Oncology and Translational Medicine
DOI 10.1007/s10330-016-0164-6

August 2016, Vol. 2, No. 4, P145–P149

REVIEW ARTICLE

# Molecular subtypes of colorectal cancer: Evaluation of outcomes and treatment

**Weijing Sun (✉)**

University of Pittsburgh School of Medicine, University of Pittsburgh Medical Center (UPMC), University of Pittsburgh Cancer Institute (UPCI), Pittsburgh, USA

## Abstract

Colorectal cancer (CRC) is a biologically heterogeneous disease with diverse clinical outcomes and responses to treatment. In the past two to three decades, a major effort has focused on classifying colorectal cancer subtypes based on causation, etiology, gene expression profiles, different pathways, and translational data from clinical trials. The goal is to uncover prognostic and predictive factors for outcomes in patients with colorectal cancer and to guide therapeutic approaches and management for the improvement of overall survival. Significant advances have been achieved in this area. However, tremendous work is still needed to accomplish the goal of better understanding intratumoral heterogeneity and the influence of the colonic environment, among other facets of colorectal cancer.

Received: 8 June 2016
Revised: 9 July 2016
Accepted: 25 July 2016

**Key words:** colorectal cancer (CRC); molecular subtype; evaluation

Colorectal cancer (CRC) is the third most common cancer in the United States and the world (95 300 new cases in the United States and 1.4 million worldwide in 2016) [1, 2]. Patient survival and treatment options are still largely dependent on TNM stage at the time of diagnosis, even though we know colorectal cancer is a biologically heterogeneous disease that develops via distinct pathways involving alternative combinations of genetic and epigenetic factors [3]. Subtypes of colorectal cancer based on molecular and pathway profiles with defined prognostic markers would predict individual patient outcomes more precisely and therefore better inform on appropriate therapeutic intervention, especially targeted therapy. Various methods have been attempted and different directions taken to achieve this goal. Some approaches are as simple as focusing on the implication of defects in a single oncogene or tumor suppressor gene or assessing the consequences of a limited combination of gene mutations [4]. Other approaches are based on morphological characteristics, clinical and molecular features [5], or gene expression-based data classification [6].

To date, the microsatellite instability (MSI)-H phenotype has demonstrated the most robust prognostic role in terms of improved survival in stages II and III CRC patients [7]. Most single gene mutation markers have modest prognostic or predictive value, except Braf and Kras mutations. The Braf mutation ($BRAF^{V600E}$) has been associated with poorer survival in CRC. Kras and Nras mutations are associated with resistance to epidermal growth factor receptor (EGFR) targeted therapy [4].

'Pathway' based CRC subtype classification has been proposed because of the distinctive association of 'serrated polyps/adenocarcinoma' with MSI-H, CpG island methylation phenotype (CIMP) with Braf mutation; as well as the predictive and potential prognostic value of Kras [5]. At the same time, it has been also been learned that as a somatic genetic disease that is generally sporadic in nature, pathogenesis is influenced by the local colonic environment as well as the genetic background of the individual patient.

There have been many attempts to find consensus in classification of subtypes of CRC based on causation, etiology, gene expression profiles, different pathways, and translational data from clinical trials. Such efforts are geared towards revealing prognostic and predictive factors for patient outcomes and to guide therapeutic approaches and management to eventually improve overall survival. However, no universal subclassification has been agreed upon because of the various views and opinions of different groups of investigators and experts [6, 8–11]. Overall,

✉ Correspondence to: Weijing Sun. Email: sunw@upmc.edu

the proposed models are similar and based on the types and frequency of genetic alterations, epigenetic modifications, and molecular pathways.

## Genetic alterations and epigenetic modifications

Whole genome sequencing has confirmed research findings from the past three decades on the genetic and epigenetic abnormalities underlying CRCs. CRCs are formed through the accumulation of genetic and epigenetic events, which include gain-of-function defects as well as loss-of-function defects of selected tumor suppressor genes. It is suggested that approximately 25 different genes are commonly affected by somatic mutations in CRCs, with tumor suppressor genes outnumbering oncogenes by about four to one [9, 12]. Events conferring growth advantage are considered 'driver' mutations and the remaining mutations are called 'passenger' mutations that are the result of genomic chaos and random events with no clear effects on the disease process. Only two to eight driver gene alterations are found in a typical sporadic CRC. It is important to know that driver mutations in one CRC patient may differ from those in another patient [9, 13]. Additional genetic and epigenetic events are acquired in progeny cells beyond those inherited from parental cells.

