Colorectal cancer carcinogenesis: a review of mechanisms

Kanwal Tariq, Kulsoom Ghias
Department of Biological and Biomedical Sciences, Aga Khan University, Karachi 74800, Pakistan

ABSTRACT
Colorectal cancer (CRC) is the second most common cancer in women and the third most common in men globally. CRC arises from one or a combination of chromosomal instability, CpG island methylator phenotype, and microsatellite instability. Genetic instability is usually caused by aneuploidy and loss of heterozygosity. Mutations in the tumor suppressor or cell cycle genes may also lead to cellular transformation. Similarly, epigenetic and/or genetic alterations resulting in impaired cellular pathways, such as DNA repair mechanism, may lead to microsatellite instability and mutator phenotype. Non-coding RNAs, more importantly microRNAs and long non-coding RNAs have also been implicated at various CRC stages. Understanding the specific mechanisms of tumorigenesis and the underlying genetic and epigenetic traits is critical in comprehending the disease phenotype. This paper reviews these mechanisms along with the roles of various non-coding RNAs in CRCs.

KEYWORDS
Colorectal cancer; chromosomal instability; microsatellite instability; non-coding RNA mismatch repair

Introduction
Globally, colorectal cancer (CRC) is the second most common cancer in women (614,000 cases per year) and the third most common in men (746,000 cases per year). The incidence rates are higher in more developed countries (737,000 cases per year) than in less developed ones (624,000 cases per year). However, mortality is higher in the latter (52% of total deaths), which indicates poor survival. In 2015, the GLOBOCAN online analysis tool has predicted 61,228 new CRC cases for Asia. Accordingly, 25,816 of these cases are associated with people who are less than 65 years old.

Mechanisms of carcinogenesis
CRCs can arise from one or a combination of three different mechanisms, namely chromosomal instability (CIN), CpG island methylator phenotype (CIMP), and microsatellite instability (MSI). According to Fearon, the classical CIN pathway begins with the acquisition of mutations in the adenomatous polyposis coli (APC), followed by the mutational activation of oncogene KRAS and the inactivation of the tumor suppressor gene, TP53. Aneuploidy and loss of heterozygosity (LOH) are the major players in CIN tumors, which not only constitute most of the sporadic tumors (85%) but also involve familial adenomatous polyposis cases associated with germline mutations in the APC gene. The CIMP pathway is characterized by promoter hypermethylation of various tumor suppressor genes, most importantly MGMT and MLH1. This hypermethylation is often associated with BRAF mutation and microsatellite instability. The MSI pathway involves the inactivation of genetic alterations in short repeated sequences. This activation occurs in CRCs in DNA mismatch repair (MMR) genes, and is a hallmark condition in familial Lynch syndrome (LS), which also appears in ~15% of the sporadic CRC cases. In addition, the hypermethylation of the MMR genes may lead to MSI. This mechanism is often associated with the CIMP pathway. MSI tumors are often associated with proximal colon and poor differentiation but better prognosis. Moreover, the three mechanisms often overlap in CRC. Figure 1 illustrates important molecular, genetic, and epigenetic changes with respect to disease progression. A detailed account of which is provided in the sections that follow.

Chromosomal instability
Chromosomal instability is associated with 65%-70% of sporadic CRCs. This pathway comprises aneuploidy, which is an imbalance in the chromosome number, and LOH. Defects in chromosomal segregation, DNA damage repair,
and telomere function along with specific mutations in certain oncogenes and tumor suppressor genes may be responsible for such instability. The mechanisms underlying genetic instability and other causative mutations in certain genes are reviewed elsewhere. This review discusses updated findings relevant to these mechanisms.

Aneuploidy arises because of defects in the mitotic checkpoint, which cause chromosome mis-segregation. Mutational upregulation or downregulation of various mitotic checkpoint players, such as hRod, hZwilch, hZw10, Ding, budding uninhibited by benzimidazoles (Bub) R1, centromere-associated protein E (CENP-E), and mitotic arrest deficient (MAD) 1, can result in CIN. A deleterious increase in aneuploidy may also lead to cell death. Other centromere protein genes, such as CENP-A and CENP-H, may also be overexpressed in CRCs, which leads to mislocalizations on non-centromeric chromatin positions.

