For reprint orders, please contact: reprints@futuremedicine.com

The genetic and epigenetic landscape of early-onset colorectal cancer

C Richard Boland^{*1} , Ajay Goel² & Swati G Patel³

¹Professor of Medicine, University of California San Diego School of Medicine, CA, USA

²Professor & Chair, Department of Molecular Diagnostics & Experimental Therapeutics, Director, Biotech Innovations, Beckman Research Institute, City of Hope Comprehensive Cancer Center, Monrovia, CA 91016, USA

³Assistant Professor of Medicine-Gastroenterology, University of Colorado School of Medicine, Rocky Mountain Regional Veterans Affairs Medical Center, Aurora, CO, USA

*Author for correspondence: Tel.: +1 626 281 0541, ext. 80541; crboland@health.ucsd.edu

Colorectal cancer (CRC) in individuals under the age of 50 is a problem that is increasing in USA and around the world. In this review, we discuss the degree to which early-onset (EO)CRC may be due to unsuspected Lynch syndrome or other inherited germline variants that predispose to cancer, describe the known somatic genetic alterations in EO tumors and discuss alterations in DNA methylation. Approximately 20% of EOCRCs can be attributed to identifiable germline mutations in genes that cause familial cancer syndromes. A variety of other genetic/epigenetic alterations have also been reported. We conclude that this is a heterogeneous problem, that requires a comprehensive analysis of genetic/epigenetic signatures to better understand EOCRC. Various subsets of EOCRCs must be analyzed individually for clues regarding the etiologies and possible specific therapies for this disease.

First draft submitted: 19 March 2020; Accepted for publication: 12 June 2020; Published online: 28 October 2020

Keywords: early-onset colorectal cancer • epigenetics • familial cancer • genetics • LINE-1 hypomethylation • Lynch syndrome

This issue of *Colorectal Cancer* is focused upon the topic of early-onset colorectal cancer (EOCRC). As is discussed in the chapter by Murphy [1] and the editorial by Siegel [2], since the 1980s, there has been a consistent trend with an increase in the incidence of colorectal cancer (CRC) in people under age 50, while at the same time, the incidence and mortality have been falling for those over 50 in USA [3,4]. This trend has been confirmed in non-US populations as well [5]. This counterintuitive observation has not been accompanied by a mechanistic biological explanation, although theories have been offered to account for this. It has been proposed that changes in the diet, microbiome, physical activity or perhaps unapparent environmental toxic exposures might be responsible for the rise in incidence of EOCRC; however, there is no integrated analysis of these tumors that provide a coherent narrative for this disturbing epidemiological trend. EOCRC patients tend to present with more advanced stage disease, to have a higher proportion of rectal cancers, to have unique pathological features, to occur more frequently in African-Americans and to have positive family histories compared with older-onset CRC patients [6]. Many of these features are discussed in the accompanying articles in this issue. This paper explores the cellular, genetic and epigenetic features of EOCRC in an attempt to understand the mechanistic basis of this disease.

Genetic hypotheses to explain the mechanistic basis of EOCRC

To date, there are no comprehensive studies that explore differences in environmental exposures, microbiome characteristics or other factors that appear to be involved in the genesis of CRC in general compared with EOCRC. However, there have been multiple articles that report the cellular and genetic/epigenetic features of CRC in general and EOCRC specifically. These observations do not harmonize in such a way to explain the disease. Our approach has been to take these observations one by one to determine what features might lead to an explanation of why some young people develop CRC.

Lynch syndrome & microsatellite instability in EOCRC

The first hypothesis developed to explain EOCRC goes back to the discovery of microsatellite instability (MSI) in the context of Lynch syndrome in 1993 [7,8]. The hypothesis was that some proportion of EOCRC was due to Lynch syndrome – but the ability to find germline mutations for the culprit Lynch syndrome genes was in its infancy, so surrogate methods were needed. What made this hypothesis testable was the discovery of a specific mutational signature for Lynch syndrome – MSI – which can be identified in any cellular system that has lost its DNA mismatch repair (MMR) activity. Table 1 summarizes a series of studies that examined microsatellite instability in CRC cohorts.

In 1995, Liu *et al.* reported that MSI was present in 58% of 31 CRC patients who were ≤ 35 years old, versus 17% of CRC patients between 36 and 55 years of age, and 12% of all 158 patients > 35 years old [9]. Using an insensitive early-stage assay for germline mutations, they documented Lynch syndrome in just 5 of 12 patients tested. It is possible that this referral center (Johns Hopkins, MD, USA) may have had a bias for young CRC cases from familial clusters. Additionally, the diagnostic criteria for MSI had not yet been codified, as MSI-high was not completely defined until 1998 [18]. However, this was a starting place that led to the concept that some proportion of young patients with CRC may have had Lynch syndrome. The facile germline diagnosis of that disease was not achieved for another 15 years.

Capozzi *et al.* from northern Italy in 1999 reported data on 47 tumors that were studied for MSI, immunohistochemistry (IHC) for two of the DNA MMR proteins (*MSH2* and *MLH1*, but not *MSH6* nor *PMS2*), and somatic mutations in a number of known genetic targets of MSI. Again, the MSI panel used predated the development of a consensus panel of PCR-based marker microsatellites. The cohort was a heterogeneous group of patients with EOCRC including those with known or suspected Lynch syndrome. They demonstrated MSI in 33% of the suspected Lynch syndrome patients and 21% in the EOCRC patients [10].

Gryfe *et al.* published the first large, multicenter study of MSI in young patients (≤ 50 years old) including 607 patients from Ontario, Canada in 2000 [11]. The tumors were analyzed using the newly developed consensus panel of MSI markers [18], but not IHC for the DNA MMR proteins, and there was no statement of whether they excluded patients with strong family histories of CRC. They discovered MSI (specifically MSI-high) in 17% of the EOCRCs, and this was associated with a significant survival advantage independent of tumor stage at diagnosis, and a decreased likelihood of regional or distant metastasis [11].

An article by Naidoo *et al.* from South Africa in 2000 looked at CRCs in patients < 35 years, compared them with CRC patients > 50 years and reported MSI-high as well as MSI-low, which somewhat confused the issue. The younger cohort demonstrated MSI-high in 31.3% of cases versus 12% in the older cohort. They also reported allelic imbalance (characteristic of chromosomal instability or CIN) in 9–27% of CRCs; MSI and CIN were observed simultaneously in some cases, indicating that overlapping forms of genomic instability could occur in these tumors [12]. A report by Fernebro in 2002 from Sweden utilizing a PCR assay for MSI and IHC for *MSH2* and *MLH1* proteins suggested that only 3/30 (10%) patients with rectal cancer before age 50 had defective MMR activity [13]. In 2005, Losi *et al.* reported that 14/71 (20%) of cases of CRC in patients < 45 years old had MSI and abnormal IHC for three of the MMR genes (*MLH1*, *MSH2*, *MSH6*), that chromosomal instability (CIN) was common in the early-onset cases, and that somatic mutations in *APC*, *β -catenin* and *TP53* were more common in non-MSI early-onset tumors [14].

