

81730_Auto_Edited.docx

Name of Journal: *World Journal of Biological Chemistry*

Manuscript NO: 81730

Manuscript Type: REVIEW

Molecular genetics of early-onset colorectal cancer

Early-onset colorectal cancer

Olivia Marx, Marc Mankarious, Gregory Yochum

Abstract

Early-onset colorectal cancer (EOCRC) has been rising in global prevalence and incidence over the past several decades. Environmental influences, including generational lifestyle changes and rising obesity, contribute to these increased rates. While the rise in EOCRC is best documented in western countries, it is seen throughout the world, although EOCRC may have distinct genetic mutations in patients of different ethnic backgrounds. Pathological and molecular characterizations show that EOCRC has a distinct presentation compared with later-onset colorectal cancer (LOCRC). Recent studies have identified DNA, RNA, and protein-level alterations unique to EOCRC, revealing much-needed biomarkers and potential novel therapeutic targets. Many molecular EOCRC studies have been performed with Caucasian and Asian EOCRC cohorts, however, studies of other ethnic backgrounds are limited. In addition, certain molecular characterizations that have been conducted for LOCRC have not yet been repeated in EOCRC, including high-throughput analyses of histone modifications, mRNA splicing, and proteomics on large cohorts. We propose that the complex relationship between cancer and aging should be considered when studying the molecular underpinnings of EOCRC. In this review, we summarize current EOCRC literature, focusing on sporadic molecular alterations in tumors, and their clinical implications. We conclude by discussing current challenges and future directions of EOCRC research efforts.

Key Words: Early-onset colorectal cancer; later-onset colorectal cancer; mutations; oncogenes; molecular characteristics; transcriptomics

Marx O, Mankarious M, Yochum G. Molecular genetics of early-onset colorectal cancer. *World J Biol Chem* 2022; In press

Core Tip: Early-onset colorectal cancer (EOCRC) has a considerably different clinical presentation and genetic profile compared with later-onset colorectal cancer (LOCRC).

Furthermore, molecular alterations in EOCRC tumors differ in patients from separate geographical locations and distinct ethnic groups. Small human cohorts and the lack of a suitable mouse model system limit EOCRC studies, however, several actionable clinical targets and biomarkers specific to EOCRC have been identified. In this review, we discuss molecular alterations in EOCRC tumors at the DNA, RNA, and protein levels, and suggest future work to examine how these changes contribute to EOCRC pathogenesis.

INTRODUCTION

Early-onset colorectal cancer is a growing global issue

Cancers of the colon and rectum are the third most commonly found in both men and women globally^[1]. Colon cancer screenings have increased early detection in patients over the age of 50 years old and have contributed to the overall decline in global rates of colorectal cancer (CRC)^[2, 3]. However, the population of early-onset CRC (EOCRC) patients, under 50 years old, has been steadily rising over the past several decades^[4, 5], and by 2030, the rate of early-onset colon and rectal cancer is expected to increase by 27.7%, and 46.0%, respectively^[6]. Unfortunately, 10-15% of CRC patients are diagnosed before the age of average-risk screening recommendation (before 2018, 50 years old)^[7]. Due to a lack of screenings and a delay in the diagnosis of younger patients, EOCRC is often detected at advanced stages, reducing the chances of long-term survival^[8]. Many studies have shown that EOCRC is molecularly distinct from later-onset CRC (LOCRC), or CRC diagnosed after the age of 50 years old. Compared with LOCRC, EOCRC has a differing frequency of oncogenic mutations^[9], increased prevalence of mucinous and signet (poorly differentiated) histology^[10], a more distal location^[4], and exhibits a distinct DNA methylation profile^[11, 12]. Despite aggressive treatment of EOCRC patients, their overall survival is worse compared to those with LOCRC^[7, 13].

EOCRC risk factors

There is no clear cause for most EOCRC cases, although environmental risk factors are likely key contributors to cancer development. Lifestyle factors such as smoking, unhealthy diet, obesity, and alcohol consumption increase the risk of developing CRC early on^[1]. In the U.S., EOCRC has a strong birth cohort effect, implicating generational lifestyle changes in the development of EOCRC^[3, 14].