The centrosome-associated kinases, Aurora (AURK) and Polo-like (Plk), are also implicated in CRCs. AURK overexpression causes arrested mitosis and multi-nucleation. This kinase also has a positive association with the degree of instability in CRCs and activates Plk1 overexpressed in 73% of primary CRCs and associated with poor prognosis. Similarly, Aurora-B regulates histone modification, chromosome segregation, and cytokinesis. Aurora-B is also overexpressed and associated with advanced stage tumors. The overexpression of Aurora-C prevents the activation of the mitotic checkpoint leading to tumorigenesis.

Disruptions in the DNA damage pathway and excess telomere breakage can lead to chromosomal instability. The damaged DNA is repaired by four mechanisms, namely base excision repair (BER), double-stranded break repair (DSBR), mismatch repair (MMR), and nucleotide excision repair (NER). The polymorphisms in BER-associated genes XRCC1, OGG1, and MUTYH have been correlated with reduced oxidative DNA damage repair efficiency in CRC. Similarly, the MRE11 gene involved with the DSBR contributes to genomic instability. DNA damage can activate two different signaling pathways, namely ataxia telangiectasia mutated-checkpoint kinase 2 (ATM-Chk2) and ataxia telangiectasia and Rad3-related-checkpoint kinase 1 (ATR-Chk1). Inactivating mutations in Chk2 along with AURKA

![Diagram](image-url)
overexpression results in increased microtubule assembly rates as noted in 73% of CRCs. Excess telomere breakage is responsible for the transition from adenoma to carcinoma in CRCs. A comparison with adjacent normal tissues has also shown that shorter chromosomes are found in both adenomatous polyps and carcinomas. These results imply that telomere shortening initially causes instability. However, mutations in the APC gene and an increase in the telomerase activity drive metastasis.

Accordingly, LOH also leads to CIN. LOH is the loss of one of the parental alleles, which is caused by mitotic nondisjunction, recombination between two homologous or non-homologous chromosomes, or a deletion or gene conversion event. In CRCs, LOH is frequently observed in at least five different loci (i.e., 1, 5, 8, 17, and 18). Previous studies have linked various CRC forms with losses at 1p, 4q, 5q, 8p, 9q, 11q, 14q, 15q, 17p, 18p, and 18q. These findings are reviewed elsewhere. Mutations are also observed in these locations. Inactivating the mutations in the tumor suppressor APC gene (5q21) is considered to be a carcinogenesis-initiating event. This process is usually followed by the activation of KRAS mutations (12p12). This step only promotes tumor progression in combination with the APC mutations. Mutations in other genes, such as TP53 (17q13), PIK3CA (3q26), and TGF-β (3p22), are also acquired. These carcinogenesis-promoting mutations are further described in the sections that follow.

Mutational landscape in chromosomal instability

Mutations in the APC activate the Wnt signaling pathway by increasing β-catenin levels. β-catenin is translocated to the nucleus and enhances the transcription of various oncogenes with T-cell factor (TCF) transcription factors. High β-catenin levels are noted in gastrointestinal tumors.

APC inactivation may occur through germline and somatic mutations or the hypermethylation of its promoter. Germline mutations occur in a small fraction of cases. Somatic mutations occur in 72% of sporadic tumors, 30%-70% of sporadic adenomas, and 5% of dysplastic aberrant crypt foci (ACF). Hypermethylation-led inactivation occurs in about 12% of the primary carcinomas and adenomas.

Around 75% of CRCs have mutations or LOH in the APC gene. Most of these mutations are clustered in the mutation cluster region between codons 1282 and 1581. However, a complete inactivation is not required. Mutations sufficient for tumorigenesis differ between the proximal and distal CRCs. The effects of APC restoration in mice are demonstrated on tumor regression by the conversion of cancer cells back to normal, which indicates a similar possibility in humans. Mutations in other genes of this pathway, particularly in β-catenin, may also lead to CIN. These mutations are found in 48% of CRCs without APC mutations. β-catenin activates a set of 162 Wnt pathway target genes in a colon cancer cell line. However, no conclusions could be drawn for their effect on disease prognosis.