Goel *et al.* from USA in 2010 examined this issue in 75 patients < 50 years old in a single-institution cohort in which all cases of familial adenomatous polyposis, known Lynch syndrome and patients with more than one first-degree relative with CRC were excluded to help reduce bias [15]. In this study, both IHC and MSI testing were used to study the tumors, but germline DNA was not available for genetic diagnosis. The study determined that abnormal IHC or MSI was present in 21% of the tumors. Again, most EOCRC was not due to Lynch syndrome. It is clear that not all CRCs with MSI (15% of all CRCs) are accompanied by germline mutations in DNA MMR genes, as most are related to nonfamilial methylation of the promoter of the DNA MMR gene *MLH1* [19]. Additionally, some of the patients with MSI or abnormal IHC may have double somatic mutations in MMR genes in their tumors [20]. It is also likely that some others have cryptic germline mutations in the DNA MMR genes that are not detectable using current diagnostic methods, as many have a family history that suggests a genetic basis, and there were no *BRAF* mutations in this cohort. Furthermore, the majority of these tumors (72%) occurred in the distal colon, and the IHC highlighted that there was an unexpected over-representation of involvement by the MMR genes *MSH6* and *PMS2* [15].

Table 1. Lynch syndrome and early-onset colorectal cancer.

Study, year	Cohort characteristics	Testing performed	MSI results, n (%)	IHC results, n (%)	Germline results, n	Ref.
Liu <i>et al.</i> , 1995	Excluded HNPCC [†] , FAP, IBD	4 or 5 MSI markers '4 MMR genes' tested in germline	<35: 18/31 (58) 36–55: 8/46 (17) >55: 11/112 (11)	Not performed	3 <i>MSH2</i> 2 <i>MLH1</i>	[9]
Capozzi <i>et al.</i> , 1999	Amsterdam Suspected HNPCC [‡] EOCRC [§]	6–8 MSI markers <i>MLH1</i> , <i>MSH2</i> germline	Amsterdam: 8/11 (73) Suspected HNPCC: 4/12 (33) EOCRC: 5/24 (21)	Not performed	<i>MLH1</i> or <i>MSH2</i> Amsterdam: 6/9 Suspected HNPCC: 1/10 EOCRC: 2/24	[10]
Gryfe <i>et al.</i> , 2000	All <50	5–10 MSI markers	MSI-H: 102/607 (17) MSI-L: 20/607 (3) MSS: 485/607 (80)	Not performed	Not performed	[11]
Naidoo <i>et al.</i> , 2000	<35 >50	5 MSI markers	<35: MSI-L 6/32 (19); MSI-H 10/32 (31) >50: MSI-L 22/50 (44); MSI-H 6/50 (12)	Not performed	Not performed	[12]
Fernebro <i>et al.</i> , 2002	Rectal cancer <50	6 MSI markers <i>MLH1</i> , <i>MSH2</i> MMR IHC	MSI-H 3/30 (10)	3/30 (10) absent <i>MSH2</i> 0/30 (0) absent <i>MLH1</i>	Not performed	[13]
Losi <i>et al.</i> , 2005	<45	MSI <i>MLH1</i> , <i>MSH2</i> MMR IHC	MSI-H: 14/71 (20)	6/14 (43) absent <i>MLH1</i> or <i>MSH2</i>	Not performed	[14]
Goel <i>et al.</i> , 2010	<50	5 MSI markers <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i> MMR IHC	MSI-H: 13/75 (17) MSI-L: 3/75 (4) MSS: 59/75 (79)	16/75 (21) absent MMR	Not performed	[15]
Giraldez <i>et al.</i> , 2010	≤50 36 with FDR or SDR CRC 8 meeting Amsterdam II	5 MSI markers <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i> MMR IHC Germline testing based on IHC	MSI-H: 15/140 (11)	20/140 (14) absent MMR	4 <i>MLH1</i> 1 <i>MSH2</i> 6 <i>MSH6</i> 4 biallelic <i>MUTYH</i>	[16]
Chang <i>et al.</i> , 2012	<40	5 MSI markers <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i> MMR IHC	MSI-H: 4/13 (31) MSS: 1/13 (8)	5/75 (7) absent <i>MLH1</i> / <i>PMS2</i> 5/75 (7) absent <i>MSH6</i> 2/75 (3) absent <i>MSH2</i> / <i>MSH6</i> 1/75 (1) absent <i>PMS2</i>	Not performed	[17]

[†] HNPCC defined as patients meeting Amsterdam Criteria (three affected members in the family, two consecutive generations, one diagnosis under age 50).

[‡] Those with suspected HNPCC, but not meeting Amsterdam criteria.

[§] Those with no family history of CRC.

EOCRC: Early-onset colorectal cancer; FAP: Familial adenomatous polyposis; FDR: First-degree relative; HNPCC: Hereditary nonpolyposis colorectal cancer; IBD: Inflammatory bowel disease; IHC: Immunohistochemistry; MMR: Mismatch repair; MSI: Microsatellite instability; MSI-H: Microsatellite instability-high; MSI-L: Microsatellite instability-low; MSS: Microsatellite stable; SDR: Second-degree relative.

In a paper the same year by Giraldez *et al.*, a multicenter collaborative study from Spain of EOCRC patients (≤ 50 years old) reported similar results, except that germline DNA was available and the *MutYH* gene was also studied [16]. In this cohort of 140 patients, 15% of the EOCRC cases demonstrated abnormalities in IHC or MSI, germline mutations in DNA MMR genes were identified in 8%, and 3% had biallelic *MutYH* germline mutations. One case of somatic methylation of *MLH1* was shown and 4% had *BRAF* mutations. *KRAS* mutations were observed in 39/140 cases (28%) and was significantly associated with male sex. *MSH6* was the most prevalent MMR gene germline variant (40% of the MMR-deficient tumors), 25% of those tumors did not show MSI and 4/6 were in the distal colon. Thus, MSI is easily missed in this group using the typical tumor screening. Therefore, some number of genetic alterations were identified, but as many as 85% had something else.

In 2012, Chang *et al.* reported that in 75 CRC patients ≤ 40 years old collected between 2000 and 2010, 17% reported evidence of defective DNA MMR activity, and a small proportion had germline mutations in two other genes known to cause familial CRC – *APC* and *TP53* [17].