Based on the commonalities among CRCs, the disease has been grouped into hypermutated (approximately 16% of sporadic CRCs) with mutation rates of > 12 per $10^6$ bases and nonhypermutated (approximately 84% of sporadic CRCs) with mutation rates of < 8.24 per $10^6$ bases based on TCGA data [9]. The median number of non-silent mutations is 728 in tumors of the hypermutated group compared to 58 in tumors of the nonhypermutated group. The etiology for hypermutated tumors is driven largely by the presence of MSI-H and CIMP resulting from defects in the MMR (DNA mismatch repair, MMR-D) genes hMSH6, hMSH2, hMSH3, and, hMLH3, as well as POLE (DNA polymerase ε). Hypermethylation of the hMLH1 promoter should also be noted. Most hypermutated CRCs have mutations in genes that contain intrinsic coding microsatellites. Nonhypermutated CRCs are more frequently associated with somatic copy-number alterations with more chromosomal or subchromosomal changes including either gains (1q, 7p and q, 8p and q, 12q, 13q, 19q, and 20p and q) or deletions (18p and q: 66% with SMD4; 17p and q: 56% with TP53; 1p, 4q, 5q, 8p, 14q, 15q, 20p, and 22q). Chromosomal region 10p25.2 is commonly involved (*FHIT, RBFOX1, WWOX, SMAD4, APC, PTEn, SMAD3,* and *TCF7L2*) as well as segment amplifications (*USP12, CDK8, KLF5, HNF4A, WHSC1L1, MYC, ERBB2,* and *IGF2*) [9].

Although hypermutated and nonhypermutated CRCs progress through different sequences of genetic events, there is some overlap of affected pathways. For example, APC is mutated in both groups, consistent with its role as a gatekeeper mutation. Alterations of MYC transcriptional targets are also noticed in both groups. Consistent activation of Wnt, RAS, PI3K signaling, inactivation of TGFβ signaling, and inactivation of TP53 function are demonstrated in both groups [9, 13].

Besides the direct genetic and epigenetic analysis of CRC tissue, single nucleotide polymorphisms (SNPs) from blood of CRC patients have also been examined as potential biomarkers. However, these genome-wide association studies are unable to determine the cause or mechanism of tumor initiation, progression, and/or metastasis. MicroRNA (miR) mutations and polymorphisms may also have profound effects on tumor behavior and offers potential therapeutic options [14].

## Molecular pathways

The simplified model of normal-adenoma-carcinoma-metastasis sequence has been established and accepted by investigators for many years in understanding and evaluating CRC initiation and processing. With recent molecular biology analysis, various molecular subtypes have been established and discussed [6, 8–11]. Although there is some disagreement among them, they are all based on three identified molecular pathways: CIN (chromosomal instability), MSI-H (microsatellite instability-high), and CIMP (CpG Island Methylator Phenotype). These molecular pathways may dictate the timing and process of tumor initiation, progression, and metastasis with distinguishes in epidemiology, mutational events, and immune response, therefore, treatment approaches could be vary as well.

The CIN pathway is affected in approximately 85% of CRC cases and is the most common and the first described molecular pathway in CRC. However, the mechanisms leading to CRC are still unclear but are thought to include extensive copy number of somatic mutations throughout the genome, which results in aneuploidy tumors with nonhypermutated adenomas [13, 15, 16]. Loss of APC and TP53 appears sufficient for generation of significant aneuploidy, particularly when additive with SMAD4 and mutant KRAS [17]. APC is a part of the Wnt signaling pathway, which regulates cytoplasmic levels of β-lactenin and is related to cellular proliferation as a tumor suppressor gene. Wnt signaling is deregulated in 93% of all nonhypermutated CRCs with APC being the most commonly mutated component (81%). KRAS mutation as the oncogenic activator is the next most common event. KRAS is a part of the ERBB/KRAS/BRAF/MAPK signaling axis. Mutant KRAS protein causes acceleration of tumor proliferation. In nonhypermutated CRCs, the prevalence

of KRAS mutation is 41% and overall active mutations of KRAS, NRAS, or BRAF is approximately 55% [9]. Clinical evidence demonstrated that patients with mutations in KRAS, NRAS, or BRAF have poorer outcomes compared to those with wild type genes. PIK3CA, the catalytic subunit of the mitogenic PI3K complex, controls levels of phosphatidylinositol triphosphate and is antagonized by PTEN. Mutations of PIK3CA are found in 18% of nonhypermutated CRCs. Alteration of these two pathways is found in about 33% of CRCs. Therefore, inhibiting both pathways simultaneously may be beneficial in clinic management of CRC [9].