Point mutations in codons 12, 13, and 61 of exons 2 and 3 of KRAS activate the enzyme increasing RAS signaling. This enzyme is mutated in 30%-40% of CRCs and 60%-90% of hyperplastic or non-dysplastic ACF. Mutations in KRAS are not precursory unlike APC mutations. Mutations in codon 12 have higher chances for lymph node metastasis associated with the advanced form of the disease. The activated RAS further activates the Raf-MEK-ERK and PI3K/AKT/PKB pathways or Ras small GTPases. PI3K is first activated by Ras and then by the receptor tyrosine kinases as inactive in the inactivation of PI3K by KRAS knockdown. However, most CRC cases with PI3K mutations also carry mutations in KRAS. The activated PI3K further activates AKT1 and AKT2, which then enhances tumor growth by promoting epithelial to mesenchymal transition (EMT). Loss-of-function mutations in PTEN, which is a tumor suppressor and an antagonist of the PI3K/AKT pathway, induce AKT-regulated metastasis in CRCs. PI3K/AKT and PI3K/AKT pathways often converge to activate a cap-dependent translation through survivin knockdown, which can inhibit metastasis. Therefore, targeting both pathways together is more clinically effective.

The tumor suppressor transcription factor 53 (TP53) is located on chromosome 17, which is activated under stress. TP53 targets cell cycle inhibitors such as GADD45 and 14-3-3 and pro-apoptotic factors including BAX, KILLER (DR5), FAS (APO1), and PIG3. However, p53 is dysfunctional in the majority of human tumors. Certain tumors show a gain-of-function mutation in p53, which results in mutated proteins, notably the mutp53 isoform. This isoform causes chronic transcription factor NF-κB activation in mice models, which enhances inflammation and accelerates tumorigenesis that finally result in invasive carcinoma. This activation is independent of the loss of wild-type p53. Mutp53 may also inactive the tumor suppressor RasGAP and disabled 2 interacting protein, which makes the cancer cells more responsive toward inflammatory cytokines. Paradoxically, this increases the cancer cells' invasive behavior but reduces their aggressiveness.
transactivation domain of p53 under normal conditions is bound and ubiquitinated by MDM2, E3-ubiquitin ligase, and MDM4. MDM2 can also bind and inactivate the mutp53 isoform. A comparison with the adjacent normal tissues shows that mRNA and protein overexpression of the spliced isoform MDM2-B is observed in CRCs. This overexpression is mainly correlated with the mutp53 protein as MDM2-B binds to MDM2, thereby allowing mutp53 accumulation in CRCs. The loss-of-function of p53 is reported in 44.9% of colorectal adenoma and 42.22% of single and 43.75% of multiple primary CRCs. The downregulation of p53 via IL-6-dependent rRNA transcription enhancement leads to the development of EMT specific changes that are also seen in ulcerative colitis patients.

Cyclooxygenase-2 (COX-2) is overexpressed in 43% of adenomas and 86% of carcinomas. The deletion of a single allele causes 34% reduction, whereas a complete knockdown causes 86% reduction in the number of intestinal polyps in APC knockout mice. Increased levels of both COX-2 and prostaglandin E2 (PGE2), which is the enzymatic product of COX-2, are found in CRCs. Another study found 51 SNPs in HPGD, SLCO2A, and ABCC4 aside from Cox-2, which code for 15-PGDH (COX-2 antagonist), MRP4 (Cox-2 transporter), and PGT (Cox-2 transporter), respectively. Seven of these SNPs are important. Four are validated for gene-gene interaction.

The loss of 18q chromosomal region has been correlated with poorer survival rate in stage II CRCs. The region carrying the deleted in colorectal cancer gene (a dependence factor) is more frequently lost in advanced cancers. The mutations in similar to mothers against decapentaplegic (SMAD) homologproteins, namely SMAD2, SMAD3, SMAD4 and SMAD7, which are regulators of TGFβ signaling, have been associated with CRCs.

In 1997, the National Cancer Institute held a workshop where a five-marker MSI panel was validated. This panel included two mononucleotide markers, namely BAT25 and BAT26, and three dinucleotide markers, namely D5S346, D2S123, and D17S250. Tumors with instability in ≥30% of markers are called MSI-high (MSI-H). Those with instability in <30% are called MSI-low (MSI-L), while those without microsatellite instability are called MSI stable (MSS). Mutations in MLH1, MSH2, MSH6, and PMS2 have been associated with a risk of developing LS. By the age of 70, an estimated risk for developing LS of 38% and 31% is observed in males and females, respectively. An additional cumulative risk of 41%, 48%, and 12% for mutated hMLH1, MSH2, and MSH6, respectively, have been found.