Several recent studies have demonstrated the association between obesity and metabolic disorders with the development of EOCRC^[15-17]. Tang *et al* found that EOCRC patients had a worse metabolic profile, with higher levels of triglycerides and lower levels of HDL-C compared with LOCRC patients^[18]. Molecular links between obesity, metabolic disorders, and colorectal cancer have been suggested, including the promotion of intestinal stem cell populations^[19, 20], increased insulin resistance, adipocyte levels, and inflammation^[21]. How EOCRC risk factors affect clinical presentation is still under investigation. One aspect of EOCRC clinical presentation of particular interest is tumor location^[22].

Differences in left- and right-sided EOCRC

Over half of pre-malignant polyps in EOCRC are found in the distal colon and rectum^[22, 23], and this has prompted calls for screening sigmoidoscopy at an earlier age than current guidelines, which were recently changed from 50 to 45^[24]. While left-sided colon cancer is more predominant in EOCRC, right-sided EOCRC is associated with lower overall survival compared to left-sided EOCRC (44% vs. 61%)^[25]. Several factors have been implicated in the difference in survival between right-sided and left-sided CRC. During embryonic development, the proximal colon originates from the midgut while the distal colon originates from the hindgut. This developmental difference may impact cancer cell origins as well as the metastatic potential of tumors due to differences in vascularization. Additionally, several microbiota changes have been characterized between the proximal and distal colon which may play a role in oncogenesis^[26, 27]. Proximal colonic tumors also have distinct histopathological features as they tend to be more mucinous with microsatellite instability and mismatch repair

deficiency compared to distal tumors^[28, 29]. These distinctions may indicate unique molecular drivers of distal and proximal EOCRC tumors.

Common molecular drivers of colorectal adenocarcinoma

The multi-step progression from normal colonic mucosa to adenoma to colorectal cancer was first described in 1990 by Fearon and Vogelstein^[30]. In this model, an APC-inactivating mutation is an initiating event, followed by KRAS-activating mutations driving adenoma development^[30]. Further studies found that the malignant transformation of adenomas was driven by additional mutations in the TGF- β , PIK3CA, and TP53 pathways (Figure 1)^[31]. Sessile serrated polyps act as a precursor to up to one-third of CRCs and are thought to arise through mechanisms distinct from canonical APC mutation-driven polyps^[32-34]. Instead, serrated polyps are thought to develop from a BRAF mutation and are also often characterized by DNA hypermethylation^[32]. Mutations to BRAF, KRAS, and TP53 are also found in other cancers and the multi-step progression to adenocarcinoma may not strictly follow the canonical or serrated pathways described. While most sporadic LOCRCs can be categorized by deregulation of canonical Wnt/ β -catenin/APC signaling or serrated BRAF mutation pathways^[34] (Figure 1), it is less clear how sporadic EOCRCs develop.

CRC is often categorized into consensus molecular subtypes (CMS). At the genomic level, CRC can be categorized as microsatellite instable (MSI, often caused by a defect in mismatch repair genes) or chromosome instable (CIN). At the transcriptomic level, there are four main CMS. CMS1 is associated with immune JAK-STAT activation, microsatellite instability, and hypermutated tumor DNA^[35]. CMS2 is associated with canonical Wnt/MYC activation, CMS3 is characterized by metabolic alterations, and CMS4 is associated with epithelial-mesenchymal transition (EMT) and immunosuppression^[35]. CMS1 tumors are more often considered CIMP-high (CpG Island Methylator Phenotype-high, characterized by genome-wide hypermethylation). These CMS1 tumors also often have BRAF(V600E) mutations and are associated with sessile serrated adenomas^[35]. RSPO fusions and RNF43 mutations are often seen in

hypermethylated CMS1 CRC^[35], and Yan *et al* also identified these alterations in a subset of EOCRC tumors^[36].