Thus, between 1995 and 2010, it was a reasonable hypothesis that undetected Lynch syndrome might be a significant cause of EOCRC. However, this accounted for no more than 21% of the problem, and MSI was more prevalent in the very young cohorts (i.e., <35 years old) than in those 35–50 years old. What is important to keep in mind is that Lynch syndrome tumors are associated with a significantly better prognosis than CRC in general and especially in EOCRC. Therefore as we study this problem in the future, it is essential to identify those cases of EOCRC with MSI, abnormal IHC or germline variants in DNA MMR genes and remove them from the cohort under study, since the inclusion of these cases will only serve to confuse our analysis of the rest.

Genetic panel testing for germline variants in EOCRC

Advancing technology has made it possible to directly assay for germline variants in DNA MMR genes in virtually all CRC patients. Table 2 summarizes a series of studies describing germline variants in EOCRC patients.

Limburg *et al.* reported in 2011 germline analyses on 195 patients <50 years old from three familial CRC registries [21]. They extracted lymphocyte DNA and sequenced the *MLH1*, *MSH2* and *MSH6* Lynch syndrome genes, and found ‘deleterious’ or ‘suspected deleterious’ germline variants in 11 patients (6%, 95% confidence intervals [CI] 3%–10%). Another 12 patients had what were ‘likely neutral’ sequence variants, and 14 had ‘variants of uncertain significance’. (These descriptions are the conventional terms outlined by the American College of Medical Genetics used to estimate the impact of the sequence variants.) These CRC patients were highly selected based upon their family histories, in addition to the early age of onset of their cancers. The challenge with direct DNA sequencing is that not all variants can be identified, and sequence variants can be difficult to interpret.

In 2015, Mork *et al.* from MD Anderson Cancer Center (TX, USA) looked for germline variants in 193 patients ≤ 35 years old who were referred for genetic counseling and testing between 2009 and 2013 [22]. All tumors were tested for DNA MMR deficiency (MSI and IHC testing) and directed genetic tests were mainly performed in commercial laboratories. In this very young, highly selected cohort, 66 (34%) had an identifiable hereditary cancer syndrome, including 22 with Lynch syndrome variants, 22 with MMR deficiency but no identifiable germline variant in a DNA MMR gene (sometimes called Lynch-like syndrome [27]), 16 with FAP, two with constitutional mismatch repair deficiency (CMMRD, caused by biallelic DNA MMR gene variants), two with biallelic *MutYH* gene variants (*MutYH*-associated polyposis or MAP), and one with a germline variant in *TP53* (Li-Fraumeni syndrome). Of interest, 19% of the patients did not have a family history typical of the genetic disease for which they were diagnosed. These were very young patients and there was a high degree of referral bias, but this represents one end of the spectrum of possibilities in adolescents and young adults with CRC.

In 2017, Yurgelun *et al.* from Dana Farber Cancer Institute in Boston (MA, USA) reported a study of germline testing using a locally developed panel of 25 genes associated with familial cancer syndromes in 1,058 consecutive CRC patients of all ages [23]. The cohort was taken from a registry of consecutively enrolled patients not otherwise selected for age, family history or other factors (Figure 1). This group reported that 105 (10%, CI: 8–12%) of the patients had a convincing germline variant in a familial cancer gene, including the expected high penetrance CRC genes (Lynch syndrome genes, *APC*, *MutYH*, the hamartomatous polyp genes, etc.), as well as germline variants in unexpected genes including the *BRCA*-related family of genes and other moderate penetrance genes not necessarily associated with familial CRC. The median age at first diagnosis of CRC in this cohort was 56 years, and 336/1,058 (32%) were <50 years old. In the early-onset subgroup, 47 (14%) had at least one germline variant, of which 21 (45% of that group) had early-onset Lynch syndrome, and 5 each had variants in a polyposis syndrome gene, *BRCA1/2* or *ATM*.

Table 2. Germline variants in early-onset colorectal cancer.

Study	Cohort	MSI-H and/or absent MMR n patients (%)	Genes tested	Pathogenic or likely pathogenic variants Overall n patients (%) Germline variants found (n)	Variants of uncertain significance	Ref.
Limburg <i>et al.</i> 1997–2002 Multi-center US, Canada, Australia	<50: n = 195 mean age 43	18/155 (11.6) [†]	<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i>	11/195 (5.6) <i>MLH1</i> (3), <i>MSH2</i> (6), <i>MSH6</i> (2)	14 (7.2)	[21]
Mork <i>et al.</i> 2009–2013 Single-center US	<35: n = 193 Mean age 29	45/173 (26.0) [†]	Phenotype driven	44/193 (22.8) <i>APC</i> (13), biallelic <i>MUTYH</i> (2), <i>CMMRD</i> (2), <i>TP53</i> (1), monoallelic <i>MUTYH</i> (1), <i>MLH1</i> (6), <i>MSH2</i> (5), <i>MSH6</i> (2), <i>PMS2</i> (1), <i>EPCAM</i> (1)	6 (3.1)	[22]
Yurgelun <i>et al.</i> 2008–2014 Single-center US	<50: n = 336 Median age 56	70/572 (12.2) [‡]	25-gene panel	52 mutations in 47 patients/336 (14.0) <i>MLH1</i> (13), <i>MSH2</i> (4), <i>MSH6</i> (3), <i>PMS2</i> (2), <i>APC</i> (3), <i>BRCA1</i> (2), <i>BRCA2</i> (4), <i>APC</i> I1307K (4), biallelic <i>MUTYH</i> (3), monoallelic <i>MUTYH</i> (6), <i>BRIPI1</i> (1), <i>ATM</i> (6), <i>CHEK2</i> (1)	479/1260 (38.0) [§]	[23]
Pearlman <i>et al.</i> 2013–2016 Multi-center US	<50: n = 450 Mean age 43	MSI-H: 49 (10.8) MMR deficient: 48 (10.7)	25-gene panel	75 variants in 72 patients/450 (16.0) <i>MLH1</i> (13), <i>MSH2</i> (17), <i>MSH6</i> (2), <i>PMS2</i> (6), <i>APC</i> (6), <i>APC</i> I1307K (4), biallelic <i>MUTYH</i> (4), monoallelic <i>MUTYH</i> (8), <i>SMAD4</i> (1), <i>BRCA1</i> (2), <i>BRCA2</i> (4), <i>ATM</i> (4), <i>CHEK2</i> (1), <i>PALB2</i> (2), <i>CDKN2A</i> (1)	145 (32.2) [¶]	[24]
Stoffel <i>et al.</i> 1998–2015 Single-center US	<50: n = 430 Mean age 40	MSI-H: 41/202 (20.3)	124 or 67 gene panel	79/315 (25.1) <i>MSH2</i> (25), <i>MLH1</i> (24), <i>MSH6</i> (5), <i>PMS2</i> (2), <i>APC</i> (10), <i>MUTYH</i> (8), <i>SMAD4</i> (2), <i>BRCA1</i> (1), <i>TP53</i> (1), <i>CHEK2</i> (1)	21/315 (6.7)	[25]
Toh <i>et al.</i> 2014–2016 Single-center Singapore	<50: n = 88 Mean age 41	MMR deficient: 0/69 (0) [#]	64-gene panel, excluding MMR genes	12/88 (13.6) <i>APC</i> (4), monoallelic <i>MUTYH</i> (2), <i>ATM</i> (1), <i>BRCA2</i> (1), <i>NTHL1</i> (1), <i>PALB2</i> (1), <i>WRN</i> (2)	236/88 [¶]	[26]