More than 1,500 germline variants have been found in the MMR genes along with promoter methylation, somatic deletions, or point mutations. The germline hypermutations in MLH1 and MSH2 may also increase susceptibility. Furthermore, deletions in EPCAM, which is a gene upstream of MSH2, encompass the termination signal, thereby reading through MSH2 and silencing it in colonic tissues. Moreover, a correlation between MSH2 promoter methylation and EPCAM deletion is observed in both CRC tissues and adjacent normal ones. Large variations are also observed in MLH1 and MSH2 genes although they are not detected by conventional exon-specific methods such as screening. Deletions have been confirmed in 27% of patients; 12% of these deletions belong to those lacking MLH1 and 56% to those lacking the MSH2 expression. A 10Mb inversion has been identified in MSH2 in samples lacking the MSH2 protein. Other related genes must be tested in a similar manner.

The somatic mutations in the MMR genes and in EPCAM, POLE, and POLD1 have also been associated with CRCs. A recent study has found somatic mutations in at least two of these genes in almost 70% of patient tumors. In another study, two acquired mutations in MLH1 and MSH2 have been observed in 52% (13/25) of the patients studied.

The hypermethylation of the MLH1 promoter in MSI-H sporadic CRCs is found in 83%-100% of tumors. The same finding is obtained in 15% of LS cases, which makes differentiation from sporadic CRCs difficult. In such cases, the V600E mutation in v-Raf murine sarcoma viral oncogene homolog B1 (BRAF) gene eliminates the possibility of LS.

### Microsatellite instability

Microsatellite instability occurs because of inactivating mutations in the DNA mismatch repair genes that are responsible for correcting DNA replication errors. The important components of the DNA mismatch repair system are ATPases hMSH2, hMSH6, hMSH3, hMLH1, hPMS2, hPMS1, and hMLH3. The germline mutations that may render these proteins dysfunctional can predispose to cancer as in the case of LS. MSI is found in 15% of colorectal cancers, with only 3% associated with LS. The rest are sporadic cases caused by the hypermethylation of the MLH1 gene promoter.

### CpG island methylator phenotype

Global DNA hypomethylation and localized promoter hypermethylation are common epigenetic events that occur
in cancer. Hypomethylation refers to a marked global decrease in methylation on cytosine bases that is observed in hyperplastic and adenomatous polyps and carcinomas. Hypomethylation in the repetitive DNA sequences, such as in satellite regions, can lead to genomic instability. Furthermore, loss of imprinting or promoter demethylation could reactivate the retrotransposons. The demethylation of the long interspersed nuclear element-promoter has been suggested as an early event. However, demethylation has also been observed in the normal colonic mucosa of the same patients. The loss of imprinting of insulin-like growth factor 2 (IGF2) is seen in almost 40% of CRC tumors, which leads to microsatellite instability in younger patients.

The CpG Island (CGI) hypermethylation in the promoter region results in the transcriptional inactivation of genes that have tumor suppressive roles. Mutations in the BRAF gene are considered to be an early event in the CIMP tumors. The BRAF V600E mutation is strongly correlated with MLH1 hypermethylation and has a frequency of 20.3% in unselected and 18.7% in sporadic cases. Another study has confirmed MLH1 hypermethylation in 80% of MSI-H sporadic CRC, with loss of expression and without a known germline mutation in the MMR genes.

The hypermethylation of other gene promoters (e.g., CDKN2A) has led to the development of a commonly used five-marker gene panel. This panel includes MLH1, CDKN2A, methylated in tumors 1 (MINT1), MINT2, and MINT31 and has been extensively studied. This classic panel is used to term tumors as CIMP-high (CIMP-H) if two or more of these promoters are hypermethylated. A new panel has also been designed from 195 different CIMP markers comprising CACNA1G, IGF2, SOCS1, RUNX3, and NEUROG1. Using which, CIMP positive tumors are defined as those with three or more of these panel gene promoters methylated. Studies comparing the two panels show various results. The new panel accurately predicts CIMP1 (positive) but not CIMP2 (CIMP low) tumors. Another study has concluded that the new panel underperformed in determining the clinicopathological features of the tumors. Accordingly, the classic panel could better predict the clinical outcomes. However, another study has tested eight markers. Five of these markers belong in the old panel. MLH1, CDKN2A, and CRABP1 have been added. Tumors with six or more methylated promoters are identified as CIMP-H. MLH1, RUNX3, IGF2, and CACNA1G have been considered to be the most specific and sensitive out of these eight markers. This set creates yet another panel.