Differing frequencies of CMS in EOCRCs compared with LOCRCs have been identified. Willauer and colleagues showed that patients under 40 were more likely to exhibit CMS1 or CMS2, with CMS1 being considerably more prevalent in EOCRC compared with the average of most CRCs, though no differentiation between sporadic and hereditary mutations was made^[37]. Therefore, the increased prevalence of Lynch syndrome, a hereditary mismatch repair deficiency syndrome, in younger patients^[3] could contribute to this observation. In fact, MSI/CIMP-high tumors were associated with Lynch syndrome in young patients, whereas in older patients, they were associated with BRAF mutations^[38]. As sporadic EOCRC patients are less likely to have Lynch syndrome^[39], Kirzin *et al* found that sporadic EOCRC tumors ($n = 39$) were less likely to have a CpG island methylator phenotype, which is associated with CMS1^[40]. Limitations to sample size necessitate future studies to examine the CMS of sporadic EOCRCs compared with LOCRCs.

Many excellent reviews explain the molecular subtypes and mutations in CRC^[34, 35, 41]. Likewise, many excellent reviews have focused on the different clinical presentations and outcomes of EOCRC and LOCRC patients, with some mention of prevalent molecular distinctions of EOCRC^[4, 6, 42]. However, reviews that focus primarily on the molecular characterizations of EOCRCs are limited. There has been a rise in EOCRC literature over the past five years that we aimed to summarize here. In this review, we searched the literature for early- and young- onset colorectal cancer, along with specific searches for molecular genetics, DNA methylation, histone alterations, transcriptomics, splicing, proteomics, ethnic disparities, and biomarkers. Relevant papers were selected for discussion. Here, we summarize key findings of the molecular genetic underpinnings of EOCRC thus far.

GENETIC STUDIES OF EOCRC

DNA mutations associated with EOCRC

Hereditary mutations often lead to CRC at a younger age in comparison with sporadic mutations. Compared with LOCRC, EOCRC patients have an increased polygenic risk score based on profiling single nucleotide polymorphisms (SNPs)^[9]. Although approximately 30% of EOCRC cases report a family history of related cancers, only an estimated 10-20% have known genetic risk factors like familial adenomatous polyposis (FAP), Lynch syndrome, or inflammatory bowel disease (IBD)^[3, 13, 43-45]. Therefore, ¹our understanding of the genetic and molecular pathways that drive carcinogenesis in most patients is far from complete.

Sporadic EOCRCs, patients with no family history, have been shown to have different mutational profiles compared with LOCRCs (Table 1). Notably, many studies have found a significant decrease in the prevalence of *APC* and Wnt pathway mutations in EOCRC compared with LOCRC^[36, 45-47], with the exception of the β -catenin gene, *CTNNB1*, which is mutated in more EOCRCs compared with LOCRCs (Figure 2)^[36, 37]. Interestingly, a recent study by Yan *et al* used organoids to demonstrate the heterogeneity between EOCRC patients, identifying some with *APC* mutations and others with *RSPO* fusions, which render the cultures hypersensitive to Wnt withdrawal^[36].

c-MYC (*MYC*) is a key oncogenic target of the Wnt/ β -catenin pathway that is often deregulated in CRC^[48]. Copy number variations (CNV) of *MYC* are seen in 8-15% of CRCs^[49-51], however, we recently reported that 35% of EOCRC tumors from a 21-patient cohort had an increase in *MYC* copy number^[52]. In addition, Pan *et al* reported increased *MYC* copy number in younger CRC patients^[53], however, another study found no association between *MYC* copy number and age^[49]. A previous study by Berg *et al* showed that EOCRC tumors had significantly more gene copy number alterations overall compared with LOCRC tumors^[54]. Alterations to *MYC* regulatory genes have also been identified in *MYCBP2*^[55], an E3 ubiquitin ligase that may regulate *MYC* transcription^[56], and *FBXW7*^[46, 55], a tumor suppressive ubiquitin ligase that mediates degradation of *MYC*, among other oncoproteins, though more work is needed to

determine the functional effect of these mutations^[57]. Together, these findings provide evidence for a Wnt-independent increase in MYC activity in EOCRC.

