with incidence rates 1.89-fold (95% CI 1.65, 2.51) and 2.98-fold (95% CI 2.29, 3.87) higher among individuals born in 1980–1984 and 1990–1994, respectively, compared to those born between 1950 and 1954 [16].

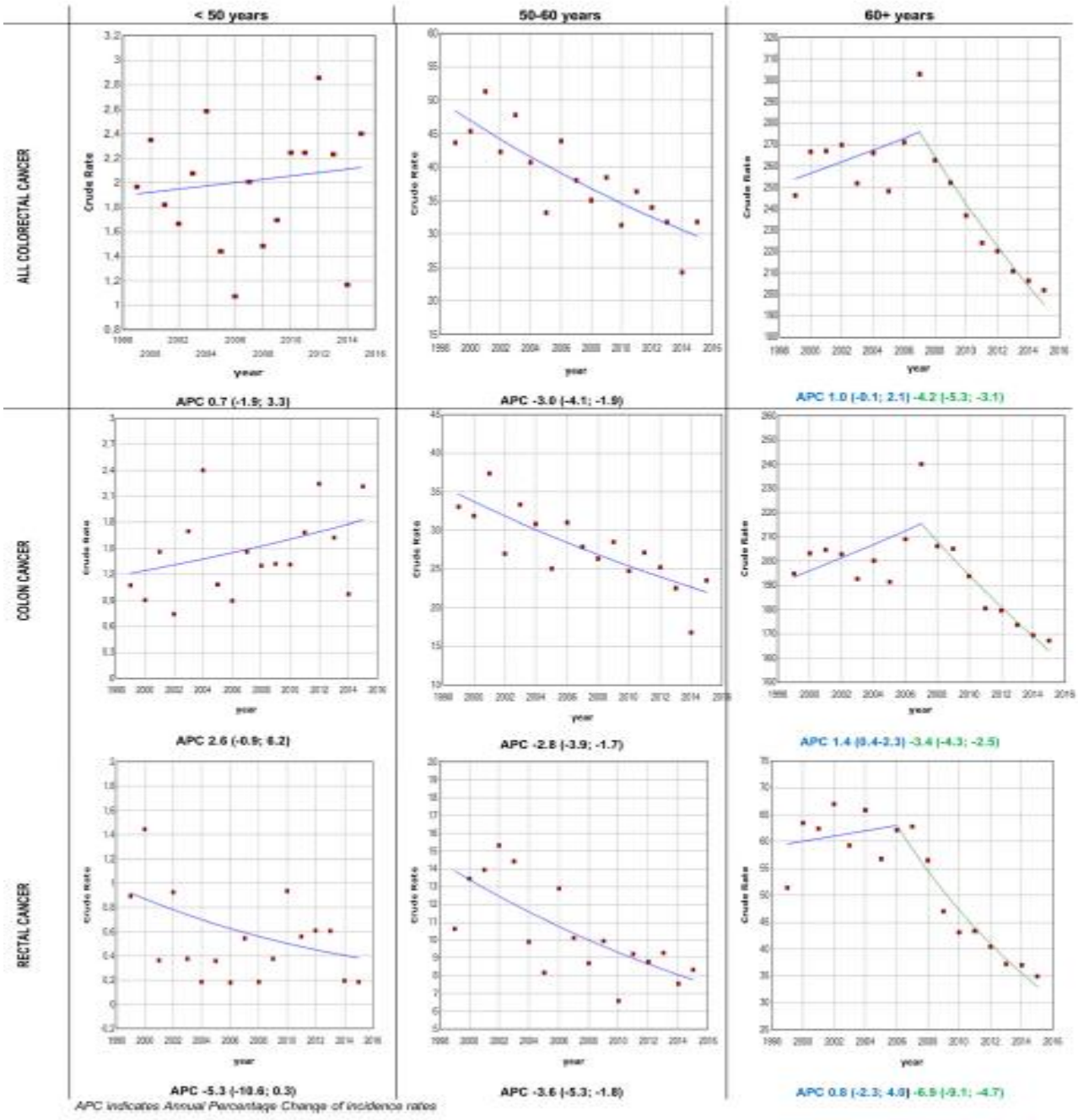


Figure 3: Trends in age specific annual incidence rates per 100000 for CRC (upper panel), colon cancer (middle panel) and rectal cancer (lower panel) [5]

It was estimated that approximately 10% of all new diagnoses of CRC are represented by eoCRCs and that in the next 10 years, 25% of young onset CRCs will be rectal cancers and 10–12% colon cancers; moreover, an increase in CRC–related mortality has also been described among younger patients in contrast to the steady decline in the incidence of loCRCs and related mortality over the past two decades. This prompted the U.S Preventive Services Task Force and the American Cancer Society to decrease the recommended age to initiate CRC screening from 50 years to 45 years [17-22].

1.1.2 Pathophysiology of CRC

CRC is considered a heterogeneous disease in terms of pathophysiological pathway of carcinogenesis, prognosis and treatment. Three different pathways of colorectal carcinogenesis are recognized: the chromosomal instability (CIN) pathway, the microsatellite instability (MSI) pathway, and the CpG island methylator phenotype (CIMP) pathway.

The majority of CRCs stem from the *CIN pathway*, accounting for approximately 65%–70% of sporadic CRCs. The CIN pathway is characterized by alterations of whole or large portions of chromosomes, leading to imbalances in chromosome number (aneuploidy), sub-chromosomal genomic amplifications, and frequent loss of heterozygosity (LOH), that in turn result in the over-activation of growth pathways and/or down-regulation of apoptotic pathways. As described in 1990 by Fearon and Volgelstein, this colorectal carcinogenesis begins with inactivating mutations in the adenomatous polyposis coli (APC) tumor suppressor gene (effector in WNT pathway), typically observed in adenomatous polyps, and is followed by activating mutations of KRAS (receptor tyrosine kinase signaling). The progression to colorectal adenocarcinomas occurs through the accumulation of further inactivating mutations in SMAD4 (involved in TGF-beta signaling) and TP53 (responsible for cell cycle control) [23,24] (Fig. 4).

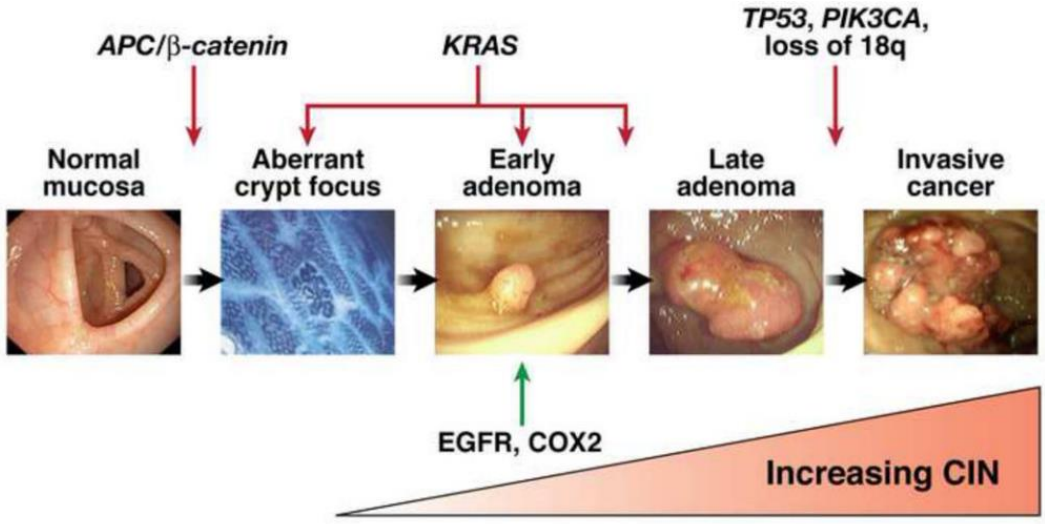


Figure 4: Multistep genetic model of colorectal carcinogenesis [25]

The *MSI pathway*, characterized by alterations in the number of microsatellites which are short repeated sequences spread throughout the genome, arises from a deficiency in the DNA mismatch repair (MMR). This MMR deficiency can be associated with: (i) germline pathogenic variants of MMR genes (MLH1, MSH2, MSH6 and PMS2); (ii) biallelic hypermethylation of the MLH1 promoter, resulting in MLH1 inactivation; (iii) double somatic mutations in MMR genes [26–28].

The 3rd pathway is the *CIMP path*, which exhibits gene silencing due to CpG islands hypermethylation at the promoter regions of several tumor suppressor genes. This pathway

is associated with the serrated colorectal carcinogenesis, with proximal colonic location, female gender, old age at diagnosis, poor histology and BRAF mutations [29,30].
These 3 pathways are not mutually exclusive. CRC can occasionally exhibit features of multiple pathways. To address this challenge and facilitate the correlation between CRC cell phenotype and clinical behavior for targeted treatments, the CRC Subtyping Consortium unified six independent molecular classification systems [31–36], based on gene expression data, into a single consensus system with four groups, called Consensus Molecular Subtypes (CMS 1-4) [37], based also on epigenomic, transcriptomic, microenvironmental, genetic, and clinical characteristics of the tumors (Fig. 5-6).

	CMS1	CMS2	CMS3	CMS4
Alternate Name	Microsatellite Instability Immune	Canonical	Metabolic	Mesenchymal
Primary Characteristics	Hypermutated, microsatellite unstable and strong immune activation	Epithelial, marked WNT and MYC signaling activation	Epithelial and evident metabolic dysregulation	Prominent TGF–β activation, stromal invasion and angiogenesis. +/− WNT
Incidence	14%	37%	13%	23%
Genomic Associations	MSI, high mutation count, low copy number	Chromosomal instability (CIN), low-moderate mutation count and copy number	CIN, moderate mutation count, low-moderate copy number	CIN, low mutation count, high copy number
Precursor Lesions	Serrated (low TGFβ microenvironment)	Tubular adenoma	Tubulovillous adenoma with serrated features[21]	Serrated (high TGFβ microenvironment)
Epigenomic Associations	High methylation	Low methylation	Moderate methylation	Low methylation
Transcriptomic Pathways	Immune activation, JAK-STAT activation, Caspases	WNT targets, MYC activation, EGFR or SRC activation, VEFG or VEGFR activation, Integrin activation, TGFB activation, IGF and IRS2 activation, HNF4a, HER2 and cyclin upregulation	DNA damage repair, Glutaminolysis, Lipidogenesis, Cell cycle	Mesenchymal activation, complement activation, immunosuppression, integrins
Stroma-Immune Microenvironment	Few CAF, Highly immunogenic, large immune infiltrate, tends towards adaptive immune response	Very few CAF, Poorly immunogenic, Tends toward innate immune response	Few CAF, Highly immunogenic, tends toward adaptive immune response	Many CAF, inflamed, tends toward innate immune response, Epithelial to Mesenchymal Transition
Associated mutations	MSH6, RNF43, ATM, TGFBR2, BRAF, PTEN	APC, KRAS, TP53, PIK3CA	APC, KRAS, TP53, PIK3CA	APC, KRAS, TP53, PIK3CA

Figure 5. Consensus Molecular Subtypes of Colorectal Cancer (CMS1-4): molecular characteristics [38]

	CMS1	CMS2	CMS3	CMS4
Stage at Diagnosis				
(%)	12	13	17	8
I	44	40	41	33
II	40	39	37	47
III	4	8	5	12
IV				
Grade				
1	15	22	20	9
2	40	73	68	72
3	45	5	12	19
Histopathologic Associations	Solid, trabecular, mucinous features	Tubular	Papillary	Prominent desmoplasia, stroma
Age (years)	69	66	67	64
Sex	44% M, 56% F	58% M, 42% F	53% M 47% F	55% 45%
Location	Proximal	Distal	Mixed	Distal

Figure 6. Consensus Molecular Subtypes of Colorectal Cancer (CMS1-4): clinical associations [38]

1.1.3 Clinical and histopathologic characteristics

eoCRC exhibits a distinct anatomical localization, histopathology, and clinical presentation compared to loCRC [39]. Approximately 70% of eoCRC manifest in the left colon, primarily in the rectum, followed by the left colon (sigmoid and descending colon) [3,40–43]. Conversely, loCRCs occur with comparable frequencies in the proximal colorectum.
eoCRCs typically show a higher proportion of signet ring and mucinous histology [3,40,44], along with microsatellite instability (MSI-H) which is strongly associated with poor tumor differentiation [45]. In eoCRCs, MSI-H tumors are more frequently associated with germline pathogenic variants in mismatch repair (MMR) genes [46–48], as opposed to epigenetic silencing of the MLH1 MMR gene, which is prevalent in sporadic MSI-H tumors observed in individuals with loCRC. [49].
eoCRCs are often diagnosed at advanced TNM stages, typically stages III–IV [50–53]. This presentation with a more advanced stage of disease has been attributed to prolonged symptoms duration at diagnosis and delayed diagnosis, ranging from 7 to 9 months, as compared to loCRC [50,54,55]. However, recent evidence suggest that it might not solely be explained by diagnostic delays [50,56]. Indeed, findings from one study indicate that stage III/IV eoCRCs often present with alarming symptoms that prompt rapid endoscopic evaluation compared to stage I/II eoCRCs [50].

eoCRC typically presents with hematochezia (46%), iron deficiency anemia (13.0%), and weight loss (10.0%) [57–62]. Hematochezia and iron deficiency anemia carry hazard ratios

of 10.66 and 10.81 for eoCRC, respectively, with a greater risk observed in men and individuals aged 40–49 compared to those under 30 years old [63].

Patients with eoCRC referring hematochezia had more rectal cancers, compared with those presenting with iron deficiency anemia (38% vs 20%, respectively). A case-control study in which 40% of eoCRCs were rectal cancers, weight loss of 5 kg within 5 years was linked to increased odds of eoCRC [59]. Additional symptoms at diagnosis include abdominal pain, abdominal distention, changes in bowel habits, and fatigue [41,42,55,58,64,65]. In a single-center study carried out in Italy on a cohort of 54 eoCRCs and 494 loCRCs, 25% eoCRCs presented with symptoms at diagnosis, in contrast to only 8.8% of loCRCs (P = 0.01) [41]. This is the reason why, the recent international guidelines on the management of eoCRC (DIRECt) [1] state that symptoms and signs that should prompt evaluation for eoCRC include (but are not limited to) any of the following: hematochezia, unexplained iron deficiency anemia, or unexplained weight loss.

1.1.4 Diagnosis and treatment

As reported in the DIRECt guidelines, colonoscopy is recommended for the diagnostic evaluation of individuals with hematochezia, unexplained iron deficiency anemia, or unexplained weight loss. Colonoscopy should be complete to the cecum and of high quality and should be ideally within 30 days after referral to a healthcare professional [1].
The use of different diagnostic techniques for symptomatic young individuals remains a subject of debate. Recent studies have reported that fecal immunochemical test (FIT) performs well in both symptomatic and asymptomatic subjects under 50y [66–68]. Nonetheless, a positive FIT always requires a subsequent colonoscopy, potentially leading to diagnostic delay and possibly with an increased risk of advanced-stage disease [69,70]. Consequently, FIT is not recommended for symptomatic patients. However, using FIT to triage young patients exhibiting low-risk symptoms, such as changes in bowel habits or abdominal pain, could be considered. Conversely, for individuals presenting high-risk symptoms like hematochezia, unexplained iron deficiency anemia, or unexplained weight loss, diagnostic colonoscopy remains the preferred approach [62].

In terms of treatment, the latest guidelines from the National Comprehensive Cancer Network (NCCN) advocate identical treatment plans for eoCRC and loCRC, whether in curative or palliative contexts [71].
Segmental colon resection with adequate lymph node dissection is the standard of care for patients with non-metastatic colon cancer, even though eoCRCs carry a slightly higher risk (1.4%) of metachronous CRC compared with loCRCs (0.6%) [72]. That risk of metachronous colon cancer is significantly higher and increases with time in eoCRCs with Lynch syndrome, in whom the surgical options include both segmental and extended resection (eg, total abdominal colectomy with ileorectal anastomosis). In these syndromic patients, even without proved difference in survival for segmental versus extended resection, the latter is associated with decreased risk of metachronous CRC [73,74].

Colonic resection with adequate lymph node dissection is encouraged as the initial approach for all patients with locally advanced disease. It is also encouraged, when feasible, for patients with metastatic disease with improved 5-year overall survival of 32–46% [75].
The current guidelines from DIRECt recommend similar systemic treatment for eoCRC and loCRC [1]. Considering that disease recurrence affects 20% of patients with stage II colon and rectal cancer, and 35% of those with stage III disease, the primary aim of adjuvant treatment is to shorten the therapy duration to minimize the risk of treatment-related side effects [76-78]. While the routine use of adjuvant chemotherapy in stage II colon cancer is not recommended [79], the standard adjuvant therapy for stage III colon cancer involves the combination of fluorouracil plus leucovorin (LV5FU2) with oxaliplatin (FOLFOX) or capecitabine and oxaliplatin (CAPOX) [76]. Adjuvant FOLFOX has been linked to a 7.5% absolute reduction in the risk of recurrence and a 4.2% absolute reduction in the risk of death in stage III eoCRC. On the other hand, patients with MSI-H tumors have a better prognosis and do not benefit from a fluoropyrimidine alone [80].

Neoadjuvant chemoradiotherapy with fluorouracil as a radiation sensitizer prior to surgical resection has been established as the standard treatment approach for locally advanced rectal cancer, resulting in enhanced locoregional control and patient compliance, and reduced toxicity [81-82]. The standard practice for surgical intervention involves a total mesorectal excision, using either a low anterior resection or an abdominoperineal resection with a permanent end colostomy, performed through an open, laparoscopic, or robotic approach. Nevertheless, rectal resection remains a procedure associated with significant morbidity, with postoperative complications reported in 34–58% of patients globally [83,84]. Considering this morbidity, Habr-Gama and colleagues [85] were the first to introduce the non-operative management for patients with rectal cancer experiencing clinical complete response after neoadjuvant chemoradiotherapy. The Authors have demonstrated that 26.7% of patients with resectable distal rectal cancer treated with neoadjuvant chemoradiotherapy had a clinical complete response and underwent observation only; of these patients, 2.8% developed an endoluminal recurrence and 4.2% experimented distant metastasis. 5-year overall survival was 88% in the resection group versus 100% in the observation group and disease-free survival was 83% in the resection group versus 92% in the observation group [85].

1.2 Non-modifiable endogenous risk factors

1.2.1 Family history

Family history represents a strong risk factor for eoCRC, increasing the relative risk to 4.21 (95% CI 2.61–6.79) [86]. This risk factor accounts for approximately 20–30% of eoCRC. [47,48,87–90].
There is a consensus that individuals with a family history of CRC should undergo more intensive surveillance than the general population, starting at an earlier age. In particular,

having at least 2 first-degree relatives (FDRs) with CRC and/or at least 1 FDR with CRC diagnosed before the age of 50–60 years are associated with a significant increase of CRC risk. Therefore, screening colonoscopy should start at 40 years (or 10 years before the youngest CRC), allowing prevention of up to 16% of eoCRC [89,91–95].
Hence, the recently published DIRECTt guidelines on eoCRCs emphasize the significance of family cancer history in evaluating risks associated with both syndromic and non-syndromic CRC. They recommend that a comprehensive family history be routinely obtained from all individuals [1]. Validated risk assessment instruments, such as the Colon Cancer Risk Assessment Tool and the PREMM5, can facilitate the collection of family history data and aid in identifying patients who may benefit from germline genetic testing [96,97]. Specifically, the PREMM5 tool enables the determination of the probability of a pathogenic variant (PV) or likely pathogenic variant (LPV) in a Lynch syndrome (LS) gene.
Nevertheless, the DIRECt guidelines recommend that all patients with eoCRC should be offered multi-gene panel germline genetic testing and genetic counseling for those with a positive germline finding. Genetic testing should be performed before treatment to maximize clinical utility, if feasible and without delaying treatment. Germline genetic testing for eoCRCs should include at a minimum: APC, BMPR1A, EPCAM, MLH1, MSH2, MSH6, MUTYH, POLD1, POLE, PMS2, PTEN, SMAD4, STK11, and TP53. If available and not cost-prohibitive, it should also include the following genes that could change clinical management: BRCA1, BRCA2, ATM, CHEK2, PALB2, BRIP1, BARD1, CDKN2A, CDH1, RAD51C, and RAD51D. Finally, it should also include the following genes associated with polyposis and CRC: AXIN2, GREM1, MLH3, MSH3, MBD4, NTHL1, RNF43, and RPS20 [1].

1.2.2 Hereditary Cancer Syndromes

With the introduction of next-generation sequencing (NGS), multigene panel testing has been implemented across various cohorts of cancer patients, including those diagnosed with eoCRC, revealing that 16–20% of eoCRC can be attributed to hereditary CRC syndromes [47,88,98] (Fig. 7).
Indeed, a study enrolling 450 eoCRCs at 51 hospitals in Ohio and using a multigene panel of 25 genes identified germline PVs in 16% of eoCRC; approximately half of them had LS [47]. It is noteworthy that one third of the mutated patients did not meet guideline-based criteria for genetic testing.
Another study involving 759 patients with eoCRC showed a 17.5% prevalence of germline PVs [42].

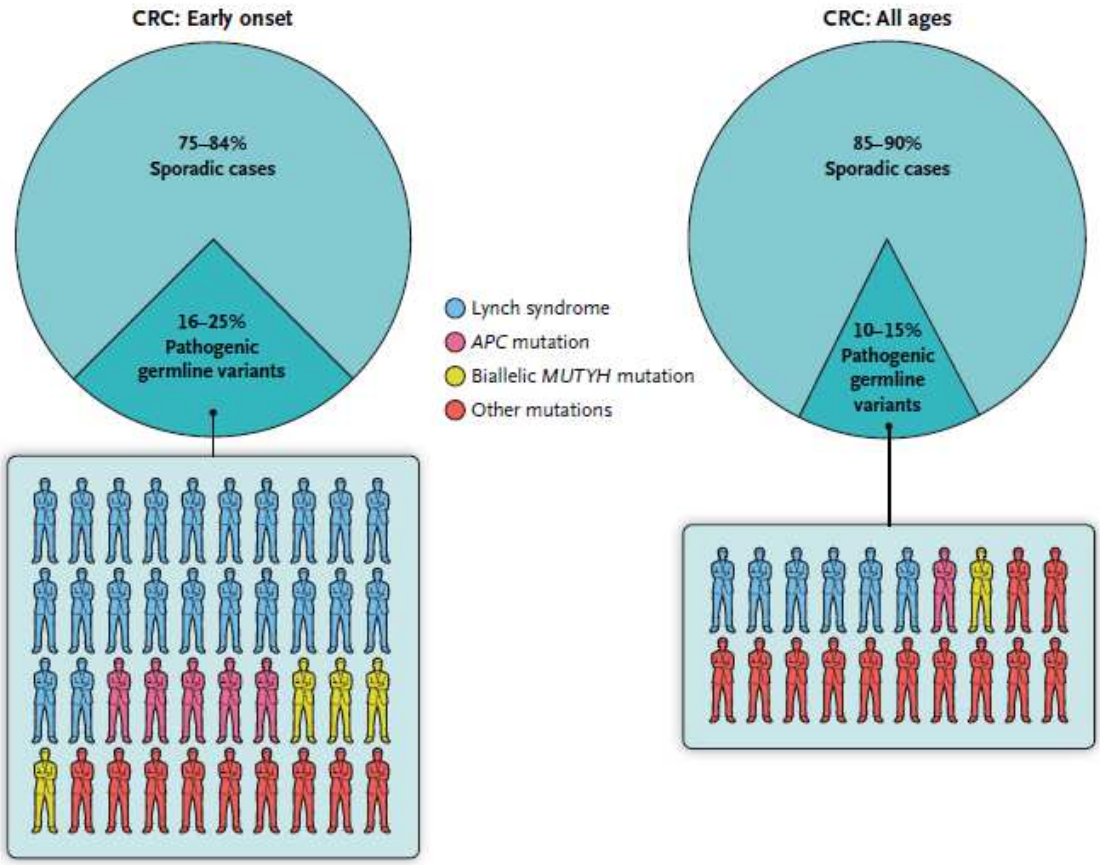


Figure 7: Prevalence of germline pathogenic variants in eoCRCs [99]

Among eoCRC patients, 2%–16% have LS, and up to 14% have PV/LPVs in other cancer susceptibility genes [46–48,98,100,101].

LS is the most common genetic disease among eoCRCs and is associated with PVs/LPVs of the DNA-mismatch repair (MMR) genes (MLH1, MSH2, MSH6, PMS2), as well as EPCAM 3' deletions. It is associated with a lifetime CRC risk of 40–80% and a mean age of CRC onset of 44-52 years. The PVs in MMR genes result in dMMR and MSI-H, being of therapeutic importance in patients with metastatic colorectal cancer, since those with dMMR or MSI-H are candidates for first line treatment with an immune checkpoint inhibitor [102].

Colorectal polyposis syndromes account for 2%–3% of eoCRC and include familial adenomatous polyposis (FAP, associated with PV/LPV in APC), MUTYH-associated polyposis (MAP, associated with biallelic PV/LPV in MUTYH), juvenile polyposis (JP, associated with PV/LPV in SMAD4, BMPR1A), and Peutz-Jeghers syndrome (PJS, associated with PV/LPV in STK11). Patients with FAP have approximately 100% risk of CRC, which results near to 69% in those with attenuated FAP (AFAP) [103,104]. In MAP, CRC risk is around 19% by age of 50 y and increases to 43% by age of 60 y [105,106]. Finally, PJS subjects carry a 39% risk of CRC [106,107], while JP a 10–38% lifetime risk of CRC [103,108].

While studies on the newer genes associated with polyposis and CRC (GREM1, POLE, POLD1, AXIN2, MSH3, MLH3, MBD4, RNF43, and RPS20) are still scant, research on other highly actionable and high-penetrance genes that were not previously associated with CRC (TP53, BRCA1, BRCA2, and PALB2) has shown a higher prevalence than some of the known

polyposis genes [109–111]. There is also emerging evidence that ATM may be a CRC susceptibility gene [112].
The majority of hereditary CRC syndromes exhibit variability in terms of penetrance and expressivity [93,113,114] that, at least partially, is associated with different dietary habits [115–117]. Only a few preliminary studies have explored this hypothesis [118–120] with underwhelming results as shown in two studies involving LS cohorts [119,120]. CRC risk is not uniformly distributed among the four MMR genes [121–123] and diet could potentially explain disease variability in those genes associated with a lower risk of CRC. Therefore, there is still a need for well-designed and adequately powered studies in this area.

1.3 Modifiable exogenous risk factors

As hereditary gastrointestinal tumor syndromes contribute to only a small portion of eoCRCs, it's essential to explore the exogenous risk factors to gain a deeper understanding of eoCRC pathogenesis. Risk factors such as alcohol consumption, physical activity, red and processed meat, and adherence to a Western diet are widely recognized as loCRC risk factors [124–130]. However, to date, a very small number of well-designed studies have investigated dietary, lifestyle, and anthropometric risk factors for eoCRC and its precursors [59,131–137].

1.3.1 Diet

Imperiale et al. conducted the first multi-center retrospective case–control study [138] evaluating the association of dietary habits and advanced colorectal neoplasia (ACRN) in North America. The research used the 1998 block food-frequency questionnaire, a validated food-frequency questionnaire (FFQ) analyzing 28 nutrients, including total calories, daily fat consumption in grams per day, folate consumption in micrograms per day, percent of calories from fat, protein, and carbohydrates. Twenty ACRN, of which 11 eoCRCs, were compared with 54 age-matched controls without observing differences in any of the 28 nutrients of the questionnaire (Fig. 8).
Another multi-center case–control study carried out on an Italian/Swiss population of 329 eoCRCs and 1361 age-matched controls [134] used a validated FFQ to assess the usual consumption of 78 foods [139]. The study revealed a significant increase in eoCRC risk associated with high intake of processed meat, with an OR 1.56 for high tertile of intake compared to the lowest one. Conversely, high consumption of vegetables (OR 0.4), citrus fruit (OR 0.61) and fish (OR 0.78) significantly reduced eoCRC risk (Fig. 8). A high tertile intake of red meat, bread and cereals, fruit, and olive oil did not reach statistical significance.
On the other hand, Archambault et al., in their analysis of data from 13 population-based studies encompassing 3767 eoCRCs and 4049 age- and sex-matched controls, have demonstrated that red meat is a risk factor for eoCRC (OR 1.10) [140], while no significant differences were observed regarding fruit, vegetable, processed meat, and total dietary fiber

intake. However, a significant reduction of eoCRC risk associated with being vegetarian, consuming a non-high fat diet and rice/rice powder was reported by a Pakistani single-center case–control study [141] evaluating few dietary and alcohol habits as non-quantitative data on 74 eoCRCs and 148 age- and gender-matched controls (Fig. 8).

Chang et al. [142], in a retrospective case–control study recruiting 175 eoCRCs, showed that higher consumption of sugary drinks (7 per week; OR 2.99), desserts (3–6 per week; OR 2.28), fast food (2 per week; OR 1.84) and canned food (3 per week; OR 1.70), along with a higher Western-like dietary pattern score (OR 1.92) were linked with an increased eoCRC risk. Conversely, there were no significant differences in the consumption of fruits, vegetables, high-fiber foods, red meat, or processed meat between the two cohorts.

The first multicenter prospective cohort study using a quadrennial FFQ was performed by Zheng et al. [137]. They aimed at evaluating the association of dietary pattern with the risk of early-onset high-risk adenomas (eoHRA), considered as eoCRC precursor. A total of 375 eoHRAs were identified during colonoscopy surveillance within the Nurses' Health Study II (NHS II), with the highest quintile of Western diet as eoHRA risk factor (OR 1.67) and the highest quintile of prudent diet as a protective factor (OR 0.69) (Fig. 8). Indeed, two dietary patterns, termed the 'healthy' pattern (rich in fruits, vegetables, whole grains or legumes, fish, and low-fat dairy products) and the 'unhealthy' or 'Western dietary' pattern (high in red and processed meat, sugary drinks, refined grains, and desserts), have been identified as exogenous factors influencing the prevention or predisposition to eoCRC [143]. The Western diet stands out as one of the most significant risk factors for loCRC [141,144,145], and it is also associated with the development of high-risk rectal adenomas later in life if started during adolescence [146]. Conversely, the Mediterranean diet has shown a protective effect against CRC development [147,148], a finding supported by evidence from the Italian segment of the EPIC cohort [149].

Another prospective study interrogating the NHS II female cohort was conducted by Hur et al. [150]. In contrast to Zheng et al. [137], they examined the relative risk of eoCRC associated with sugar-sweetened beverages (SSBs) during both adulthood and adolescence (13–18 years), observing a doubled risk of eoCRC in women who consumed 2 servings of SSBs per day during adulthood, with a 16% higher risk for each serving/day increase. The risk of eoCRC increased by 32% for each serving/day increment of SSBs during adolescence. Conversely, the research found that replacing each serving/day of SSB in adulthood with coffee, low-fat milk, or total milk was associated with a 17–36% lower risk of eoCRC (Fig. 8). This result is consistent with findings reported by three meta-analyses of observational studies, which described the protective effect of milk and dairy products on loCRC [151,152]. Calcium is the main protective nutrient of dairy products, It possibly act through binding to secondary bile acids and ionized fatty acids, thus reducing their carcinogenic effects on the colorectal epithelium [153] and promoting both differentiation in normal cells and apoptosis of transformed cells through cell signaling modulation [154].