† Only reported MMR.

‡ Reported as MSI-H and/or MMR.

§ Includes total cohort of patients (data on only <50 cohort not available).

¶ Some patients had more than one variant.

Excluded patients with MMR-deficient tumors.

CMMRD: Constitutional mismatch repair deficiency; MMR: Mismatch repair; MSI-H: Microsatellite instability-high; US: United States.

Figure 1. Germline variants (n) by age and penetrance. (A) Colorectal variants from [24]. **(B)** Noncolorectal variants from [24]. **(C)** Colorectal variants from [23]. **(D)** Noncolorectal variants from [23]. High-penetrance colorectal genes included *MLH1*, *MSH2*, *MSH6*, *PMS2*, *APC*, biallelic *MUTYH*, *SMAD4*. Moderate-penetrance colorectal genes included *APC I1307K*, monoallelic *MUTYH*. High-penetrance noncolorectal genes included *BRCA1*, *BRCA2*, *CDKN2A*, *PALB2*, *TP53*. Moderate penetrance noncolorectal genes included *ATM*, *CHEK2*, *BARD1*, *BRIP1*, *NBN*.

The Y-axis represents number of patients with a variant. Some patients had more than one variant; the higher-penetrance variant is represented.

This was followed by two papers with a similar format that included only patients <50 years old. Pearlman *et al.* from Ohio State reported a study in 2017 of 450 patients collected prospectively from 51 hospitals in Ohio with CRC before age 50. They used MSI and IHC testing in addition to germline testing using their panel of 25 cancer susceptibility genes using next-generation sequencing, and essentially the same genes used by the Dana Farber group [24]. In this study, germline variants were observed in 72 patients (16% of the group), and 48 patients (11%) had MMR-deficient tumors. About half of the patients with germline variants had Lynch syndrome (51% of that group), but a substantial number had germline variants in the *BRCA* gene family, just as had been reported in the Dana Farber study [23].

In 2018, Stoffel *et al.* from the University of Michigan (MI, USA) reported data on 430 CRC patients <50 years old who had been referred for genetic evaluation between 1998 and 2015 [25]. Targeted DNA sequencing was directed by the clinical features, and the diagnostic power of germline testing gradually evolved over the 17-year period of study. Ten percent of the group had evidence of an MMR-deficient tumor, and 79 of 315 (25%) who underwent germline testing had a sequence variant in a familial cancer-associated gene, of which 56 (71%) had Lynch syndrome. Another 21 patients had variants of uncertain significance.

A study from Singapore by Toh *et al.* in 2018 studied 88 EOCRCs (<50 years old, median age 41 years) using whole exome sequencing [26]. They reported that in addition to finding germline variants in the *APC* and *MutYH*

genes, six patients had pathological germline variants in the DNA repair genes *ATM*, *BRCA2*, *PALB2*, *NTHL1* and *WRN*, similar to what had been reported in the gene panel studies earlier. This underscores the necessity to look for defects in homologous recombination genes when analyzing tumors for somatic alterations.

Two other studies searched for novel candidate germline variants in EOCRC. de Voer *et al.* from the Netherlands performed whole exome sequencing on blood from 55 patients with CRC <45 years of age, and determined potentially damaging variants in at least three individuals in two genes: *PTPN12* and *LRP6* [28]. Jansen *et al.* from the USA performed whole exome sequencing on nine patients from three families with a dominant pattern of cancer risk inheritance and no germline variant in a known syndromic gene, and observed sequence variants in 17 candidate genes not previously linked with familial cancer syndromes, including *NOTCH2*, *CYP11B1* and *RAB25* [29].

Therefore, the use of broad diagnostic genetic panels and searches for novel familial cancer genes revealed that the suspected CRC genes account for a relatively small proportion of EOCRC, whereas most cases fall outside of those categories, and to date, no genetic explanation is available for most of them.

Somatic alterations in EOCRC

Another approach to solving the possible mechanistic basis of cancer is to study the patterns of somatic mutations in the tumor DNA. It was the discovery of MSI that led to the linkage of Lynch syndrome to the DNA MMR genes, but there are other mutational signatures that can be linked to either specific germline variants or specific environmental exposures [30–32].

In 2007, Hong *et al.* published an analysis of gene expression using microarrays in early-onset CRCs (≤ 50 years old) from Singapore [33]. They studied 12 cases of EOCRC, and reported consistent upregulation of seven genes in the tumor tissue compared with normal colons from healthy controls, including: *CYR61*, *UCHL1*, *FOS*, *FOSB*, *EGR1*, *VIP* and *KRT24*, and validated these findings in six additional cases. In 2014, Kirzin *et al.* compared molecular and clinical findings from 39 cases of sporadic CRC in 39 patients ≤ 45 years old versus 36 CRC patients >60 years old [34]. After excluding the cases with MSI (Lynch syndrome and Lynch-like syndrome), the early-onset cases were found to be more distally located, more likely to have synchronous metastases, to have a positive family history for CRC, to have a wild-type *BRAF* gene and the absence of a methylator phenotype, and to have evidence of CIN.

In 2017, Perea *et al.* from Madrid reported the frequent somatic deletion of the *NOMO-1* gene in EOCRC and homozygous losses of *NOMO-1* in CRCs that demonstrated microsatellite stability (MSS) [35]. A follow-up clinical and molecular analysis from the Perea group in 2019 reported that right-sided EOCRCs tended to show clinical features of Lynch syndrome and less CIN, whereas the later-onset CRCs showed a higher degree of chromosomal copy number variation (CNV) and CIN. In the EOCRC group, 76% were in the left colon or rectum. Interestingly, none of the rectal cancers in the early-onset group showed MSI, so presumably none of those represented Lynch syndrome [36].