A recent study has also combined markers with different functions and analyzed their methylator phenotype including DNA repair gene MGMT, tumor suppressors (i.e., CDKN2A, HHTF, GATA5, ID4, and TSLC1), metastasis suppressors (i.e., CDH4, CDH13, and TIMP3), apoptosis-related genes (i.e., CACNA1G, HRK, and RSASF1A), and angiogenesis inhibitors (i.e., TSP1). This study classifies the carcinomas as CIMP-H, CIMP-low (CIMP-L), and CIMP-normal (CIMP-N) if more than six, less than six, or none of the genes are methylated, respectively.

Hypemerthylated promoters are associated with BRAF mutations. A total of 759 hypomethylated regions are found. Accordingly, 96% of these regions occur in BRAF mutant tumors. Out of these, 229 regions are localized in the promoter regions enhancing five different pathways, namely the Wnt signaling, hedgehog signaling, bZIP transcription regulation, PI3 kinase, and IGF-protein kinase B signaling pathways. Early Wnt signaling activation has been attributed to APC mutations in CIN tumors. However, promoter hypermethylation of the Wnt antagonists suggests a role in the Wnt activation at later stages. Hypermethylation is noted in seven gene promoters in the normal to adenoma transition, and in four of these seven genes from adenoma to carcinoma.

Similarly, ten eleven translocation 1 (TET1) methylation is an early event linked to BRAF mutations in CIMP+ tumors and polyps. The TET family proteins regulate demethylation by catalyzing the conversion of 5-methylcytosine to 5-hydroxymethylcytosine. Consequently, the inactivation results in hypermethylation leading toward CIMP.

The aberrations in chromatin-remodeling genes, such as ATP-dependent chromatin remodelers, chromodomain helicase 7 (CHD7) and CHD8, may also be associated with CIMP tumors. These mutations may lead to chromatin structure modification and deregulation, which contributes to CIMP.

Clinical implications of the disease molecular mechanisms in CRC

CIN, MSI, and CIMP often overlap in molecular tumor subtypes, which have considerable prognostic implications. In one study, tumors were initially classified as MSI or MSS tumors based on 509 CRC cases. The latter group was further classified as CIN-only, CIMP-only, CIN+CIMP, and triple negatives. As expected, MSI tumors had the lowest frequency for APC and KRAS mutations, the second lowest for p53 mutations, and the highest for BRAF V600E mutations. By contrast, CIN-only tumors showed the
The highest frequency for p53 mutations and the lowest for BRAF V600E mutations.

The patient survival outcome was recently associated with specific molecular subtypes classified based on the MSI, CIMP, BRAF-mutation, and KRAS-mutation status. The MSI-H tumors (types 1 and 5) had the highest five-year disease specific survival (89.5% and 93.1%), followed by the MSI-L/MSS tumors (type 4; 82.5%) without CIMP or BRAF and KRAS mutations, and the tumors with only KRAS mutations (type 3; 72.4%). The tumors with CIMP and BRAF mutations had the worst survival (type 2; 49.2%)127. CIMP was associated in another study with poor disease free survival and overall survival rate irrespective of the MSI status128. Another six-subtype classification (C1-C6) was proposed based on the mutational landscape. The C1, C5, and C6 tumors frequently had chromosomal instability, TP53 mutations, and were distally located lacking the mutator phenotype. The other three subtypes had proximally located tumors often associated with CIMP. Furthermore, C2 had BRAF mutations and deficient MMR, while C3 had KRAS mutations129.