In addition to differences in the prevalence of Wnt pathway/MYC mutations, many studies reported a decrease in BRAF (V600E) mutations (Table 1)^[37, 40, 58], although Xu *et al* noted an increase in BRAF (V600E) in EOCRC compared with LOCRC tumors^[45]. This finding has clinical implications, as BRAF (V600E) mutant tumors are associated with CIMP high status, have a worse prognosis, and respond differently to cancer treatments^[59]. The decrease in APC and BRAF mutations in EOCRC indicates that a higher percentage of EOCRC tumors may not follow the canonical or serrated carcinogenesis pathways commonly observed in LOCRC (Figure 1)^[37]. Overall, many key oncogenes and tumor suppressors are differentially mutated in EOCRC compared with LOCRC, which may impact cancer progression and prognosis (Table 1, Figure 1). As new mutations unique to EOCRC are uncovered, more work is needed to determine how such mutations affect gene activity.

Epigenetic modifications in EOCRC

In addition to DNA mutations, studies have suggested that EOCRC has a distinct DNA methylation profile from LOCRC^[11, 12, 60]. DNA methylation regulates gene expression and has been implicated in CRC^[61]. Methylation of the long non-coding RNA LINE-1 is often thought to represent global DNA methylation^[62]. Studies suggest that DNA is overall hypomethylated in EOCRC compared with intermediate or later-onset CRC^[11, 43, 60]. Epigenetic modifications may also be detectable in the blood, and a study by Walters *et al* found ³hypermethylation of DNA repetitive elements, including LINE-1, in white blood cells from EOCRC patients^[63]. While previous studies examined global DNA methylation, a recent high-throughput study by Joo *et al* identified 234 differentially methylated regions unique to EOCRC tumors^[12]. The authors then compared EOCRC DNA methylation patterns to those which occur upon age related methylomic drift in the normal mucosa. They suggest that EOCRC tumors more rapidly accumulate cancer-related methylomic drift compared to intermediate or later onset CRC tumors, though it

remains unclear when this drift occurs during cancer progression^[12]. More work is needed to assess DNA methylation over time and within patient-matched tumors and normal mucosa to better understand how age-related DNA methylation changes contribute to EOCRC.

In addition to DNA methylation, histone methylation and acetylation are associated with both aging and CRC^[61], however, there have been few studies on histone modifications in EOCRC. A study in 2015 found that high levels of H3K27me3 were associated with a better prognosis in younger CRC patients and a worse prognosis in older CRC patients^[64]. DNA and histone modifications are a natural part of aging, however, how these impact gene expression, cancer progression and drug response remains to be elucidated.

EOCRC transcriptomics

Transcriptome analysis is a comprehensive tool to identify deregulated signaling pathways in cancer^[65, 66]. When applied to human tissues, this approach considers both genetic and environmental factors that contribute to the profile of deregulated gene expression on a per-patient basis. Several studies have analyzed the transcriptomic profile of EOCRC; however, many studies are limited by sample size and availability of patient-matched control samples.

Deregulation of mRNA targets of the Wnt/ β -catenin pathway has been demonstrated in EOCRC, though at a lower frequency compared with LOCRC^[36]. We have recently published transcriptome analyses implicating deregulated *MYC*, and its downstream targets, in EOCRC^[52]. The proto-oncogene c-MYC (*MYC*) is upregulated in the intestines of obese individuals^[67] and has been suggested to control obesity-mediated metabolic dysfunction in the intestines^[68], though Ellegaard *et al* found no relationship between *MYC* expression and BMI^[69]. Interestingly, our recent transcriptomic study found increased *MYC* expression in the EOCRC tumors of a subset of patients who were obese, suggesting a distinct tumor gene expression profile in obese and non-obese patients^[52]. We did not find significant deregulation of the Wnt/ β -

catenin hallmarks of cancer in our EOCRC tumors compared with adjacent normal tissue. The age-associated role of *MYC* in CRC is supported by another recent study that implicated overexpression of c-MYC, along with the lncRNA *WNTRLINC1* and the gene *ASCL2*, in younger colon cancer patients^[70].