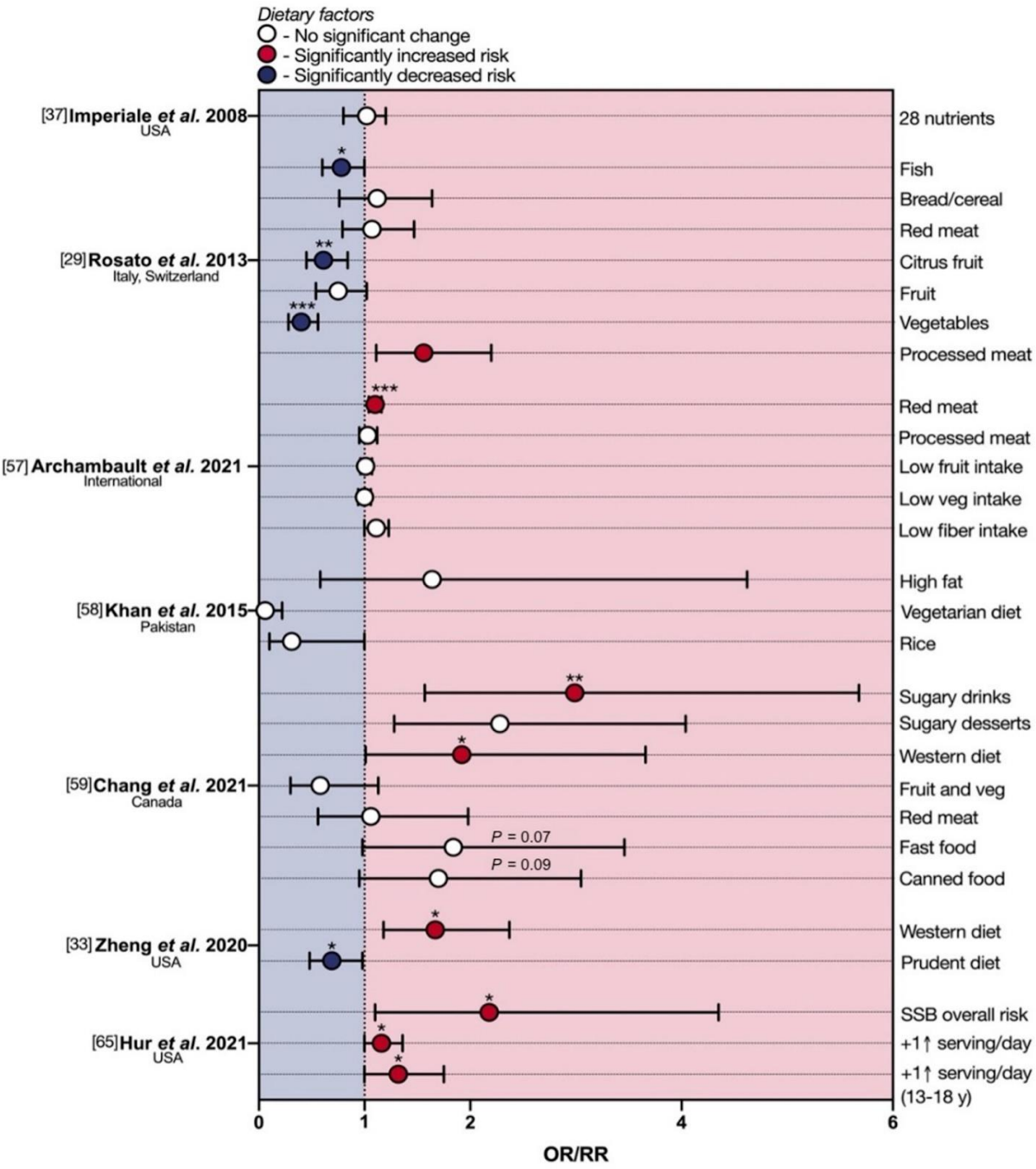


Figure 8. Protective and deleterious effects of diet in early-onset colorectal cancer. eoCRC—early-onset colorectal cancer; Met–meta-analysis; OR—odds ratio; RR—relative risk; SSB—sugar-sweetened beverage. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ [155]

Regarding the deleterious effect of alcohol on eoCRC (Fig. 9), few studies analyzed alcohol habits using standardized measurement units, while the remaining ones used the unit of "any alcohol intake".

Rosato et al. [134], showed a significant increase in eoCRC risk with alcohol consumption of 14 drinks per week (OR 1.56); Kim et al. [132] confirmed alcohol intake of 20 g per day as an exogenous risk factor for ACRN (including 14 cases of eoCRC) in the 30–39 y group (OR 1.34); Archambault et al. [140] suggested that eoCRC may be significantly associated with heavier alcohol use of more than 28 g per day of alcohol (OR 1.25).

On the other hand, the detrimental effect of "any alcohol intake" was reported by a population-based cohort study on 5710 eoCRCs aged 25–49 from the Explorys Database (OR 2.46) [60].
Moreover, two meta-analysis were performed on eoCRC populations, producing results consistent with those of nine meta-analyses of observational studies on loCRCs [156–164]. The first meta-analysis, performed by Breau and Ellis on a cohort of young-onset colorectal adenomas and cancer (yCRAC) diagnosed in young adults under 50 [165], suggested an association between advanced yCRAC and alcohol intake with a pooled OR of 1.46, even though the three studies [132,166,167] employed different criteria to define excessive alcohol consumption.
The second meta-analysis, conducted by O'Sullivan et al. and including 14 studies, confirmed the enhanced risk of eoCRC with higher alcohol consumption compared to abstinence, with a relative risk of 1.71 [86].
In contrast to the aforementioned findings, Imperiale et al. discovered no disparities in alcohol intake between ACRN cases and controls [138], a conclusion echoed by Chang et al. [142] (Fig. 9). Glover et al. conducted a retrospective analysis of data from 26 healthcare systems across 50 US states, revealing no correlation between eoCRC risk and alcohol abuse (not specified further) [168].
Intestinal dysbiosis is considered the putative mechanism of colorectal carcinogenesis associated to chronic alcohol abuse. Ethanol has been observed to decrease the abundance of Bacteroidetes and Firmicutes, while increasing Proteobacteria and Actinobacteria, thus resulting in intestinal hyperpermeability, increased translocation of Gram-negative endotoxins, and systemic inflammation [169,170]. Moreover, gut dysbiosis may intensify ethanol oxidation, leading to increased levels of intracellular acetaldehyde contributing to carcinogenesis [171].

The existing literature on the involvement of diet and alcohol in the pathogenesis of eoCRC consists mostly of retrospective observational studies marked with significant heterogeneity. Many studies encompassed not only a population of eoCRC but also included precursors of eoCRC. Most of the aforementioned studies were conducted in the US, on small patient samples or exclusively focused on either male or female populations. Only a limited number of studies explored the presence of hereditary gastrointestinal syndromes, thus hindering the assessment of possible interactions with genetic predispositions. Furthermore, some authors concentrated solely on specific dietary and lifestyle factors and used different, non-shared questionnaires that primarily assessed selected food groups, without considering cooking, processing, and storage techniques.

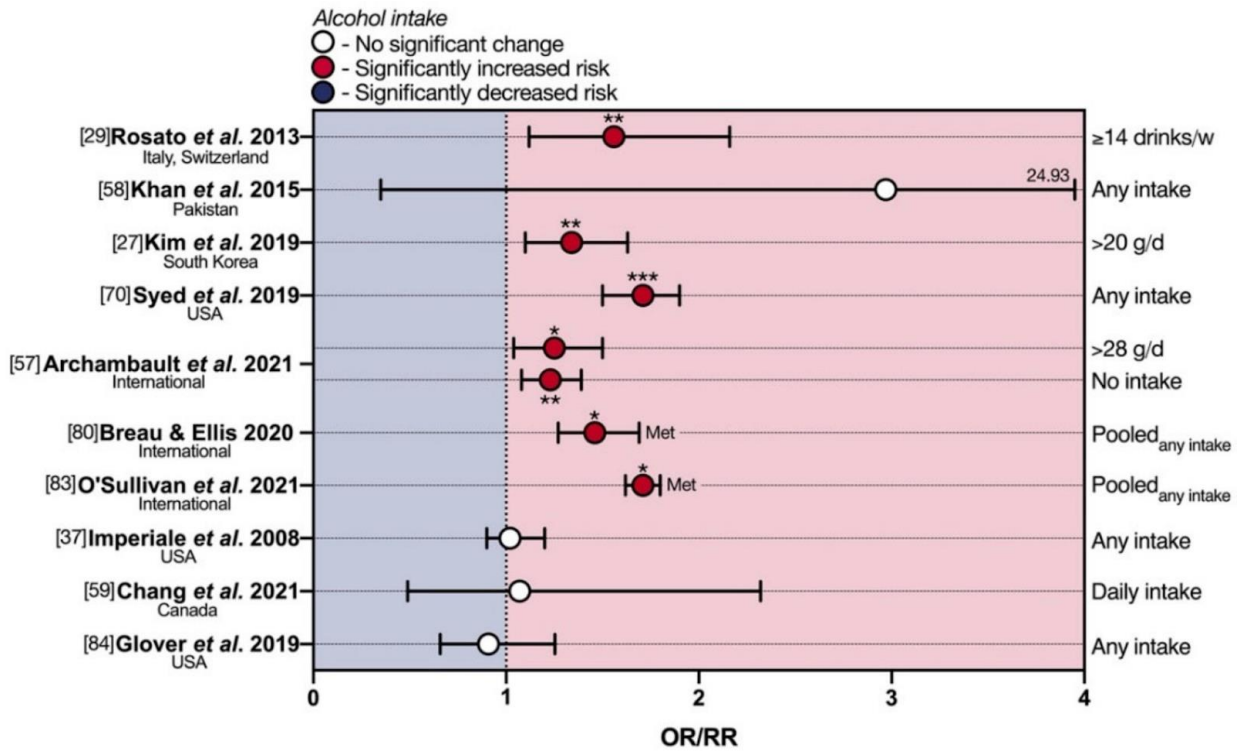


Figure 9. Deleterious effects of alcohol in early-onset colorectal cancer. eoCRC—early-onset colorectal cancer; Met—metaanalysis;OR—odds ratio; RR—relative risk. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ [155]

1.3.2 Physical activity

It is well known that reduced physical activity, leading to a sedentary lifestyle, increases the likelihood of loCRC [172,173], primarily through gut dysbiosis [174,175]. Research has demonstrated that regular exercise increases the Bacteroidetes-to-Firmicutes ratio in rat models [176,177], while moderate physical activity has been associated with a greater presence of beneficial bacterial species in active women, including Faecalibacterium prausnitzii, Roseburia hominis, and Akkermansia muciniphila [174].

Conversely, the literature available on physical activity as an exogenous risk factor for eoCRC is more limited. Moreover, physical activity or sedentary lifestyle were evaluated using different activity or inactivity indexes among all the available studies.

Nguyen et al. [133] analyzed weekly TV viewing time. They observed a significant 1.69-fold rise of eoCRC RR within the NHS II cohort when individuals reported weekly TV viewing times of ≥14 h. Additionally, there was a marginally significant elevation in RR of 1.12 for weekly TV viewing between 7.1 and 14 hours. This association was more pronounced for rectal cancer (RR 2.44 for weekly TV viewing times of ≥14 h) (Fig. 10). Furthermore, individuals with BMI ≥ 25, those engaging in less physical activity (<15 metabolic equivalents of task-hours per week), and ever-smokers displayed an elevated risk of eoCRC. A population-based cohort study [178] examined leisure-time physical activity at age 20 or later, showing that the absence of such activity was associated with areas of elevated mortality from eoCRC among US women.

Chang et al. [142] investigated both sedentary time (hours per day) and physical activity (hours per week). They found that spending more time inactive (10 h per day compared to less than 5 h per day) was associated with a significantly enhanced risk of eoCRC with an OR 1.93 and engaging in less physically activity (< 3.5 h per week) showed a tendency towards a higher risk of eoCRC, although this association did not reach statistical significance.

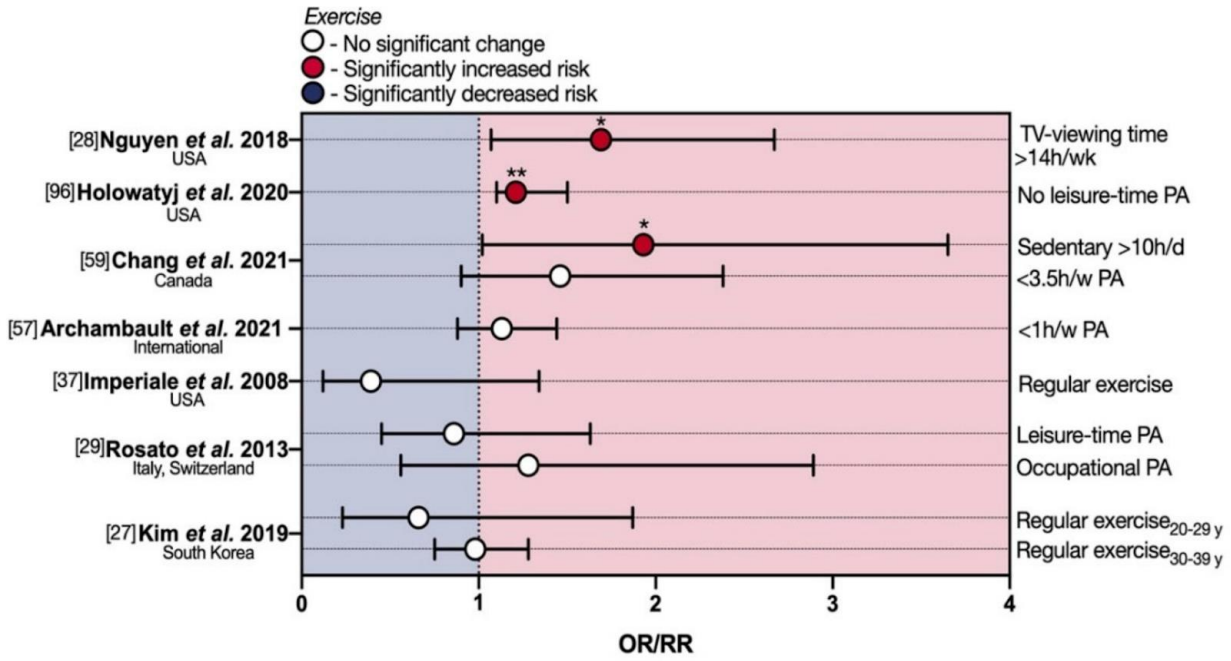


Figure 10. Protective and deleterious effects of physical activity in early-onset colorectal cancer. eoCRC—early-onset colorectal cancer; OR—odds ratio; PA—physical activity; RR—relative risk. * $p < 0.05$; ** $p < 0.01$ [155].

Archambault et al. [140] analyzed moderate/vigorous physical activity (MVPA), without findings significant differences in eoCRC with MVPA <1 h/w.

Imperiale et al. [138] evaluated regular exercise, without providing further details. The Authors observed a reduction in eoCRC risk for active individuals (OR 0.39) (Fig. 10), although this reduction was not statistically significant, likely due to the small sample including 20 ACRN, with 11 of these classified as eoCRCs.

An Italian/Swiss case–control study assessed occupational and leisure-time physical activity (hours per week) both at 30–39 years, as time spent mainly sitting, mainly standing, intermediate, heavy, strenuous PA. They failed to identify a statistically significant reduction in eoCRC risk [134] (Fig. 10).

Finally, a South Korean multi-center retrospective cross-sectional study produced similar non-significant results [132], with an OR for ACRN of 0.66 for regular exercise defined as moderate or vigorous physical activity 3 times per week in the 20–29 years cohort and an OR of 0.98 in the 30–39 years cohort.

1.3.3 Obesity

Obesity is a widely recognized risk factor for different diseases, including loCRC [172]. It is involved in colorectal carcinogenesis through different mechanisms: hyperinsulinism and metabolic syndrome [179,180], changes in the intestinal microbiota [181-182], epigenetic alterations [183,184].

While the systematic literature regarding the association between diet, physical activity, and the onset of eoCRC is still limited, several high-quality studies on overweight/obesity as risk factors for eoCRC have been published in recent years (Fig. 11), showing a clear trend toward an association between excess body weight and eoCRC.

One of the first was published in 2016 by Kim et al., who retrospectively assessed body mass index (BMI) in 564 cases of ACRN, of which 25 were eoCRCs [131]. Patients with BMI ≥ 25 kg/m2 showed a significant 1.23-fold increase in multivariate OR for ACRN.

These findings were later confirmed in a homogeneous cohort of eoCRCs [132], in which having a BMI ≥ 25 kg/m2 in both the 20–29 y age group (OR 2.46) and 30–39 y age group (OR 1.33) was linked with a significant increase in eoCRC risk. Furthermore, abdominal obesity was identified as the primary contributor to this risk with an OR of 2.26.

Liu et al. [136] investigated eoCRC risk associated with increased BMI in the NHS II, categorizing BMI into two risk groups: 'overweight' (BMI 25–29.9 kg/m2) and 'obese' (BMI 30 kg/m2). They found a significantly elevated eoCRC risk in both groups, with RR of 1.37 and 1.93, respectively. The Authors also demonstrated that weight gain after the age of 18, either between 20–39.9 Kg or over 40 Kg, were associated with significant increases in eoCRC risk, with fold increases of 1.65 and 2.15, respectively.

Syed et al. [60], analyzing patients from the Explorys database, confirmed the association between BMI over 30 kg/m2 and eoCRC, demonstrating a significant 2.88-fold increase in multivariate OR. Similar results were obtained by Sanford et al. [185] and Glover et al. [168], who found, respectively, a 1.39-fold and 1.82-fold increased eoCRC risk for obese patients with BMI over 30 kg/m2.

A 2020 cohort study by Hussan et al. [186] reported an increase in young adults with obesity in the 20–49 year age group undergoing CRC resections.

Additionally, Chen et al. investigated metabolic syndrome in eoCRC [187], showing a significant 1.25-fold increase in multivariate OR for eoCRC.

It was also observed that a BMI over 30 kg/m2 was a greater risk factor for eoCRC localized at colon site (OR 1.56) [188].

Three meta-analyses have been published on this risk factor. The first meta-analysis by Breau and Ellis [165] included four studies enrolling advanced yCRAC [132,166,167,189], reporting a pooled OR of 1.26 for BMI over 25 kg/m2. Conversely, the one conducted by Li et al. [190], analyzing six studies [60,191–195] enrolling eoCRCs, found a 1.38-fold increased risk of eoCRC in patients with BMI ≥ 25 kg/m2 compared to normal weight individuals, with a higher risk observed when BMI was above 30 kg/m2. The last meta-analysis, comprising 14 studies involving eoCRCs [86], confirmed BMI ≥ 30 kg/m2 as a risk factor in this young population (RR 1.54).

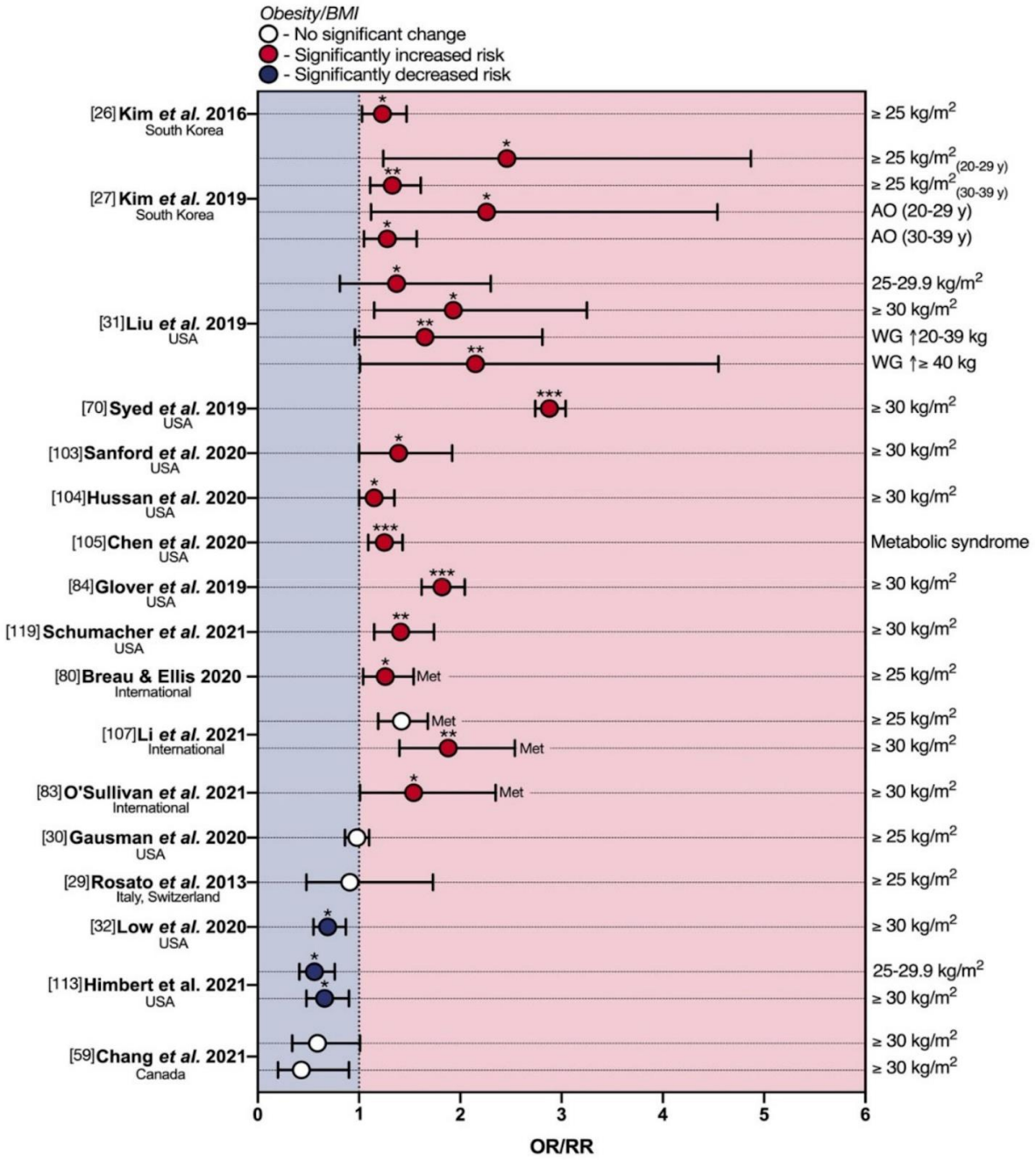


Figure 11. Protective and deleterious effects of obesity in early-onset colorectal cancer. AO—abdominal obesity; BMI—Body mass index; eoCRC—early-onset colorectal cancer; Met—meta-analysis; OR—odds ratio; RR—relative risk; SSB—sugar sweetened beverage; WG—weight gain. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ [155].

On the other hand, a US single-center retrospective case–control study [135] failed to demonstrate a statistically significant difference in the multivariate OR for BMI above 25 kg/m2 (0.98- fold) compared to age-matched healthy controls. Furthermore, the Italian/Swiss study by Rosato et al. [134] reported a slight reduction in the risk of eoCRC with a BMI above 25 kg/m2, but this reduction was not statistically significancant (OR 0.91) (Fig. 11). Similarly, a US multi-center retrospective case–control study identified a significant 0.69-fold decrease in the multivariate OR for eoCRC among both overweight and obese subjects and surprisingly a significant 1.87-fold increase in eoCRC risk among underweight

individuals [59]. In a prospective cohort study comparing eoCRCs with loCRCs [196], young patients affected by CRC were found to be less likely overweight/obese compared to loCRCs (OR 0.56 and 0.66 respectively). Finally, Chang et al. [142] confirmed a significant inverse correlation between obesity and eoCRC, both in early adulthood and 2 years before diagnosis (OR of 0.43 and 0.59, respectively).

1.3.4 Smoking

As obesity, smoking is a widely recognized significant risk factor involved in loCRC development [197,198]. Increasing evidence also indicate an association with eoCRC (Fig. 12), even using different measurement unit.
Various studies have investigated smoking status, defining it solely as either smoking or never smoking, without considering the number of cigarettes smoked or the number of pack-year.
Khan et al. [141], in a population of 148 CRC patients with a mean age of 41.47 ± 15.48 years among whom 22 were categorized as CRC under 25 years old, observed a significantly increased risk associated with smoking in the entire patient cohort (OR 2.12).
Similar results were described in eoCRC populations. Indeed, Syed et al. [60] reported a significant 2.46-fold increase in eoCRC risk for smokers; Sanford et al. [185] demonstrated a significant 1.51-fold change in OR for current/former smokers; Low et al. [59] showed a modest but significant 1.10-fold increase in eoCRC risk for current smoking, and a significant 0.82-fold decrease in OR for former smoking; Glover et al. [168], in a population of eoCRC under 40 years, reported an OR of 2.675 for smoking, not further specified. Whilst smoking trended as a risk factor (RR 1.35) according to O'Sullivan et al., the correlation did not reach significance [86].
Kim et al. obtained comparable findings in their two studies; however, their research was conducted on ACRN cohorts. In the first study [131] they found a statistically significant association between ACRN and current smoking (OR 1.37). The second one [132] demonstrated a significantly increased multivariate risk of ACRN for current or former smokers only in the 30–39 y group (OR 1.30). Conversely, Krigel et al. failed to demonstrate current smoking as a risk factor for eoCRC in a small sample of 48 ACRN [62].
The meta-analysis by Breau and Ellis on yCRAC [165] analyzed current and regular smoking, demonstrating an association with advanced yCRAC, with a pooled OR of 1.56.
Chang et al. [142] in addition to finding no differences in eoCRC risk between ever- vs. never-smoking, evaluated also smoking habits in pack-years. They observed that smokers in the first tertile of pack-years had a significantly increased risk compared to never smokers (OR 1.94).

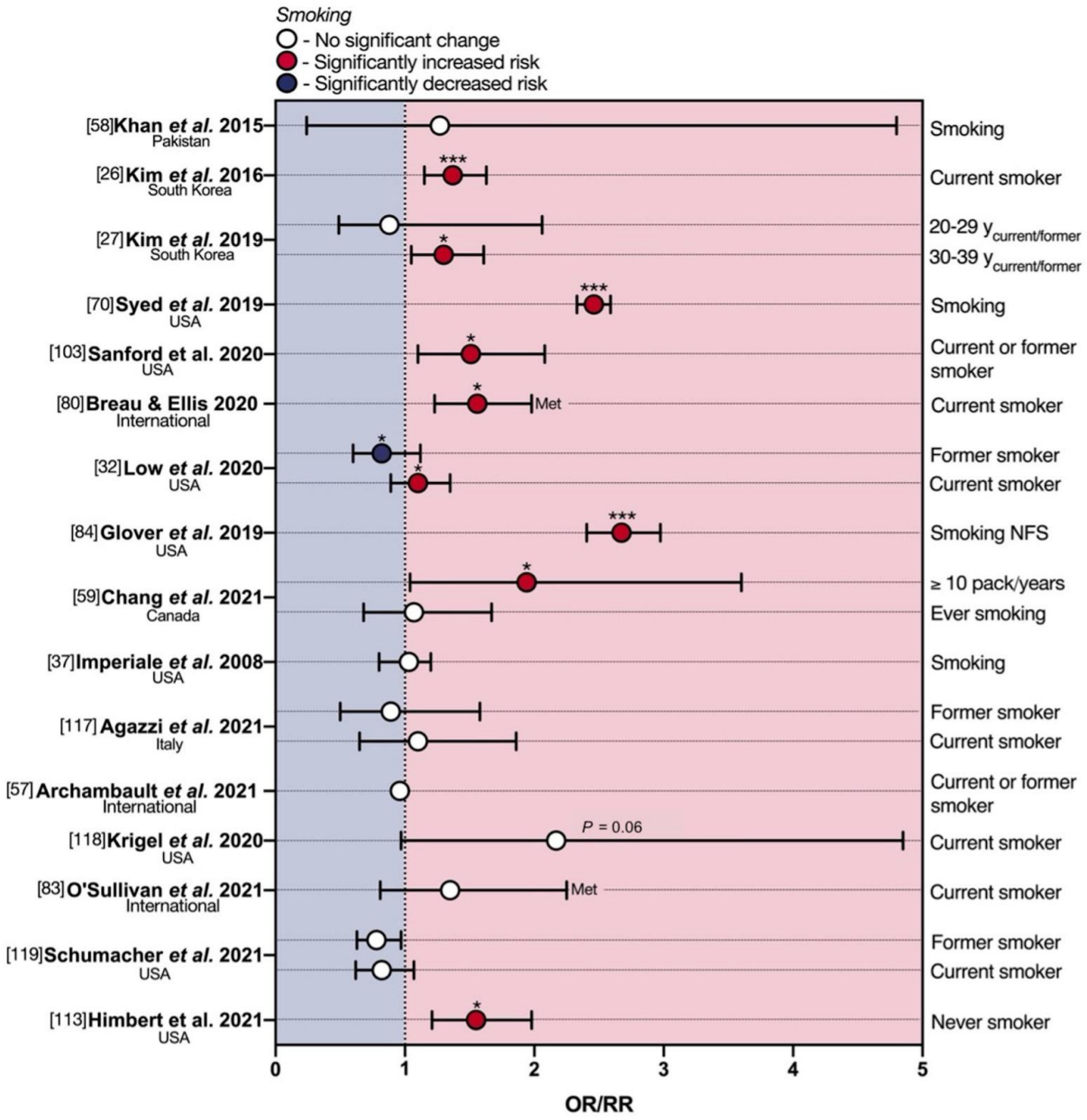


Figure 12. Protective and deleterious effects of smoking in early-onset colorectal cancer.
eoCRC—early-onset colorectal cancer; Met—meta-analysis; NFS—not further specified; OR—odds ratio; RR—relative risk. * $p < 0.05$; *** $p < 0.001$ [155].

On the other hand, Imperiale et al. failed to demonstrate any significant effect of smoking, even if not further specified, on the risk of eoCRC [138]. Agazzi et al. reported that both current and former smoking were not significantly associated with adenomas and eoCRC, even after accounting for bias related to the pooling of eoCRC and adenomas cases [199]. Similarly, Schumacher et al. [188] reported no significant association with eoCRC for the two cohorts of current and former smokers [188].

Archambault et al. [140] also evaluated smoking habits in pack-years among current and former smokers, confirming that there were no significant differences between eoCRCs and healthy age and sex-matched controls.

Finally, Himbert et al. showed that eoCRCs were significantly more likely to be never smokers compared to loCRCs (OR 1.55) [196].

1.4 Epigenetics and CRC

Epigenetics, firstly described by the biologist Conrad H. Waddington in 1942, is defined as heritable alterations in gene expression that do not result from permanent changes in the DNA sequence. CRC grows as a result of the gradual accumulation of genetic and epigenetic alterations in precursor lesions, such as adenomatous and serrated lesions. Epigenetic modifications play a pivotal role in the pathogenesis of several cancers, including CRC [200], bridging the gap between certain CRC cases, specific gene expression patterns and the lack of genetic alterations.

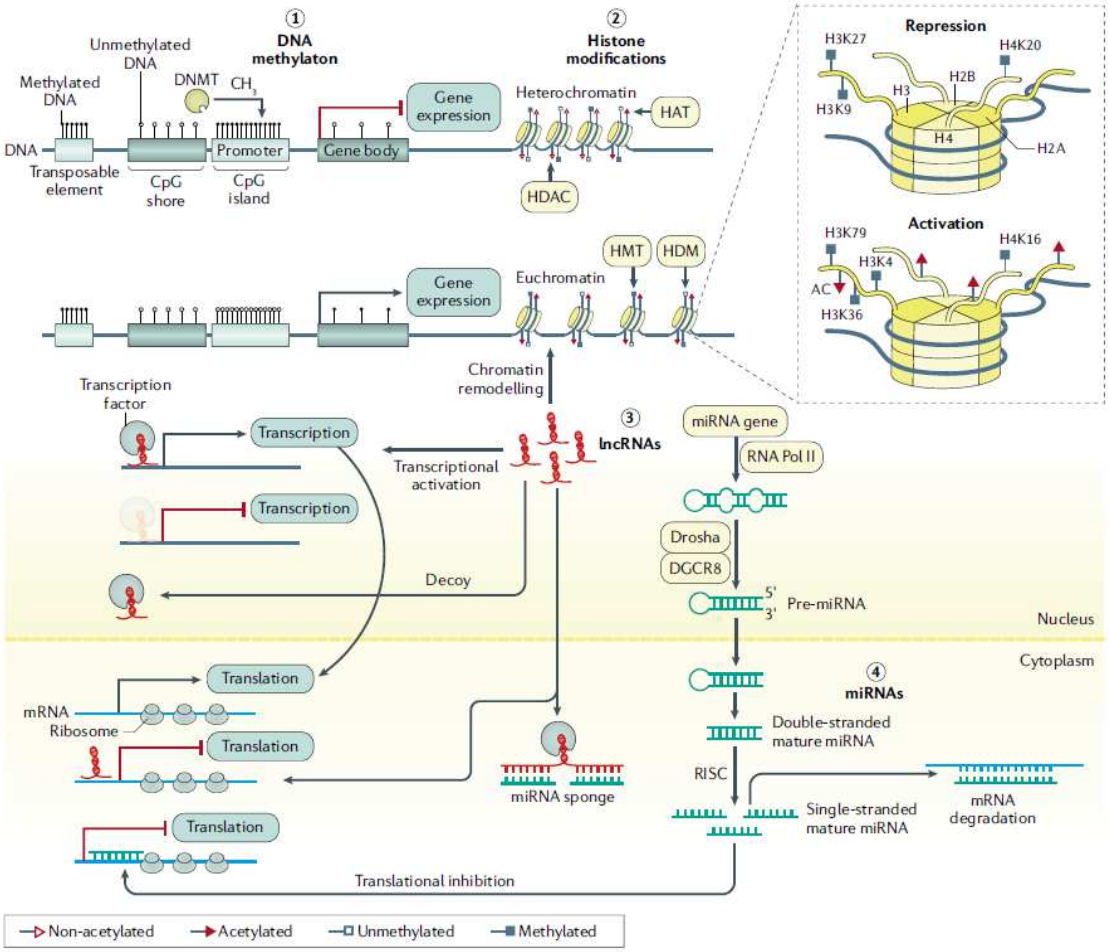


Figure 13. Principles of epigenetics [203]
The discovery of epigenetic changes has not only improved our understanding of CRC pathophysiology but has also allowed the discovery of new disease biomarkers and therapeutic targets.

The most important epigenetic changes involved in carcinogenesis include aberrant DNA methylation, abnormal histone modifications and altered expression of non-coding RNAs

(ncRNA), including microRNAs (miRNAs) and long ncRNA (lncRNAs). Global hypomethylation has been demonstrated to induce chromosomal instability (CIN) in CRC [201], while miRNAs act by preventing protein expression and influencing numerous cancer related pathways at the post-transcriptional level, impacting all stages of CRC [202] (Fig. 13). These epigenetic alterations operate in a coordinated fashion to influence gene expression. The identification of epigenetic changes has not only improved the comprehension of colorectal carcinogenesis but has also facilitated the discovery of novel biomarkers and therapeutic targets.