Classifying tumors according to the particular mutations present is necessary in determining the treatment regimen to be offered. A combination of either 5-fluorouracil (5-FU) or capecitabine is usually offered with irinotecan or oxaliplatin as the main line of treatment. Targeted therapies such as vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR) inhibitors have recently been found effective as both first and second lines of treatment in CRC. However, evidence suggests that anti-EGFR treatment is effective only in tumors lacking codon 12 and 13 mutations in KRAS. The treatment may even be detrimental when used with oxaliplatin in patients with KRAS-mutated tumors as shown in clinical trials130–134. Accordingly, anti-EGFR inhibitors are contraindicated in patients with mutant KRAS metastatic disease as per the guidelines of the American Society for Clinical Oncology135. The BRAF V600E mutation correlated with poor disease specific survival also confers resistance to anti-EGFR therapies even in the presence of wild-type KRAS127,136,137. Similarly, the presence of microsatellite instability has been shown to be predictive of failure of the standard, first-line 5-fluorouracil treatment138. Other genetic and epigenetic biomarkers also have significant implications for CRC diagnosis and treatment. These biomarkers have been previously described in detail in the literature139.

Non-coding RNAs in colorectal cancer

Only 2% of the human genome comprises protein coding genes. The remaining 98% region is transcribed into non-protein-coding RNAs. A variety of RNAs such as long non-coding RNA (lncRNA), microRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA) are present in eukaryotes, which are mainly classified depending on their sizes and functions.

MicroRNAs

MicroRNAs are small (20-25 nucleotides), single-stranded, non-coding RNAs that negatively regulate gene expression140. They exert action by binding to the 3'UTR sequence of a target mRNA transcript. Binding is done by base-pairing to a partially complementary region, which leads to either transcript degradation or transcriptional inhibition140,141. MicroRNAs can act either as tumor-suppressing miRNAs inactivating the oncogenes or as oncogenic miRNAs (oncomiRs) inactivating the tumor suppressor genes in cancer142,143. MicroRNA expression is frequently altered in cancers as they lie in unstable genomic regions. The miRNAs shown to be particularly important in CRCs are discussed in the sections that follow.

MiR-143 and miR-145

MiR-143 and miR-145 are concomitantly downregulated in most CRCs as both are found on 5q33. Their downregulation is an early event and occurs even before the APC gene aberrations144. The two miRNAs function as oncosuppressors in a coordinated fashion, that is, they either target the same genes or different genes regulating the same pathway145. MiR-143 expression is inversely related to KRAS expression and has inhibitory effects on KRAS in Lovo cells146,147. IGF-I receptor is another miR-143 target, and the overexpression of which inhibits proliferation, angiogenesis, and migration through the PI3K/AKT/HIF-1/VEGF pathway. MiR-143 overexpression also increases sensitivity to the chemotherapeutic drug, oxaliplatin148. MiR-143 and -145 also target cluster of differentiation 44 (CD44), Kruppel-like factor 5 (KLF5), and BRAF.

Lethal-7 (let-7) family and miR-18a*

Lethal-7 (let-7) family and miR-18a* have tumor suppressive roles influencing the RAS pathway. The let-7 expression is negatively correlated with RAS and c-myc expression149 and hypothesized to regulate the KRAS expression through TP53.
However, no such association has been proven\(^{150}\). An increase in the let-7i expression is observed in relation to polyamine depletion, suggesting a role of polyamines in regulating the let-7 family. The high mobility group A2 (\(HMGA2\)), which is another let-7 target gene, is downregulated in a polyamine-dependent manner\(^{151}\). MiR-18a\(^*\) also negatively regulates the KRAS gene. The KRAS-dependent increase in cell proliferation and anchorage dependence is observed in colon adenocarcinoma HT-29 cells transfected with anti-miR-18a\(^*\)\(^{152}\).

**MiR-200 family**

The miR-200 family has a metastasis inhibitory role via the regulation of ZEB1 and ZEB2, which are E-cadherin repressors. MiR-200 downregulation is seen in invading cells of adenocarcinoma cells, which indicates a strategic decrease in EMT. However, its expression is restored when regaining the epithelial phenotype\(^{153}\). The miR-200c/141 cluster is also overexpressed in liver metastasis when regaining the epithelial phenotype\(^{154}\). MiR-200b indirectly upregulates the KRAS gene by repressing its inhibitor (i.e. PTPN12). Notably, the chemotherapeutic drug 5-FU upregulates miR-200b\(^{155}\).