In addition to *MYC*, studies comparing EOCRC and LOCRC have found enrichment of cell signaling, apoptosis/inflammation, proliferation, adhesion, and development^[40]. The recent success in cancer immunotherapies has prompted an interest in examining the immune profiles of CRC^[71], however few studies have interrogated immune response in EOCRC. A recent study highlights the importance of aging and tumor immune response, showing that aging-related gene ontology sets were enriched in CRC tissues compared with normal tissues and this signature was higher in tumors with high immune infiltration^[72]. Profiling approximately 40 tumors from each late- and early-onset CRC patients, Gardner *et al* found that three immune genes *SAA1*, *C7*, and *CFD*, have deregulated expression in EOCRC primary tumors compared with LOCRC tumors^[73]. Changes in expression of these genes were shown to alter the tumor immune microenvironment and are associated with intestinal inflammation^[73]. Another study identified age-associated changes in tumors compared with normal tissues and found enrichment of ⁴ the nuclear factor erythroid 2 Like 2 (*NRF2*) oxidative stress-response in the tumors of younger patients compared with older patients (Figure 2)^[74]. The tumor immune microenvironment is a complex system that involves many different cell types. Immune studies in EOCRC are limited by using a homogenized tumor population for bulk RNA-sequencing instead of examining alterations at a single cell level.

Most transcriptomic studies of EOCRC have focused on mRNA, however, there is increasing evidence for the relevance of microRNAs (miRNAs) in cancer. miRNAs are short (>20nt) RNA transcripts that generally function to bind and repress a specific target mRNA. Two notable miRNA studies have been performed for EOCRC, the first by Nakamura and colleagues examined miRNAs from tumors and normal samples and found a seven-miRNA panel that was upregulated in EOCRC ($n = 42$), but not LOCRC

($n = 370$), in tumor *vs* normal tissues, a (Table 2)^[75]. An earlier study using microarray analyses of a Turkish EOCRC cohort identified downregulation of miR-143, miR-125b, and upregulation of miR-106a in tumors *vs* normal tissues, although no comparison with LOCRC was preformed^[76]. While these miRNAs have been suggested as biomarkers for EOCRC, limited sample sizes, lack of patient-matched controls, and a lack of functional studies leaves their roles in cancer progression unclear.

As with most EOCRC studies, transcriptomic analyses of EOCRC are limited by sample size and availability of quality sequencing data from tumors and patient matched normal control samples. In addition to changes in transcript abundance, RNA-sequencing can provide information about alternative polyadenylation and splicing events, which can alter protein structure and function. Alternative polyadenylation is associated with cellular proliferation and cancer^[77]. It serves to alter the 3' UTR length, which can affect miRNA regulation of the target mRNA in many cancers including CRC^[78]. Unfortunately, to our knowledge, there have been no studies on alternative splicing or polyadenylation events in EOCRC. However one study did find a POLE mutation that may be associated with aberrant splicing in EOCRC^[79]. As aberrant alternative splicing has been implicated in both CRC^[80] and aging^[81], we propose that examining EOCRC specific splicing events would uncover novel insight into disease pathogenesis. Additional post-translational modifications to mRNA, lncRNA, tRNA, and rRNA, such as methylation, have been associated with CRC but remain unexplored in EOCRC^[82-84].

In addition to post-translational modifications and miRNA analysis, another type of transcriptomic analysis that is gaining popularity is single cell RNA-sequencing (scRNA-seq), which can be used to identify gene regulation in the different cell types involved in cancer. Single cell transcriptomics has been used to analyze the age-associated transcriptome in cancers including CRC^[85]. Saul *et al* found that many aging- and senescence- associated genes were generally upregulated in cancers, including CRC^[85]. These authors also found that CRC displayed distinct populations of epithelial cells with elevated age-related gene expression, underscoring the importance of

examining age-related differences in CRC at a single-cell level^[85]. Understanding immune infiltration and stem cell populations is crucial for developing cancer treatments that reduce the risk of relapse. Yan *et al* performed scRNA-seq of EOCRC organoids and showed differing stem cell populations in response to Wnt media supplementation for six different EOCRC and LOCRC tumoroids with different underlying mutations^[36]. Expanding single cell RNA-sequencing of EOCRCs would further elucidate information about disease progression that is specific to distinct cell populations in younger patients.