1.4.1 DNA methylation

The most important DNA methylation process involves the transfer of a methyl group (CH3) at the C5 position of the cytosine at CpG dinucleotides by DNA methyltransferases (DNMTs), yielding 5-methylcytosine [204].
DNA methylation changes encompass DNA hypomethylation in tipically unmethylated regions of the genome, and DNA hypermethylation occurring in CpG islands of gene promoters (Fig. 13). The latter is linked with the suppression of tumor suppressor genes within cancer cells, including CDKN2A, MLH1 and APC genes, consequently promoting cancer progression [205–209].
Genome-wide hypomethylation constitutes an early event in colorectal carcinogenesis, manifesting across various stages of disease, from early adenomas to adenocarcinomas and distant metastases [210,211]. It has been linked to proto-oncogene activation in CRC. When occurring at the level of antisense promoters located downstream in repetitive elements such as long interspersed element 1 (LINE-1), which are typically silenced under physiological conditions [211], it triggers the activation of these LINE-1 elements. Consequently, they act as retrotransposons through a 'cut- and-paste' mechanism, inserting themselves in distant fragile sites and inducing genomic instability. Considering that up to 17% of the human genome comprises LINE-1 elements, their hypomethylation serves as a proxy for global DNA hypomethylation and is associated with eoCRC and poor prognosis, thereby making LINE-1 a potentially significant biomarker [212,213].

1.4.2 Histone modifications

In non-dividing cells, DNA is naturally wound around histones, forming nucleosomes. These nucleosomes are further combined with other nuclear proteins to construct chromatin. Each histone protein has a tail abundant in lysine and arginine, serving as sites susceptible to post-translational alterations. These modifications can impact gene expression, thus contributing to both physiological processes and cancer development [214].
Genetic mutations associated with histone modifications involve different histone modifiers, including histone deacetylases (HDACs) - histone acetyltransferases (HATs) and histone methyltransferases (HMTs) - histone demethylases (HDMs), catalyzing post-translational modification of the histone tails (Fig. 13).

HDACs and HATs are responsible for histone acetylation and deacetylation, respectively, and were associated with CRC pathogenesis [215]. Histone acetylation neutralizes the positive charge on the histone tails, thus reducing the interaction DNA-histones and the compaction status of chromatin [216]. Therefore, when hyperacetylation occurs at the level of histones associated with proto-oncogenes, it activates gene expression. Conversely, when hypoacetylation involves histones linked with tumor suppressor genes, it silences these genes [217].
HMTs and HDMs are enzymes responsible for catalyzing histone methylation and demethylation, respectively. This catalytic activity determines whether gene expression is activated or repressed, depending on which lysine/arginine residues undergo methylation. Abnormal levels of HMTs or HDMs, either overexpression or underexpression, can disrupt the global histone methylation status. Consequently, this alteration may affect the expression of oncogenes or tumor suppressor genes, thereby facilitating the development or progression of cancer [218].

1.4.3 non-coding RNAs

Since their discovery in the early 1990s, it is now well known that 98% of the non-coding RNAs (ncRNAs), that can be spliced after transcription without being translated into proteins, play crucial roles in regulating gene expression in the context of both normal physiological development and the pathogenesis of virtually all diseases including CRC pathogenesis [219,220]. NcRNAs can be categorized into two groups based on their size: (i) small ncRNAs with less than 200 nucleotides, including microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and small nucleolar RNAs (snoRNAs); (ii) long non-coding RNAs (lncRNAs) with more than 200 nucleotides (Fig. 13) [221].

1.4.3.1 microRNAs

miRNAs are short (18–25 nucleotides in length), single stranded RNAs that act as post-transcriptional repressors by binding to complementary sequences in the 3'-untranslated regions (UTRs) of their target mRNA, thus regulating the translation of more than 60% of protein-coding genes, including those involved in cell proliferation, differentiation and apoptosis. In detail, after the miRNA gene's transcription by RNA polymerase II, the double-stranded pri-miRNA is processed to the pre-miRNA by Drosha–DGCR8in the nucleus and then translocated to the cytoplasm by exportin 5. Subsequently the RNase III enzyme DICER cuts the hairpin loop, resulting in a double-stranded miRNA–miRNA. Finally, the RNAinduced silencing complex (RISC) mediates the interaction of one of the miRNA's strands with the target mRNA, leading either to translational inhibition or mRNA degradation (Fig. 13) [202].
In 2002 Croce et al. described for the first time the role of miRNAs in cancerogenesis, reporting a reduction of miRNA-15 and miRNA-16 expression in patients with chronic lymphocytic leukemia [222]. Since then, a great number of miRNAs have been discovered to be deregulated in other malignancies, including CRC [223]. They act either by regulating specific individual target mRNAs, or as broad regulators of gene expression mediating the

expression of hundreds of genes simultaneously. Their expression can be altered through different genetic alterations, such as point mutations, deletions, amplifications and translocations, as well as through DNA hypermethylation and hypomethylation [224,225]. miRNAs result mostly overexpressed in different cancers, although they can be both upregulated or downregulated in tumor tissues. Finally, they can either inhibit the expression of tumor suppressor genes (oncogenic miRNAs), or of oncogenes (tumor-suppressive miRNAs, ts-miRNAs) [226].

Distinct deregulated miRNAs, regulating all relevant signaling pathways, have been discovered acting during each step of the canonical and serrated colorectal carcinogenesis (Fig. 14):

- miRNA143 and miRNA31: the activation of the RAS–RAF–MEK pathway, which enhances the proliferation of colorectal cancer cells and reduces response to treatments, occurs through the downregulation of miRNA-143 [227] and/or upregulation of miRNA-31 [228]. Therefore, miRNA-143 is not only a promising biomarker for early diagnosis but could also be useful as a potential anti-cancer drug in patients with KRAS activating mutations who are resistant to anti-EGFR therapy.
- miRNA-21: the most frequently overexpressed miRNA in CRC, favors cancer progression by downregulating the expression of phosphatase and tensin homologue (PTEN), preventing phosphatidylinositol-3,4,5-trisphosphate (PIP3) dephosphorylation and hyperactivating the phosphoinositide 3-kinase (PI3K) – AKT pathway. This, in turn, contributes to cell cycle progression, invasion and metastasis [229]. miRNA-21 can also reduce apoptosis by downregulating the programmed cell death protein 4 (PDCD4) [230].
- miRNA-34a: in physiological conditions, in case of DNA damage, p53 keeps the cell cycle at the G1-S checkpoint to allow DNA repair or to induce apoptosis if repair is not possible and, in a positive feedback loop, increases the expression of miRNA-34a, that in turn enhances p53 activity [231]. Moreover, miRNA-34a directly targets mothers against decapentaplegic homologue 4 (SMAD4), a key effector in transforming growth factor- β (TGFβ) signaling. Thus, the downregulation of miRNA-34a associated with colorectal cancerogenesis induces a reduction of p53 activity and consequently of DNA repair and enhances TGFβ signaling resulting in epithelial–mesenchymal transition (EMT) and tumor cell invasion [232,233].
- miRNA29a: in CRC cells, miRNA-29a decreases the expression of E- cadherin in epithelial cells. This leads to a loss of contact inhibition, which induces cell growth, migration and invasion via β-catenin–T cell transcription factor (TCF) signaling [234].
- miRNA-135: directly downregulates adenomatous polyposis coli (APC), which, under normal conditions, is responsible for β-catenin proteolysis, resulting in downstream activation of the WNT–β- catenin pathway [235].
- miRNA126: When tumors grow rapidly, hypoxia stimulates the formation of new blood vessels through angiogenesis, which is crucial for tumor survival and is regulated by the vascular endothelial growth factor (VEGF) pathway. VEGF is a direct target of miRNA-126, which reduces neo-angiogenesis. However, miRNA126 is

downregulated in CRC [236], leading to neo-angiogenesis and development of metastasis in both tissue samples and serum from patients with CRC [237,238]

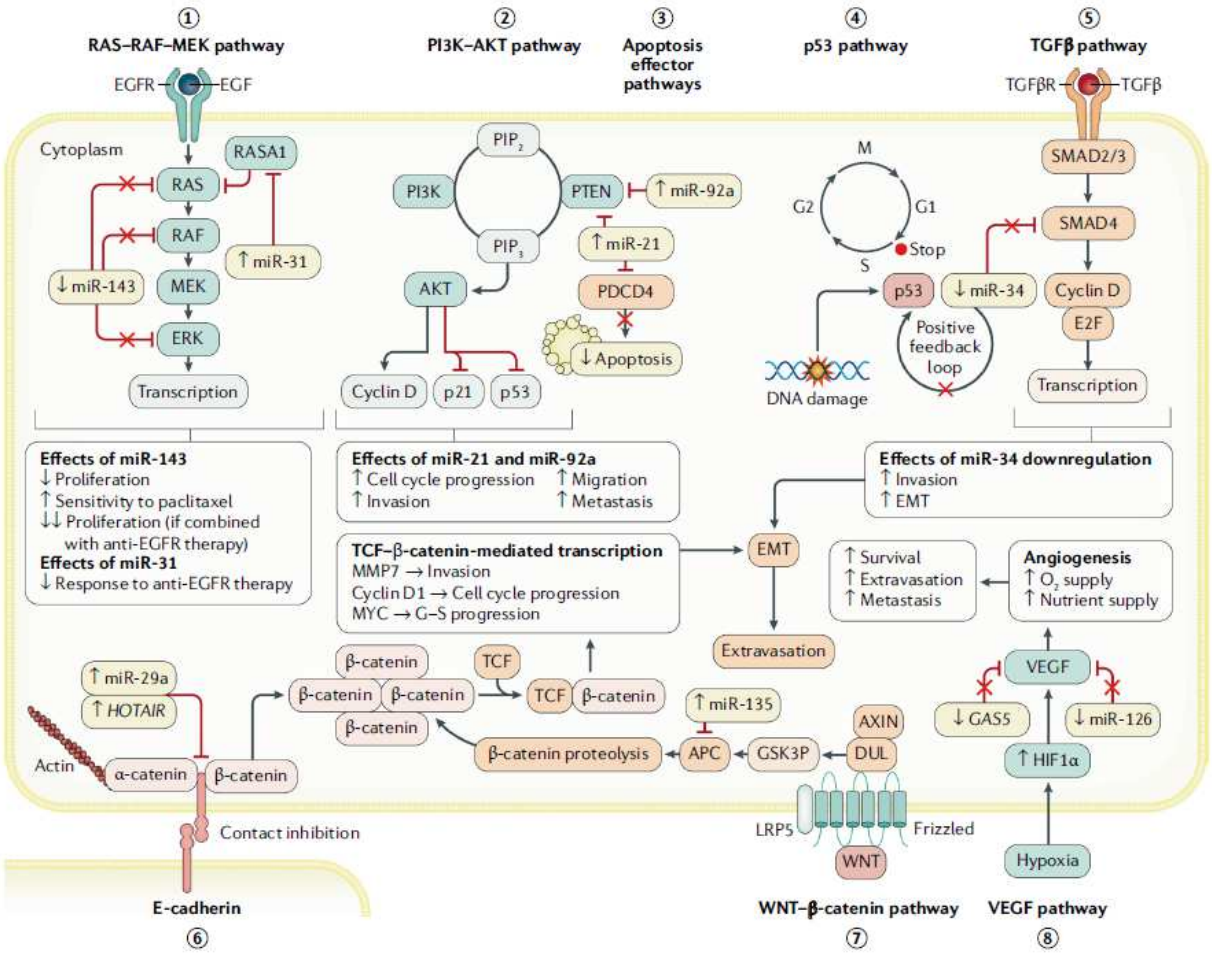


Figure 14. Role of miRNAs and lncRNAs in regulating important signaling pathways relevant to CRC [203]

Given their stability in different biological samples such as tissue, blood and stools (the small size and the hairpin-loop structure protect them from RNase degradation) and the availability of routine laboratory techniques for their identification and quantification (such as microarrays and quantitative reverse transcription PCR [RT-qPCR]), miRNAs represent attractive biomarker candidates for early diagnosis and prognosis of CRC as well as prediction of cancer recurrence in order to personalize surveillance and/or assign more aggressive and targeted therapies. For this reason, in the past decade a great number of studies investigating miRNAs in CRC has been performed.

The first study on comprehensive miRNA expression profiling was conducted by Ng et al [239], demonstrating that high expression of miR-92a and miR-17-3p on tissue and plasma samples could discriminate CRCs from healthy controls (sensitivity 64% and 89%; specificity 70% and 70%; AUC 0.72 and 0.89 for each miRNA, respectively). Upregulation of miRNA-92a in both tissue and blood was also associated with shorter overall survival, showing its prognostic potential as biomarker [240].

Another oncogenic miRNA, considered one of the promising non-invasive biomarkers for early CRC diagnosis, is miRNA-21, which was identified as differentially expressed in 30 CRCs tissues compared to the adjacent normal mucosa [241]. These results were subsequently confirmed in an independent validation set of plasma samples (20 CRCs and 20 healthy controls) with 90% sensitivity and specificity. Another recent study [242] identified miRNA-21 as differentially expressed also in patients with advanced adenomas (sensitivity 91.1% and 81.1%; specificity 81.1% and 76.7%; AUC 0.92 and 0.81, for cancer and adenoma detection, respectively). In a meta- analysis of 16 studies conducted between 2010 and 2014 that included more than 1000 patients with CRC [243], the overall sensitivity and specificity of miRNA-21 for early CRC diagnosis was 64% and 85%. miRNA-21 was also reported to be a good prognostic biomarker in both tissue and blood samples [243,244] and the most frequently reported miRNA predictive biomarker for response to treatment in at least three different clinical settings, including response to neoadjuvant chemoradiotherapy in advanced rectal cancer and response to both neoadjuvant and adjuvant chemotherapy in advanced CRC [245–247].

Using a unique approach involving small-RNA sequencing in 48 pairs of frozen CRC tissue samples and ten different CRC cell lines, Sun et al. [248] identified miRNA-21, miRNA-143, miRNA-148a, miRNA-194, miRNA-192, miRNA-200b, miRNA-200c, miRNA-10b, miRNA-26a and miRNA-145 as the top ten differentially dysregulated miRNAs in CRC; miRNA-21 and miRNA-143 were the two most abundantly expressed and had key pathophysiological roles in this malignancy.

It soon became clear that a signature of two or more miRNAs could be potentially more accurate as diagnostic biomarkers compared to the use of a single miRNA. Indeed, a study by Liu et al. [249] showed an excellent performance of a serum- based 4-miRNA signature (miRNA-21, miRNA-29a, miRNA-92a and miRNA-125b) to diagnose CRC, with an AUC of 0.95, a sensitivity of 85% and a specificity of 99%. However, this study included only 85 patients and lacked an independent validation cohort. Another 2-miRNA panel (miRNA-223 and miRNA-92a) was analyzed in more than 200 blood samples from CRC patients showing good accuracy with a sensitivity of 97%, a specificity of 75% and an AUC of 0.91 for CRC detection [250]. Similarly, in a 2019 study including almost 300 CRCs, a plasma- based 6-miRNA signature (miR-19a, miR-19b, miR-15b, miR-29a, miR-335 and miR-18a) accurately differentiated healthy controls from patients with CRC and advanced adenomas, with an AUC of 0.92 and a sensitivity and specificity of 85% and 90%, respectively [251].

Conversely, studies analyzing the role of signatures of two or more miRNAs as biomarkers of early recurrence, survival and treatment response remain limited. miRNA-31, as well as miRNA-143 and miRNA-145 (which are usually co-expressed), have been reported as good tissue biomarkers for prediction of response to different treatments, whereas plasma based miRNA-106a has been reported to be predictive of response to adjuvant chemotherapy in metastatic CRC [252].

1.4.3.2 long non-coding RNAs

Long non-coding RNAs (lncRNAs) can influence gene and protein expression through different molecular mechanisms [253] (Fig. 13). They can enhance or repress transcription by recruiting transcription factors or by decoying transcription factors and preventing their recruitment to transcriptional start sites, respectively. lncRNAs can restore translation by 'sponging' miRNAs that would otherwise prevent translation of their corresponding mRNA. They can also directly inhibit translation.
Through those mechanisms, lncRNAs are involved in cell proliferation, differentiation, apoptosis and stem cell self-renewal, having roles in many cancer-related pathways, including colorectal cancerogenesis, such as the WNT, epidermal growth factor receptor (EGFR), transforming growth factor- β (TGFβ) and p53 signaling pathways [254,255].

AIMS

To define the endogenous and exogenous risk factors associated with such an increase in eoCRC incidence and to unravel the pathogenesis of eoCRC, we formulated the following hypothesis and focused on the subsequent aims:

i. *Hypothesis:* 15-20% eoCRCs are caused by germline pathogenetic variants (PVs), the most frequent of mismatched repair genes causing Lynch syndrome, and a negative family history does not exclude hereditary cancer syndromes.
Aim: to evaluate the association of germline pathogenetic variants and family history of CRC with eoCRC.

ii. *Hypothesis:* After performing a systematic review [155] on the role of diet and lifestyle factors associated with eoCRC, we understood that only a scant literature is still available on these risk factors, particularly regarding dietary risk factors and physical activity. Most studies are small, observational, retrospective, heterogeneous, focused exclusively on peculiar dietary and drinking habits of single countries without analyzing cooking, processing, and storage techniques, or used different activity or inactivity indexes. On the other hand, a growing number of recent, high-quality studies appears to demonstrate a consistent association between obesity, tobacco smoking, alcohol abuse and eoCRC even if none of the studies used international indexes like pack years or alcohol units.
Preliminary results: We performed a single-center case-control study enrolling 47 eoCRCs and 71 HCs. We analyzed body mass index, smoking habits, physical activity, eating and drinking habits through non-detailed questions, covering the previous 5 years. We found that fresh meat ($p = 0.003$), processed meat ($p < 0.001$), dairy products ($p = 0.013$), and smoking ($p = 0.0001$) were significantly associated with eoCRC compared to controls, while other variables did not differ significantly between the two groups [256].
Aims:
   - To develop a unique and shared semi-quantitative food frequency questionnaire (SQFFQ) able to accurately describe dietary and drinking habits of eoCRCs and healthy controls of different countries at global level that will be involved in the future DEMETRA study (international case-control study evaluating the association of dietary, lifestyle and anthropometric factors with eoCRC of countries with different eoCRC incidence).
   - To validate the SQFFQ, making data obtained from different dietary questionnaires comparable.

iii. *Hypothesis:* MicroRNAs (miRNAs) perform a variety of biological functions and, most importantly, they regulate gene expression at the transcriptional level. By virtue of their ability to finely regulate cellular processes, cancer cells exploit this feat to their

advantage to develop more aggressive traits. Therefore, given the uniquely aggressive nature of eoCRC, which has a unique tendency to recur more commonly than loCRC, the overarching hypothesis of this project is that eoCRC increases its aggressiveness and has a higher likelihood of disease recurrence because of biological processes that are intrinsically regulated and fine-tuned by microRNAs. We hypothesize that miRNAs may be differentially expressed in surgical specimens collected from patients who will develop recurrent vs. non-recurrent CRC in the following five years. Selection of the optimal post-treatment surveillance regimen for the appropriate patient subgroups, remains the most challenge in managing patients with stage I-III eoCRC.

Aim: to identify candidate miRNAs that were differentially expressed in eoCRC patients with and without recurrence and define which patients could benefit most from more aggressive surveillance

MATERIALS AND METHODS

3.1 Genetic risk assessment

All individuals consulted for eoCRC at IRCCS San Raffaele Scientific Institute, a tertiary academic medical center in Milan (Italy), were considered eligible. Based on the recommendations initially from the US Multi-Society Task Force on CRC and finally from the Delphi Initiative for Early-Onset Colorectal Cancer (DIRECt) International Management Guidelines [1], we defined eoCRC patients as those diagnosed between 18 and 49 years of age.
Patients with eoCRC were prospectively enrolled from January 2020 to December 2023 at the GASTRO PER ME (GASTROintestinal PERsonalized MEdicine) outpatient clinic, a multidisciplinary program dedicated to gastrointestinal tumor syndromes.
All patients with eoCRC were offered next-generation sequencing analysis (NGS; analyzed genes: APC, BMPR1A, BRCA1, BRCA2, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, PALB2, PMS2, PTEN, SMAD4, STK11, TP53) and multiplex ligation-dependent probe amplification (MLPA; analyzed genes: BRCA1, BRCA2, CDH1, MLH1, MSH2, MSH6, PMS2, PALB2) for genes associated with CRC, after receiving tailored information about genetic testing and providing written informed consent. Germline test results, as part of clinical procedures, were revealed to study participants or their physicians for clinical decisions.
Their family history of cancers and clinicopathological data were collected in a clinical research database, after written informed consent. This study was reviewed and approved by the IRCCS San Raffaele Scientific Institute Institutional Review Board (Protocol BIOGASTRO/2011, Version n. 2, 17/10/2013).

3.2 DEMETRA project

An international, multicenter, retrospective case-control study called DEMETRA (Diet obEsity sMoking Epigenetics geneTics biomaRkers physical Activity) was designed to evaluate the associations of dietary, lifestyle and anthropometric factors between eoCRCs and HCs in countries with increasing vs stable/decreasing eoCRC incidence.
This study was reviewed and approved by the IRCCS San Raffaele Scientific Institute Institutional Review Board (Protocol DEMETRA 2020.001, Version 2.0, 01/09/2022) and is registered on ClinicalTrials.gov with ID NCT05732623.

The DEMETRA project comprises three phases:

i. The development of an online platform [accessible at the link: https://demetraproject.it/admin/login.php], designed to be easily accessible for young participants via smartphone or PCs, even from the comfort of their homes. This platform includes a newly created, unique and detailed semi-quantitative food frequency questionnaire (SQFFQ), tailored for this study, to be shared among

different countries, along with preexisting validated questionnaires assessing smoking habits, physical activity and inactivity, and anthropometric factors.

ii. The validation study for the SQFFQ

iii. The international case-control study.

3.2.1 Development of the online DEMETRA Platform

The online platform was divided in two sections:

i. one completed by doctors, concerning *clinical data*:
   - date of eoCRC diagnosis, symptoms at diagnosis, eoCRC localization, type of surgery and date (if performed), chemoterapy and radiotherapy (if performed), histological diagnosis, eoCRC TNM classification and stage, vital status and duration of follow-up, family history of CRC and other cancers (uterus, ovary, stomach, small intestine, urinary tract/bladder/kidney, bile ducts, brain, pancreas, skin tumors), type of germline pathogenetic variant (if performed).

ii. one completed by patients, concerning:
   - *anthropometric data and lifestyle habits*: date of birth, sex, race, Ashkenazi jew, ethnicity, country, education, rural/urban area, weight (kg)/height (m)/BMI (kg/m2) at the time of eoCRC diagnosis and at 18 years old, waist circumference (cm), smoking status (tobacco, e-cigarette, e-liq) at the time of eoCRC diagnosis and at 18 years old, home blood pressure levels (mmHg), fasting blood glucose (mg/dl), sitting time, moderate-to-vigorous physical activity (MVPA), regular consumption of aspirin/NSAID, calcium and folate supplements, oral contraceptive agents, post-menopausal hormones and years of consumptions,
   - *Dietary and drinking* habits through the ad hoc designed and shared SQFFQ, developed in collaboration with the epidemiologists involved in the EPIC study and colleagues from the other countries in order to include international dietary habits. To obtain an accurate description of the dietary and drinking habits of eoCRCs and healthy controls of different countries at global level that will be involved in the future international case-control study, called DEMETRA study, different dietary assessment instruments were applied to capture the wide range of food and drinks characterizing the different populations at global level. The SQFFQ is a semi-quantitative tool that investigates the usual consumption of 329 foods, grouped into 61 food groups and classified using the same criteria (groups and subgroups) as the EPIC Italy study [257]. The SFFQ explores dietary habits over the past year. In addition to frequency of consumption, the tool investigates the portions habitually consumed using validated photographs, household measures, and standard units where appropriate. Information collected will also concern types of seasoning, and methods of cooking. In addition to the groups of foods and beverages typically consumed in most countries, the SFFQ includes questions about dietary habits typical of few countries (e.g. consumption of reindeer meat, etc.) as well as emerging dietary habits in that young age group (e.g., high-protein foods, sushi, etc.).

Hence, to obtain an accurate description of the dietary and drinking habits of eoCRCs and healthy controls of different countries at global level that will be involved in the future international case-control study, different dietary assessment instruments were applied to capture the wide range of food and drinks characterizing the different populations at global level. A special software was developed to analyze responses and link them to food composition tables in order to provide a nutritional breakdown of individual and collective diets.

3.2.2 The SQFFQ's Validation study

The validation of the SQFFQ, developed to investigate dietary and drinking habits of young adults from different countries, will consist of two phases:

- Internal validation, aimed at assessing the repeatability of the SQFFQ, in which the same volunteer will compile the SQFFQ twice;
- External validation, with the aim of making data obtained from different dietary questionnaires comparable, in which the SQFFQ will be compared with the gold standard, represented by the 4-days food diary.

3.2.2.1 Internal validation: SQFFQ's repeatability

Repeatability was evaluated by administering the SQFFQ twice, 3 weeks apart, to the same subject, the second time under similar conditions and in the same season of the year.
A sample of 30 young adults under 50 years (mean age 35.1 y ± sd 7.71; 63,3% females, 33.3% males, 3% preferred not to answer) was recruited for internal validation to faithfully represent the population of eoCRCs involved in the upcoming case-control study and to reflect their habits.
To evaluate repeatability, the measurement error for each food group X was estimated as the percentage change between the estimates of food consumption for the same individual. Designating the first administration of the SQFFQ as Q1 and the second administration as Q2, the measurement error (E), expressed as a percentage, was computed using the following formula:

$$E\ \% = \frac{(\mathrm{XQ1} - \mathrm{XQ2})}{\mathrm{XQ2}}$$

Afterwards, the agreement between the two measurements for each food group X was measured with Cohen's kappa coefficient.
The tertiles of the distribution of each food group X were calculated. For each food group X, the measurement of the quantity consumed was considered concordant if X_{Q2} and X_{Q1} belonged to the same tertile and non-concordant otherwise. The Cohen's kappa was then calculated using the following formula:

$$K = \frac{\text{observed concordant proportion} - \text{concordant proportion by chance}}{1 - \text{ concordant proportion by chance}}$$

Cohen's kappa values were evaluated according to Landis and Koch [258]:

- Equal to 0: The observed agreement is the same as that expected by chance.
- Between 0.01 and 0.20: Slight agreement between the two measurements.
- Between 0.21 and 0.40: Fair agreement between the two measurements.
- Between 0.41 and 0.60: Moderate agreement between the two measurements.
- Between 0.61 and 0.80: Substantial agreement between the two measurements.
- Greater than 0.80: Almost perfect agreement between the two measurements.

3.2.2.2 External validation: SQFFQ's validity

In this phase, the SQFFQ will be validated against the reference method and gold standard, represented by a 4-days food diary in a cohort of 100 subjects from the same age group as the internal validation. The reference method consists of four food records compiled over four consecutive days, with two recorded during the week and the other two during the weekend.

The 4-days food diary, containing all the instructions for correct completion, will be distributed on Thursdays to 5 subgroups of 20 volunteers each. Volunteers will complete the food diary, providing detailed descriptions of all food and drink consumed over the following 4 days: Friday, Saturday, Sunday, and Monday. On the subsequent Tuesday and Wednesday, the compiled 4-days food diaries will be submitted to dietitians, who will contact the participants via Zoom or in-person for an interview. The dietitians will assess the proper completion of the diary, seeking additional details on what was reported by the volunteers (e.g., the type of milk used – skimmed or not skimmed, fully or partially skimmed, if not specified; the type of sugar used in the coffee – brown or white sugar or natural/artificial sweetner, if not specified; etc). During the interview, a food atlas [259] with standardized photos of foods will be provided to quantify the portions of all food consumed, along with the recipes.
The day after completing the 4-days food diary, volunteer will fill out the online SQFFQ after receiving detailed instructions from the researchers.
Qualitative and quantitative data extracted from the 4-days food diary will be entered into the RedCap system and statistically compared with data extracted from the online SQFFQ.

3.3 Genome-wide discovery and identification of a miRNA signature

3.3.1 Patients cohorts

This project involved 177 patients collected in a large, multicenter, international global effort to represent a diverse population that would comprehend American, European, and Asian patients in similar parts.
Samples were collected from a retrospective, multicenter repository of formalin-fixed Paraffin-Embedded (FFPE) specimens from patients with stage I-III eoCRC and were divided into three distinct cohorts. All patients underwent standard endoscopic and surgical treatments, and the tumor staging was evaluated according to the TNM grading system (seventh edition) and Japanese guidelines, according to their nationality.
Patients with inflammatory bowel disease and hereditary CRC syndromes were excluded.

The biomarker *Discovery cohort* comprised 20 FFPE samples from patients with stage II-III eoCRC, enrolled at three Japanese institutions (Kumamoto University, Tokyo Medical and Dental University [TMDU], and University of Tokyo), and consisted of 10 recurrent and 10 non-recurrent eoCRCs with a median follow-up of 41.4 months (IQR, 21.75-61.1).

Two clinical independent, non-overlapping, and ethnically distinct cohorts were used assay training and validation. The first clinical cohort (herein referred to as *Training cohort*) consisted of 88 FPPE samples from patients with stage I-III eoCRC, enrolled at Hospital Clinic, Barcelona, Spain, and consisted of 25 recurrent and 63 non-recurrent eoCRCs. The median progression-free survival of the entire training cohort was of 67.5 months, and it was shorter for the recurrent group (14 months) vs. the non-recurrent group (92.9 months). The second clinical cohort (herein referred to as *Validation cohort*) included 69 FFPE samples from patients with stage I-III eoCRC enrolled at four Japanese Institutions (Kumamoto University, TMDU, National Cancer Center Hospital, and University of Tokyo), and consisted of 9 recurrent and 60 non-recurrent eoCRCs.

Clinicopathologic parameters of the three cohorts are provided in Table 1.

The study was carried out in accordance with the Declaration of Helsinki and was approved by the ethical committee of each center. A written informed consent was obtained from all patients for their willingness to participate in this study.

Table 1. Clinicopathologic parameters of the three cohorts				
		Discovery cohort n (%)	**Training cohort n (%)**	**Validation cohort n (%)**
N.		20	88	69
Age	**≤ 45 years**	10 (50)	50 (57)	35 (51)
	40-50 years	10 (50)	63 (72)	51 (74)
Sex	**Males**	7 (35)	46 (52)	36 (52)
	Females	13 (65)	42 (48)	33 (48)
Tumor location	**Proximal**	5 (25)	26 (30)	50 (72)
	Distal	15 (75)	62 (70)	19 (28)
T stage	**T1**	0 (0)	7 (8)	11 (16)
	T2	0 (0)	16 (18)	11 (16)
	T3	14 (70)	49 (56)	31 (45)
	T4	6 (30)	16 (18)	16 (23)
Lymph node invasion	**Yes**	12 (60)	51 (58)	47 (68)
	No	8 (40)	37 (42)	22 (32)
Vascular invasion	**Yes**	11 (55)	33 (38)	33 (48)
	No	9 (45)	10 (11)	35 (51)
	Unavailable	0 (0)	45 (51)	1 (1)
Lymphatic invasion	**Yes**	13 (65)	28 (32)	24 (35)
	No	7 (35)	16 (18)	45 (65)
	Unavailable	0 (0)	44 (50)	0 (0)
Grade of differentiation	**Well-moderate**	16 (80)	74 (84)	59 (86)
	Poor	4 (20)	9 (10)	4 (6)
	Unavailable	0 (0)	5 (6)	6 (8)
CEA	**Low < 5 ng/mL**	14 (70)	58 (66)	0 (0)
	High ≥ 5 ng/mL	6 (30)	10 (11)	0 (0)
	Unavailable	0 (0)	20 (23)	69 (100)
Adjuvant chemotherapy	**Yes**	8 (40)	17 (19)	17 (25)
	No	10 (50)	48 (55)	49 (71)
	Unavailable	2 (10)	23 (26)	3 (4)

3.3.2 Nucleic acid isolation and miRNA expression analysis

The following techniques were used, thanks to the collaboration with the Department of Molecular Diagnostics & Experimental Therapeutics, Beckman Research Institute of City of Hope, USA:

- *Nucleic acid extraction*. Total RNA was isolated from FFPE samples using AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Valencia, CA, USA), as per manufacturer's instruction.
- *High-throughput genome-wide small RNA sequencing*. Construction of next-generation sequencing libraries for miRNA from FFPE was performed using a modified protocol for the Truseq Small RNA Kit (Illumina) with 200 ng total RNA input. The quality of individual libraries was assessed using a High Sensitivity DNA Kit (Agilent). Libraries were size selected individually (~148 nt) by gel electrophoresis using a Pippin HT instrument (Sage Science). Efficiency of size selection was assessed using a High Sensitivity DNA Kit. Libraries were equimolar-pooled; pooled libraries were quantitated via qPCR using a KAPA Library Quantification Kit, Universal (KAPA

Biosystems) prior to sequencing on an Illumina HighSeq 2500 with single-end 35-base read lengths at an average of 10 million reads per sample. For preprocessing, Illumina small RNA-seq 3' adapters were trimmed using cutadapt software. Post-trimming, all retained sequences were confirmed to contain high-quality scores and peaks concentrated at 22 nt, representing miRNAs.