**MiR-34 family**

The miR-34 family prevents metastasis by inhibiting the SNAIL expression, which is an EMT-inducing transcription factor\(^{156}\). MiR-34a repression is required for interleukin-6 (IL-6) induced metastasis. The IL-6 receptor (IL-6R) mediates the activation of STAT-3 transcription factor by IL-6. This activation represses miR-34a, thereby facilitating EMT. However, the activation of p53 results in the inhibition of IL-6R, which is a direct miR-34a target\(^{157}\). Controversially, mutation rs4938723 in the promoter region of primary miR-34b/c is associated with a decreased CRC risk\(^{158}\). The study has several limitations. Therefore, a more thorough analysis of the mutation on the transcriptional and translational levels can unravel the mechanisms involved.

**MiR-17-92 cluster**

The miR-17-92 cluster includes miR-17, miR-18a, miR-19a, miR-19b-1, miR-20a, and miR-92a-1. This cluster is induced by transcription factors c-myc and E2F3\(^{159}\). All six miRNAs are significantly overexpressed in colorectal tumors. The higher expression is noted in adenoma to adenocarcinoma progression. A copy number gain at the locus is associated with the overexpression of these miRNAs except for miR-18a.

### Impact of mutations

Particular CRC mutations may also regulate microRNA expression. For example, inactivating mutations in the \(APC\) increases miR-135b expression by stabilizing \(\beta\)-catenin\(^{160}\). Moreover, epigenetic changes (e.g. DNA methylation) result in the downregulation of miRNAs. Three of the five downregulated tumor-suppressing miRNAs are restored in the CRC cell lines after treatment with DNA methyltransferase (DNMT) and histone deacetylase (HDAC) inhibitors. A comparison with the adjacent normal mucosa shows higher methylation observed in primary tumors. The miR-9-1 methylation is associated with lymph node metastases\(^{161}\). In another study, miR-9-1 and -34c have significantly reduced expression, while miR-34b has also been hypermethylated along with these two microRNAs\(^{162}\). Several SNPs in the genes coding for miR-146a, miR-196a2, miR-149, and miR-499 are tested. A significantly reduced expression of miR-499 is observed with the GG genotype for rs3746444G>A.

### Therapeutic biomarkers

As previously reviewed, microRNAs serve as potential biomarkers for therapeutic outcomes\(^{163}\). These factors are important in predicting the response to particular therapies. MicroRNAs act as positive biomarkers indicating sensitivity to therapy or as negative biomarkers signifying resistance to treatment and assisting in choosing the appropriate treatment regimen (Table 1)\(^{164–179}\).

### Long non-coding RNAs

LncRNAs are a heterogenous group of more than 200 nucleotides. LncRNAs lack the open reading frames (ORF) larger than 100 amino acids in length\(^{180}\). These RNAs are categorized depending on their location and function\(^{181}\). They are becoming increasingly important in cancer and metastasis studies because they play important roles in chromatin remodeling and transcriptional and posttranscriptional gene expression regulation\(^{182}\). The role of the five most common lncRNA, namely HOTAIR, MALAT-1, CCAT-1 and -2, and H-19, in colorectal carcinogenesis is reviewed in this paper.

**Hox transcript antisense intergenic RNA (HOTAIR)**

The 2.2 kb long lncRNA is present on the \(HOXC\) locus. This lncRNA binds to the polycomb repressive complex 2 (PRC2) at 5’ end resulting in the H3 lysine-27 (H3K27) methylation that leads to the repression of tumor suppressing gene \(HOXD\)
and lysine-specific demethylase1 (LSD1) at 3’ end causing H3K4 demethylation. The HOTAIR expression in CRCs is lower in cancerous tissues than in their normal counterparts. The CRC cells overexpressing HOTAIR are more invasive. Moreover, its suppression decreases the invasion. A higher HOTAIR expression increases the potential of liver metastasis because it cooperates with PRC2 in maintaining the mesenchymal phenotype.

Prostate cancer-associated ncRNA transcripts-1 (PCAT-1)
PCAT-1 is located on position 8q24 and is overexpressed up to almost 1.5 fold in cancerous versus normal tissues. This overexpression is associated with distant metastasis and influences the overall survival rate. However, the mechanism of action of PCAT-1 in CRCs is still not understood.