In 2019, a multicenter study of 18,218 CRC cases were subjected to targeted next-generation genomic sequencing of 3769 exons from 403 cancer-related genes, and of 47 introns commonly rearranged in cancer tissues. The patient groups were divided into an early-onset group (1420 cases <40 years old), an intermediate-aged group (3,248 cases 40–49 years old) and an older-onset group (13,550 cases ≥ 50 years old) [37]. Generally, the genomic alterations were similar between the younger and older cohorts, with a few notable differences. *TP53* and β -*catenin* alterations were more common in the youngest cohort, whereas *APC*, *KRAS*, *BRAF* and *FAM123B* variants more frequent in the older cohort. Unfortunately, in spite of considerable efforts from multiple groups, it has been difficult to find any actionable data from these analyses of somatic mutations that might help understand the pathogenesis or guide the treatment of EOCRC.

Aneuploidy, copy number variants, & MACS in EOCRC tumors

Most cancers show some type of genomic instability, and the commonest type is characterized by aneuploidy and large amounts of CNV throughout the genome (CIN), which leads to the deletion of some genes and the amplified expression of others. However, some CRCs have been observed that have minimal evidence of CNV using flow cytometry or comparative genomic hybridization (CGH), making them near-diploid, and also show no evidence of MSI. These tumors have been termed ‘microsatellite and chromosomally stable’ or MACS. In a report by Chan *et al.* from Hong Kong in 2001, tumors with the MACS phenotype were reported in 14/22 EOCRCs (64%, mean

age 33 years) versus 2/16 late-onset CRCs (13%, mean age 68 years) [38]. This would imply that some other form of genomic instability may be at work in some tumors.

The higher frequency of MACS CRCs was confirmed in a prospectively collected series of 84 microsatellite-stable (MSS) EOCRCs (≤ 50 years old) and compared with 90 MSS later-onset CRCs (> 65 years old) from the Mayo Clinic, reported by Boardman *et al.* in 2007 [39]. The MSS tumors in the EOCRC group were more frequently diploid (46%) than the older-onset group (26%, $p = 0.006$), and were more frequently late-stage (42 vs 18%, $p = 0.02$) and located in the rectum (50 vs 23%, $p = 0.04$). It is not clear what pathological process is driving MACS tumors, or whether this is reflective of the distal colorectal location of these tumors.

In a publication from Madrid in 2016, Arriba *et al.* from Perea's group reported that CNVs occurred at different chromosomal locations in EOCRCs versus later-onset CRCs, leading to the stratification of the tumors into two 'clusters' of CIN [40]. This group then looked for CIN (using array CGH), and tested for MSI and the CpG island methylator phenotype (CIMP) in CRCs from patients ≤ 45 years old [41]. 13 of 81 tumors (16%) demonstrated MSI (7 of which were shown to have Lynch syndrome) and there was only one case with a methylated *MLH1* promoter (but there was no mutation in *BRAF*), indicating the relative frequencies of those abnormalities in this early-onset cohort.

Epigenetic alterations in EOCRC

Given the challenges in finding simple explanations for the biological basis of EOCRC by analyses for MSI, CIN, CIMP and somatic mutational patterns, it was reasonable to look for alterations in DNA methylation that might be involved in these cancers.

Epigenetics is the study of heritable changes in gene expression not caused by an alteration in the DNA sequence [42]. This is commonly mediated by DNA methylation or modifications in histone structure. Differentiated cells use only a small proportion of the genes in its genome, and permanent silencing gene expression by methylating CpG (C-phosphodiester-G) sequences in the promoters of genes leads to the formation of heterochromatin in which the promoter is inaccessible to transcription factors. This methylation is stably passed on to the cellular progeny. In addition to using methylation to silence genes not needed as cells enter into differentiated states, cells also use methylation to silence noncoding DNA, which makes up perhaps 98.5% of the human genome [43]. A considerable portion of the noncoding DNA is made up of tandem repeats including long terminal repeats (LTRs), which make up $\sim 45\%$ of the human genome, long and short interspersed nuclear elements (LINEs and SINEs), which make up another $\sim 25\%$ and shorter tandem repeats such as satellites and microsatellites (the targets of MSI) [44–46].

The most prevalent of these repetitive sequences are *LINE-1*, which make up $\sim 18\%$ of the human genome [47,48]. *LINE-1* sequences are retrotransposons and mobile elements that, if expressed, permit our genome to rearrange itself [49]. *LINE-1* is comprised of a UTR, two open reading frames, and a 3'-UTR. The 5'-UTR contains a CpG island (like many promoters), and this is usually heavily methylated in normal cells, preventing transcription of the DNA. The frequency of *LINE-1* methylation in human DNA makes it a useful surrogate for estimating global hypomethylation, particularly in cancer.

In 2010, a multi-center study by Goel *et al.* used *LINE-1* methylation initially to study familial clusters of CRC that were not Lynch syndrome (i.e., the CRCs had no MSI). These familial clusters of CRC met the Amsterdam Criteria for familiarity [18], but when Lynch syndrome could not be documented, the families were called 'familial CRC Type X' [50]. These non-Lynch syndrome familial clusters of CRC may be pertinent to the EOCRC problem, as the age of onset of cancer is younger in this group than in totally sporadic CRC.

This group identified that the CRCs in 22 Type X cases had significantly lower levels of *LINE-1* methylation (60%) than 21 tumors from Lynch syndrome patients (66%, $p = 0.015$), 92 sporadic tumors with MSI (67%, $p = 0.001$) and 46 sporadic MSS CRCs (65%, $p = 0.009$) [51]. Global hypomethylation was confirmed using a panel of 5 'local' methylation markers, and the tumors were considered to have a low methylation index (MI) if 0–3 of the local CIMP markers were methylated and a high MI if 4–5 markers were methylated. All 22 (100%) of the Type X CRC cases had a low MI, versus 96% of the sporadic MSS tumors, 91% of Lynch syndrome tumors, and 33% of the sporadic MSI tumors [51]. The MI in the sporadic MSI CRCs was anticipated since most of these come from a CIMP background [52,53], but the level of hypomethylation in the non-Lynch syndrome familial tumors was unanticipated, and raised the possibility of epigenetic factors in certain subgroups of CRC.

In 2012, this same multi-center collaborative group looked at *LINE-1* hypomethylation in EOCRC tumors [54]. Two independent cohorts totaling 188 cases of EOCRC (≤ 50 years old), 89 cases of sporadic MSS later-onset CRCs (> 50 years old), 46 sporadic MSI CRCs, 20 Lynch syndrome CRCs, and 20 samples of normal colonic

tissues were studied for MMR protein expression and MSI, *LINE-1* methylation, and other genetic analyses. Both cohorts of EOCRC showed significantly lower levels of *LINE-1* methylation compared with the other three cohorts and the normal colonic DNA (which had the highest level of *LINE-1* methylation ($p < 0.001$)). Survival was significantly worse in the groups with *LINE-1* methylation $<65\%$ ($p = 0.039$) (Figure 2).