- *Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) assays.* In the Training and Validation cohorts, miRNA expression was assessed by quantitative reverse transcription PCR (RT-qPCR) on a StepOne Real-Time PCR System (Applied Biosystems, Foster City, CA) using the following miRNA Taqman probes purchased from ThermoFisher Scientific (assay ID): hsa-let-7g-5p (#002282), hsa-miR-125b-2-3p (#002158), hsa-miR-125b-5p (#000449), hsa-miR-142-3p (#000464), hsa-miR-15b-3p (#002173), hsa-miR-30e-5p (#000421), hsa-miR-365a-3p (#001020), hsa-miR-410-3p (#001274), hsa-miR-654-3p (#002239), and hsa-miR-99a-5p (#000435). The relative expression of target miRNAs was determined and normalized using snRNA U6b by $2\text{-}\Delta^{Ct}$ and were further log2 transformed.

3.3.3 Statistical analysis

3.3.3.1. Candidate miRNA selection

Preprocessed reads were aligned to a human reference genome (human genome build 38) and annotated using GENCODE miRNA annotation. miRNAs that were differentially expressed between patients with vs. without recurrence of eoCRC were identified using DESeq2. Volcano plots, representing the differential gene expression versus the statistical significance of said difference, were generated with the "EnhancedVolcano" package in R (V1.20.0, Bioconductor release V.3.18 [260]), and miRNAs were selected based on a $|\log_2(\text{FoldChange})| \geq 0.5$ in the expression level and a p value < 0.05. Subsequent biomarker prioritization was performed using Cox-LASSO regression, by eliminating parameters with a coefficient of 0. The remaining candidates were then progressively ranked based on their individual AUC values and the top 10 performing miRNAs were selected as the final candidates. For further analysis on the expression patterns of these miRNAs, additional analyses were performed, including the following.

3.3.3.2. Candidate miRNA sequencing data analysis

Ridgeline plots produce partially overlapping line plots to visualize changes in distribution between groups. They were generated with "ggridges" in R (V.0.5.5 [261]) to visualize the differences in gene expression profiles for the top 10 performing candidate miRNAs. The expression values were initially transformed via a z-normalization with the scale function of R (Formula 1) before plotting. (**Formula 1)**

$$Z = \frac{Xi - \mu}{\sigma}$$

Heatmaps were generated in R with the "pheatmap" package (V.1.0.12 [262]) using an unsupervised approach for column clustering ("hclust" hierarchical clustering) with the Ward D2 method. Unlike Ward D, the Ward D2 criterion values are on a scale of distances, whereas the Ward D method employs a scale of distances squared [263]. This approach, in R, minimizes the energy distance between cluster groups (Formula 2).

(**Formula 2**)

$$e(A,B) = \frac{n1n2}{n1+n2}\left(\frac{2}{n1n2}\sum_{i=1}^{n1}\sum_{j=1}^{n2}|a_i - b_j| - \frac{2}{n_1^2}\sum_{i=1}^{n1}\sum_{j=1}^{n1}|a_i - a_j| - \frac{2}{n_2^2}\sum_{i=1}^{n2}\sum_{j=1}^{n2}|b_i - b_j|\right)$$

Visual processing was performed with the Viridis package of R, using the "magma" option to improve readability and create perceptually-uniform graphs (V.0.6.4 [264]). Hazard ratios of recurrence were computed fitting a cox proportional hazard model on the expression data, with confidence intervals computed with the coxph function of the survival package in R (V. 3.5 [265]) (Formula 3).

(**Formula 3**)

$$HR = \frac{P}{1-P}$$

Forrest plots were then generated in the ggplot2 environment using the ggforestplot package [266] and, for ease of visualization, the confidence intervals were visually limited to a maximum value of 15 and a minimum value of 0.05.

3.3.3.3. Machine learning approach (XGBoost)

EXtreme Gradient Boosting (XGB) is a scalable, distributed, highly efficient and versatile, gradient-boosted decision tree machine learning method that provides parallel tree boosting [267]. XGB currently represents the leading machine learning library for regression, classification, and ranking problems, thanks to its ability to perform parallel computation on a single machine, a feature that increases its efficiency by over 10 fold, compared to other gradient boosting methodologies. Moreover, XGB creates a very large number of decision trees, all of which contribute to the final model performance and its ability to predict an event accurately. The advantage of this method, compared to others, lies in its extreme versatility, which provides XGB with the capacity to draw meaningful conclusions even from minimal differences between the independent variables that predict the outcome of interest. XGB works by optimizing the objective function at iteration *t* to minimize the following parameter (formula 4)

(**Formula 4**)

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n1} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

At its core, XGBoos works as a function of functions: the loss function "*l*" is a function of CART learners, a sum of the current tree plus the previous additive trees. More specifically, the loss function "*l*" uses a Taylor approximation such that

(**Formula 5**)

$$f(x) \approx f(a) + f'(a)(x - a)$$

Where f(x) represents the loss function "*l*", "a" is the predicted value at the initial step (or at *t-1* for subsequent iterations), and (*x-a*) is the new learner we need to apply at the subsequent step *t.* Using this approach, XGB can progressively write the objective loss function as a series of new added learners. By progressively increasing the number of learner functions, XGB can effectively represent a greedy learning approach. When a second/order Taylor approximation is then applied to the aforementioned formula, we obtain formula 6

(**Formula 6**)

$$\bar{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t)$$

After the first learner is built, the next builders needs to improve on the quality of its predecessor learner model *t-1*. This is accomplished by measuring the quality of a tree structure "q" with a scoring function (Formula 7)

(**Formula 7**)

$$\bar{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\Sigma_{i \in Ij} g_i)^2}{\Sigma_{i \in Ij} h_i + \lambda} + \gamma \mathrm{T}$$

The scoring formula 7 returns the minimum loss value for a given tree structure. In other words, the original loss function "*l*" is evaluated by using the optimal weight values and, for any given tree structure, we can calculate the optimal weights in leaves.

In practical terms, each tree learner is built in three subsequent steps, which the XGBoost package can conveniently perform in a time-efficient manner. In the first step, the algorithm starts with a single root, which contains all the training dataset. It will then use the loss function to evaluate each possible split loss reduction over all features and values per value. At the end of this second step, each possible split is evaluated by the gain function (formula 8). The split with the highest gain metric is then selected out of the entire landscape of splits. If the best-performing, highest-gain split is greater than zero, the branch may be kept and the process repeated to grow another sub-branch. If the gain is negative, then the branch is stopped from growing and that part of the decision tree is trimmed.

(**Formula 8**)

$$GAIN = LOSS_{(father\ instances)} - [LOSS_{(right\ branch)} + LOSS_{(left\ branch)}]$$

However, an algorithm that removes a branch only in negative gain circumstances is generally referred to as an "exactly greedy algorithm". Such an approach has the tendency to over-fit the model to the training cohort, and therefore there are some measures that can be taken to limit overfitting (discussed below). This allowed for a successful independent validation.

3.3.3.4. XGB model building, parameters hyper-tuning, and application

For this specific project, we used the XGBoost package in R (v1.7.6.1) [267] to predict a dichotomous outcome of 5-year recurrence-free survival. The independent (predictor) variables were the log2-transformed Δ^{Ct} values of the 10 candidate microRNAs. The XGBoost was allowed to learn for a maximum of 5000 iterations or until no further additional gain was possible.
The first feature optimization to limit overfitting concerns the learning rate parameter "eta" (or "ε"). Eta regulates the step size at which the optimizer function updates the weights. Lower learning rates result in slower and more accurate updates, therefore limiting overfitting, and for this project ε was set at 1%. The second parameter to optimize represents the "max_depth" function, which controls the maximum depth of the trees in the model (that is to say, how many branches). While XGBoost allows for a maximum of 8 for this parameter, this greedy approach often leads to overfitting, therefore a max depth of 4 was applied to this model. The third parameter, "subsample", refers to the fraction of *observations* (ie, patients) used for each tree, which was optimized at 75%. The fourth parameter, "colsample_bytree" controls the fraction of *features* (ie, miRNAs) used in each tree, which represents the only parameter we kept at 100%. The fifth parameter, "gamma" represents the most aggressive parameter optimization to limit overfitting. Gamma represents the gain threshold against which each tree is measured. As explained above, XGBoost eliminates branches whose gain is negative. To limit overfitting, one can decide to use a more aggressive trimming strategy where only branches with a higher gain are kept and the others are removed. We approached this project highly pruning strategy by setting gamma=5. Finally, performance monitoring during training for early stopping was set up using the "eval_metric" parameter equal to "auc".

After fitting the XGBoost to the classification problem, the XGBoost package allows to retrieve some feedback in terms of how the model was constructed and how each independent variable contributed to the overall model structure. This allows to explain and visualize the inner workings of the machine learning algorithm by virtue of three main metrics: gain, cover, and frequency. Gain refers to the relative contribution of the corresponding feature (ie, miRNA) to the model, after measuring the contribution of that feature to each tree in the model. The model then returns the percentage gain that is contributed to the model by each feature and ranks the miRNA by virtue of their percentage contribution to the model. A higher gain metric implies that that feature (miRNA) is more important for generating the final prediction. In fact, gain represents the improvement in accuracy that is brought by a feature when it sits on a branch. The cover metrics, instead, pertains to the quality and applicability of the split. Coverage represents the relative number of observations (patients) that are related to that feature (miRNA). For example, given 100 observations, 10 features, and 10 trees, if feature 1 is used to decide the leaf node for 50, 15, and 10 observations in three trees, the metric for said feature will be 50+15+10=75. This calculation will be repeated for all the features and the cover will be expressed as a percentage for all the features' cover metric. MicroRNAs with higher cover values positively

affect a higher number of patients. Finally, frequency is a simpler way to measure gain, as it counts the number of times a feature is used in all the generated trees, and therefore it is of lesser importance than gain. Gain, cover, and frequency metrics were obtained from the XGBoost package in R. To visually represent these features, we plotted the gain, cover, and frequency metrics as radar plots, using the radarchart function of the "fmsb" package in R (V.0.7.5) [268].

Although gain, cover, and frequency explain the importance of each feature, they do not provide a deeper insight into how each feature influences the model prediction for each individual observation (ie, patient). In fact, machine learning models such as XGBoost are so complex, powerful, and accurate, that they often become similar to black boxes where it becomes virtually impossible to understand how such predictions were made. In an effort to reach machine learning explainability and interpretability, SHAP values (SHapley Additive exPlanations) measure how much each featrure contributes to the model's prediction because they symulatenously inform you on (1) which features are the most important, (2) how they affect the final outcome, and (3) they do so for each patient. SHAP values originate from a game theory approach to machine learning, where each feature is considered as a player and is assigned an importance value representing its contribution to the model's output. Therefore, SHAP values provide a more nuanced approach to explain the inner working of a model and showcase the importance and contribution of each feature to the final model. Features with a positive value positively impact the prediction, while those with negative values have a negative impact, and the magnitude is a measure of how strong the effect is. SHAP values have several advantages compared to other methods as tools for interpreting machine learning algorithms. First, SHAP values are additive, herefore the contribution of each feature to the final model is computed independently. Second, SHAP values have local accuracy, which means that they add up to the difference between the expected model output (the true patient status) and the actual output (the machine learning prediction) for any given input. Third, SHAP values are zero for missing values and for values of irrelevant importance, which makes sure that missing data does not distort the interpretation of data (ie, SHAP values do not have issues with missingness). Finally, SHAP values remain consistent across the model, unless a feature changes. SHAP values were obtained using the shap.prep function of the "SHAPforxgboost' package for R (V 0.1.3. [269]) and then plotted as a beeswarm plot and, for selected patients, as forceplots to further explain how each feature contributed to the final prediction. Finally, we drew a collapsed view of all the decision trees by collapsing the decision forest into a single decision tree using the "xgb.plot.multi.trees" function of the XGBoost package in R. This decision tree provides an estimation of how much each feature contributes to the split that occurs at each node. As such, the collapsed decision tree is not accountable for individual decisions, but it provides a birds-eye view of where each feature contributes to the model in most trees.

3.3.3.5 Training and validation cohort results processing

The performance of the XGBoost model in both training and validation cohorts was

benchmarked on the area under the receiver operator curve (AUC) using the pROC package in R (V. 1.18.5 [270]) Sensitivity, specificity, and accuracy were evaluated after threshold optimization based on Youden index using the cutpointr package in R (V 1.1.2 [271]).
These are specifically calculated as follow:

- Sensitivity = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- Positive predictive value = TP / (TP + FP)
- Negative predicted value = TN / (TN + FN)
- Correct classification rate = (TP + TN) / (TP + TN + FP + FN)

To plot the distribution of the XGB-derived predictions, we created raincloud plots using the "RaincloudPlots" package in R (V2 [272]) and waterfall plots using the ggbarplot function of ggplot2 in R.

To further understand the ability of our model to perform optimal classification of recurrent vs. non-recurrent cases in both the training and validation cohorts, we plotted the Youden index at each cutpoint using the plot_metric function of the cutpointr package. Finally, we measures the prediction error of our machine learning algorithm using bootstrap aggregating to calculate the out-of-bag estimated Youden (plot function of cutpointr package). This procedure involves subsampling with replacement.

3.3.3.6 Survival evaluation

Recurrence free survival data were initially analyzed using the swimplot package in R (V 1.2.0 [273]). Kaplan-Meier survival probability curves were drawn using the "survival" package in R (V 3.5 [274]). P values for survival differences were analyzed with log rank test. Similary, cumulative hazard plots were generated in R with the 'survivial' package and p values calculated with a log rank test. The majority of the patients received a follow-up of at least up to 5 years (median 85.4, 95% confidence interval [CI] 68.87-104.75), unless they died.

3.3.3.7 Decision curve analysis and net benefit analysis

To evaluate the population impact of adopting a risk prediction instrument into clinical practice, we implemented a series of decision curve strategies using the "DecisionCurve" package in R (V 1.4 [275]). For Net Benefit Analysis plots, we estimated the net benefit curves with bootstrapped confidence intervals using an opt-in policy for a case-control study design. A model based on our miRNA signatures, derived from the XGB approach, was tested against three possible scenarios: a test-all approach (where every patient with stage I-III eoCRC would receive additional endoscopic surveillance after treatment for the foreseeable future), a test-non approach (where patients would not be redirected to a more intensive endoscopic surveillance approach) and, finally, a clinical approach (where patients would be redirected to an intensive endoscopic surveillance approach based on their clinic-

pathological features). This clinical approach was derived by fitting a logistic regression model based on six features (age, biological sex (male, female), tumor localization (right, left), tumor grade (high, intermediate, low), vascular invasion (yes, no), and lymphatic invasion (yes, no)). We then plotted the Net Benefit curves of the four models to estimate which model would provide the greatest clinical effects in a population. Finally, since we wanted to estimate how the model would affect clinical practice, we estimated how many patients with stage I-III eoCRC would be considered high-risk vs. how many future recurrent cases would be identified at progressively higher high-risk thresholds. These graphs of clinical impact were built in R using the "plot_clinical_impact" function of the "DecisionCurve" package. For stage-specific subanalysis, we evaluated the net benefit of each model for each stage.

RESULTS

4.1 Endogenous risk factors

To evaluate the association of family history of CRC with eoCRC and to assess the prevalence of germline PVs through NGS-based multigene panels testing and MLPA, genetic counseling was offered to all 110 eoCRCs initially recruited. Five eoCRCs refused to undergo genetic testing and were subsequently excluded from the study.

A total of 105 eoCRC individuals were enrolled, with a mean age at diagnosis of 41.2 ± 6.7 years; 48.6% were females, and 51.4% were males.

It was found that 20% of eoCRC carried a germline PV of genes known to be associated with CRC (Tab. 2). Specifically, 13 eoCRCs had PVs of MMR genes (12.4%) responsible of LS, 3 eoCRCs had BRCA1-2 PVs (2.9%) responsible of Hereditary breast and ovarian cancer syndrome, 4 eoCRCs had MUTYH PVs of which 3 were heterozygous and 1 homozygous, with the latter responsible of MAP (3.8%). Additionally, 1 eoCRC individual had a PV in the ATM gene (0.9%), and 1 had an SDHAF2 PV (0.9%); one patient exhibited mosaicism of PVs in MSH2/MUTYH genes.

Overall, 71.4% of eoCRCs didn't have a family history of CRC. 19% of eoCRCs reported having a first degree relative (FDR) with CRC and 12.4% had a second degree relative (SDR) with CRC; three patients had both a FDR and SDR with CRC and were all LS patients.

In detail, 33.3% mutated eoCRCs had a FDR with CRC (mean age 46 ± 12.6 y) and 38.1% had a SDR with CRC (mean age 50 ± 17.3 y). Conversely, 15.5% non-mutated eoCRCs referred a FDR with CRC (mean age 66.4 ± 8.3 y) and 6% reported a SDR (mean age 69.8 ± 7.5 y).

When comparing mutated-eoCRCs with non-mutatetd eoCRCs, no statistically significant differences were found in terms of age at diagnosis or sex, location of CRC, presence of FDR with CRC (Tab. 2). Mutated eoCRC differed significantly from non-mutated eoCRCs in terms of SDRs with CRC ($p < 0.001$).

Table 2. Patients' data and risk factors

		Overall (n/%)	eoCRC with PV (n/%)	eoCRC without PV (n/%)	P-value
N.		105	21	84	/
Age at diagnosis median [IQR]		43.00 [38.00, 47.00]	40.00 [36.00, 46.00]	43.50 [39.00, 47.00]	0.116
Sex	**F**	51 (48.6)	9 (42.9)	42 (50)	0.630
	M	54 (51.4)	12 (57.1)	42 (50)	
Relatives with CRC	**FDR**	20 (19)	7 (33.3)	13 (15.5)	0.116
	SDR	13 (12.4)	8 (38.1)	5 (6)	**<0.001**
Germline PVs			21 (20)	/	
eoCRC site	**Right colon**	34 (32.4)	9 (42.9)	25 (29.8)	0.400
	Left colon	34 (32.4)	7 (33.3)	27 (32.1)	
	Rectum	37 (35.2)	5 (23.8)	32 (38.1)	

4.2 Exogenous risk factors: DEMETRA project

4.2.1 Internal Validation: SQFFQ's repeatability

To evaluate the repeatability of the SQFFQ developed to investigate dietary habits in a population of young adults, the SQFFQ was administered twice, 3 weeks apart, to a sample of 30 young adults under 50 years.
Table 3 shows the calculated measurement error (E), expressed as a percentage, between the consumption estimates from the two administrations (Q1 and Q2) of the SQFFQ. For 40 out of 67 food groups, the difference between the two SFFQs is less than 20%. The greatest percentage differences were found, as expected, for food groups with sporadic consumption (snacks, barbecue sauces, ketchup, mayonnaise) or consumed in minimal quantities (spices).

Table 3. Measurement error (E%)

		1st completion of SQFFQ				2nd completion of SQFFQ				E	E%
		Q_1				Q_2					
Food group	**code**	**Mean (g)**	**Std. Dev.**	**Min**	**Max**	**Mean (g)**	**Std. Dev.**	**Min**	**Max**	**Q_2-Q_1**	**Q_2-Q_1/Q_2**
Potatoes	0101	11,99	9,14	1,57	31,50	11,95	8,10	-	27,40	0,04	0,00
Sweet potatoes	0102	5,41	6,98	-	19,83	4,89	6,82	-	27,40	0,52	0,11
Vegetables	0200	7,60	5,11	0,23	20,89	7,45	5,99	0,03	26,72	0,16	0,02
Leafy vegetables (except cabbages)	0201	39,70	38,04	4,18	171,86	33,45	22,35	3,18	104,92	6,25	0,19
Other vegetables	0202	40,63	34,93	4,20	151,15	45,54	34,76	5,78	138,31	-4,91	-0,11
Root vegetables	0203	25,02	20,44	1,57	67,13	32,34	26,04	3,47	122,95	-7,33	-0,23
Cabbages	0204	19,82	20,00	-	69,40	21,19	23,16	-	112,70	-1,37	-0,06
Onion, garlic	0207	14,76	13,82	-	57,29	14,81	11,33	-	46,13	-0,05	-0,00
Stalk vegetables, sprouts	0208	13,40	15,00	-	62,79	15,72	20,92	-	112,70	-2,31	-0,15
Legumes	0301	15,16	14,11	-	62,86	12,89	11,59	-	34,29	2,27	0,18
Citrus fruits	040101	45,74	48,68	-	208,41	52,34	54,34	2,68	207,84	-6,60	-0,13
Fruits	040103	195,57	195,94	17,51	962,50	222,13	243,38	8,29	1.062,50	-26,56	-0,12
Nuts and seeds (+ nut spread)	0402	5,46	4,42	-	19,09	5,38	4,89	-	18,00	0,07	0,01
Milk	0501	153,25	153,84	-	386,68	149,75	152,23	-	386,00	3,50	0,02
Milk beverages	0502	-	-	-	-	-	-	-	-	-	-
Yogurt, thick fermented milk	0503	33,32	46,56	-	158,21	30,85	46,75	-	158,21	2,47	0,08
Cheese	0505	31,53	26,12	-	124,58	29,87	24,24	-	102,18	1,66	0,06
Dairy creams	0507	-	-	-	-	-	-	-	-	-	-
Pasta	0602	57,05	40,12	-	160,00	57,16	38,73	0,28	160,00	-0,11	-0,00

Filled pasta	060201	14,19	15,90	-	65,22	12,42	14,03	-	70,59	1,76	0,14
Rice	060202	15,02	11,37	-	29,57	11,47	10,42	-	36,43	3,56	0,31
Pasta like cereal-based products (not 100% cereal)	060204	2,49	6,85	-	29,57	4,84	9,21	-	36,43	-2,35	-0,49
White bread	06030101	123,68	230,71	-	1.188,00	115,95	225,45	-	1.188,00	7,74	0,07
Non-white bread	06030102	60,11	100,43	-	396,00	72,82	157,46	-	784,00	-12,71	-0,17
Breakfast cereals	0604	12,45	15,72	-	56,25	14,14	19,83	-	80,22	-1,69	-0,12
Salty biscuits, aperitif, biscuits, crackers	0605	3,83	12,61	-	58,33	6,58	23,81	-	122,50	-2,75	-0,42
Red meat	0701	40,71	51,95	3,08	276,00	29,18	26,91	-	98,57	11,53	0,40
White meat	0702	40,75	30,19	-	141,00	41,25	30,80	-	140,64	-0,50	-0,01
Game and offal meat	0703	2,38	7,47	-	39,43	3,55	14,52	-	79,29	-1,16	-0,33
Processed meat	0704	20,12	20,25	-	73,29	18,88	21,28	-	73,29	1,24	0,07
Fish	0801	47,33	37,97	-	163,43	44,18	33,04	-	99,65	3,15	0,07
Crustaceans, clams	0802	4,97	5,60	-	25,71	4,14	5,34	-	25,71	0,83	0,20
Egg	0901	18,06	14,86	-	51,43	17,13	13,62	-	51,43	0,92	0,05
Liquid egg	090101	3,69	16,15	-	85,71	6,99	22,00	-	85,71	-3,30	-0,47
Other fats	1000	2,08	2,54	-	12,01	1,84	3,25	-	17,60	0,24	0,13
Oils	1001	9,04	5,63	-	21,10	9,77	6,20	-	22,69	-0,74	-0,08
Butter	1002	0,80	1,27	-	4,57	0,57	0,96	-	4,29	0,23	0,40
Margarine	1003	0,00	0,02	-	0,08	0,09	0,29	-	1,43	-0,08	-0,97
Mix butter margarine	100302	-	-	-	-	-	-	-	-	-	-
Sugar, honey, jam, syrup	1101	8,26	12,82	-	55,99	7,81	12,31	-	54,99	0,44	0,06
Chocolate, candy bars, paste, confetti/flakes	1102	15,06	23,94	-	87,50	9,84	12,71	-	42,86	5,23	0,53
Ice cream	110501	11,12	24,02	-	116,67	8,73	15,33	-	75,00	2,40	0,27
Cakes, pies, pastries, puddings (non-milk based)	1201	13,91	22,05	-	111,81	14,49	24,03	-	118,13	-0,58	-0,04
Dry cakes, biscuits	1202	3,81	6,77	-	32,22	2,30	3,09	-	13,13	1,50	0,65
Arbonated/soft/isotonic drinks, diluted syrups	1302	8,64	18,14	-	69,14	13,23	30,41	-	150,00	-4,60	-0,35
Coffee	130301	66,61	39,49	-	150,00	69,75	43,19	-	150,00	-3,14	-0,04
Decaffeinated coffee	13030101	8,93	34,24	-	150,00	17,14	60,97	-	300,00	-8,21	-0,48
Tea	130302	55,98	118,02	-	525,00	42,68	73,93	-	300,00	13,30	0,31
Water	1304	1.516,67	516,68	500,00	2.500,00	1.450,00	562,48	500,00	2.500,00	66,67	0,05

Wine	1401	53,13	77,48	-	294,64	64,27	87,40	-	285,71	-11,13	-0,17
Beer	140301	97,04	241,34	-	1.272,86	103,58	182,65	-	848,57	-6,55	-0,06
Spirits, brandy	1404	1,15	2,61	-	11,43	1,58	3,10	-	14,29	-0,43	-0,27
Cocktails, punches	1407	10,17	22,87	-	117,86	12,11	29,25	-	160,71	-1,93	-0,16
Sauces	150100	3,20	3,03	0,06	14,64	2,11	1,72	0,10	5,84	1,10	0,52
Tomato sauces	150101	14,64	12,14	0,58	59,73	14,14	11,28	0,07	53,89	0,50	0,04
Ketchup	15010101	0,30	1,03	-	5,43	0,12	0,43	-	1,99	0,18	1,55
Mayonnaises and similars	150103	0,38	2,11	-	11,54	0,05	0,25	-	1,36	0,34	7,50
Condiments	1504	1,74	2,34	-	8,36	1,83	2,27	-	8,97	-0,09	-0,05
Miscellaneous	17	0,15	0,60	-	3,00	0,10	0,55	-	3,00	0,05	0,50
Soya products	1701	11,33	55,55	-	303,29	11,33	55,55	-	303,29	-	0
Other dietetic products	170200	1,21	3,93	-	15,00	0,71	2,94	-	15,00	0,50	0,70
Snacks	1703	5,96	13,95	-	67,50	4,40	8,14	-	39,38	1,56	0,35
Pizza	170303	35,79	22,99	3,80	130,29	36,93	28,90	2,53	130,29	-1,15	-0,03
Non-dairy creams, creamers	1704	0,17	0,80	-	4,33	0,36	1,70	-	9,29	-0,19	-0,53
Sweet food spreads	170401	0,30	0,75	-	3,93	0,96	2,24	-	10,00	-0,66	-0,69

The results of the agreement test between the 2 administrations of the SQFFQ obtained from the calculation of Cohen's kappa coefficient are shown in Table 4. Cohen's kappa values generally indicated good agreement. Insufficient agreement, with Cohen's Kappa values < 0.21, was observed for only three food groups (snacks, sauces, mayonnaises) out of the 61 examined. Agreement levels were as follows: 12 food groups showed fair/sufficient agreement (Cohen's kappa 20-40), 23 foods exhibited good/moderate agreement (Cohen's kappa >40-60), and 22 food groups demonstrated high substantial agreement (Cohen's kappa >60).

Table 4. Results of the agreement test

Variable	Label	Agreement	Expected Agreement	Cohen's Kappa	Prob>Z
0101	Potatoes	53,33%	33,33%	0,30	0,010
0102	Sweet potatoes	83,33%	37,78%	0,73	-
0200	Vegetables	73,33%	33,33%	0,60	-
0201	Leafy vegetables (except cabbages)	73,33%	33,33%	0,60	-
0202	Other vegetables	56,67%	33,33%	0,35	0,003
0203	Root vegetables	80,00%	33,33%	0,70	-
0204	Cabbages	76,67%	33,33%	0,65	-
0207	Onion,garlic	63,33%	33,33%	0,45	0,000
0208	Stalk vegetables,sprouts	60,00%	33,33%	0,40	0,001
0301	Legumes	80,00%	39,44%	0,67	-
040101	Citrus fruits	80,00%	33,33%	0,70	-
040103	Fruits	66,67%	33,33%	0,50	-

0402	Nuts and seeds (+ nut spread)	60,00%	33,33%	0,40	0,001
0501	Milk	73,33%	34,22%	0,59	-
0503	Yogurt, thick fermented milk	90,00%	33,33%	0,85	-
0505	Cheese	60,00%	33,33%	0,40	0,001
0602	Pasta	66,67%	33,33%	0,50	0,000
060201	Filled Pasta	80,00%	33,33%	0,70	-
060202	Rice	60,00%	34,22%	0,39	0,001
060204	Pasta like cereal-based products (other than cereals)	56,67%	42,00%	0,25	0,016
06030101	White bread	66,67%	33,33%	0,50	0,000
06030102	Non-white bread	63,33%	33,33%	0,45	0,000
0604	Breakfast cereals	70,00%	33,33%	0,55	-
0605	Salty biscuits, aperitif, biscuits, crackers	80,00%	67,78%	0,38	0,017
0701	Red meat	70,00%	34,67%	0,54	-
0702	White meat	73,33%	32,89%	0,60	-
0703	Game and offal meat	90,00%	62,44%	0,73	-
0704	Processed meat	66,67%	33,33%	0,50	0,000
0801	Fish	70,00%	33,33%	0,55	-
0802	Crustaceans, clams	70,00%	33,33%	0,55	-
0901	Egg	83,33%	36,22%	0,74	-
090101	Liquid egg	93,33%	81,78%	0,63	0,000
1000	Oil	53,33%	33,33%	0,30	0,010
1002	Butter	66,67%	34,67%	0,49	0,000
1101	Sugar, honey, jam, syrup	80,00%	33,33%	0,70	-
1102	Chocolate, candy bars, paste, confetti/flakes	76,67%	33,33%	0,65	-
110501	Ice cream	60,00%	33,33%	0,40	0,001
1201	Cakes, pies, pastries, puddings (non milk based)	73,33%	33,33%	0,60	-
1202	Dry cakes, biscuits	66,67%	33,56%	0,50	0,000
1302	Carbonated/soft/isotonic drinks, diluted syrups	73,33%	41,56%	0,54	-
130301	Coffee	76,67%	39,56%	0,61	-
13030101	Decaffeinated coffee	96,67%	84,67%	0,78	-
130302	Tea	73,33%	33,33%	0,60	-
1304	Water	53,33%	30,22%	0,33	-
1401	Wine	96,67%	33,33%	0,95	-
140301	Beer	63,33%	33,44%	0,45	0,000
1404	Spirits, brandy	73,33%	48,89%	0,48	0,001
1407	Cocktails, punches	66,67%	34,67%	0,49	0,000
150100	Sauces	43,33%	33,33%	0,15	0,123
150101	Tomato sauces	70,00%	33,33%	0,55	-
15010101	Ketchup	93,33%	76,67%	0,71	-
150103	Mayonnaises and similars	93,33%	93,56%	-0,03	0,575
1504	Condiments	53,33%	33,33%	0,30	0,010
17	Miscellaneous	96,67%	90,44%	0,65	0,000
1701	Soya products	100,00%	87,56%	1,00	-
170200	Other dietetic products	96,67%	84,67%	0,78	-
1703	Snacks	46,67%	33,33%	0,20	0,060
170303	Pizza	83,33%	45,00%	0,70	-
1704	Non-dairy creams, creamers	96,67%	84,67%	0,78	-
170401	Sweet food spreads	63,33%	33,33%	0,45	0,000

Therefore, the ad hoc designed SQFFQ provides a reasonably repeatable measure of dietary intake and can be used to assess the dietary and drinking habits of volunteers in this age

group. Most staple foods in the Italian diet, including pasta, fruit, vegetables, legumes, eggs, meat, coffee, and tea, are well estimated by the SQFFQ. However, challenges and difficulties persist, as well-described in the literature, particularly in estimating the consumption of foods consumed sporadically (such as snacks) or in small quantities such as spices.
Definitive conclusions will be drawn after the completion of the ongoing external validation, involving 100 volunteers from the same age group. In this phase, the SQFFQ is going to be validated against the gold standard, represented by a 4 days-food diary.