Metastasis-associated lung adenocarcinoma transcript-1 (MALAT-1)
The 6918bp to 8441bp region is located on chromosome 11q13 and is upregulated in both SW480 CRC cells and primary cancer tissues. Mutations have also been identified in this fragment. The higher expression of MALAT-1 is correlated with cell proliferation, migration, and invasion and observed in metastatic tumors. MALAT-1 promotes tumor metastasis in nude mice by inhibiting the polypyrimidine tract binding protein 2 (PTBP2)/splicing factor proline/glutamine-rich (SFPQ) complex and releasing the oncogene PTBP2. Furthermore, both MALAT1 and PTBP2 are overexpressed. The higher expression is associated with higher invasion and metastatic potential. In vivo studies indicate PRKA kinase anchor protein 9 (AKAP-9) as a target gene of MALAT-1. Its upregulation is correlated with MALAT-1 tumor promoting activity.

Colon cancer-associated transcript (CCAT1)
CCAT1 is upregulated in CRCs with a 100-fold higher expression in the HCT116 cell line and more than 30-fold higher expression in tumor versus normal colon cells. This higher expression is also associated with regional and distant liver metastasis. CCAT1-L is found on 8q24.21, upstream of MYC, and acts as an enhancer in CRC cell lines. The knockdown of CCAT1-L can reduce MYC levels. CCAT1-L assists in maintaining chromatin looping by modulating the binding of CTCF to the MYC locus.

Colon cancer-associated transcript (CCAT2)
CCAT-2 is another IncRNA located on 8q24, which encompasses rs6983267 SNP and is involved in chromosomal instability that results in increased proliferation and metastasis in MSS tumors. The G allele has been associated with higher susceptibility to CRC compared to the T allele with ratios of 1.4 and 1.27 for homozygotes and heterozygotes, respectively. This difference in alleles affects its binding efficiency to the transcription factor 7-like 2 (TCF7L2), through which it upregulates MYC, miR-17-5p, and miR-20a. This difference also further enhances Wnt activity and may form a feedback loop as a downstream target of Wnt.

### Table 1: Effect of microRNA expression on colorectal cancer treatment

<table>
<thead>
<tr>
<th>MicroRNA</th>
<th>Expression</th>
<th>Effect on treatment</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let-7</td>
<td>↑</td>
<td>Increased sensitivity to EGFR*-targeted treatment</td>
<td>164-167</td>
</tr>
<tr>
<td>MiR-126</td>
<td>↓</td>
<td>Decreased sensitivity to capecitabine and oxaliplatin (XELOX)</td>
<td>168-169</td>
</tr>
<tr>
<td>MiR-31</td>
<td>↑</td>
<td>Decreased sensitivity to 5-fluorouracil</td>
<td>170</td>
</tr>
<tr>
<td>MiR-192/ miR-215</td>
<td>↑</td>
<td>Increased resistance to 5-fluorouracil</td>
<td>171</td>
</tr>
<tr>
<td>MiR-148a</td>
<td>↓</td>
<td>Poor sensitivity to oxaliplatin and oxaliplatin plus 5-fluorouracil</td>
<td>172</td>
</tr>
<tr>
<td>MiR-21</td>
<td>↓</td>
<td>Poor sensitivity to 5-fluorouracil</td>
<td>173, 174</td>
</tr>
<tr>
<td>MiR-129</td>
<td>↑</td>
<td>Increased sensitivity to 5-fluorouracil</td>
<td>175</td>
</tr>
<tr>
<td>MiR-19b</td>
<td>↑</td>
<td>Increased response to 5-fluorouracil</td>
<td>176</td>
</tr>
<tr>
<td>MiR-34a</td>
<td>↑</td>
<td>Resistance to 5-fluorouracil</td>
<td>177</td>
</tr>
<tr>
<td>MiR-143</td>
<td>↑</td>
<td>Increased response to 5-fluorouracil</td>
<td>178</td>
</tr>
<tr>
<td>MiR-203</td>
<td>↑</td>
<td>Resistance to oxaliplatin</td>
<td>179</td>
</tr>
</tbody>
</table>

* EGFR: Epidermal growth factor receptor. Arrows indicate upregulated or downregulated expression.

**References**