EOCRC Proteomics

Proteomic studies have advanced cancer treatments by identifying therapeutic targets^[86]. While the proteomic signature of CRC has been established^[86, 87], few studies have profiled the proteome of EOCRC^[74, 88]. Gong *et al* recently published a paper using mass spectrometry to identify age-associated differential expression of proteins in tumors compared to adjacent normal tissues. The authors found an age-associated proteomic signature in CRC tumors, which included MYC, E2F, and mTORC1 targets, and proteins controlling the G2M checkpoint, DNA repair, and unfolded protein response (UPR) pathways expressed at higher levels in older CRC patients. Overall, 208 proteins were found to positively correlate with age and only 20 negatively correlated with age. Many of these proteins reside in pathways that are targetable with known cancer drugs, supporting the potential use of specific cancer treatments for different ages of CRC patients^[88]. For example, the proteins PIN1, ROCK1 and ANXA5 are expressed higher in EOCRC and are targetable by FDA-approved drugs or drugs in clinical trials^[88]. While this study demonstrated the difference in proteomic signatures in younger and older CRC tumors, it was limited by sample size of 50 total patients with young, intermediate, or older onset CRC^[88].

Another recent study by Holowatyj *et al* found no significant differences (FDR q-value < 0.05) in the plasma proteome of younger-onset ($n = 11$) compared with older-onset ($n = 45$) CRCs using an antibody microarray platform to detect 206 inflammatory

proteins. An increased sample size may shed light on interesting targets, as the authors found that the cancer-related proteins BRCA2, PTEN, WNT5B, and WNT7A, among others, had a fold change around two ($P < 0.05$) in EOCRC *vs* LOCRC serum^[74]. While to our knowledge, no other proteome-wide studies have assessed EOCRC, some studies have identified individual proteins that are uniquely expressed in EOCRC tumors. For example, overexpression of the ALDH1/ALDH1A1 protein has been identified in most EOCRC tumors compared with LOCRCs (Figure 2)^[89, 90], and an increase of β -catenin in the nucleus and cytoplasm of EOCRC compared with more membrane staining in LOCRC was shown *via* immunostaining^[40].

Protein modifications such as glycosylation^[91], ubiquitination^[92], phosphorylation, and acetylation^[93] are associated with colorectal cancer, but age-related characterizations remain limited. A recent study found that an increase in glycosylated hemoglobin in the serum of younger non-diabetic adults correlated with increased risk for CRC^[94], though no studies could be found that focused on post-translational modifications within EOCRC tumors. In addition to changes in protein modifications and expression changes, disruptions to protein folding are common in cancers, eliciting the UPR, which promotes cancer cell survival^[95]. Indeed, our previous work showed enrichment of the UPR gene set in EOCRC tumors compared with adjacent control samples^[52].

EOCRC in non-western countries

While many studies focus on profiling EOCRC in western countries, the incidence of EOCRC is also increasing in many Asian countries such as Korea, Thailand, Japan, India, and the city of Hong Kong^[96, 97]. While India reports one of the lowest rates of CRC incidence in the world^[96], over half of sporadic rectal cancers in this country are diagnosed in patients under 50 years old^[98-100]. In addition, studies from Indian cohorts found that under half of early-onset sporadic rectal cancer (EOSRC) tumors exhibit a Wnt signature, the most common driver of CRC, indicating distinct tumor drivers in this population^[99, 101]. Tumors without Wnt signaling showed increased activation of

calcium/ nuclear factor of activated T-cells signaling compared to EOSRC with high Wnt signaling^[99]. Molecular studies in Indian EOCRC patients have also found a decrease in *KRAS* mutations^[101] and deregulation of MAPK and PI3K/AKT pathways^[102] compared with LOCRC patients. Whether EOSRC in Indian patients is molecularly distinct from Western or Caucasian patients, from whom most available CRC data were collected, remains unclear.

A recent study by Xu and colleagues compared germline mutations in a Western Caucasian EOCRC cohort to a Chinese EOCRC cohort and found the Chinese cohort had significantly less hereditary syndromes, with no germline *APC* mutations (mutated in 13% of western cohort) or *BRCA1*, *SMAD4*, or *CHEK2* mutations, while these genes were mutated in 16% of western cohort patients under 50 years old (330 Chinese and 430 Caucasian)^[45]. Another recent molecular study in China examined clinical information for 947 EOCRC and 3521 LOCRC and found that EOCRCs were more likely to have a family history of cancer, higher TNM stage, and higher three-year overall survival, but also lower three-year disease-free survival^[103]. EOCRCs were also more likely to have defective mismatch repair^[103], though it is unclear whether this is a product of Lynch syndrome or sporadic mutations.