LINE-1 hypomethylation is a feature of many types of cancer, and can be potentially dangerous because expression of otherwise silenced genes contained within these *LINE-1* elements could lead to genomic instability [49]. The 5'UTR of *LINE-1* has two promoter elements that are aligned in opposite directions. One of these is the sense promoter that regulates expression of the retrotransposon-related genes including an endonuclease, a reverse transcriptase and an RNA-binding protein. The antisense promoter aligned in the opposite direction, however, creates an alternative start site for transcription, and an analysis of the human genome suggests that over 275,000 alternative transcription start sites are involved in *LINE-1* sequences [55]. Therefore, maintaining the methylated states of these promoters is critical.

A study by Hur *et al.* in 2014 examined human tumors and cultured CRC cells, and explored the genetic consequences of *LINE-1* hypomethylation in CRC [56]. The group examined 77 CRCs that had matched samples of normal colonic mucosa and liver metastases, again looking at *LINE-1* methylation as a surrogate for global methylation status. *LINE-1* methylation was significantly lower in the liver metastases than in the primary tumor (64 vs 66%, $p < 0.001$). In addition, the expression of three oncogenes known to be embedded in *LINE-1* intronic sequences were studied at the RNA and protein level. There was significant hypomethylation of the promoters associated with the oncogenes *MET* (<0.001), *RAB3IP* ($p = 0.05$) and *CHRM3* ($p = 0.01$) in the metastatic tumor tissues compared with the matching primary CRC. *LINE-1* methylation had a significant inverse correlation with expression of *MET*, *RAB3IP* and *CHRM3* oncogenes at the mRNA level, and was inversely related to the level of local promoter methylation. Immunohistochemistry documented expression of *MET* at the protein level in the metastatic tissues. Among the 72 matched pairs of primary and metastatic tumors, 22 (31%) showed overexpression of *MET* protein in the liver metastasis, 44 (61%) showed expression in both the primary and metastatic site and in 6 (8%) there was lower expression of *MET* in the metastasis compared with the primary site.

The MSS CRC cell lines CaCo2, SW480, and SW620 all demonstrated significant global *LINE-1* hypomethylation compared with CRC cell lines with MSI (HCT116, RKO and SW48). Additionally, the CRC cell line SW620 had been derived from a lymph node metastasis in the patient whose primary tumor gave rise to the CRC line SW480. *LINE-1* methylation was significantly less in the metastatic cell line than in the line developed from the primary tumor. Additionally, local methylation of the *MET* promoter in the MSS cell lines was very low ($\sim 2\%$) compared with that in the MSI CRC cell lines HCT116 (82%), RKO (83%) and SW48 (32%), all highly significant.

These studies suggest that a subset of CRCs shows a significant degree of global hypomethylation, which is linked to hypomethylation of the promoters of three oncogenes, and their expression in tumor tissues. *LINE-1* hypomethylation is significantly more apparent in familial clusters of non-Lynch syndrome CRC (Type X) [51], cohorts of EOCRC [54], is associated with a higher CRC mortality [54], and is associated with the inappropriate expression of three oncogenes embedded in these elements, *MET*, *RAB3IP* and *CHMR3* [56]. It has also been reported that *LINE-1* hypomethylation is more pronounced in adenomatous polyps associated with a synchronous CRC [57].

Finally, *de novo* constitutional *MLH1* epimutations have been described rarely in patients with EOCRC [58]. In these instances, neither parent shows methylation of *MLH1*, but this epimutation is identified in all tissues of the affected patient (i.e., blood, hair follicles, cheek swabs), with deletion of the wild-type allele in the tumor, leading to MSI. There has been a report of a cancer family with two linked single nucleotide variants in the *MLH1* gene (one in the promoter, c.-27C>A, and one near the transcription start site (85G>T)). This combination leads to constitutional methylation of *MLH1*, which is erased in spermatozoa, but then recurs in the somatic cells [59].

Use of mutational signatures to determine the mechanistic basis of EOCRC

How will the biological basis of the rising incidence of EOCRC finally be resolved? Two genetic diseases linked to familial CRC were solved by linking a specific mutational signature in the tumor tissues to germline variants in DNA repair processes ultimately identified to be responsible for the disease. The first of these was Lynch syndrome, in which the MSI phenotype was linked to germline variants in DNA MMR genes [7,60]. The second instance was when familial clusters of adenomatous polyps of the colon could not be linked to the expected germline variants in the *APC* gene. Somatic mutational analysis of the polyps revealed that the polyps all contained unique *APC* variants,

Figure 2. LINE-1 hypomethylation in subsets of colorectal cancer. (A) Significant hypomethylation is seen in both cohorts of patients with EOCRC, from Argentina and Spain, compared with methylation in the normal colon, cohorts of older-onset CRC (one with MSI and the other without MSI, or MSS), and tumors from patients with Lynch syndrome [54]. **(B)** Patient survival in CRCs with LINE-1 hypomethylation (lower line) was significantly worse than in those patients who do not have this epigenetic alteration (upper line) [54]. EOCRC: Early-onset colorectal cancer; MSI: Microsatellite instability; MSS: Microsatellite stability.

and there was a recurring signature involving G>T alterations in the mutated genes. This led the investigators in the United Kingdom to suspect involvement of the base excision repair system in this disease, which then led to the discovery of biallelic germline variants in the *MYH* (*MutYH*) gene [61,62].

Subsequently, large numbers of cancers from many sites have been analyzed for mutational signatures. An international consortium of investigators (the Pan-Cancer Analysis of Whole Genomes or PCAWG of the International Cancer Genome Consortium – ICGC), and The Cancer Genome Atlas (TCGA), have sequenced and analyzed signatures from 4,645 whole-genome and 19,184 exome sequences in a wide variety of cancers. Initially, about 30 single base substitution (SBS) signatures were identified [63], but the advancing technologies have identified many more signatures that involve more complex alterations than SBS [30]. Importantly, many of the signatures have not yet been linked to a specific form of genomic instability or environmental insult (although many have) [32].