4.3 miRNA signature to predict recurrence in stage I-III eoCRCs

4.3.1 Discovery phase

To identify a miRNA signature for stratification of patients with low- and high-risk stage I and III eoCRC, and therefore to identify differentially expressed miRNAs in eoCRC patients with and without recurrence, we performed genomewide, high-throughput, small RNA-sequencing in FFPE-derived RNA in the Discovery cohort of patients with stage II-III eoCRC (n=20; 10 recurrent and 10 non-recurrent).
These sequencing efforts initially led to the identification of 35 miRNAs out of 268 that were significantly and differentially expressed between patients with eoCRC experiencing post-curative intent surgery recurrence and those without recurrence. In particular, we found 18 up-regulated and 17 down-regulated miRNAs in eoCRCs experiencing recurrence compared to eoCRC without recurrence (Fig. 15A).
Subsequently, we performed univariate LASSO-based Cox regression analysis and AUC-based evaluation which yielded a panel of 10 best-performing target miRNAs, including: hsa-let-7g-5p, hsa-miR-125b-2-3p, hsa-miR-125b-5p, hsa-miR-142-3p, hsa-miR-15b-3p, hsa-miR-30e-5p, hsa-miR-365a-3p, hsa-miR-410-3p, hsa-miR-654-3p, and hsa-miR-99a-5p (Tab. 5).

Table 5. Selected microRNA genes

miRNA name	Base Mean	Log2 Fold Change	P value	AUC
hsa-miR-125b-5p	1835.618443	1.125609884	7.73E-05	0.939393939
hsa-miR-654-3p	125.6082911	0.786300883	0.001165211	0.939393939
hsa-miR-410-3p	107.4597691	0.650584929	0.001871584	0.939393939
hsa-miR-30e-5p	8030.774299	-0.537393936	0.001996273	0.919191919
hsa-miR-365a-3p	71.48834822	0.822557382	0.003058388	0.919191919
hsa-let-7g-5p	5637.914671	-0.343753675	0.004480106	0.919191919
hsa-miR-125b-2-3p	67.66862698	1.522615732	0.000232049	0.898989899
hsa-miR-99a-5p	770.1452165	1.400822254	0.000373016	0.898989899
hsa-miR-142-3p	807.4770031	-1.178800775	0.000151859	0.888888889
hsa-miR-100-5p	4853.758414	0.996673131	0.00102211	0.878787879
hsa-miR-15b-3p	114.9609699	-0.738961969	0.008867044	0.878787879

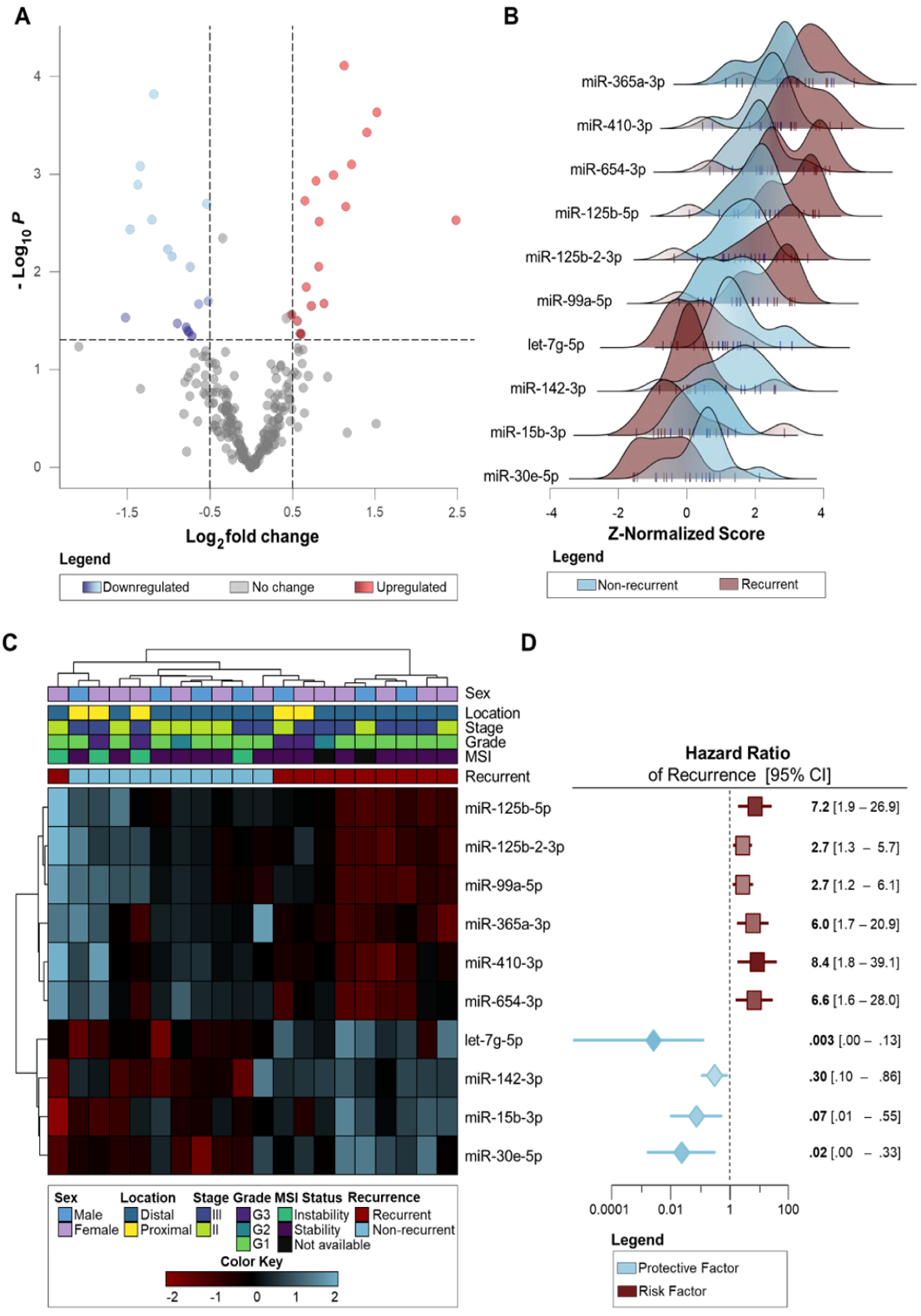


Figure 15: Discovery phase.

(A) Volcano plot showing differentially expressed miRNA between eoCRCs experiencing post-curative intent surgery recurrence and those without recurrence. Red: overexpression in recurrent cases; blue: overexpression in non-recurrent cases; color grading follows significance; (B) Ridgeline plot of the 10 best-performing miRNA showing difference in expression between patients with and without recurrent eoCRC (red and blue, respectively); (C) Heatmap with unsupervised clustering showing separation between eoCRC experiencing recurrence and those not experiencing recurrence; (D) Hazard ratio of recurrence for each individual miRNA: squares represent a higher risk of recurrence, diamonds represent a higher likelihood of non-recurrence.

This panel of candidate miRNAs showed significant differences in expression between eoCRC patients with and without recurrence: the hsa-miR-365a-3p, hsa-miR-410-3p, hsa-miR-654-3p, hsa-miR-125b-5p, hsa-miR-125b-2-3p and hsa-miR-99a-5p resulted up-regulated in eoCRCs experiencing recurrence, while hsa-let-7g-5p, hsa-miR-142-3p, hsa-miR-15b-3p, and hsa-miR-30e-5p were found up-regulated in eoCRCs without recurrence (Fig. 15B). Unsupervised clustering of the expression levels of the candidate miRNA revealed that the only clinically relevant factor co-segregating with the unsupervised clustering was the development of recurrence in the five years following curative intent surgery (Fig. 15C). The heatmap shows a clear separation between eoCRC experiencing recurrence and those not experiencing recurrence. We finally used this panel of 10 miRNAs to assess the hazard ratio (HR) of recurrence for each individual miRNA (Fig. 15D), demonstrating that hsa-miR-365a-3p (HR 6.0; 95% confidence interval (CI) 1.7-20.9) , hsa-miR-410-3p (HR 8.4; 95% CI 1.8-39.1), hsa-miR-654-3p (HR 6.6; 95% CI 1.6-28.0), hsa-miR-125b-5p (HR 7.2; 95% CI 1.9-26.9), hsa-miR-125b-2-3p (HR 2.7; 95% CI 1.3-5.7) and hsa-miR-99a-5p (HR 2.7; 95% CI 1.2-6.1) carry the higher risk of eoCRC recurrence. Conversely, hsa-let-7g-5p (HR 0.003; 95% CI 0.00-0.13), hsa-miR-142-3p (HR 0.3; 95% CI 0.10-0.86), hsa-miR-15b-3p (HR 0.07; 95% CI 0.01-0.55), and hsa-miR-30e-5p (HR 0.02; 95% CI 0.00-0.33) are associated with a higher likelihood of non recurrence.

4.3.2 Training and validation phases

To validate the results found during the discovery phase and to evaluate the diagnostic potential of the 10-miRNA signature in identifying eoCRC patients at high-risk of recurrence, we examined its performance in FFPE tissues of other two clinical cohorts: the Training and Validation cohorts.

The majority of the patients received a follow-up of at least up to 5 years (median 85.4, 95% confidence interval [CI] 68.87-104.75), unless they died first.

First, we analyzed the differential expression of the 10 miRNAs signature using qRT-PCR assay in a *Training cohort* of 88 FPPE samples from Spanish patients with stage I-III eoCRC (24 recurrent and 63 non-recurrent eoCRCs). Using XGBoosting, the performance of this miRNAs panel was trained in the first FFPE cohort. The resulting learning model (Fig. 16A) was able to perform robustly and predict accurately the development of recurrence based on a hyper-selected panel of only 9 microRNAs, a slight improvement compared to the 10 miRNA panel. More specifically, miR-99 appeared to be non-contributory to the model architecture because its expression levels could not further refine the accuracy of the prediction nor provide a gain for the model in any leaf (Figure 16B). The optimized XGBoost-based 9-miRNAs risk-assessment model demonstrated a high accuracy in predicting recurrence in stage I-III eoCRC patients in the training cohort with an AUC value of 0.90 (95% CI 83-95%) (Figure 16C) with a Youden index of 64.9% (CI95%, 55%-82%), an accuracy of 81.8% (77-93%), sensitivity 84.0% (65-96%), specificity 81.0% (72-98%). The raincloud plot (Figure 16D) shows the distribution of XGB prediction scores between eoCRC

patients with and without recurrence development in the years following curative intent surgery.

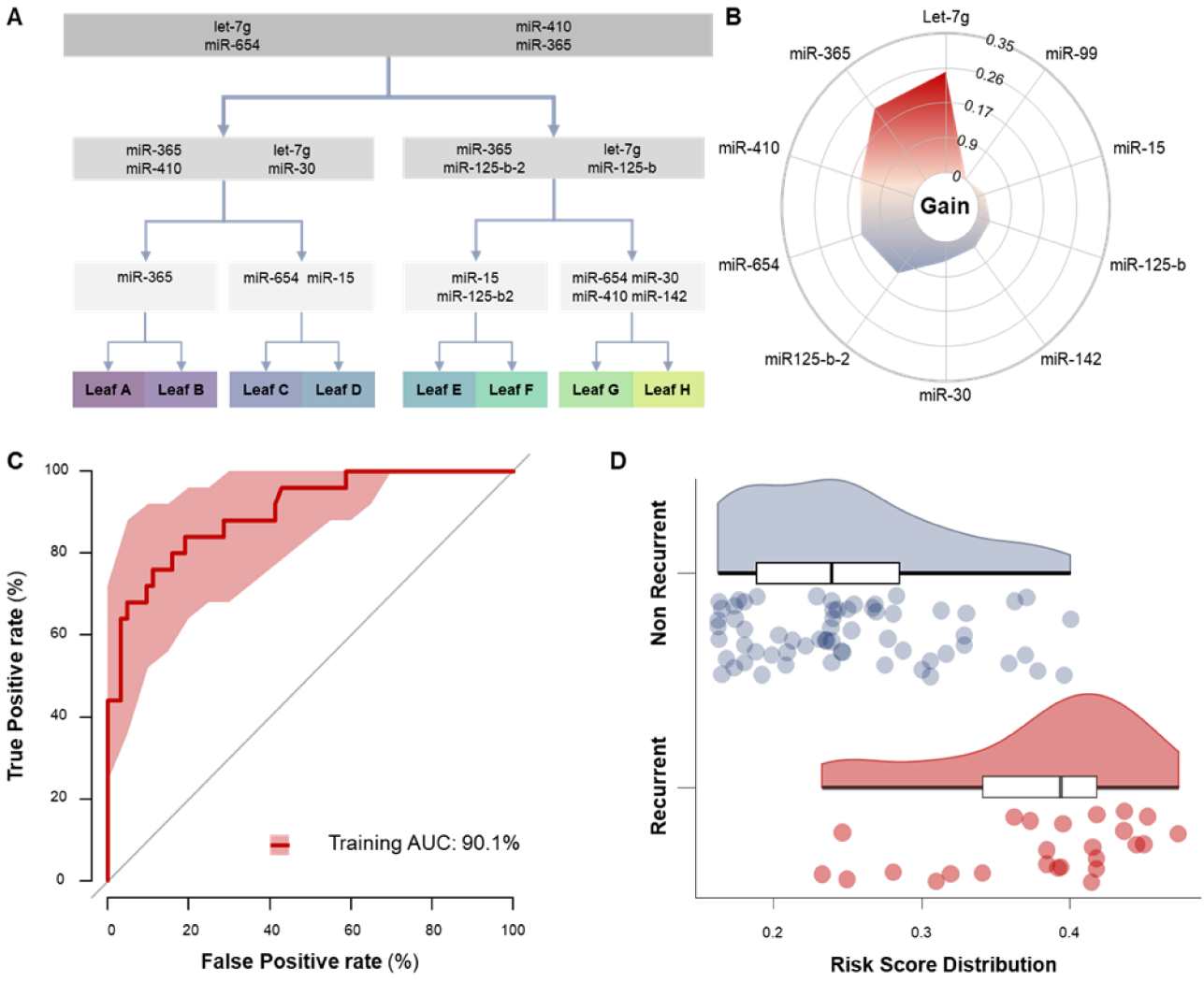


Figure 16: Architecture, Functionality, and Performance of the Learning Algorithm (XGBoost) in the Training Cohort.

(A) Decision tree showing the collapsed view of the decision forest; (B) Gain represents the improvement in accuracy brought by each microRNA to the branches it is on (C) AUROC in the training cohort; (D) Raincloud plots with superimposed whisker and box plot demonstrating the distribution of the risk scores between recurrent cases (red) and non-recurrent cases (blue).

To better understand the performance of the learner algorithm, we evaluated each feature individually in terms of gain, cover, and frequency metrics. We could consistently observe that the features provided the most contribution to the model was let-7g, followed consistently by iR-365, and miR-410 (Figure 17A, 17B, and 17C).

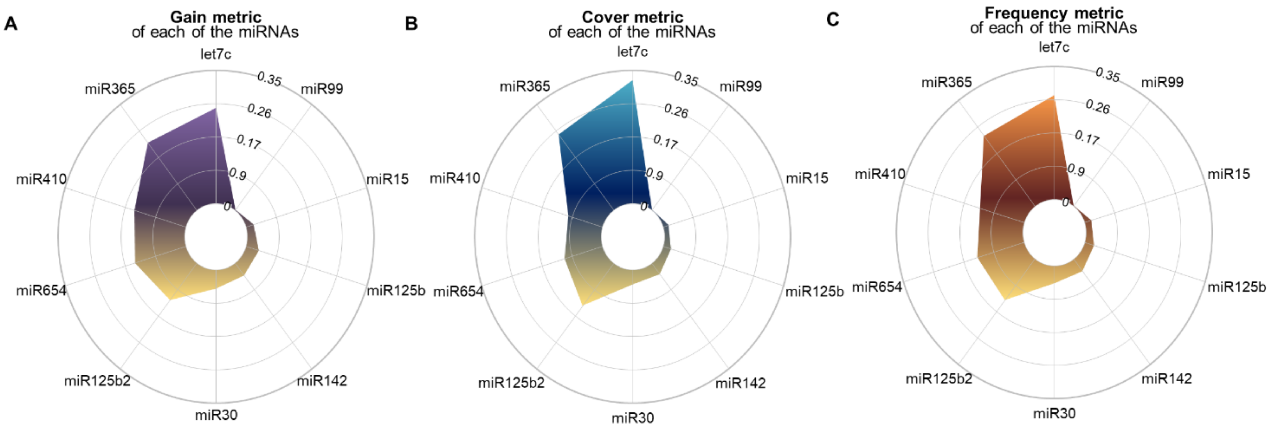


Figure 17: Feature Values

(A) Gain associated with each model feature; (B) Cover associated with each model feature; (C) Frequency associated with each model feature

Because gain, cover and frequency provide an aggregated measure of a feature's performance, rather than its workings at the patient level, we evaluated the model inner architecture and structure using a game-theory derived approach (SHAP, fig. 18A).

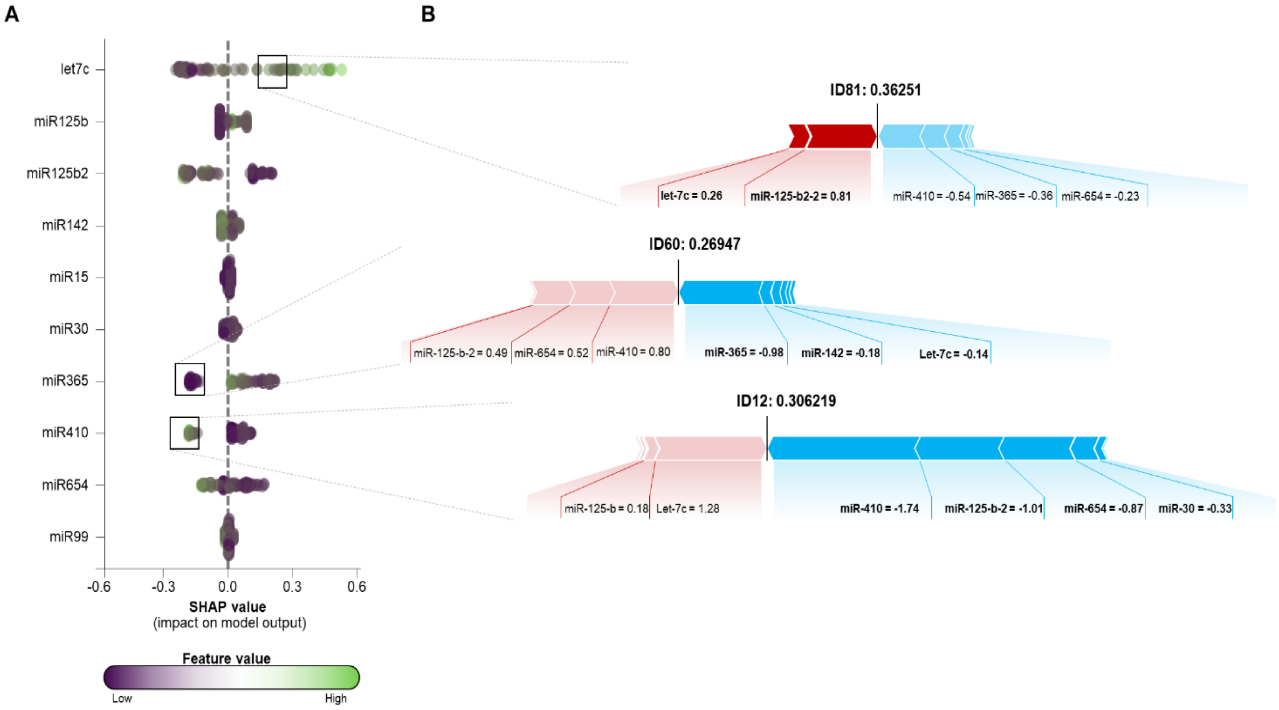


Figure 18: SHAP Value Model Evaluation

(A) Beesworm plot plotting the SHAP value of each feature for each patient included in the training cohort; (B) SHAP Value Force plots indicate the convergence of positive and negative predictors towards the final XGB prediction which favors the development of recurrence (top force plot), or the lack of it (middle and bottom force plots)

SHAP values demonstrate that most patients have let-7g values that consistently separate from 0 in both directions, which demonstrate that let-7g not only contributes to the overall model, but also to practically all patients to a relatively large extent. The second- and third-highest gains were contributed to by miR-365 and miR-410, respectively, whose SHAP

values were extremely informative in the negative direction, with almost all patients being well separated from the 0-value line, but fewer patients being separated from it in the positive side. We then interrogated the inner workings our model at the patient level by means of force plots, which revealed that, for the few patients with low let-7g SHAP values, the greatest contribution to the model prediction came from miR-125-b2-2, miR-365, or miR-410 (Fig. 18B). The three patients here portrayed are representative examples.

We then evaluated the survival characteristics of the XGB model and observed that our 9-miRNA model can discriminate effectively between recurrent and non-recurrent cases up to 20 years after surgical resection (Fig. 19A), with most cases occurring in the first few years after surgery. While the entire training cohort had a median recurrence free survival (mRFS) of 67.5 months (IQR: 21.9-111.5), there were statistically significant differences between patients classified as high-risk vs. low-risk. Patients predicted to be at a high risk of recurrence by the XGB model had a statistically significant cumulative hazard of disease recurrence than those classified as low-risk ($p<0.001$, Fig. 19B). Likewise, patients with a high-risk 9-miRNA signature a significantly shorter mRFS versus low-risk patients (24.4 months, CI95%: 14.6 – NA, versus median not reached, $\chi^2 = 38.8$, $p<0.0001$) (Fig. 19C).

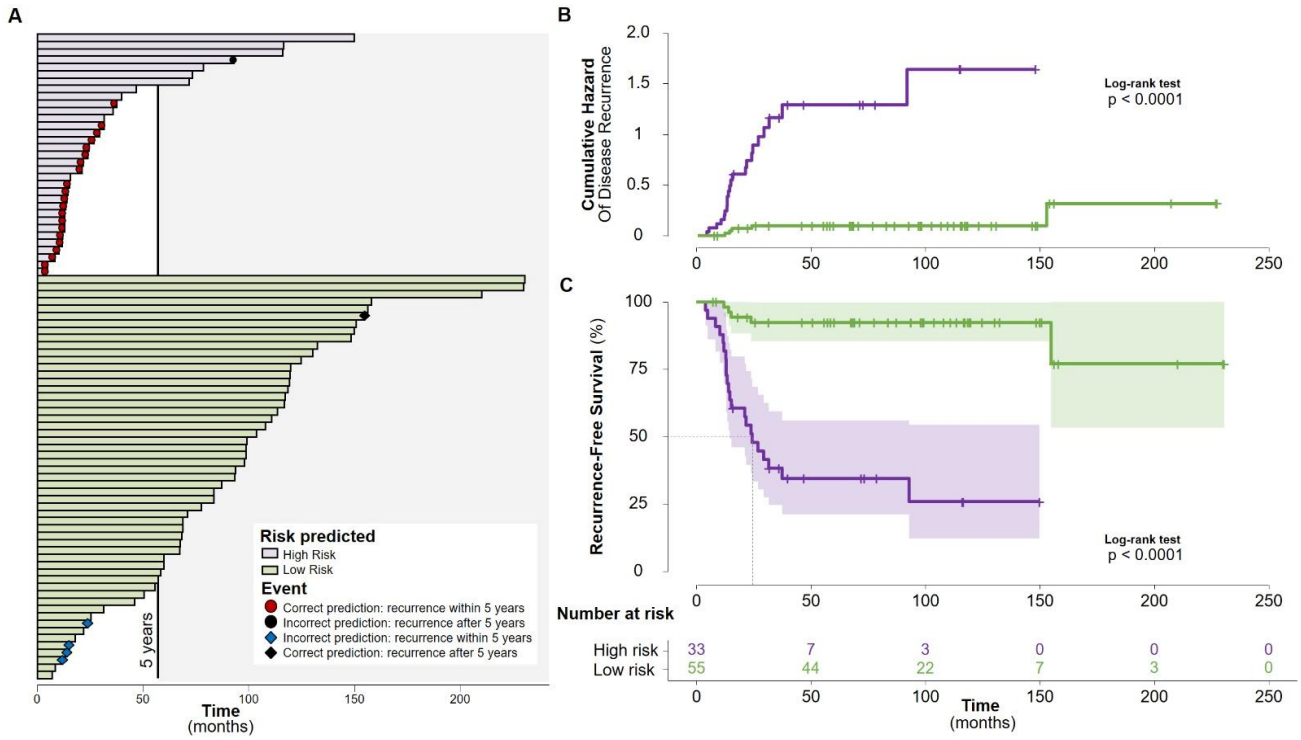


Figure 19: Survival curves

(A) Swimmers plot demonstrates that the learning algorithm can discriminate between recurrent and non-recurrent cases up to 20 years after surgical resection; (B) Cumulative hazard of disease recurrence in patients predicted to be at a high risk of disease recurrence (purple) vs. low risk (green); (C) Kaplan-Meier curve demonstrating the probability of disease recurrence between patients classified as high risk (purple) and low risk (green).

In view of the encouraging results of the tissue-based 9-miRNA panel in discriminating between recurrent and non-recurrent eoCRC in the training cohort, the robustness and

accuracy of our risk-assessment model were assessed in a separate and independent cohort of 69 FFPE samples from Japanese patients with stage I-III eoCRCs (9 recurrent and 60 non-recurrent), the *Validation cohort* (Fig. 20).

The XGBoost-based risk-assessment model, incorporating 9-miRNAs, exhibited good accuracy in predicting recurrence among stage I-III eoCRC patients in the validation cohort, achieving an AUC value of 77.3% (95% CI 67.0-87.0%; Fig. 20A), with a sensitivity of 100% (88-100%), specificity of 62.9% (55-82%), accuracy 66.7% (59.0-83.0%), and Youden index of 62.9% (54-73%). Despite the substantial differences between the two cohorts, these results confirm that the findings from the training cohort are transferrable to an independent, differently composed, and ethnically distinct cohort of patients, a finding that speaks of the general reliability and robustness of the approaches employed to limit overfitting. Even with notable differences between the two cohorts, such as varying proportions of recurrent vs. non-recurrent cases and differences in ethnicity, the Waterfall plot represented in Figure 20B demonstrates that our 9-miRNAs risk-model has high sensitivity even in the ethnically distant and independent validation cohort.

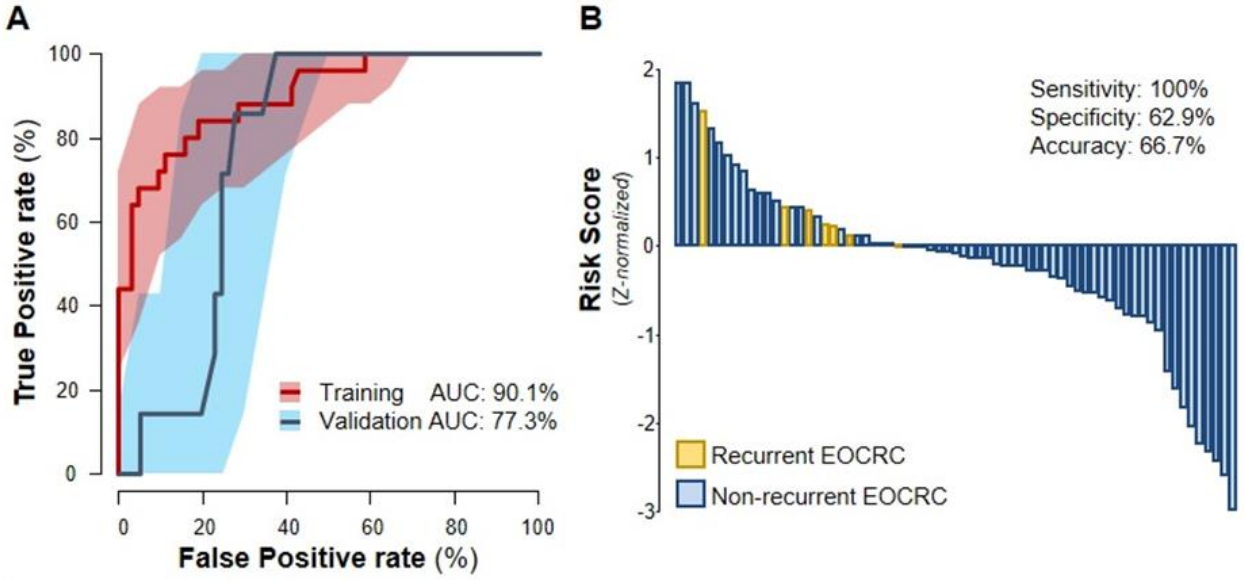


Figure 20: Validation Analyses

(A) ROC curved of the training and validation cohorts; (B) Waterfall plot representative of the Z-normalized risk scores observed in the validation cohort between eoCRC cases with and without a recurrence (gold and blue, respectively)

Even under such constrains, the Kaplan-Meier analyses revealed that over 10 years of follow-up, all patients who developed a recurrence were correctly classified as high risk, while also capturing the majority of non-recurrent cases as low risk (62.9%), for an overall $\chi^2 = 9.6$, $p<0.002$ (Fig. 21A). We finally demonstrated a remarkable stability of the trained 9miRNA risk assessment model, as showed by the comparison of the Youden index obtained for the training and validation cohort (Fig. 21B), both reaching a value > 0.5. Finally, 1000 Bootstrap demonstrated that both the training and validation cohorts have a peak Youden index > 60%, further validating the stability of the model in patients of both Caucasian

(Training cohort) and Asian (Validation cohort) ancestry (Fig. 21C).

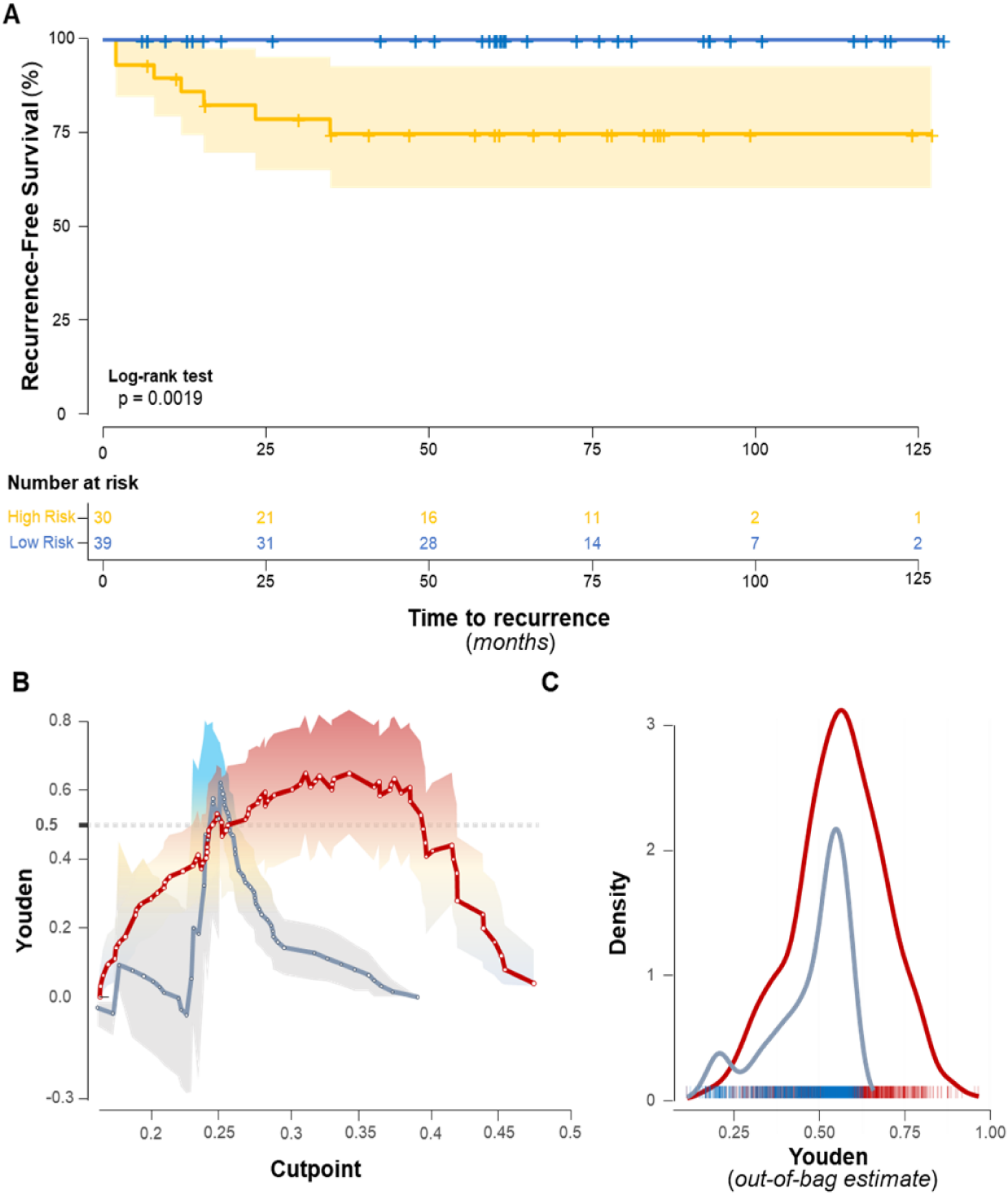
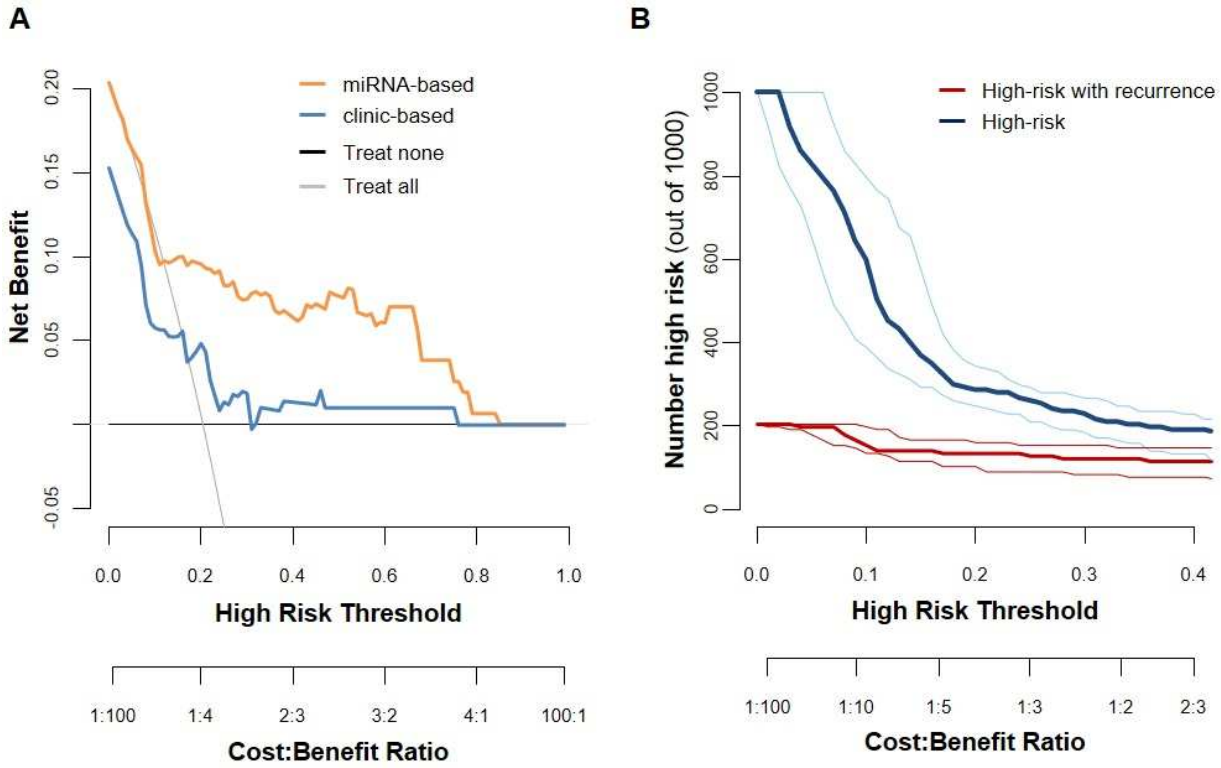


Figure 21: Survival Characteristics and Model Stability in the Validation Cohort

(A) Kaplan-Meier recurrence-free survival curves in the independent Asian cohort between patients classified as high-risk (yellow) and low risk (blue); (B) Comparison of the Youden index obtained for the training (red) and validation (blue) cohort demonstrates remarkable stability of the trained model; (C) 1000 Bootstrap demonstrated that both the training (red) and validation (blue) cohorts have a peak Youden index at about .65, further validating the stability of the model in patients of both Caucasian and Asian ancestry.