While data on CRC age-of-onset are available from many European and Asian countries, limited information on CRC epidemiology is available from countries in Africa^[96]. New data identified an increased prevalence of CRC, with most African countries where data is available, reporting an average age of CRC diagnosis between 43-46 years old^[104]. One study compared EOCRC in Nigerians and African Americans (AA) and found that over 60% of Nigerian CRC patients were diagnosed before the age of 50 years old, compared with 13.2% of AA^[105]. The authors identified many differences between the two populations, where Nigerian EOCRCs were younger and had more rectal cancers^[105]. Unfortunately, the demographic patterns of EOCRC in black individuals remain severely understudied^[105].

In the United States, a clear racial disparity in CRC diagnosis and treatment exists^[106], where the median age of CRC diagnosis is 68 for whites and 64 for blacks^[1]. A

study by Galadima *et al* found that young African Americans had higher CRC incidence compared with young individuals of other races^[107]. In addition, EOCRC rates in the United States are highest among Indigenous and black Americans^[108]. Unfortunately, non-Hispanic blacks with EOCRC have a significantly worse 5-year survival than their white counterparts^[109, 110]. Previous studies have demonstrated ethnicity-specific differences in underlying CRC mutations^[111], but few have focused on EOCRC, especially in people of African descent. One study did find a decrease in the prevalence of *APC* mutations and an increase in gene methylation in an AA EOCRC cohort compared to the mostly white CRC dataset provided by The Cancer Genome Atlas (TCGA)^[112].

Overall, the incidence of CRC is increasing in young patients on a global scale, likely due to dietary and lifestyle changes across the world^[96]. EOCRC may present differently and have different mutations in different populations around the world, likely owing to both lifestyle and genetic differences^[45, 99, 105, 111]. Therefore, it is crucial to increase our understanding of unique EOCRC drivers and to translate this knowledge to improve clinical outcomes for patients worldwide.

EOCRC Biomarkers

¹ The majority of EOCRCs are diagnosed between the ages of 40-49^[113], leading to the American Cancer Society lowering the recommended CRC screening age from 50 to 45 in 2018^[24]. However, several concerns remain on the cost/benefit analysis of this decision^[3], indicating a crucial need for cost-effective early screening options.

While colonoscopy remains the gold standard of CRC screenings, blood and fecal tests are cheaper and less invasive options. Blood-based microRNA and DNA methylation biomarkers have been shown to accurately identify EOCRC^[114]. ⁶ There is currently one FDA-approved blood-based CRC screening test, Epi proColon®, which measures methylation of the gene *SEPT9* in cell-free DNA in serum. A recent study showed that methylation of *SEPT9* could accurately distinguish EOCRC patients from healthy controls, indicating that this test is effective for younger patients^[114]. Another

study suggested that the DNA repetitive elements LINE-1, Sat2, and Alu are hypermethylated in the white blood cells isolated from EOCRC patients, providing an additional potential methylation biomarker signature^[63]. In addition to DNA methylation, microRNA (miRNA) expression is gaining popularity as a potential blood-based cancer biomarker. A recent study identified a miRNA signature of four miRNAs that could distinguish both EOCRC and LOCRC serum from healthy controls (Table 3)^[75]. Serum expression of inflammatory genes has also been suggested to identify EOCRC patients, and one study found that the chemokine *CXCL12* has lower expression in younger compared to older patients^[74].

Another minimally invasive screening option is a fecal test, such as CologuardTM, which measures methylation of the genes *BMP3* and *NDRG4* and assesses samples for *KRAS* mutation. Cologuard also includes a fecal immunohistochemical test (FIT), which measures human globin, or blood, in the stool (Figure 2). Studies of whether these biomarkers can detect EOCRC are limited, though recent studies have found no significant difference in these markers within CRC tumors in younger and older patients^[115-117]. Therefore, while Cologuard and FIT may be effective to detect EOCRC in fecal samples, additional studies are required before such recommendations can be made.