Conclusion

In conclusion, EOCRC is increasing while later-onset CRC is decreasing in incidence and mortality in the US and much of Europe. Early-onset tumors are most likely to occur in the rectum and rectosigmoid colon. The mechanistic basis for this phenomenon is unknown. Known hereditary gastrointestinal syndromes account for fewer than 20% of these tumors, so other factors are at work. Some of these tumors show hypomethylation of *LINE-1* sequences, and there are some hints in this process that might explain the worse clinical outcome. Since these patients are developing these tumors prior to the usual onset of CRC screening, novel preventive and screening approaches will be necessary.

Current knowledge gaps & future perspective

In spite of recent progress, at this time we know that no more than 20% of EOCRC can be traced to germline mutations, and the rest are not well understood. There is evidence that some of the cases of EOCRC have an epigenetic ‘signature’ (*LINE-1* hypomethylation), which is associated with a worse outcome, but we have no insight into the mechanistic basis of that alteration. None of what we know about the genetic/epigenetic landscape informs us regarding the earlier onset, the rising incidence, nor the clinical features such as the clustering of tumors in the distal colon and rectum.

The next steps in solving the biological basis of EOCRC would be to assemble a large cohort of cases, exclude those with identifiable hereditary syndromes (which only serve to confuse the analysis), and determine how many identifiable mutational signatures are present in the rest of the cancers. There is unlikely to be a sharp change in the mutational signatures in CRC at age 50, but there may be an identifiable change in the frequency of certain signatures that can ultimately be traced to some novel environmental exposure or shift in the gut microbiome. Only when we understand how these tumors form will we be able to contemplate appropriate interventions.

Executive summary

- Early-onset colorectal cancer (EOCRC) is rising in incidence, and there is not currently an explanation for how this has occurred.
- One hypothesis would be that much of EOCRC is due to unsuspected hereditary CRC syndromes, but those make up no more than ~20% of the total.
- One should screen EOCRCs for Lynch syndrome using immunohistochemistry or microsatellite instability testing.
- It is reasonable to perform multigene platform testing in EOCRC patients, whether or not there is a suspicious family history.
- A heterogeneous variety of somatic genetic alterations have been reported, but none is clinically actionable.
- Reduced methylation of *LINE-1* sequences is more common in EOCRC tumors than in older-onset CRCs or in Lynch syndrome tumors.
- Reduced methylation of *LINE-1* sequences is associated with a significant reduction in survival in EOCRC.

Financial & competing interests disclosure

This research was funded by NIH grant number R01 CA072851 (Familial and Early-Onset Colorectal Cancer). CR Boland is a consultant for Ambry Genetics. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Murphy C, Sanford N, Dharwadkar P. Early-onset colorectal cancer: two sides to the story? *Colorectal Cancer* doi:10.2217/crc-2020-0016 (2020).
2. Siegel R. Early-onset colorectal cancer: when hoof beats are zebras. *Colorectal Cancer* doi:10.2217/crc-2020-0013 (2020).
3. Siegel RL, Miller KD, Jemal A. Colorectal cancer mortality rates in adults aged 20 to 54 years in the United States, 1970–2014. *JAMA* 318(6), 572–574 (2017).
- **Best describes the increase in incidence of early-onset colorectal cancer (EOCRC).**
4. Siegel RL, Fedewa SA, Anderson WF *et al.* Colorectal cancer incidence patterns in the United States, 1974–2013. *J. Natl Cancer Inst.* 109(8), 1–6 (2017).
5. Siegel RL, Torre LA, Soerjomataram I *et al.* Global patterns and trends in colorectal cancer incidence in young adults. *Gut* 68(12), 2179–2185 (2019).
6. 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.* 15(5), 728–737 e723 (2017).
7. Aaltonen LA, Peltomäki P, Leach FS *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science* 260(5109), 812–816 (1993).
8. Peltomäki P, Lothe RA, Aaltonen LA *et al.* Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res.* 53(24), 5853–5855 (1993).
9. Liu B, Farrington SM, Petersen GM *et al.* Genetic instability occurs in the majority of young patients with colorectal cancer. *Nat. Med.* 1(4), 348–352 (1995).
10. Capozzi E, Della Puppa L, Fornasari M, Pedroni M, Boiocchi M, Viel A. Evaluation of the replication error phenotype in relation to molecular and clinicopathological features in hereditary and early onset colorectal cancer. *Euro. J. Cancer* 35(2), 289–295 (1999).
11. Gryfe R, Kim H, Hsieh ET *et al.* Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N. Engl. J. Med.* 342(2), 69–77 (2000).
12. Naidoo R, Tarin M, Chetty R. A comparative microsatellite analysis of colorectal cancer in patients <35 years and >50 years of age. *Am. J. Gastroenterol.* 95(11), 3266–3275 (2000).
13. Fernebro E, Halvarsson B, Baldetorp B, Nilbert M. Predominance of CIN versus MSI in the development of rectal cancer at young age. *BMC Cancer* 2, 25 (2002).
14. Losi L, Di Gregorio C, Pedroni M *et al.* Molecular genetic alterations and clinical features in early-onset colorectal carcinomas and their role for the recognition of hereditary cancer syndromes. *Am. J. Gastroenterol.* 100(10), 2280–2287 (2005).
15. Goel A, Nagasaka T, Spiegel J, Meyer R, Lichliter WE, Boland CR. Low frequency of Lynch syndrome among young patients with non-familial colorectal cancer. *Clin. Gastroenterol. Hepatol.* 8(11), 966–971 (2010).
- **Used immunohistochemistry and microsatellite instability and found that only 21% of EOCRC had defective DNA mismatch repair activity.**
16. Giraldez MD, Balaguer F, Bujanda L *et al.* MSH6 and MUTYH deficiency is a frequent event in early-onset colorectal cancer. *Clin. Cancer Res.* 16(22), 5402–5413 (2010).
- **The work was done collaboratively with reference 16.**
17. Chang DT, Pai RK, Rybicki LA *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.* 25(8), 1128–1139 (2012).
18. Boland CR, Thibodeau SN, Hamilton SR *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 58(22), 5248–5257 (1998).
19. Kane MF, Loda M, Gaida GM *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* 57(5), 808–811 (1997).
20. Haraldsdóttir S, Hampel H, Tomsic J *et al.* Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. *Gastroenterol.* 147(6), 1308–1316.e1301 (2014).
21. Limburg PJ, Harmsen WS, Chen HH *et al.* Prevalence of alterations in DNA mismatch repair genes in patients with young-onset colorectal cancer. *Clin. Gastroenterol. Hepatol.* 9(6), 497–502 (2011).
22. Mork ME, You YN, Ying J *et al.* High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. *L. Clin. Oncol.* 33(31), 3544–3549 (2015).