To estimate the clinical significance of the miRNA panel, decision curve analysis (DCA) was performed (Fig. 22A). The DCA curve revealed that a surveillance strategy based on our 9-miRNA panel achieved a higher net benefit in comparison with a surveillance strategy that would be aggressive for all patients, no patient, or based on clinical characteristics. This clinical approach was derived by fitting a logistic regression model based on six features (age, biological sex (male, female), tumor localization (right, left), tumor grade (high, intermediate, low), vascular invasion (yes, no), and lymphatic invasion (yes, no)). We then plotted the Net Benefit curves of the four models to estimate that the 9-miRNA model provided the greatest clinical effects in a population starting at a high-risk threshold corresponding to 0.12, that is to say, a cost:benefit ratio of 3:22. This implies that, adhering to the cost:benefit threshold value of roughly 1:4, a surveillance strategy based on our miRNA panel would classify 45.2% (95% CI: 33.8-74.5%) of eoCRC patients as being high-risk, with a true-positive rate of 68.8% (95% CI 62.5-93.8%) (Fig. 22B).
In particular, we observed a net benefit of the surveillance based on our microRNA signature compared to the clinical-based surveillance especially in stage I and II high risk eoCRC, representing the subgroups of patients that could benefit more from more aggressive post-treatment follow-up strategies (Fig. 22C-E).



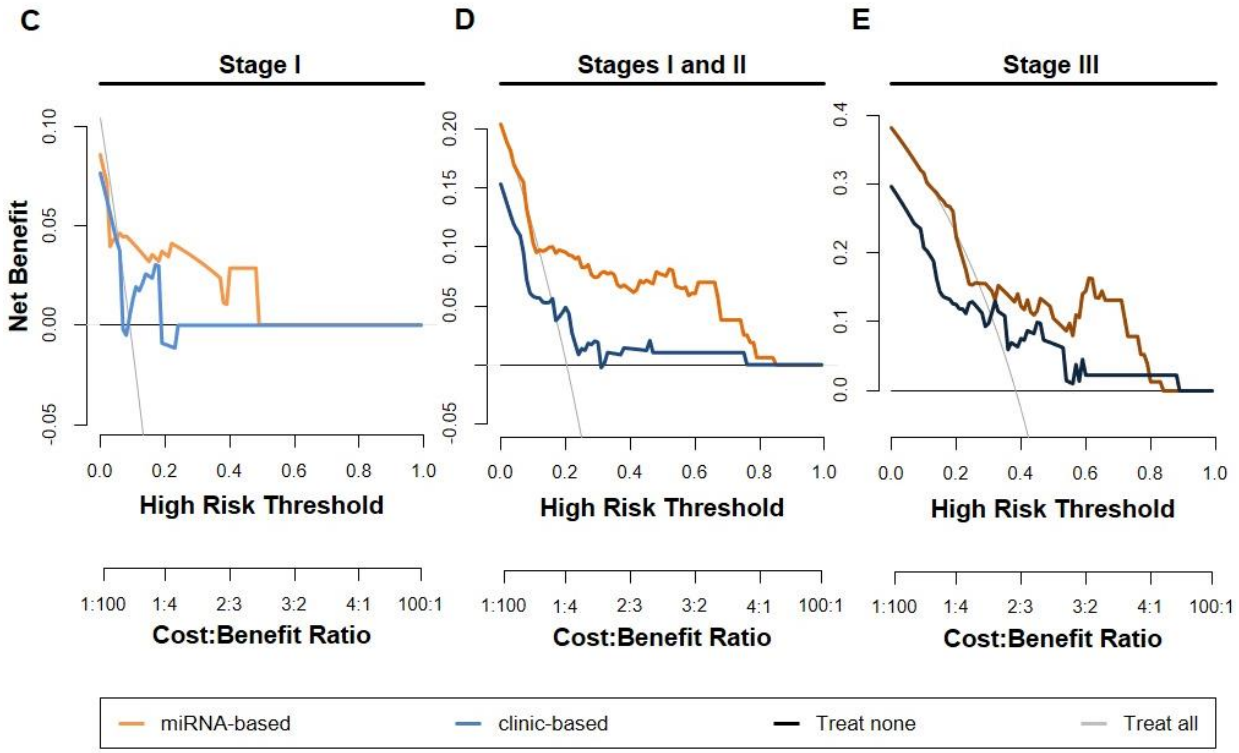


Figure 22: Decision Curve Analysis

(A) Unstandardized Net Benefit and cost: benefit analysis of a surveillance strategy based on our microRNA panel vs. clinical characteristics vs. treat all vs. treat none approaches; (B) Clinical Impact of a strategy based on a microRNA panel; (C-E) Net benefit analysis and cost: benefit analysis of our microRNA-based strategy vs. clinical factors for EOCRC diagnosed in stage I (C), a compound cohort of both stages I and II (D), and stage III (E)

DISCUSSION

5.1 Genetic risk assessment

In this doctoral project, we firstly aimed at evaluating the association of family history of CRC with eoCRC and assessing the prevalence of germline PVs through NGS-based multigene panels testing and MLPA.
We demonstrated that 71.4% of eoCRCs did not have a family history of CRC. Conversely, 19% of eoCRCs reported having a FDR with CRC and 12.4% had a SDR with CRC; three patients had both a FDR and SDR with CRC and were all Lynch patients.
Several studies are present in the literature concerning family history of CRC in young patients with eoCRC with similar results.
In a case series of 94 patients with eoCRC diagnosed before the age of 30 years, 57% did not report any family history of CRC, and less than 5% fulfilled the Amsterdam II criteria [276]. In a registry-based retrospective study of 5710 eoCRC and 11800420 age-matched healthy controls, eoCRCs more commonly had a family history of cancer (OR 11.66; 95% CI 10.97–12.39), gastrointestinal malignancies (OR 28.67; 95% CI 26.64–30.86), and polyps (OR 8.15; 95% CI 6.31–10.52). This was confirmed when eoCRCs were compared with loCRC (OR for family history of cancer 1.78; 95% CI, 1.67–1.90; OR for gastrointestinal malignancies 2.36; 95% CI 2.18–2.55; OR for polyps 1.41; 95% CI, 1.08– 1.20) [60].
In a retrospective, case-control study of 253 eoCRC and 232 loCRC, patients with eoCRC more frequently reported a family history of CRC (25% vs 17%; P = 0.03) or having a hereditary cancer syndrome (7% vs 1%; P < 0.01) [50]. Another retrospective study of 107 eoCRC and 139 loCRC, showed similar results in terms of family history of CRC (30% eoCRC vs 16% loCRC, P = 0.02) [277]. Finally, also a Dutch retrospective, registry-based, case-control study (521 patients with CRC at <40 years and 15000 with CRC at ages 66–75), described a more marked family history of CRC in eoCRC cohort of patients (24.1% vs 12.4%; P < 0.0001) [61].

We also observed that 33.3% mutated eoCRCs had a FDR with CRC at a mean age of 46 ± 12.6 y, while 15.5% non-mutated eoCRCs referred a FDR with CRC at a mean age 66.4 ± 8.3 y.
There is a consensus that having at least 2 first-degree relatives with CRC and/or at least 1 first-degree relative diagnosed with CRC before the age of 50–60 years are associated with a significant increase in risk for CRC. In these situations, screening colonoscopy starting at 40 years (or 10 years before the age at diagnosis of the youngest affected relative) is usually recommended. Indeed, a recent study showed that up to 16% of eoCRC could be prevented [89] if colonoscopy was performed at the age recommended by guidelines based on family history [91–95].
Therefore, as recently stated in the DIRECTt guidelines on eoCRCs, our results confirm that family history of CRC can inform risk assessment for both syndromic and non-syndromic CRC [1]. A thorough family history should be routinely collected for all individuals with eoCRC

through validated risk assessment tools to identify patients who would benefit more from germline genetic testing and from dedicated colonoscopy screening timelines. Among these risk assessment tools the Colon Cancer Risk Assessment Tool and the PREMM5 are the most important [96,97]. The PREMM5 tool can be used to determine the likelihood of a PV/LPV in a LS gene.

We finally found that 20% of eoCRC carried a germline PV of genes known to be associated with CRC. In detail, 12.4% eoCRC carried a germline PV of mismatch repair genes responsible of Lynch syndrome, 2.9% of BRCA1-2 responsible of Hereditary breast and ovarian cancer syndrome, 3.8% of MUTYH of which 3 were heterozygous and 1 homozygous, with the latter responsible of MAP. Moreover 0.9% eoCRC had a PV of ATM and 0.9% of SDHAF2. One patient exhibited mosaicism of PVs in MSH2/MUTYH genes.
Thirteen studies are available in literature on prevalence of PV/LPVs in cancer susceptibility genes in individuals with eoCRC. NGS revealed that the prevalence of PVs in cancer genes is 9.0%–35% among patients with eoCRC. The prevalence of LS was variable from 0% to 18.3%. The prevalence of other, non-LS, hereditary predisposition PV/LPV ranged from 2.3% to 23.1% (Tab. 6).

Table 6. Articles Providing the Prevalence of PV/LPV in Cancer Susceptibility Genes Among Early-Onset Colorectal Cancer Patients (modified from [1]).

Article	Age (y)	N. genes	PV/LPV	Cohort	Prevalence of Lynch syndrome (%)	Prevalence of PV/LPV in non-Lynch syndrome cancer genes (%)	Overall prevalence of PV/LPV in cancer susceptibility genes (%)
Laduca 2020 [100]	<50 subset	5–49	362	4017	5.3	4.7	9.0
Jiang 2020 [278]	<50 subset	14	47	261	15.7	2.3	18
Zhunussova 2019 [98]	<50	94	20	125	2.4	13.6	16
You 2019 [279]	<50 mts	46	10	67	2.9	12	14.9
Mork 2019 [101]	≤ 35	>1	24	136	18.3	11.1	29.4
AlDubayan 2018 [112]	<50 subset	54	5	35	0	14.3	14.3
Stoffel 2018 [48]	< 50	>1	85	430	13.5	6.5	20
Pearlman 2017 [47]	< 50	25	72	450	8	8	16
DeRycke 2017 [280]	< 50	36	88	333	13.5	12.9	26.4
Chubb 2016 [281]	≤ 55	WES	158	1006	11	4.7	15.7
Mork 2015 [46]	< 35	>0	67	193	11.9	23.1	35
Toh 2018 [282]	< 50	64	12	88	0	13.6	13.6
Yurgelun 2017 [109]	<50 subset	25	40	336	6.3	5.6	11.9

Estimating the impact of PVs in eoCRC predisposition remains an active field of research. With massive use of multigene panel testing, some recent trials have reported a significant

number of unexpected diagnoses. These include PVs in BRCA1/2, ATM, TP53, CDKN2A, CDH1, as well as other genes listed in Table 7.
However, as stated in the DIRECt guidelines, we agree with the recommendation that all patients with eoCRC should be offered multi-gene panel germline genetic testing before treatment to maximize clinical utility, when feasible. Germline genetic testing for eoCRC should include at a minimum: APC, BMPR1A, EPCAM, MLH1, MSH2, MSH6, MUTYH, POLD1, POLE, PMS2, PTEN, SMAD4, STK11, and TP53. Where available and not cost-prohibitive testing should also include the following genes that are reasonably prevalent in CRC and change clinical management: BRCA1, BRCA2, ATM, CHEK2, PALB2, and possibly, but less prevalent, BRIP1, BARD1, CDKN2A, CDH1, RAD51C, and RAD51D. Genetic testing should also include the following genes associated with CRC or polyposis: AXIN2, GREM1, MLH3, MSH3, MBD4, NTHL1, RNF43, and RPS20 [1].

Table 7. Data from all published articles regarding the prevalence of PV/LPV in each gene among eoCRC (modified from [1])

Gene	Positive	Prevalence (%)
Colorectal cancer genes:		
Lynch sd. Genes (MLH1, MSH2, MSH6, PMS2, EPCAM)	551	7.7
APC	93	1.3
Biallelic MUTYH	47	0.7
SMAD4	11	0.2
BMPR1A	6	0.09
STK11	2	0.03
PTEN	2	0.03
GREM1	1	0.04
AXIN2	-	-
POLE/POLD1	4	0.1
Other actionable cancer genes		
BRCA1/2	50	1.2
CHEK2	56	0.8
ATM	29	0.7
TP53	14	0.2
PALB2	7	0.2
BRIP1	5	0.1
CDKN2A	3	0.09
CDH1	6	0.09

5.2 SQFFQ's repeatability

Given that hereditary gastrointestinal tumor syndromes only contribute to a small portion of eoCRCs, it is essential to explore exogenous risk factors to gain a deeper understand eoCRC pathogenesis. Alcohol intake, physical activity, red and processed meat, and a Western dietary pattern are well demonstrated loCRC risk factors [124–130]. On the contrary, there have been relatively few studies on eoCRCs and their precursors, with the majority being case–control studies and only a handful being prospective studies that have analyzed dietary, lifestyle, and anthropometric risk factors [59,131–137]. Most of those studies were small, heterogeneous, focused exclusively on peculiar dietary and drinking habits of single countries without analyzing cooking, processing, and storage techniques.

Therefore, to comprehensively evaluate the association of dietary, lifestyle and anthropometric factors with eoCRC at global level, we designed an international case-control study, called DEMETRA. This study will involve populations from different European and American countries, characterized by large variations in dietary habits and eoCRC risk. These countries differ markedly in terms of climate, physical geography, history and wealth, with corresponding variations in diet, eating behaviors, cooking and storage methods. To overcome this issue, our 2nd aims were to develop a unique and shared SQFFQ able to accurately describe dietary and drinking habits of eoCRCs and healthy controls of different countries at global level that will be involved in the future DEMETRA study and to validate the SQFFQ.

To reach this aim we designed a shared, online, detailed SQFFQ investigating the usual consumption of 329 foods, grouped into 61 food groups, over the past year. The SQFFQ contains 25 sections: fruit, vegetables, legumes, red meat, white meat, game and offal, processed meat, fish, shellfish, milk and vegetables substitutes, yogurt and other dairy and non-dairy fermented products, cheese, bread, eggs, sweets/desserts/snacks, sugar, coffee, tea, drinks (sugary drinks, wine, beer, spirits), protein powders, pasta, rice, other cereals and pseudocereals, pizza, cooking fats and sauces. Each section analyzes food consumption through 8-9 questions, concerning:

(i) Food portion: the quantity of most of the consumed food was assessed by means of the respondent's selection of a food portion image; each set of images included 6 food portions. Conversely, drinks and milk were quantified with the aid of pictures of standard units such as glasses and cups; fruit was evaluated in standard unit (1 unit = one apple, one orange, 2 tangerine, 2 apricots, 2 medlars or ½ mug of berries or grapes [1 mug = 250 ml] or 1 slice of watermelon or 2 slices of pineapple); bread was quantified using number of slices and loaves.

(ii) Frequency of consumption, expressed as the number of times a given food item was consumed (Rarely [<1 time per month], 1-3 times per month, once a week, 2-4 times per week, 5-6 times per week, once a day, twice a day, 3-4 times per day, >4 times per day). For fruits, vegetables, processed meat, fish, shellfish, cheese, bread, sweet/dessert/snacks, pasta, and rice we also evaluated the relative consumption of different subtype (never, sometimes, half the time, often, always);

(iii) If the food was organic and consumed according to the seasonality

(iv) Types of seasoning

(v) Methods of cooking and storage: questions about food processing methods are of note. Indeed, food processing can have a marked influence on the quality and safety of foods, particularly because processing often involves the addition of various chemicals and cooking at high temperatures. These high-temperature cooking methods have been shown to be a source of mutagens and carcinogens such as heterocyclic amines and polycyclic hydrocarbons [283]. These substances induce breast tumors in rats [284] and have been implicated by dietary epidemiology studies as increasing the risk of breast [283,285,286] and colon cancer [287] in humans.

A software was finally developed to analyze responses and link them to food composition tables in order to provide a nutritional breakdown of individual and collective diets.

Once the online platform was completed, the SQFFQ was validated for repeatability by administering it twice, 3 weeks apart, to a sample of 30 young adults under 50 years (Internal Validation). Agreement levels, represented by the calculation of Cohen's kappa coefficient, were as follows: 12 food groups showed fair/sufficient agreement (Cohen's kappa 20-40), 23 foods exhibited good/moderate agreement (Cohen's kappa >40-60), and 22 food groups demonstrated high substantial agreement (Cohen's kappa >60). Therefore, the ad hoc designed SQFFQ provides a reasonably repeatable measure of dietary intake and can be used to assess the dietary and drinking habits of volunteers in this age group. Most staple foods in the Italian diet, including pasta, fruit, vegetables, legumes, eggs, meat, coffee, and tea, are well estimated by the SQFFQ. However, challenges and difficulties persist, as well-described in the literature, particularly in estimating the consumption of foods assumed sporadically (such as snacks) or in small quantities such as spices.
Definitive conclusions will be drawn after the completion of the ongoing External validation, involving 100 volunteers from the same age group. In this phase, the SQFFQ will be validated against the gold standard, represented by a 4 days-food diary followed by a dietary recall. Once validation is complete, the international, multicenter, case-control study will start to evaluate the associations of diet, lifestyle and anthropometric factors with eoCRC comparing patients from countries with different incidence of eoCRC.

5.3 miRNA signature as prognostic biomarker

The detection of precursor lesions and early-stage CRC in average-risk, asymptomatic individuals is the goal of screening. Colonoscopy is still considered the gold standard for CRC screening due to its potential to both detect and remove precursor lesions. However, colonoscopy remains an invasive and expensive procedure, hampered by possible complications. Conversely, the non-invasive screening tests, fecal occult blood test (FOBT) and fecal immunochemical test (FIT), have a lower sensitivity and specificity than colonoscopy, at least for precursor lesions. This highlights the need for novel, simple and non-invasive strategies for the detection of precursor lesions and early-stage CRC [288].

5-year relative survival rates range from 93.2% for stage I, 82.5% for stage II, 59.5% for stage III, and 8.1% for stage IV [289]. Post surgery, 30% of stage II and 50% to 60% of stage III CRC develop a recurrence within 5 years [290]. Although there is general agreement that adjuvant chemotherapy in patients with stage III improves patient survival [78,291], the use of such treatments in stage II remains debatable due to lack of risk stratification for identifying true high-risk patients [79,292]. Current NCCN guidelines recommend adjuvant chemotherapy for patients with high-risk stage II CRC, where the risk

is primarily defined by the clinicopathologic features such as tumor size, number of lymph nodes investigated, degree of differentiation, tumor perforation, bowel obstruction, and lymph vascular invasion [79]. However, several studies have highlighted the inadequacy of these pathologic features in identifying such high-risk patients, providing a potential explanation for the lack of clinical benefit from adjuvant therapy in these patients [293,294]. Furthermore, a significant proportion of stage III suffer from adverse effects of adjuvant chemotherapy [289]. Moreover, the current TNM classification system for CRC staging is not completely adequate for prognosis and clinical decision-making, particularly for intermediate CRC stages (II-III) [295]. Therefore, there is also the need of biomarkers able to identify patients at higher risk of recurrence (prognostic biomarkers), as well as patients who would benefit from chemotherapy, immunotherapy and/or targeted therapy (predictive biomarkers).

CRC develops through a stepwise accumulation of genetic and epigenetic alterations in precursor lesions eventually progressing to adenocarcinoma. Hence, epigenetic changes might be potential diagnostic and prognostic biomarkers for cancers.

While biomarkers based on DNA methylation have already been commercialized, miRNAs are the most promising group of potential future biomarkers for CRC by virtue of their ability to resist RNAase-mediated degradation and their intact expression in a variety of human samples including FFPE.

In the past decade, the number of studies investigating miRNAs in loCRC has increased exponentially [239,242,296–299]. One of those systematic studies was performed by Kandimalla et al. [298]. They used three independent genome-wide miRNA expression profiling datasets for biomarker discovery (n. 158) and in silico validation (n. 109 and n. 40) to identify a miRNA signature for predicting tumor recurrence in patients with stage II-III loCRC. Subsequently, this signature was trained and validated in retrospectively collected independent patient cohorts of fresh-frozen (n 127, cohort 1) and FFPE (n. 165 cohort 2 and n 139 cohort 3). They identified an 8-miRNA signature that significantly predict cancer recurrence: *miRNA-744, miiRNA-429, miRNA-362, miRNA-200b, miRNA-191, miRNA-30c2, miRNA-30b, and miRNA-33a*. This miRNA signature has superior predictive power over clinicopathologic risk determinants and currently available commercial assays.

For a limited number of miRNAs - miR-21[243], miR-224[300], miR-106a[301], miR-29a[302] and miR-92a[303] - the first meta-analyses are now available in loCRC populations. However, none of these biomarkers has met the key requirements for adoption in the clinical setting at the present time, such as cohorts of more than 1000 individuals, inclusion of prospective studies and comparison with established screening or diagnostic methods.

Studies evaluating the role of miRNAs as biomarkers in eoCRC populations are still very scant, mostly performed on small eoCRC cohorts and without independent validation cohorts. Ak et al. [304] evaluated the expression profiles of 38 different miRNAs associated with CRC through RT-PCR in tumors and surgical margin tissue samples from a small Turkish cohort of 40 sporadic eoCRCs. The expression of *miRNA-106a* was found to be upregulated,

and *miRNA-143 and miRNA-125b* levels were found to be downregulated in tumor tissues compared with the normal tissues.
Liu et al. [305] analyzed mRNA and miRNA profiles of 3 cohort of sporadic eoCRC and sporadic loCRC. The expression of dystrophin (DMD) was found to be downregulated and that of miRNA-31-5p upregulated in eoCRC, compared with adjacent peritumoral tissue and loCRC. eoCRCs with low DMD expression had significantly poorer overall survival, cancer specific survival and recurrence free survival, possibly implying it as a prognostic biomarker.
In 2022 Nakamura et al. [306] were the first who systematically established a circulating 4-miRNA signature able to robustly identify patients with eoCRC. They firstly analyzed a large, publicly available, ncRNA expression profiling dataset (GSE115513, n. 42 FFPE from stage I/II eoCRC, N. 370 from loCRC, n. 62 normal mucosal samples from young adults <50y, n. 587 normal mucosal samples from adults ≥50y), identifying a first tissue-based signature of 7-miRNAs found upregulated in eoCRCs (AUC 0.82, 95% CI 0.73-0.91, $p < 0.01$, sensitivity 0.72, specificity 0.84 for detection of eoCRC). This first tissue-based 7-miRNA signature found in the discovery phase, was then tested in plasma samples: 4 miRNAs (*miR-193a-5p, miR-210, miR-513a-5p, and miR-628–3p*) out of 7 were found to be expressed in blood samples in the discovery phase. Subsequently, the performance of the 4-miRNAs panel was examined by qRT-PCR in blood samples from 2 large, independent clinical cohorts. The 4-miRNAs yielded an AUC of 0.92 (95% CI 0.85–0.96) for identification of eoCRC in blood samples of the training cohort (n. 72 eoCRC, n. 45 healthy donors <50y) and an AUC of 0.88 (95% CI 0.82–0.93) in the validation cohort (n. 77 eoCRC, 65 non diseased controls). Moreover, the decreased expression of miRNAs in post-surgery plasma samples confirmed their tumor specificity, possibly implying it as a diagnostic biomarker as a non-invasive assay for young population screening.

Our study represents a step forward in this direction. Indeed, we performed a systematic and comprehensive biomarker discovery that let us to identify a 9-miRNA tissue-based signature that is highly robust in the identification of eoCRCs at higher risk of recurrence who would benefit most from more aggressive post-treatment surveillance.
Through a five-layer approach we developed a simple, inexpensive, and clinically feasible test that may be seamlessly transitioned into clinical practice immediately after completion of this project. The first phase of the study (Discovery) consisted in the systematic interrogation and profiling of miRNA expression levels in 20 FFPE samples of stage II-III eoCRCs that did (n 10) or did not (n 10) develop recurrence in five years following curative-intent surgery. In phase two (Assay development), we performed several bioinformatic analyses and identified the 10 best candidate miRNAs that were differentially expressed in eoCRCs with and without recurrence, thus providing the highest discriminatory power between the two groups. The hsa-miR-365a-3p, hsa-miR-410-3p, hsa-miR-654-3p, hsa-miR-125b-5p, hsa-miR-125b-2-3p and hsa-miR-99a-5p resulted up-regulated in eoCRCs experiencing recurrence, while hsa-let-7g-5p, hsa-miR-142-3p, hsa-miR-15b-3p, and hsa-miR-30e-5p were found down-regulated in eoCRCs with recurrence.
In the third phase, the performance of this first 10-miRNAs panel was trained in the FFPE training cohort (88 FFPE from stage I-III eoCRC who received curative-intent surgery; 24

recurrent and 63 non-recurrent). Because there is no unique and universally accepted normalized miRNA, we employed several bioinformatic approaches to rigorously establish the ideal candidate based on intra- and inter-group expression stability. In the Assay Training phase, we optimized and trained an advanced machine learning algorithm (XGBoost) to predict the development of eoCRC recurrence based on RT-qPCR data and performed several interrogations to the model to understand its functioning. The resulting learning model was able to robustly and accurately predict the development of recurrence based on a hyper-selected panel of only 9 microRNAs: *hsa-let-7g-5p, hsa-miR-125b-2-3p, hsa-miR-125b-5p, hsa-miR-142-3p, hsa-miR-15b-3p, hsa-miR-30e-5p, hsa-miR-365a-3p, hsa-miR-410-3p, hsa-miR-654-3p*. This high accuracy in predicting recurrence in the training cohort is shown by the AUC value of 0.90 (95% CI 83-95%) with a Youden index of 64.9% (CI95%, 55%-82%), accuracy of 81.8% (77-93%), sensitivity 84.0% (65-96%), specificity 81.0% (72-98%). The feature providing the most contribution to the model was let-7g, followed consistently by iR-365, and miR-410. This 9-miRNA panel can discriminate effectively between recurrent and non-recurrent cases up to 20 years after surgical resection: patients predicted to be at a high risk of recurrence by the XGB model had a statistically significant cumulative hazard of disease recurrence than those classified as low-risk ($p<0.001$).

Finally, we performed an independent validation of our assay in a distinct and ethnically different validation cohort of 69 FFPE of stage I-III eoCRCs who received curative-intent surgery for eoCRC (9 recurrent and 60 non-recurrent). It should be highlighted that we decided to employ our test in particularly dire circumstances to truly establish its ability to capture the unique nature of eoCRC across diverse populations. First, we decided to test our model in an independent cohort. Second, such an independent cohort consisted of patients of a substantially different ethnic background, compared to the training cohort. It has been demonstrated that ethnicity plays a substantial role in determining one's miRNA expression profile. Because we intended to generate a model as widely applicable as possible, we tested it in a very genetically different population to assess whether the test would be replicable. Finally and most importantly, we tested our model in a population comprising very few cases developing a recurrence (9/60). The results here presented would have probably benefitted from a more balanced validation cohort, because each misclassification would have impacted less on the overall performance. However, it should be highlighted that patients with eoCRC are at an increased risk of recurrence compared to loCRC, but their overall recurrence risk remains overall low. Therefore, we applied it to an independent cohort of patients where the overall recurrence prevalence was low, and the ethnicity was substantially different from the training cohort. The XGBoost-based risk-assessment model, incorporating 9-miRNAs, exhibited good accuracy in predicting recurrence among stage I-III eoCRCs also in the validation cohort, achieving an AUC value of 0.77 (95% CI 67.0-87.0%), with a sensitivity of 100% (88-100%), specificity of 62.9% (55-82%), accuracy 66.7% (59.0-83.0%), and Youden index of 62.9% (54-73%). Despite substantial differences between the two cohorts, these results confirm that the findings from the training cohort are transferrable to an independent, differently composed, and ethnically distinct cohort of patients, a finding that speaks of the general reliability and robustness of

the approaches employed to limit overfitting. To truly gauge the effects that this assay would have in a real-world scenario, we performed several decision-curve analyses. The DCA curve revealed that a surveillance strategy based on our 9-miRNA panel achieved a higher net benefit in comparison with a surveillance strategy that would be aggressive for all patients, no patient, or based on clinical characteristics (age, biological sex (male, female), tumor localization (right, left), tumor grade (high, intermediate, low), vascular invasion (yes, no), and lymphatic invasion (yes, no)). Finally, we observed a net benefit of the surveillance based on our 9-miRNA signature compared to the clinical-based surveillance especially in stage I and II high risk eoCRC, representing the subgroups of patients that could benefit more from more aggressive post-treatment follow-up strategies.

CONCLUSIONS AND FUTURE PERSPECTIVES

6.1 Genetic risk assessment

In our cohort of eoCRC, 20% of patients carried a germline PV of genes known to be associated with CRC. In detail, 12.4% eoCRC carried a germline PV of mismatch repair genes responsible of Lynch syndrome, 2.9% of BRCA1-2 responsible of Hereditary breast and ovarian cancer syndrome, 3.8% of MUTYH of which 3 were heterozygous and 1 homozygous, with the latter responsible of MAP. Moreover 0.9% eoCRC had a PV of ATM and 0.9% of SDHAF2. Although 71.4% of eoCRCs did not have a family history of CRC, 19% of eoCRCs reported having a FDR with CRC and 12.4% had a SDR with CRC.
Therefore, as recently stated in the DIRECTt guidelines on eoCRCs, all patients with eoCRC should be offered multi-gene panel germline genetic testing before treatment to maximize clinical utility, when feasible. Germline genetic testing for eoCRC should include at a minimum: APC, BMPR1A, EPCAM, MLH1, MSH2, MSH6, MUTYH, POLD1, POLE, PMS2, PTEN, SMAD4, STK11, and TP53. Where available and not cost-prohibitive testing should also include the following genes that are reasonably prevalent in CRC and change clinical management: BRCA1, BRCA2, ATM, CHEK2, PALB2, and possibly, but less prevalent, BRIP1, BARD1, CDKN2A, CDH1, RAD51C, and RAD51D. Genetic testing should also include the following genes associated with CRC or polyposis: AXIN2, GREM1, MLH3, MSH3, MBD4, NTHL1, RNF43, and RPS20 [1]. Family history of CRC should inform risk assessment for both syndromic and non-syndromic CRC [1]. Therefore, a detailed family history of cancer should be always collected for all eoCRC through validated risk assessment tools to identify patients who would benefit more from germline genetic testing and from dedicated colonoscopy screening timelines.