Gene expression biomarkers within tumors have also been suggested to serve as prognostic indicators. miR-31-5p was found to be uniquely overexpressed in sporadic EOCRC tumors *vs* normal samples, while it was not overexpressed in LOCRC. Moreover, the miR-31-5p target, *DMD*, was also shown to be decreased in EOCRC tumors, and this change in expression correlated with a worse prognosis (Table 3)^[118]. In addition to genetic, transcriptomic, and proteomic alterations, biomarkers of the gut microbiome have also been suggested to identify EOCRC, as EOCRC has been shown to have a distinct microbiome compared with LOCRC^[119]. Microbiota studies in CRC and EOCRC are outside the scope of this review but are nicely summarized by Abdullah *et al*^[120]. Overall, a limited number of studies have shown that common CRC screening options may apply to EOCRC as well. miRNA and DNA methylation biomarkers have

been proposed to help identify EOCRC (Table 3), however additional studies with larger sample sizes and clinical validations are required.

Currently, clinical practice guidelines do not differentiate the treatment of EOCRC vs. LOCRC^[121]. However, CRC treatment is dependent on tumor mutational profiling, and thus, the lower frequency of BRAF^[37, 40, 58] and KRAS^[101, 122] mutations in EOCRC means the mutation-specific treatments will be less commonly used in EOCRC patients. The efficacy of these drugs in EOCRC has not been directly studied but the mechanism is likely very similar to LOCRC.

CONCLUSION

Several clinical features have been associated with EOCRC. Approximately 75% of sporadic cases occur in the 40–49-year-old age group, with 55–80% of EOCRCs occurring in the distal colon or rectum^[23, 123, 124]. The increasing rate of EOCRC has been predominated by an increasing rate of the distal colon and rectal cancer, with individuals born circa 1990 having double and quadruple the risk of colon and rectal cancer, respectively, compared to those born circa 1950^[125]. While many strides have been made to understand alterations at the DNA, RNA, and protein levels that contribute to EOCRC, questions remain on how EOCRC patients should be treated compared with their older counterparts. Currently, young patients are more likely to be treated, or overtreated, with systemic chemotherapy, but have similar clinical outcomes compared with older patients^[13, 107, 126].

As the number of EOCRC cases continues to rise globally, there is a critical need to optimize cancer treatment strategies, as well as to further develop non-invasive screening options to identify people at risk for EOCRC. Currently, scientific studies are limited by low sampling size, especially in non-white patients. Researchers are addressing this limitation by continuing to grow biobanks with younger and non-diseased samples. Furthermore, machine learning approaches have been suggested to increase the statistical power of limited sample sizes^[127], which could be applied to identify EOCRC risk genes or genetic loci in under-represented minorities. Another

limitation is the lack of a clear model system to test hypotheses on EOCRC. While models for EOCRC exist, as APC^{min} mice develop CRC at a young age, and HCT-116 cells are from a young patient^[128], these systems fail to recapitulate the diversity in EOCRC subtypes that are observed in patient samples. The development of stable cell lines from CRC samples generally requires transformation, altering the cellular profiles, and limiting normal controls. Organoid models are gaining popularity due to their ability to recapitulate the colonic crypt structure from both normal and tumor cells^[36, 129]. Future work will tease out the molecular mechanisms unique to EOCRC with growing biobanks and organoids as well as other innovative model systems.

2%

SIMILARITY INDEX

PRIMARY SOURCES

1	mdpi-res.com Internet	34 words — 1%
2	Aleli Jillian L. Sia, Gliceria B. Ramos, Carolyn Marie D. Legaspi, John Edward O. Tanchuco et al. "Preventive Potential of Cinnamon (Cinnamomum cassia) Extract on Aberrant Crypt Foci Formation in the Proximal and Distal Colon of Azoxymethane-induced Mice", Journal of Biological Sciences, 2013 Crossref	15 words — < 1%
3	www.wjgnet.com Internet	14 words — < 1%
4	mdpi.com Internet	13 words — < 1%
5	www.ncbi.nlm.nih.gov Internet	12 words — < 1%
6	www.springermedizin.de Internet	12 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES

EXCLUDE MATCHES

< 12 WORDS

< 12 WORDS