23. Yurgelun MB, Kulke MH, Fuchs CS *et al.* Cancer susceptibility gene mutations in individuals with colorectal cancer. *J. Clin. Oncol.* 35(10), 1086–1095 (2017).
- **This paper studied a large cohort of unselected, consecutive CRC patients and analyzed the germline for mutations in a panel of 25 known familial cancer genes.**
24. Pearlman R, Frankel WL, Swanson B *et al.* Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol.* 3(4), 464–471 (2017).
- **Studied a large number of EOCRC patients with the same gene panel used in reference 22, and confirmed the findings.**
25. Stoffel E.M, Koeppe E, Everett J *et al.* Germline Genetic Features of Young Individuals with Colorectal Cancer. *Gastroenterology* 154(4), 897–905 (2018).
26. Toh MR, Chiang JB, Chong ST *et al.* Germline pathogenic variants in homologous recombination and DNA repair genes in an Asian cohort of young-onset colorectal cancer. *JNCI Cancer Spectr.* 2(4), pky054 (2018).
27. Mensenkamp AR, Vogelaar IP, Van Zelst-Stams WA *et al.* Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. *Gastroenterology* 146(3), 643–646.e648 (2014).
28. De Voer RM, Hahn MM, Weren RD *et al.* Identification of novel candidate genes for early-onset colorectal cancer susceptibility. *PLoS Genet.* 12(2), e1005880 (2016).
29. Jansen AML, Ghosh P, Dakal TC, Slavin TP, Boland CR, Goel A. Novel candidates in early-onset familial colorectal cancer. *Familial Cancer* 19(1), 1–10 (2020).
30. Alexandrov LB, Kim J, Haradhvala NJ *et al.* The repertoire of mutational signatures in human cancer. *Nature* 578(7793), 94–101 (2020).
- **Reports the development of mutational signatures that are present in cancers, and which can be used to trace back to a possible environmental, microbiological or germline abnormality in any group of cancers.**
31. Gerstung M, Jolly C, Leshchiner I *et al.* The evolutionary history of 2,658 cancers. *Nature* 578(7793), 122–128 (2020).
32. Pan-cancer analysis of whole genomes. *Nature* 578(7793), 82–93 (2020).
33. Hong Y, Ho KS, Eu KW, Cheah PY. A susceptibility gene set for early onset colorectal cancer that integrates diverse signaling pathways: implication for tumorigenesis. *Clinical Cancer Res.* 13(4), 1107–1114 (2007).
34. Kirzin S, Marisa L, Guimbaud R *et al.* Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS ONE* 9(8), e103159 (2014).
35. Perea J, Garcia JL, Perez J *et al.* NOMO-1 gene is deleted in early-onset colorectal cancer. *Oncotarget* 8(15), 24429–24436 (2017).
36. Alvaro E, Cano JM, Garcia JL *et al.* Clinical and molecular comparative study of colorectal cancer based on age-of-onset and tumor location: two main criteria for subclassifying colorectal cancer. *Int. J. Mol. Sci.* 20(4), 968 (2019).
37. Lieu CH, Golemis EA, Serebriiskii IG *et al.* Comprehensive genomic landscapes in early and later onset colorectal cancer. *Clin. Cancer Res.* 25(19), 5852–5858 (2019).
38. Chan TL, Curtis LC, Leung SY *et al.* Early-onset colorectal cancer with stable microsatellite DNA and near-diploid chromosomes. *Oncogene* 20(35), 4871–4876 (2001).
39. Boardman LA, Johnson RA, Petersen GM *et al.* Higher frequency of diploidy in young-onset microsatellite-stable colorectal cancer. *Clin. Cancer Res.* 13(8), 2323–2328 (2007).
40. Arriba M, Garcia JL, Inglada-Perez L *et al.* DNA copy number profiling reveals different patterns of chromosomal instability within colorectal cancer according to the age of onset. *Mol. Carcinog.* 55(5), 705–716 (2016).
41. Arriba M, Garcia JL, Rueda D *et al.* Unsupervised analysis of array comparative genomic hybridization data from early-onset colorectal cancer reveals equivalence with molecular classification and phenotypes. *Neoplasia* 19(1), 28–34 (2017).
42. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. *N. Engl. J. Med.* 378(14), 1323–1334 (2018).
43. Boland CR. Non-coding RNA: it's not junk. *Dig. Dis. Sci.* 62, 1107–1109 (2017).
44. Casa V, Gabellini D. A repetitive elements perspective in polycomb epigenetics. *Front. Genet.* 3, 199 (2012).
45. Lander ES. Initial impact of the sequencing of the human genome. *Nature* 470(7333), 187–197 (2011).
46. Birney E, Stamatoyannopoulos JA, Dutta A *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146), 799–816 (2007).
47. Hohjoh H, Singer MF. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* 16(19), 6034–6043 (1997).
48. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87(5), 905–916 (1996).
49. Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science* 303(5664), 1626–1632 (2004).
50. Lindor NM, Rabe K, Petersen GM *et al.* Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA* 293(16), 1979–1985 (2005).

51. Goel A, Xicola RM, Nguyen TP *et al.* Aberrant DNA methylation in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterol.* 138(5), 1854–1862 (2010).
52. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA* 96(15), 8681–8686 (1999).
53. Goel A, Boland CR. Epigenetics of colorectal cancer. *Gastroenterology* 143(6), 1442–1460 e1441 (2012).
54. Antelo M, Balaguer F, Shia J *et al.* A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS ONE* 7(9), e45357 (2012).
- **Describes the best characterized epigenetic alteration in EOCRCs.**
55. Faulkner GJ, Kimura Y, Daub CO *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41(5), 563–571 (2009).
56. Hur K, Cejas P, Feliu J *et al.* Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* 63(4), 635–646 (2014).
57. Jiang AC, Buckingham L, Barbanera W, Korang AY, Bishesari F, Melson J. LINE-1 is preferentially hypomethylated within adenomatous polyps in the presence of synchronous colorectal cancer. *Clin. Epigenetics* 9, 25 (2017).
58. Goel A, Nguyen TP, Leung HC *et al.* *De novo* constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int. J. Cancer* 128(4), 869–878 (2011).
59. Hitchins MP, Rapkins RW, Kwok CT *et al.* Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. *Cancer Cell* 20(2), 200–213 (2011).
60. Peltomäki P, Aaltonen LA, Sistonen P *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260(5109), 810–812 (1993).
61. Sieber OM, Lipton L, Crabtree M *et al.* Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N. Eng. J. Med.* 348(9), 791–799 (2003).
62. Lipton L, Halford SE, Johnson V *et al.* Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res.* 63(22), 7595–7599 (2003).
63. Alexandrov LB, Nik-Zainal S, Wedge DC *et al.* Signatures of mutational processes in human cancer. *Nature* 500(7463), 415–421 (2013).