6.2 The International Case-Control study

The ad hoc designed SQFFQ provides a reasonably repeatable measure of dietary intake and can be used to assess the dietary and drinking habits of volunteers in this age group. Most staple foods in the Italian diet, including pasta, fruit, vegetables, legumes, eggs, meat, coffee, and tea, are well estimated by the SQFFQ. However, challenges and difficulties persist, as well-described in the literature, particularly in estimating the consumption of foods consumed sporadically (such as snacks) or in small quantities such as spices.

Definitive conclusions will be drawn upon the completion of the ongoing external validation, involving 100 volunteers from the same age group. In this phase, the SQFFQ is going to be validated against the gold standard, represented by a 4 days-food diary.

Afterwards, the international, multicenter, retrospective case-control study will be performed with the subsequent aims:

- To evaluate the associations of dietary, lifestyle (smoking habit, alcohol intake, physical activity) and anthropometric factors with the onset of eoCRC by means of an ad hoc designed and shared online questionnaire, comparing eoCRCs with healthy age- and sex-matched controls (HCs) between countries with different eoCRC incidence.
- To assess whether associations will differ among specific population subgroups (e.g., sex, ethnicity, country, smoking habits, BMI, physical activity, family history of CRC).

The case-control study will recruit:

- At least 758 patients with a recent diagnosis of eoCRC (diagnosis made within 1 years prior to enrollment). Retrospective data (dietary and lifestyle factors) related to the year prior to eoCRC diagnosis will be collected through an online platform.
- At least 1516 healthy volunteers (controls), matched by age (matching range ± 5 years) and sex with eoCRC. Healthy volunteers will be mainly enrolled among workers within the participating hospital center, followed regularly by preventive medicine. This enrollment will be carried out by e-mail invitation disseminated through the hospital's official mailing list. Retrospective data (dietary and lifestyle factors) related to the year prior to recruitment will be collected through an online platform.

All eoCRC with a diagnosis of CRC between 18 and 49 years confirmed by histology (biopsy or surgical specimen in case of surgery), of all sexes, will be included, as well as Healthy controls with negative past and present history of cancer and a negative fecal occult blood test (FOBT), or negative colonoscopy. Exclusion criteria will be represented by: CRC diagnosed at ≥ 50 years (loCRC), diseases known to predispose to CRC (personal past or recent history of inflammatory bowel disease, past history of pelvic irradiation), inability to give written informed consents and to fill in the online questionnaire.

The international study is actually involving 7 hospital centers in Italy, 7 in Finland, 2 in USA and is expected to recruit other hospital centers in Italy, USA, Germany, Spain, Norway, Netherlands, UK.

Statistical design:

Based on literature review and clinical considerations, given the lack of consensus in measuring risk factors as physical activity, alcohol, sugar drinks, meat consumption, we focused on BMI to compute sample size. BMI is routinely collected in clinical practice; it is easy to compute and well established cut-off points are provided in widely accepted guidelines. Moreover, BMI is strongly related to the other eoCRC risk factors. Due to the nature of Demetra study, we computed the sample size needed to carry out a matched case-control study with the aim of detecting a relationship between the development of eoCRC and, in particular, BMI>25 as risk factor (exposure variable). We used PASS software (v21.0.5) and the procedure "Tests for the Odds Ratio in a Matched Case-Control Design with a Binary X". Assuming a probability of exposure retrieved from the literature equal to 34.4% [26, 30], in a matched case-control study, a sample of 758 matched sets is required. Each matched set consists of 1 case and 2 controls. This sample allows to achieve 80%

power to detect an odds ratio of 1.3 calculated using conditional logistic regression with a 5% significance level. Sample allocation will account for patient recruitment capacity of each country and for country-specific prevalence. Given the 5-year eoCRC prevalence in the countries involved (obtained from the Global Cancer Observatory), the total sample size will be divided among the countries correspondingly, as shown in the table below:

	ICD	Cancer	5-year prevalence (proportion per 100000)	#cases	#controls
ITA	**C18-21**	Colorectum	18.3	102	204
GERMANY	**C18-21**	Colorectum	19.4	108	216
FINLAND	**C18-21**	Colorectum	13.9	77	155
SPAIN	**C18-21**	Colorectum	20.6	115	229
USA	**C18-21**	Colorectum	22.9	127	255
UK	**C18-21**	Colorectum	18.7	104	208
NORWAY	**C18-21**	Colorectum	22.4	125	249

Based on data distribution, parametric and nonparametric tests for two independent samples will be used for comparing cases and controls with respect to quantitative and qualitative collected variables. To evaluate the associations of dietary, lifestyle (smoking habit, alcohol intake, physical activity) and anthropometric factors with the onset of eoCRC univariate and multiple logistic regression models will be applied. Results will be expressed as odds ratio (OR) along with 95% confidence interval. Along with a standard logistic regression approach, data-mining and data-driven approaches, such as Classification and Regression Trees analysis and Bayesian Networks, will be implemented to uncover risk factors associated with the outcome while accounting for possible confounding bias typically arising in multivariate observational settings.

6.3 miRNA signature and personalized medicine

Using a systematic discovery approach, we have identified and developed a novel tissue-based miRNA signature able to distinguish patients with eoCRC stage I-III at high risk of recurrence. The 9-miRNA panel was successfully validated on tissue specimens of 2 independent cohorts.

This study establishes the fundamentals of personalized medicine for eoCRC patients with stages I-III at high risk of recurrence. In particular, as described by the decision curve analysis, we observed a net benefit of the surveillance based on our microRNA panel compared to the surveillance based on clinical factors especially in stage I and II high risk eoCRC, representing the subgroups of patients that could benefit more from more aggressive post-treatment follow-up strategies.

Prospective trials on large sample-size are necessary to confirm these results in real life.

Moreover, another objective of our project is to develop a minimally invasive tool for easily identifying these patients at high risk of recurrence in clinical settings. In the near future, we plan to validate the performance of miRNA signature in blood sample from eoCRC patients. This validation is aimed at creating an easily available blood-based assay, which

will be further validated in prospective studies to evaluate its potential as a simple prognostic test for young patients with eoCRC.

REFERENCES

1. Cavestro GM, Mannucci A, Balaguer F, Hampel H, Kupfer SS, Repici A, et al. Delphi Initiative for Early-Onset Colorectal Cancer (DIRECt) International Management Guidelines. Clin Gastroenterol Hepatol. 2023;21: 581–603.e33.
2. Ahnen DJ, Wade SW, Jones WF, Sifri R, Mendoza Silveiras J, Greenamyer J, et al. The increasing incidence of young-onset colorectal cancer: a call to action. Mayo Clin Proc. 2014;89: 216–224.
3. Chang DT, Pai RK, Rybicki LA, Dimaio MA, Limaye M, Jayachandran P, et al. Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. Mod Pathol. 2012;25: 1128–1139.
4. Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. Lancet Gastroenterol Hepatol. 2019;4: 511–518.
5. Russo AG, Andreano A, Sartore-Bianchi A, Mauri G, Decarli A, Siena S. Increased incidence of colon cancer among individuals younger than 50 years: A 17 years analysis from the cancer registry of the municipality of Milan, Italy. Cancer Epidemiol. 2019;60: 134–140.
6. Siegel RL, Torre LA, Soerjomataram I, Hayes RB, Bray F, Weber TK, et al. Global patterns and trends in colorectal cancer incidence in young adults. Gut. 2019;68: 2179–2185.
7. Vuik FE, Nieuwenburg SA, Bardou M, Lansdorp-Vogelaar I, Dinis-Ribeiro M, Bento MJ, et al. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. Gut. 2019;68: 1820–1826.
8. Zorzi M, Dal Maso L, Francisci S, Buzzoni C, Rugge M, Guzzinati S, et al. Trends of colorectal cancer incidence and mortality rates from 2003 to 2014 in Italy. Tumori. 2019;105: 417–426.
9. Zorzi M, Cavestro GM, Guzzinati S, Dal Maso L, Rugge M, AIRTUM Working Group. Decline in the incidence of colorectal cancer and the associated mortality in young Italian adults. Gut. 2020. pp. 1902–1903.
10. Gupta S, May FP, Kupfer SS, Murphy CC. Birth cohort colorectal cancer (CRC): implications for research and practice. Clin Gastroenterol Hepatol. 2023. doi:10.1016/j.cgh.2023.11.040
11. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. J Natl Cancer Inst. 2017;109. doi:10.1093/jnci/djw322
12. Murphy CC, Singal AG, Baron JA, Sandler RS. Decrease in Incidence of Young-Onset Colorectal Cancer Before Recent Increase. Gastroenterology. 2018;155: 1716–1719.e4.
13. Murphy CC, Lee JK, Liang PS, May FP, Zaki TA. Declines in Colorectal Cancer Incidence and Mortality Rates Slow Among Older Adults. Clin Gastroenterol Hepatol. 2023. doi:10.1016/j.cgh.2023.05.033

14. Zaki TA, Singal AG, May FP, Murphy CC. Increasing Incidence Rates of Colorectal Cancer at Ages 50-54 Years. Gastroenterology. 2022;162: 964–965.e3.
15. Cao Y, Nguyen LH, Tica S, Otegbeye E, Zong X, Roelstraete B, et al. Evaluation of Birth by Cesarean Delivery and Development of Early-Onset Colorectal Cancer. JAMA Netw Open. 2023;6: e2310316.
16. Rosenberg PS, Check DP, Anderson WF. A web tool for age-period-cohort analysis of cancer incidence and mortality rates. Cancer Epidemiol Biomarkers Prev. 2014;23: 2296–2302.
17. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018;68: 250–281.
18. Anderson JC, Samadder JN. To Screen or Not to Screen Adults 45-49 Years of Age: That is the Question. Am J Gastroenterol. 2018;113: 1750–1753.
19. Bailey CE, Hu C-Y, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. JAMA Surg. 2015;150: 17–22.
20. Mauri G, Sartore-Bianchi A, Russo A-G, Marsoni S, Bardelli A, Siena S. Early-onset colorectal cancer in young individuals. Mol Oncol. 2019;13: 109–131.
21. Bhandari A, Woodhouse M, Gupta S. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers. J Investig Med. 2017;65: 311–315.
22. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70: 145–164.
23. Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. Defining "chromosomal instability." Trends Genet. 2008;24: 64–69.
24. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61: 759–767.
25. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology. 2010;138: 2059–2072.
26. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010;138: 2073–2087.e3.
27. Pussila M, Törönen P, Einarsdottir E, Katayama S, Krjutškov K, Holm L, et al. Mlh1 deficiency in normal mouse colon mucosa associates with chromosomally unstable colon cancer. Carcinogenesis. 2018;39: 788–797.
28. Haraldsdottir S, Hampel H, Tomsic J, Frankel WL, Pearlman R, de la Chapelle A, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology. 2014;147: 1308–1316.e1.
29. Weisenberger DJ, Levine AJ, Long TI, Buchanan DD, Walters R, Clendenning M, et al. Association of the colorectal CpG island methylator phenotype with molecular features, risk factors, and family history. Cancer Epidemiol Biomarkers Prev. 2015;24: 512–519.
30. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ. CpG island methylator phenotype in colorectal cancer. Proceedings of the National Academy of Sciences. 1999;96: 8681–8686.

31. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. J Pathol. 2013;231: 63–76.
32. De Sousa E Melo F, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LPMH, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med. 2013;19: 614–618.
33. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10: e1001453.
34. Roepman P, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. Int J Cancer. 2014;134: 552–562.
35. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013;19: 619–625.
36. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol. 2011;29: 17–24.
37. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21: 1350–1356.
38. Thanki K, Nicholls ME, Gajjar A, Senagore AJ, Qiu S, Szabo C, et al. Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. Int Biol Biomed J. 2017;3: 105–111.
39. Stoffel EM, Murphy CC. Epidemiology and Mechanisms of the Increasing Incidence of Colon and Rectal Cancers in Young Adults. Gastroenterology. 2020;158: 341–353.
40. Yeo H, Betel D, Abelson JS, Zheng XE, Yantiss R, Shah MA. Early-onset Colorectal Cancer is Distinct From Traditional Colorectal Cancer. Clin Colorectal Cancer. 2017;16: 293–299.e6.
41. Di Leo M, Zuppardo RA, Puzzono M, Ditonno I, Mannucci A, Antoci G, et al. Risk factors and clinical characteristics of early-onset colorectal cancer vs. late-onset colorectal cancer: a case-case study. Eur J Gastroenterol Hepatol. 2021;33: 1153–1160.
42. Cercek A, Chatila WK, Yaeger R, Walch H, Fernandes GDS, Krishnan A, et al. A Comprehensive Comparison of Early-Onset and Average-Onset Colorectal Cancers. J Natl Cancer Inst. 2021;113: 1683–1692.
43. Meyer JE, Narang T, Schnoll-Sussman FH, Pochapin MB, Christos PJ, Sherr DL. Increasing incidence of rectal cancer in patients aged younger than 40 years. Cancer. 2010;116: 4354–4359.
44. Yeo SA, Chew MH, Koh PK, Tang CL. Young colorectal carcinoma patients do not have a poorer prognosis: a comparative review of 2,426 cases. Tech Coloproctol. 2013;17: 653–661.
45. Jass JR. HNPCC and Sporadic MSI-H Colorectal Cancer: A Review of the Morphological Similarities and Differences. Fam Cancer. 2004;3: 93–100.

46. Mork ME, You YN, Ying J, Bannon SA, Lynch PM, Rodriguez-Bigas MA, et al. High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. J Clin Oncol. 2015;33: 3544–3549.
47. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. JAMA Oncol. 2017;3: 464–471.
48. Stoffel EM, Koeppe E, Everett J, Ulintz P, Kiel M, Osborne J, et al. Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology. 2018;154: 897–905.e1.
49. Poynter JN, Siegmund KD, Weisenberger DJ, Long TI, Thibodeau SN, Lindor N, et al. Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. Cancer Epidemiol Biomarkers Prev. 2008;17: 3208–3215.
50. Chen FW, Sundaram V, Chew TA, Ladabaum U. Advanced-Stage Colorectal Cancer in Persons Younger Than 50 Years Not Associated With Longer Duration of Symptoms or Time to Diagnosis. Clin Gastroenterol Hepatol. 2017;15: 728–737.e3.
51. Kneuertz PJ, Chang GJ, Hu C-Y, Rodriguez-Bigas MA, Eng C, Vilar E, et al. Overtreatment of Young Adults With Colon Cancer: More Intense Treatments With Unmatched Survival Gains. JAMA Surg. 2015;150: 402–409.
52. O'Connell JB, Maggard MA, Liu JH, Etzioni DA, Livingston EH, Ko CY. Do young colon cancer patients have worse outcomes? World J Surg. 2004;28: 558–562.
53. Saraste D, Järås J, Martling A. Population-based analysis of outcomes with early-age colorectal cancer. Br J Surg. 2020;107: 301–309.
54. Ben-Ishay O, Brauner E, Peled Z, Othman A, Person B, Kluger Y. Diagnosis of colon cancer differs in younger versus older patients despite similar complaints. Isr Med Assoc J. 2013;15: 284–287.
55. Dozois EJ, Boardman LA, Suwanthanma W, Limburg PJ, Cima RR, Bakken JL, et al. Young-onset colorectal cancer in patients with no known genetic predisposition: can we increase early recognition and improve outcome? Medicine . 2008;87: 259–263.
56. Cavestro GM, Mannucci A, Zuppardo RA, Di Leo M, Stoffel E, Tonon G. Early onset sporadic colorectal cancer: Worrisome trends and oncogenic features. Dig Liver Dis. 2018;50: 521–532.
57. Ko CW, Siddique SM, Patel A, Harris A, Sultan S, Altayar O, et al. AGA Clinical Practice Guidelines on the Gastrointestinal Evaluation of Iron Deficiency Anemia. Gastroenterology. 2020;159: 1085–1094.
58. Vajravelu RK, Mehta SJ, Lewis JD, Early-age Onset Colorectal Cancer Testing, Epidemiology, Diagnosis, and Symptoms Study Group (EOCRC TrEnDS). Understanding Characteristics of Who Undergoes Testing Is Crucial for the Development of Diagnostic Strategies to Identify Individuals at Risk for Early-age Onset Colorectal Cancer. Gastroenterology. 2021;160: 993–998.
59. Low EE, Demb J, Liu L, Earles A, Bustamante R, Williams CD, et al. Risk Factors for Early-Onset Colorectal Cancer. Gastroenterology. 2020;159: 492–501.e7.

60. Syed AR, Thakkar P, Horne ZD, Abdul-Baki H, Kochhar G, Farah K, et al. Old vs new: Risk factors predicting early onset colorectal cancer. World J Gastrointest Oncol. 2019;11: 1011–1020.
61. Frostberg E, Rahr HB. Clinical characteristics and a rising incidence of early-onset colorectal cancer in a nationwide cohort of 521 patients aged 18-40 years. Cancer Epidemiol. 2020;66: 101704.
62. Krigel A, Zhou M, Terry MB, Kastrinos F, Lebwohl B. Symptoms and demographic factors associated with early-onset colorectal neoplasia among individuals undergoing diagnostic colonoscopy. Eur J Gastroenterol Hepatol. 2020;32: 821–826.
63. Demb J, Liu L, Murphy CC, Doubeni CA, Martínez ME, Gupta S. Young-onset colorectal cancer risk among individuals with iron-deficiency anaemia and haematochezia. Gut. 2020;70: 1529–1537.
64. Rajagopalan A, Antoniou E, Morkos M, Rajagopalan E, Arachchi A, Chouhan H, et al. Is colorectal cancer associated with altered bowel habits in young patients? ANZ J Surg. 2021;91: 943–946.
65. Zhu C, Ji M, Dai W, Ye C, Hu Z, Shi J, et al. Clinicopathological characteristics of chinese colorectal cancer patients under 30 years of age: implication in diagnosis and therapy. Curr Cancer Drug Targets. 2015;15: 27–34.
66. D'Souza N, Georgiou Delisle T, Chen M, Benton S, Abulafi M, NICE FIT Steering Group. Faecal immunochemical test is superior to symptoms in predicting pathology in patients with suspected colorectal cancer symptoms referred on a 2WW pathway: a diagnostic accuracy study. Gut. 2021;70: 1130–1138.
67. D'Souza N, Monahan K, Benton SC, Wilde L, Abulafi M, NICE FIT Steering Group. Finding the needle in the haystack: the diagnostic accuracy of the faecal immunochemical test for colorectal cancer in younger symptomatic patients. Colorectal Dis. 2021;23: 2539–2549.
68. Jung YS, Park CH, Kim NH, Park JH, Park DI, Sohn CI. Colorectal cancer screening with the fecal immunochemical test in persons aged 30 to 49 years: focusing on the age for commencing screening. Gastrointest Endosc. 2017;86: 892–899.
69. Corley DA, Jensen CD, Quinn VP, Doubeni CA, Zauber AG, Lee JK, et al. Association Between Time to Colonoscopy After a Positive Fecal Test Result and Risk of Colorectal Cancer and Cancer Stage at Diagnosis. JAMA. 2017;317: 1631–1641.
70. San Miguel Y, Demb J, Martinez ME, Gupta S, May FP. Time to Colonoscopy After Abnormal Stool-Based Screening and Risk for Colorectal Cancer Incidence and Mortality. Gastroenterology. 2021;160: 1997–2005.e3.
71. Benson AB 3rd, Venook AP, Cederquist L, Chan E, Chen Y-J, Cooper HS, et al. Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2017;15: 370–398.
72. Kim TJ, Kim ER, Hong SN, Chang DK, Kim Y-H. Long-Term Outcome and Prognostic Factors of Sporadic Colorectal Cancer in Young Patients: A Large Institutional-Based Retrospective Study. Medicine . 2016;95: e3641.

73. Parry S, Win AK, Parry B, Macrae FA, Gurrin LC, Church JM, et al. Metachronous colorectal cancer risk for mismatch repair gene mutation carriers: the advantage of more extensive colon surgery. Gut. 2011;60: 950–957.
74. Natarajan N, Watson P, Silva-Lopez E, Lynch HT. Comparison of extended colectomy and limited resection in patients with Lynch syndrome. Dis Colon Rectum. 2010;53: 77–82.
75. Arhin ND, Shen C, Bailey CE, Matsuoka LK, Hawkins AT, Holowatyj AN, et al. Surgical resection and survival outcomes in metastatic young adult colorectal cancer patients. Cancer Med. 2021;10: 4269–4281.
76. André T, Iveson T, Labianca R, Meyerhardt JA, Souglakos I, Yoshino T, et al. The IDEA (International Duration Evaluation of Adjuvant Chemotherapy) Collaboration: Prospective Combined Analysis of Phase III Trials Investigating Duration of Adjuvant Therapy with the FOLFOX (FOLFOX4 or Modified FOLFOX6) or XELOX (3 versus 6 months) Regimen for Patients with Stage III Colon Cancer: Trial Design and Current Status. Curr Colorectal Cancer Rep. 2013;9: 261–269.
77. Dasari A, Morris VK, Allegra CJ, Atreya C, Benson AB 3rd, Boland P, et al. ctDNA applications and integration in colorectal cancer: an NCI Colon and Rectal-Anal Task Forces whitepaper. Nat Rev Clin Oncol. 2020;17: 757–770.
78. André T, Boni C, Navarro M, Tabernero J, Hickish T, Topham C, et al. Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. J Clin Oncol. 2009;27: 3109–3116.
79. Benson AB 3rd, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. J Clin Oncol. 2004;22: 3408–3419.
80. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J Clin Oncol. 2010;28: 3219–3226.
81. Sauer R, Becker H, Hohenberger W, Rödel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004;351: 1731–1740.
82. Fernández-Martos C, Pericay C, Aparicio J, Salud A, Safont M, Massuti B, et al. Phase II, randomized study of concomitant chemoradiotherapy followed by surgery and adjuvant capecitabine plus oxaliplatin (CAPOX) compared with induction CAPOX followed by concomitant chemoradiotherapy and surgery in magnetic resonance imaging-defined, locally advanced rectal cancer: Grupo cancer de recto 3 study. J Clin Oncol. 2010;28: 859–865.
83. Jayne D, Pigazzi A, Marshall H, Croft J, Corrigan N, Copeland J, et al. Effect of Robotic-Assisted vs Conventional Laparoscopic Surgery on Risk of Conversion to Open Laparotomy Among Patients Undergoing Resection for Rectal Cancer: The ROLARR Randomized Clinical Trial. JAMA. 2017;318: 1569–1580.
84. Fleshman J, Branda M, Sargent DJ, Boller AM, George V, Abbas M, et al. Effect of Laparoscopic-Assisted Resection vs Open Resection of Stage II or III Rectal Cancer on Pathologic Outcomes: The ACOSOG Z6051 Randomized Clinical Trial. JAMA. 2015;314: 1346–1355.

85. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U Jr, Silva e Sousa AH Jr, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. Ann Surg. 2004;240: 711–7; discussion 717–8.
86. O'Sullivan DE, Sutherland RL, Town S, Chow K, Fan J, Forbes N, et al. Risk Factors for Early-Onset Colorectal Cancer: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2022;20: 1229–1240.e5.
87. Johns LE, Kee F, Collins BJ, Patterson CC, Houlston RS. Colorectal cancer mortality in first-degree relatives of early-onset colorectal cancer cases. Dis Colon Rectum. 2002;45: 681–686.
88. Pearlman R, de la Chapelle A, Hampel H. Mutation Frequencies in Patients With Early-Onset Colorectal Cancer-Reply. JAMA oncology. 2017. p. 1587.
89. Stanich PP, Pelstring KR, Hampel H, Pearlman R. A High Percentage of Early-age Onset Colorectal Cancer Is Potentially Preventable. Gastroenterology. 2021;160: 1850–1852.
90. Strum WB, Boland CR. Clinical and Genetic Characteristics of Colorectal Cancer in Persons under 50 Years of Age: A Review. Dig Dis Sci. 2019;64: 3059–3065.
91. Shaukat A, Kahi CJ, Burke CA, Rabeneck L, Sauer BG, Rex DK. ACG Clinical Guidelines: Colorectal Cancer Screening 2021. Am J Gastroenterol. 2021;116: 458–479.
92. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients From the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology. 2017;153: 307–323.
93. van Leerdam ME, Roos VH, van Hooft JE, Balaguer F, Dekker E, Kaminski MF, et al. Endoscopic management of Lynch syndrome and of familial risk of colorectal cancer: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy. 2019;51: 1082–1093.
94. Wong MCS, Chan CH, Lin J, Huang JLW, Huang J, Fang Y, et al. Lower Relative Contribution of Positive Family History to Colorectal Cancer Risk with Increasing Age: A Systematic Review and Meta-Analysis of 9.28 Million Individuals. Am J Gastroenterol. 2018;113: 1819–1827.
95. Hemminki K, Li X. Familial colorectal adenocarcinoma from the Swedish Family-Cancer Database. Int J Cancer. 2001;94: 743–748.
96. Kastrinos F, Allen JI, Stockwell DH, Stoffel EM, Cook EF, Mutinga ML, et al. Development and validation of a colon cancer risk assessment tool for patients undergoing colonoscopy. Am J Gastroenterol. 2009;104: 1508–1518.
97. Kastrinos F, Uno H, Ukaegbu C, Alvero C, McFarland A, Yurgelun MB, et al. Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. J Clin Oncol. 2017;35: 2165–2172.
98. Zhunussova G, Afonin G, Abdikerim S, Jumanov A, Perfilyeva A, Kaidarova D, et al. Mutation Spectrum of Cancer-Associated Genes in Patients With Early Onset of Colorectal Cancer. Front Oncol. 2019;9: 673.
99. Sinicrope FA. Increasing Incidence of Early-Onset Colorectal Cancer. N Engl J Med. 2022;386: 1547–1558.

100. LaDuca H, Polley EC, Yussuf A, Hoang L, Gutierrez S, Hart SN, et al. A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. Genet Med. 2020;22: 407–415.
101. Mork ME, Rodriguez A, Bannon SA, Lynch PM, Rodriguez-Bigas MA, Thirumurthi S, et al. Outcomes of disease-specific next-generation sequencing gene panel testing in adolescents and young adults with colorectal cancer. Cancer Genet. 2019;235-236: 77–83.
102. André T, Shiu K-K, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. N Engl J Med. 2020;383: 2207–2218.
103. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology. 2010;138: 2044–2058.
104. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med. 2003;348: 919–932.
105. Croitoru ME, Cleary SP, Di Nicola N, Manno M, Selander T, Aronson M, et al. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. J Natl Cancer Inst. 2004;96: 1631–1634.
106. Stjepanovic N, Moreira L, Carneiro F, Balaguer F, Cervantes A, Balmaña J, et al. Hereditary gastrointestinal cancers: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2019;30: 1558–1571.
107. Ngeow J, Heald B, Rybicki LA, Orloff MS, Chen JL, Liu X, et al. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. Gastroenterology. 2013;144: 1402–9, 1409.e1–5.
108. Brosens LAA, van Hattem A, Hylind LM, Iacobuzio-Donahue C, Romans KE, Axilbund J, et al. Risk of colorectal cancer in juvenile polyposis. Gut. 2007;56: 965–967.
109. Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. J Clin Oncol. 2017;35: 1086–1095.
110. Chang P-Y, Chang S-C, Wang M-C, Chen J-S, Tsai W-S, You J-F, et al. Pathogenic Germline Mutations of DNA Repair Pathway Components in Early-Onset Sporadic Colorectal Polyp and Cancer Patients. Cancers . 2020;12. doi:10.3390/cancers12123560
111. Jansen AML, Ghosh P, Dakal TC, Slavin TP, Boland CR, Goel A. Novel candidates in early-onset familial colorectal cancer. Fam Cancer. 2020;19: 1–10.
112. AlDubayan SH, Giannakis M, Moore ND, Han GC, Reardon B, Hamada T, et al. Inherited DNA-Repair Defects in Colorectal Cancer. Am J Hum Genet. 2018;102: 401–414.
113. Monahan KJ, Bradshaw N, Dolwani S, Desouza B, Dunlop MG, East JE, et al. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). Gut. 2020;69: 411–444.
114. Seppälä TT, Latchford A, Negoi I, Sampaio Soares A, Jimenez-Rodriguez R, Sánchez-Guillén L, et al. European guidelines from the EHTG and ESCP for Lynch syndrome: an updated third edition of the Mallorca guidelines based on gene and gender. Br J Surg. 2021;108: 484–498.

115. Coletta AM, Peterson SK, Gatus LA, Krause KJ, Schembre SM, Gilchrist SC, et al. Energy balance related lifestyle factors and risk of endometrial and colorectal cancer among individuals with lynch syndrome: a systematic review. Fam Cancer. 2019;18: 399–420.
116. Miguchi M, Hinoi T, Tanakaya K, Yamaguchi T, Furukawa Y, Yoshida T, et al. Alcohol consumption and early-onset risk of colorectal cancer in Japanese patients with Lynch syndrome: a cross-sectional study conducted by the Japanese Society for Cancer of the Colon and Rectum. Surg Today. 2018;48: 810–814.
117. van Duijnhoven FJB, Botma A, Winkels R, Nagengast FM, Vasen HFA, Kampman E. Do lifestyle factors influence colorectal cancer risk in Lynch syndrome? Fam Cancer. 2013;12: 285–293.
118. Botma A, Vasen HFA, van Duijnhoven FJB, Kleibeuker JH, Nagengast FM, Kampman E. Dietary patterns and colorectal adenomas in Lynch syndrome: the GEOLynch cohort study. Cancer. 2013;119: 512–521.
119. Brouwer JG, Makama M, van Woudenbergh GJ, Vasen HF, Nagengast FM, Kleibeuker JH, et al. Inflammatory potential of the diet and colorectal tumor risk in persons with Lynch syndrome. Am J Clin Nutr. 2017;106: 1287–1294.
120. Eijkelboom AH, Brouwer JGM, Vasen HFA, Bisseling TM, Koornstra JJ, Kampman E, et al. Diet quality and colorectal tumor risk in persons with Lynch syndrome. Cancer Epidemiol. 2020;69: 101809.
121. Dominguez-Valentin M, Sampson JR, Seppälä TT, Ten Broeke SW, Plazzer J-P, Nakken S, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. Genet Med. 2020;22: 15–25.
122. Møller P, Seppälä T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut. 2017;66: 464–472.
123. Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, 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: 1306–1316.
124. Aune D, Lau R, Chan DSM, Vieira R, Greenwood DC, Kampman E, et al. Nonlinear reduction in risk for colorectal cancer by fruit and vegetable intake based on meta-analysis of prospective studies. Gastroenterology. 2011;141: 106–118.
125. Feng Y-L, Shu L, Zheng P-F, Zhang X-Y, Si C-J, Yu X-L, et al. Dietary patterns and colorectal cancer risk: a meta-analysis. Eur J Cancer Prev. 2017;26: 201–211.
126. Barrubés L, Babio N, Becerra-Tomás N, Rosique-Esteban N, Salas-Salvadó J. Association Between Dairy Product Consumption and Colorectal Cancer Risk in Adults: A Systematic Review and Meta-Analysis of Epidemiologic Studies. Adv Nutr. 2019;10: S190–S211.
127. Petimar J, Smith-Warner SA, Fung TT, Rosner B, Chan AT, Hu FB, et al. Recommendation-based dietary indexes and risk of colorectal cancer in the Nurses' Health Study and Health Professionals Follow-up Study. Am J Clin Nutr. 2018;108: 1092–1103.

128. Bradbury KE, Murphy N, Key TJ. Diet and colorectal cancer in UK Biobank: a prospective study. Int J Epidemiol. 2020;49: 246–258.
129. Chapelle N, Martel M, Toes-Zoutendijk E, Barkun AN, Bardou M. Recent advances in clinical practice: colorectal cancer chemoprevention in the average-risk population. Gut. 2020;69: 2244–2255.
130. Wang L, Lo C-H, He X, Hang D, Wang M, Wu K, et al. Risk Factor Profiles Differ for Cancers of Different Regions of the Colorectum. Gastroenterology. 2020;159: 241–256.e13.
131. Kim JY, Jung YS, Park JH, Kim HJ, Cho YK, Sohn CI, et al. Different risk factors for advanced colorectal neoplasm in young adults. World J Gastroenterol. 2016;22: 3611–3620.
132. Kim NH, Jung YS, Yang H-J, Park S-K, Park JH, Park DI, et al. Prevalence of and Risk Factors for Colorectal Neoplasia in Asymptomatic Young Adults (20-39 Years Old). Clin Gastroenterol Hepatol. 2019;17: 115–122.
133. Nguyen LH, Liu P-H, Zheng X, Keum N, Zong X, Li X, et al. Sedentary Behaviors, TV Viewing Time, and Risk of Young-Onset Colorectal Cancer. JNCI Cancer Spectr. 2018;2: ky073.
134. Rosato V, Bosetti C, Levi F, Polesel J, Zucchetto A, Negri E, et al. Risk factors for young-onset colorectal cancer. Cancer Causes Control. 2013;24: 335–341.
135. Gausman V, Dornblaser D, Anand S, Hayes RB, O'Connell K, Du M, et al. Risk Factors Associated With Early-Onset Colorectal Cancer. Clin Gastroenterol Hepatol. 2020;18: 2752–2759.e2.
136. Liu P-H, Wu K, Ng K, Zauber AG, Nguyen LH, Song M, et al. Association of Obesity With Risk of Early-Onset Colorectal Cancer Among Women. JAMA Oncol. 2019;5: 37–44.
137. Zheng X, Hur J, Nguyen LH, Liu J, Song M, Wu K, et al. Comprehensive Assessment of Diet Quality and Risk of Precursors of Early-Onset Colorectal Cancer. J Natl Cancer Inst. 2021;113: 543–552.
138. Imperiale TF, Kahi CJ, Stuart JS, Qi R, Born LJ, Glowinski EA, et al. Risk factors for advanced sporadic colorectal neoplasia in persons younger than age 50. Cancer Detect Prev. 2008;32: 33–38.
139. Decarli A, Franceschi S, Ferraroni M, Gnagnarella P, Parpinel MT, La Vecchia C, et al. Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy. Results for specific nutrients. Ann Epidemiol. 1996;6: 110–118.
140. Archambault AN, Lin Y, Jeon J, Harrison TA, Bishop DT, Brenner H, et al. Nongenetic Determinants of Risk for Early-Onset Colorectal Cancer. JNCI Cancer Spectr. 2021;5. doi:10.1093/jncics/pkab029
141. Khan NA, Hussain M, ur Rahman A, Farooqui WA, Rasheed A, Memon AS. Dietary Practices, Addictive Behavior and Bowel Habits and Risk of Early Onset Colorectal Cancer: a Case Control Study. Asian Pac J Cancer Prev. 2015;16: 7967–7973.
142. Chang VC, Cotterchio M, De P, Tinmouth J. Risk factors for early-onset colorectal cancer: a population-based case-control study in Ontario, Canada. Cancer Causes Control. 2021;32: 1063–1083.
143. Tabung FK, Brown LS, Fung TT. Dietary Patterns and Colorectal Cancer Risk: A Review of 17 Years of Evidence (2000-2016). Curr Colorectal Cancer Rep. 2017;13: 440–454.

144. Song M, Chan AT. Environmental Factors, Gut Microbiota, and Colorectal Cancer Prevention. Clin Gastroenterol Hepatol. 2019;17: 275–289.
145. Garcia-Larsen V, Morton V, Norat T, Moreira A, Potts JF, Reeves T, et al. Dietary patterns derived from principal component analysis (PCA) and risk of colorectal cancer: a systematic review and meta-analysis. Eur J Clin Nutr. 2019;73: 366–386.
146. Nimptsch K, Malik VS, Fung TT, Pischon T, Hu FB, Willett WC, et al. Dietary patterns during high school and risk of colorectal adenoma in a cohort of middle-aged women. Int J Cancer. 2014;134: 2458–2467.
147. Castelló A, Amiano P, Fernández de Larrea N, Martín V, Alonso MH, Castaño-Vinyals G, et al. Low adherence to the western and high adherence to the mediterranean dietary patterns could prevent colorectal cancer. Eur J Nutr. 2019;58: 1495–1505.
148. Jones P, Cade JE, Evans CEL, Hancock N, Greenwood DC. The Mediterranean diet and risk of colorectal cancer in the UK Women's Cohort Study. Int J Epidemiol. 2017;46: 1786–1796.
149. Agnoli C, Grioni S, Sieri S, Palli D, Masala G, Sacerdote C, et al. Italian Mediterranean Index and risk of colorectal cancer in the Italian section of the EPIC cohort. Int J Cancer. 2013;132: 1404–1411.
150. Hur J, Otegbeye E, Joh H-K, Nimptsch K, Ng K, Ogino S, et al. Sugar-sweetened beverage intake in adulthood and adolescence and risk of early-onset colorectal cancer among women. Gut. 2021;70: 2330–2336.
151. Ralston RA, Truby H, Palermo CE, Walker KZ. Colorectal cancer and nonfermented milk, solid cheese, and fermented milk consumption: a systematic review and meta-analysis of prospective studies. Crit Rev Food Sci Nutr. 2014;54: 1167–1179.
152. Aune D, Lau R, Chan DSM, Vieira R, Greenwood DC, Kampman E, et al. Dairy products and colorectal cancer risk: a systematic review and meta-analysis of cohort studies. Ann Oncol. 2012;23: 37–45.
153. Keum N, Aune D, Greenwood DC, Ju W, Giovannucci EL. Calcium intake and colorectal cancer risk: dose-response meta-analysis of prospective observational studies. Int J Cancer. 2014;135: 1940–1948.
154. Lamprecht SA, Lipkin M. Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis. Ann N Y Acad Sci. 2001;952: 73–87.
155. Puzzono M, Mannucci A, Grannò S, Zuppardo RA, Galli A, Danese S, et al. The Role of Diet and Lifestyle in Early-Onset Colorectal Cancer: A Systematic Review. Cancers . 2021;13. doi:10.3390/cancers13235933
156. Bagnardi V, Blangiardo M, La Vecchia C, Corrao G. A meta-analysis of alcohol drinking and cancer risk. Br J Cancer. 2001;85: 1700–1705.
157. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Light alcohol drinking and cancer: a meta-analysis. Ann Oncol. 2013;24: 301–308.
158. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. Br J Cancer. 2015;112: 580–593.

159. Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. Addiction. 1999;94: 1551–1573.
160. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. Ann Oncol. 2011;22: 1958–1972.
161. Longnecker MP, Orza MJ, Adams ME, Vioque J, Chalmers TC. A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. Cancer Causes Control. 1990;1: 59–68.
162. Moskal A, Norat T, Ferrari P, Riboli E. Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. Int J Cancer. 2007;120: 664–671.
163. Wang Y, Duan H, Yang H, Lin J. A pooled analysis of alcohol intake and colorectal cancer. Int J Clin Exp Med. 2015;8: 6878–6889.
164. Zhang C, Zhong M. Consumption of beer and colorectal cancer incidence: a meta-analysis of observational studies. Cancer Causes Control. 2015;26: 549–560.
165. Breau G, Ellis U. Risk Factors Associated With Young-Onset Colorectal Adenomas and Cancer: A Systematic Review and Meta-Analysis of Observational Research. Cancer Control. 2020;27: 1073274820976670.
166. Kim JY, Choi S, Park T, Kim SK, Jung YS, Park JH, et al. Development and validation of a scoring system for advanced colorectal neoplasm in young Korean subjects less than age 50 years. Intest Res. 2019;17: 253–264.
167. Lee SE, Jo HB, Kwack WG, Jeong YJ, Yoon Y-J, Kang HW. Characteristics of and risk factors for colorectal neoplasms in young adults in a screening population. World J Gastroenterol. 2016;22: 2981–2992.
168. Glover M, Mansoor E, Panhwar M, Parasa S, Cooper GS. Epidemiology of Colorectal Cancer in Average Risk Adults 20-39 Years of Age: A Population-Based National Study. Dig Dis Sci. 2019;64: 3602–3609.
169. Bull-Otterson L, Feng W, Kirpich I, Wang Y, Qin X, Liu Y, et al. Metagenomic analyses of alcohol induced pathogenic alterations in the intestinal microbiome and the effect of Lactobacillus rhamnosus GG treatment. PLoS One. 2013;8: e53028.
170. Tsuruya A, Kuwahara A, Saito Y, Yamaguchi H, Tsubo T, Suga S, et al. Ecophysiological consequences of alcoholism on human gut microbiota: implications for ethanol-related pathogenesis of colon cancer. Sci Rep. 2016;6: 27923.
171. Tsuruya A, Kuwahara A, Saito Y, Yamaguchi H, Tenma N, Inai M, et al. Major Anaerobic Bacteria Responsible for the Production of Carcinogenic Acetaldehyde from Ethanol in the Colon and Rectum. Alcohol Alcohol. 2016;51: 395–401.
172. Robsahm TE, Aagnes B, Hjartåker A, Langseth H, Bray FI, Larsen IK. Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. Eur J Cancer Prev. 2013;22: 492–505.
173. Slattery ML. Physical activity and colorectal cancer. Sports Med. 2004;34: 239–252.
174. Bressa C, Bailén-Andrino M, Pérez-Santiago J, González-Soltero R, Pérez M, Montalvo-Lominchar MG, et al. Differences in gut microbiota profile between women with active lifestyle and sedentary women. PLoS One. 2017;12: e0171352.

175. Mach N, Fuster-Botella D. Endurance exercise and gut microbiota: A review. J Sport Health Sci. 2017;6: 179–197.
176. Evans CC, LePard KJ, Kwak JW, Stancukas MC, Laskowski S, Dougherty J, et al. Exercise prevents weight gain and alters the gut microbiota in a mouse model of high fat diet-induced obesity. PLoS One. 2014;9: e92193.
177. Matsumoto M, Inoue R, Tsukahara T, Ushida K, Chiji H, Matsubara N, et al. Voluntary running exercise alters microbiota composition and increases n-butyrate concentration in the rat cecum. Biosci Biotechnol Biochem. 2008;72: 572–576.
178. Holowatyj AN, Langston ME, Han Y, Viskochil R, Perea J, Cao Y, et al. Community Health Behaviors and Geographic Variation in Early-Onset Colorectal Cancer Survival Among Women. Clin Transl Gastroenterol. 2020;11: e00266.
179. Esposito K, Chiodini P, Capuano A, Bellastella G, Maiorino MI, Rafaniello C, et al. Colorectal cancer association with metabolic syndrome and its components: a systematic review with meta-analysis. Endocrine. 2013;44: 634–647.
180. Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care. 2012;35: 2402–2411.
181. Khan MT, Nieuwdorp M, Bäckhed F. Microbial modulation of insulin sensitivity. Cell Metab. 2014;20: 753–760.
182. Cani PD, Jordan BF. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. Nat Rev Gastroenterol Hepatol. 2018;15: 671–682.
183. Li R, Grimm SA, Chrysovergis K, Kosak J, Wang X, Du Y, et al. Obesity, rather than diet, drives epigenomic alterations in colonic epithelium resembling cancer progression. Cell Metab. 2014;19: 702–711.
184. Li R, Grimm SA, Mav D, Gu H, Djukovic D, Shah R, et al. Transcriptome and DNA Methylome Analysis in a Mouse Model of Diet-Induced Obesity Predicts Increased Risk of Colorectal Cancer. Cell Rep. 2018;22: 624–637.
185. Sanford NN, Giovannucci EL, Ahn C, Dee EC, Mahal BA. Obesity and younger versus older onset colorectal cancer in the United States, 1998-2017. J Gastrointest Oncol. 2020;11: 121–126.
186. Hussan H, Patel A, Le Roux M, Cruz-Monserrate Z, Porter K, Clinton SK, et al. Rising Incidence of Colorectal Cancer in Young Adults Corresponds With Increasing Surgical Resections in Obese Patients. Clin Transl Gastroenterol. 2020;11: e00160.
187. Chen H, Zheng X, Zong X, Li Z, Li N, Hur J, et al. Metabolic syndrome, metabolic comorbid conditions and risk of early-onset colorectal cancer. Gut. 2021;70: 1147–1154.
188. Schumacher AJ, Chen Q, Attaluri V, McLemore EC, Chao CR. Metabolic Risk Factors Associated with Early-Onset Colorectal Adenocarcinoma: A Case-Control Study at Kaiser Permanente Southern California. Cancer Epidemiol Biomarkers Prev. 2021;30: 1792–1798.
189. Jung YS, Park CH, Kim NH, Lee MY, Park DI. Impact of Age on the Risk of Advanced Colorectal Neoplasia in a Young Population: An Analysis Using the Predicted Probability Model. Dig Dis Sci. 2017;62: 2518–2525.
190. Li H, Boakye D, Chen X, Hoffmeister M, Brenner H. Association of Body Mass Index With Risk of Early-Onset Colorectal Cancer: Systematic Review and Meta-Analysis. Am J Gastroenterol. 2021;116: 2173–2183.

191. Dash C, Yu J, Nomura S, Lu J, Rosenberg L, Palmer JR, et al. Obesity is an initiator of colon adenomas but not a promoter of colorectal cancer in the Black Women's Health Study. Cancer Causes Control. 2020;31: 291–302.
192. Elangovan A, Skeans J, Landsman M, Ali SMJ, Elangovan AG, Kaelber DC, et al. Colorectal Cancer, Age, and Obesity-Related Comorbidities: A Large Database Study. Dig Dis Sci. 2021;66: 3156–3163.
193. Kantor ED, Udumyan R, Signorello LB, Giovannucci EL, Montgomery S, Fall K. Adolescent body mass index and erythrocyte sedimentation rate in relation to colorectal cancer risk. Gut. 2016;65: 1289–1295.
194. Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S. Food groups and colorectal cancer risk. Br J Cancer. 1999;79: 1283–1287.
195. Moore LL, Bradlee ML, Singer MR, Splansky GL, Proctor MH, Ellison RC, et al. BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study adults. Int J Obes Relat Metab Disord. 2004;28: 559–567.
196. Himbert C, Figueiredo JC, Shibata D, Ose J, Lin T, Huang LC, et al. Clinical Characteristics and Outcomes of Colorectal Cancer in the ColoCare Study: Differences by Age of Onset. Cancers . 2021;13. doi:10.3390/cancers13153817
197. Botteri E, Borroni E, Sloan EK, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Am J Gastroenterol. 2020;115: 1940–1949.
198. Liang PS, Chen T-Y, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. Int J Cancer. 2009;124: 2406–2415.
199. Agazzi S, Lenti MV, Klersy C, Strada E, Pozzi L, Rovedatti L, et al. Incidence and risk factors for preneoplastic and neoplastic lesions of the colon and rectum in patients under 50 referred for colonoscopy. Eur J Intern Med. 2021;87: 36–43.
200. Goel A, Boland CR. Epigenetics of colorectal cancer. Gastroenterology. 2012;143: 1442–1460.e1.
201. Suter CM, Martin DI, Ward RL. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis. 2004;19: 95–101.
202. Strubberg AM, Madison BB. MicroRNAs in the etiology of colorectal cancer: pathways and clinical implications. Dis Model Mech. 2017;10: 197–214.
203. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol. 2020;17: 111–130.
204. Kanai Y, Hirohashi S. Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state. Carcinogenesis. 2007;28: 2434–2442.
205. Ng JM-K, Yu J. Promoter hypermethylation of tumour suppressor genes as potential biomarkers in colorectal cancer. Int J Mol Sci. 2015;16: 2472–2496.
206. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology. 2015;149: 1204–1225.e12.

207. Bihl MP, Foerster A, Lugli A, Zlobec I. Characterization of CDKN2A(p16) methylation and impact in colorectal cancer: systematic analysis using pyrosequencing. J Transl Med. 2012;10: 173.
208. Cunningham JM, Christensen ER, Tester DJ, Kim CY, Roche PC, Burgart LJ, et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer Res. 1998;58: 3455–3460.
209. Liang T-J, Wang H-X, Zheng Y-Y, Cao Y-Q, Wu X, Zhou X, et al. APC hypermethylation for early diagnosis of colorectal cancer: a meta-analysis and literature review. Oncotarget. 2017;8: 46468–46479.
210. Sunami E, de Maat M, Vu A, Turner RR, Hoon DSB. LINE-1 hypomethylation during primary colon cancer progression. PLoS One. 2011;6: e18884.
211. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut. 2014;63: 635–646.
212. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, et al. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS One. 2012;7: e45357.
213. Baba Y, Yagi T, Sawayama H, Hiyoshi Y, Ishimoto T, Iwatsuki M, et al. Long Interspersed Element-1 Methylation Level as a Prognostic Biomarker in Gastrointestinal Cancers. Digestion. 2018;97: 26–30.
214. Chi P, Allis CD, Wang GG. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer. 2010;10: 457–469.
215. Gargalionis AN, Piperi C, Adamopoulos C, Papavassiliou AG. Histone modifications as a pathogenic mechanism of colorectal tumorigenesis. Int J Biochem Cell Biol. 2012;44: 1276–1289.
216. Struhl K. Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. 1998;12: 599–606.
217. Audia JE, Campbell RM. Histone Modifications and Cancer. Cold Spring Harb Perspect Biol. 2016;8: a019521.
218. Huang T, Lin C, Zhong LLD, Zhao L, Zhang G, Lu A, et al. Targeting histone methylation for colorectal cancer. Therap Adv Gastroenterol. 2017;10: 114–131.
219. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12: 861–874.
220. Kita Y, Yonemori K, Osako Y, Baba K, Mori S, Maemura K, et al. Noncoding RNA and colorectal cancer: its epigenetic role. J Hum Genet. 2017;62: 41–47.
221. Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet. 2011;12: 136–149.
222. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99: 15524–15529.
223. Slaby O, Svoboda M, Michalek J, Vyzula R. MicroRNAs in colorectal cancer: translation of molecular biology into clinical application. Mol Cancer. 2009;8: 102.
224. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet. 2009;10: 704–714.

225. Lujambio A, Calin GA, Villanueva A, Ropero S, Sánchez-Céspedes M, Blanco D, et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 2008;105: 13556–13561.
226. Svoronos AA, Engelman DM, Slack FJ. OncomiR or Tumor Suppressor? The Duplicity of MicroRNAs in Cancer. Cancer Res. 2016;76: 3666–3670.
227. Akao Y, Kumazaki M, Shinohara H, Sugito N, Kuranaga Y, Tsujino T, et al. Impairment of K-Ras signaling networks and increased efficacy of epidermal growth factor receptor inhibitors by a novel synthetic miR-143. Cancer Sci. 2018;109: 1455–1467.
228. Sun D, Yu F, Ma Y, Zhao R, Chen X, Zhu J, et al. MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASA1). J Biol Chem. 2013;288: 9508–9518.
229. Wu Y, Song Y, Xiong Y, Wang X, Xu K, Han B, et al. MicroRNA-21 (Mir-21) Promotes Cell Growth and Invasion by Repressing Tumor Suppressor PTEN in Colorectal Cancer. Cell Physiol Biochem. 2017;43: 945–958.
230. Peacock O, Lee AC, Cameron F, Tarbox R, Vafadar-Isfahani N, Tufarelli C, et al. Inflammation and MiR-21 pathways functionally interact to downregulate PDCD4 in colorectal cancer. PLoS One. 2014;9: e110267.
231. Gao J, Li N, Dong Y, Li S, Xu L, Li X, et al. miR-34a-5p suppresses colorectal cancer metastasis and predicts recurrence in patients with stage II/III colorectal cancer. Oncogene. 2015;34: 4142–4152.
232. Sun C, Wang F-J, Zhang H-G, Xu X-Z, Jia R-C, Yao L, et al. miR-34a mediates oxaliplatin resistance of colorectal cancer cells by inhibiting macroautophagy via transforming growth factor-β/Smad4 pathway. World J Gastroenterol. 2017;23: 1816–1827.
233. Rokavec M, Öner MG, Li H, Jackstadt R, Jiang L, Lodygin D, et al. IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. J Clin Invest. 2014;124: 1853–1867.
234. Tang W, Zhu Y, Gao J, Fu J, Liu C, Liu Y, et al. MicroRNA-29a promotes colorectal cancer metastasis by regulating matrix metalloproteinase 2 and E-cadherin via KLF4. Br J Cancer. 2014;110: 450–458.
235. Nagel R, le Sage C, Diosdado B, van der Waal M, Oude Vrielink JAF, Bolijn A, et al. Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer. Cancer Res. 2008;68: 5795–5802.
236. Zhang Y, Wang X, Xu B, Wang B, Wang Z, Liang Y, et al. Epigenetic silencing of miR-126 contributes to tumor invasion and angiogenesis in colorectal cancer. Oncol Rep. 2013;30: 1976–1984.
237. Yin J, Bai Z, Song J, Yang Y, Wang J, Han W, et al. Differential expression of serum miR-126, miR-141 and miR-21 as novel biomarkers for early detection of liver metastasis in colorectal cancer. Chin J Cancer Res. 2014;26: 95–103.
238. Liu Y, Zhou Y, Feng X, Yang P, Yang J, An P, et al. Low expression of microRNA-126 is associated with poor prognosis in colorectal cancer. Genes Chromosomes Cancer. 2014;53: 358–365.

239. Ng EKO, Chong WWS, Jin H, Lam EKY, Shin VY, Yu J, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut. 2009;58: 1375–1381.
240. Ke T-W, Wei P-L, Yeh K-T, Chen WT-L, Cheng Y-W. MiR-92a Promotes Cell Metastasis of Colorectal Cancer Through PTEN-Mediated PI3K/AKT Pathway. Ann Surg Oncol. 2015;22: 2649–2655.
241. Kanaan Z, Rai SN, Eichenberger MR, Roberts H, Keskey B, Pan J, et al. Plasma miR-21: a potential diagnostic marker of colorectal cancer. Ann Surg. 2012;256: 544–551.
242. Toiyama Y, Takahashi M, Hur K, Nagasaka T, Tanaka K, Inoue Y, et al. Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. J Natl Cancer Inst. 2013;105: 849–859.
243. Peng Q, Zhang X, Min M, Zou L, Shen P, Zhu Y. The clinical role of microRNA-21 as a promising biomarker in the diagnosis and prognosis of colorectal cancer: a systematic review and meta-analysis. Oncotarget. 2017;8: 44893–44909.
244. Mima K, Nishihara R, Yang J, Dou R, Masugi Y, Shi Y, et al. MicroRNA MIR21 (miR-21) and PTGS2 Expression in Colorectal Cancer and Patient Survival. Clin Cancer Res. 2016;22: 3841–3848.
245. Caramés C, Cristóbal I, Moreno V, del Puerto L, Moreno I, Rodriguez M, et al. MicroRNA-21 predicts response to preoperative chemoradiotherapy in locally advanced rectal cancer. Int J Colorectal Dis. 2015;30: 899–906.
246. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. JAMA. 2008;299: 425–436.
247. Liu K, Li G, Fan C, Zhou X, Wu B, Li J. Increased expression of microRNA-21and its association with chemotherapeutic response in human colorectal cancer. J Int Med Res. 2011;39: 2288–2295.
248. Sun G, Cheng Y-W, Lai L, Huang T-C, Wang J, Wu X, et al. Signature miRNAs in colorectal cancers were revealed using a bias reduction small RNA deep sequencing protocol. Oncotarget. 2016;7: 3857–3872.
249. Liu H-N, Liu T-T, Wu H, Chen Y-J, Tseng Y-J, Yao C, et al. Serum microRNA signatures and metabolomics have high diagnostic value in colorectal cancer using two novel methods. Cancer Sci. 2018;109: 1185–1194.
250. Chang P-Y, Chen C-C, Chang Y-S, Tsai W-S, You J-F, Lin G-P, et al. MicroRNA-223 and microRNA-92a in stool and plasma samples act as complementary biomarkers to increase colorectal cancer detection. Oncotarget. 2016;7: 10663–10675.
251. Herreros-Villanueva M, Duran-Sanchon S, Martín AC, Pérez-Palacios R, Vila-Navarro E, Marcuello M, et al. Plasma MicroRNA Signature Validation for Early Detection of Colorectal Cancer. Clin Transl Gastroenterol. 2019;10: e00003.
252. Kjersem JB, Ikdahl T, Lingjaerde OC, Guren T, Tveit KM, Kure EH. Plasma microRNAs predicting clinical outcome in metastatic colorectal cancer patients receiving first-line oxaliplatin-based treatment. Mol Oncol. 2014;8: 59–67.
253. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482: 339–346.

254. Yang Y, Du Y, Liu X, Cho WC. Involvement of Non-coding RNAs in the Signaling Pathways of Colorectal Cancer. Adv Exp Med Biol. 2016;937: 19–51.
255. Luo J, Qu J, Wu D-K, Lu Z-L, Sun Y-S, Qu Q. Long non-coding RNAs: a rising biotarget in colorectal cancer. Oncotarget. 2017;8: 22187–22202.
256. Puzzono M, Mannucci A, Di Leo M, Zuppardo RA, Russo M, Ditonno I, et al. Diet and Lifestyle Habits in Early-Onset Colorectal Cancer: A Pilot Case-Control Study. Dig Dis. 2022;40: 710–718.
257. Pala V, Sieri S, Palli D, Salvini S, Berrino F, Bellegotti M, et al. Diet in the Italian EPIC cohorts: presentation of data and methodological issues. Tumori. 2003;89: 594–607.
258. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics [Internet]. 1977; 33 (1): 159-74.
259. Turconi G, Roggi C. Atlante fotografico alimentare. Uno strumento per le indagini nutrizionali. EMSI; 2007.
260. Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. doi: 10.18129/B9. bioc. EnhancedVolcano, R package version …. 2023.
261. Wilke CO. Ridgeline Plots in "ggplot2" [R package ggridges version 0.5.5]. 2023 [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=ggridges
262. pheatmap: Pretty Heatmaps. In: Comprehensive R Archive Network (CRAN) [Internet]. [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=pheatmap
263. Szekely GJ, Rizzo ML, Others. Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. J Classification. 2005;22: 151–184.
264. Garnier S. Colorblind-Friendly Color Maps for R [R package viridis version 0.6.4]. 2023 [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=viridis
265. Therneau TM. Survival Analysis [R package survival version 3.5-7]. 2023 [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=survival
266. Ahola-Olli AV, Mustelin L, Kalimeri M, Kettunen J, Jokelainen J, Auvinen J, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia. 2019;62: 2298–2309.
267. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. pp. 785–794.
268. fmsb: Functions for Medical Statistics Book with some Demographic Data. In: Comprehensive R Archive Network (CRAN) [Internet]. [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=fmsb
269. Liu Y. SHAPforxgboost: SHAP (SHapley Additive exPlnation) visualization for "XGBoost" in "R." Github; Available: https://github.com/liuyanguu/SHAPforxgboost
270. pROC: display and analyze ROC curves in R. [cited 19 Jan 2024]. Available: https://xrobin.github.io/pROC/
271. Thiele C. Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks [R package cutpointr version 1.1.2]. 2022 [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=cutpointr

272. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit R. Raincloud plots: a multi-platform tool for robust data visualization. PeerJ Preprints; 2018 Aug. Report No.: e27137v1. doi:10.7287/peerj.preprints.27137v1
273. swimplot: Tools for Creating Swimmers Plots using "ggplot2." In: Comprehensive R Archive Network (CRAN) [Internet]. [cited 19 Jan 2024]. Available: https://CRAN.R-project.org/package=swimplot
274. Therneau T. survival: Survival package for R. Github; Available: https://github.com/therneau/survival
275. Brown M. rmda: R package to plot decision curves. Github; Available: https://github.com/mdbrown/rmda
276. Khan SA, Morris M, Idrees K, Gimbel MI, Rosenberg S, Zeng Z, et al. Colorectal cancer in the very young: a comparative study of tumor markers, pathology and survival in early onset and adult onset patients. J Pediatr Surg. 2016;51: 1812–1817.
277. Chen FW, Yang L, Cusumano VT, Chong MC, Lin JK, Partida D, et al. Early-Onset Colorectal Cancer Is Associated with a Lower Risk of Metachronous Advanced Neoplasia than Traditional-Onset Colorectal Cancer. Dig Dis Sci. 2022;67: 1045–1053.
278. Jiang T-J, Wang F, Wang Y-N, Hu J-J, Ding P-R, Lin J-Z, et al. Germline mutational profile of Chinese patients under 70 years old with colorectal cancer. Cancer Commun. 2020;40: 620–632.
279. You YN, Borras E, Chang K, Price BA, Mork M, Chang GJ, et al. Detection of Pathogenic Germline Variants Among Patients With Advanced Colorectal Cancer Undergoing Tumor Genomic Profiling for Precision Medicine. Dis Colon Rectum. 2019;62: 429–437.
280. DeRycke MS, Gunawardena S, Balcom JR, Pickart AM, Waltman LA, French AJ, et al. Targeted sequencing of 36 known or putative colorectal cancer susceptibility genes. Mol Genet Genomic Med. 2017;5: 553–569.
281. Chubb D, Broderick P, Dobbins SE, Frampton M, Kinnersley B, Penegar S, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun. 2016;7: 11883.
282. Toh MR, Chiang JB, Chong ST, Chan SH, Ishak NDB, Courtney E, et al. Germline Pathogenic Variants in Homologous Recombination and DNA Repair Genes in an Asian Cohort of Young-Onset Colorectal Cancer. JNCI Cancer Spectr. 2018;2: ky054.
283. Augustsson K, Skog K, Jägerstad M, Dickman PW, Steineck G. Dietary heterocyclic amines and cancer of the colon, rectum, bladder, and kidney: a population-based study. Lancet. 1999;353: 703–707.
284. Snyderwine EG. Mammary gland carcinogenesis by food-derived heterocyclic amines: metabolism and additional factors influencing carcinogenesis by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). Environ Mol Mutagen. 2002;39: 165–170.
285. Deitz AC, Zheng W, Leff MA, Gross M, Wen WQ, Doll MA, et al. N-Acetyltransferase-2 genetic polymorphism, well-done meat intake, and breast cancer risk among postmenopausal women. Cancer Epidemiol Biomarkers Prev. 2000;9: 905–910.
286. Zheng W, Gustafson DR, Sinha R, Cerhan JR, Moore D, Hong CP, et al. Well-done meat intake and the risk of breast cancer. J Natl Cancer Inst. 1998;90: 1724–1729.

287. Ito N, Hasegawa R, Sano M, Tamano S, Esumi H, Takayama S, et al. A new colon and mammary carcinogen in cooked food, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). Carcinogenesis. 1991;12: 1503–1506.
288. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366: 697–706.
289. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst. 2004;96: 1420–1425.
290. Manfredi S, Bouvier AM, Lepage C, Hatem C, Dancourt V, Faivre J. Incidence and patterns of recurrence after resection for cure of colonic cancer in a well defined population. Br J Surg. 2006;93: 1115–1122.
291. Haller DG, Tabernero J, Maroun J, de Braud F, Price T, Van Cutsem E, et al. Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer. J Clin Oncol. 2011;29: 1465–1471.
292. Yothers G, O'Connell MJ, Allegra CJ, Kuebler JP, Colangelo LH, Petrelli NJ, et al. Oxaliplatin as adjuvant therapy for colon cancer: updated results of NSABP C-07 trial, including survival and subset analyses. J Clin Oncol. 2011;29: 3768–3774.
293. O'Connor ES, Greenblatt DY, LoConte NK, Gangnon RE, Liou J-I, Heise CP, et al. Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. J Clin Oncol. 2011;29: 3381–3388.
294. Fang SH, Efron JE, Berho ME, Wexner SD. Dilemma of stage II colon cancer and decision making for adjuvant chemotherapy. J Am Coll Surg. 2014;219: 1056–1069.
295. Compton CC. Optimal pathologic staging: defining stage II disease. Clin Cancer Res. 2007;13: 6862s–70s.
296. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. Int J Cancer. 2010;127: 118–126.
297. Zanutto S, Ciniselli CM, Belfiore A, Lecchi M, Masci E, Delconte G, et al. Plasma miRNA-based signatures in CRC screening programs. Int J Cancer. 2020;146: 1164–1173.
298. Kandimalla R, Gao F, Matsuyama T, Ishikawa T, Uetake H, Takahashi N, et al. Genome-wide Discovery and Identification of a Novel miRNA Signature for Recurrence Prediction in Stage II and III Colorectal Cancer. Clin Cancer Res. 2018;24: 3867–3877.
299. Okuno K, Kandimalla R, Mendiola M, Balaguer F, Bujanda L, Fernandez-Martos C, et al. A microRNA signature for risk-stratification and response prediction to FOLFOX-based adjuvant therapy in stage II and III colorectal cancer. Mol Cancer. 2023;22: 13.
300. Zhang Y, Guo C-C, Guan D-H, Yang C-H, Jiang Y-H. Prognostic Value of microRNA-224 in Various Cancers: A Meta-analysis. Arch Med Res. 2017;48: 472–482.
301. Hao H, Liu L, Zhang D, Wang C, Xia G, Zhong F, et al. Diagnostic and prognostic value of miR-106a in colorectal cancer. Oncotarget. 2017;8: 5038–5047.
302. Zhi ML, Liu ZJ, Yi XY, Zhang LJ, Bao YX. Diagnostic performance of microRNA-29a for colorectal cancer: a meta-analysis. Genet Mol Res. 2015;14: 18018–18025.
303. Yang X, Zeng Z, Hou Y, Yuan T, Gao C, Jia W, et al. MicroRNA-92a as a potential biomarker in diagnosis of colorectal cancer: a systematic review and meta-analysis. PLoS One. 2014;9: e88745.

304. Ak S, Tunca B, Tezcan G, Cecener G, Egeli U, Yilmazlar T, et al. MicroRNA expression patterns of tumors in early-onset colorectal cancer patients. J Surg Res. 2014;191: 113–122.
305. Liu C, Wu W, Chang W, Wu R, Sun X, Wu H, et al. miR-31-5p-DMD axis as a novel biomarker for predicting the development and prognosis of sporadic early-onset colorectal cancer. Oncol Lett. 2022;23: 157.
306. Nakamura K, Hernández G, Sharma GG, Wada Y, Banwait JK, González N, et al. A Liquid Biopsy Signature for the Detection of Patients With Early-Onset Colorectal Cancer. Gastroenterology. 2022;163: 1242–1251.e2.