MSI-H is defined as > 30% of microsatellite markers demonstrating a frameshift mutation and is a biomarker for defective DNA MMR function in CRC [9]. DNA MMR recognizes and repairs nucleotide mismatches and mispairing during DNA replication. MSI-H is observed in approximately 15% of sporadic CRCs, consistent with the frequency of hypermutated CRCs. The defect in DNA MMR is caused by aberrant bi-allelic hypermethylation of the DNA MMR gene hMLH1 (for the most part), thereby preventing its transcription [18–20]. MSI-H CRCs accumulate mutations in driver genes with frameshift mutations and subsequently cause stop codons, which creates a truncated transcript and proteins that are neo-antigenic to the patients' immune system [21]. The other common feature of the MSI-H CRC pathway is an activating oncogenic mutation of BRAF (most commonly $BRAF^{V600E}$) in about 40% of cases via this pathway [9]. CRC via the MSI-H pathway seems to have low copy number variation and tends to be diploid with fewer TP53 (20%) and ACP (51%) mutations compared to CIN-derived CRC. Both types have a similar frequency of Wnt deregulation. Histologically, MSI-H CRCs are more likely to be poorly differentiated, contain mucin, and possess subepithelial lymphoid aggregates and intraepithelial lymphocytes due to the immune response to truncated neo-antigens produced from the epithelium. MSI-H CRCs are more commonly (approximately 70%) located proximal to the splenic flexure [22, 23]. Recent analysis showed that CRCs with MSI-H pathway involvement are less frequent among African-American populations than in Caucasians or Asians, which may partially explain the poor outcomes of CRCs in African-Americans compared with in Caucasians stage-by-stage [24].

CIMP is defined by increased or excessive epigenetic methylation of genetic loci, which contains CpG islands typically located in the promoter and upstream regulatory regions of genes. The etiology of CIMP development is less definitive with several possible mechanisms or combinations of abnormalities including DNA methyltransferase overexpression, mutations in chromatin remodeling genes (e.g., *CHD8*), mutations in *IDH1* and *TET*, or environmental exposure (e.g., tobacco use) [22, 25–28]. Based on the number of markers (*RUNX3, SoCS1, NEUROG1, CACNA1G,* and *IGF2*) positive for methylation, CIMP can be further classified as 'high' (≥ 3 markers of methylation) or 'low' (≤ 2 markers of methylation) [29]. CIMP CRCs overlap with MSI-H and CIN pathways. CIMP-H occurs as hypermutated tumors in approximately 20% of CRCs with BRAF mutation and hypermethylation of hMLH1 (for the most part). CIMP-H CRCs are most likely to manifest a serrated morphology, including sessile serrated adenomas and traditional serrated adenomas. CIMP-L occurs as nonhypermutated tumors in 20% of CRCs with some of them derived from traditional serrated adenomas with MSS and containing KRAS mutations. The CIMP pathway can be helpful for understanding pathogenesis of CRC. However, it does not appear to be a useful tool or biomarker clinically.

## Clinical evaluation

CRC subtypes/subclassifications based on distinct histopathologic and molecular alterations, as well as involved pathways, may better predict patient outcomes and will likely advance effective drug development strategies.

Recently, two large studies with more than 2000 stage III CRC patients in each revealed new and important associations between molecular alterations and patient survival [30, 31]. One study prospectively collected samples from 2720 stage III patients participating in an adjuvant chemotherapy trial (NCCTG N0147). Mutations in BRAF ($BRAF^{V600E}$) and in KRAS were tested and tumor DNA MMR status (proficiency or deficiency) was identified based on detection of MLH1, MSH2, and MSH6 proteins and methylation of the *MLH1* promoter. Findings were validated using tumor samples from a separate set of patients with stage III cancer ($n$ = 783). Based on MMR status and detection of $BRAF^{V600E}$ or mutations in KRAS (which were mutually exclusive), tumors were categorized into five subtypes: MMR-P (also as MSS or MSS-L) with $BRAF^{V600E}$ in 6.9%; MMR-P with KRAS mutations in 35%; MMR-P with no BRAF or KRAS mutations in 49%; MMR-D with $BRAF^{V600E}$ or hypermethylation of MLH1 (as the sporadic type) in 6.8%; and MMR-D with no BRAF mutation or hypermethylation of MLH1 (familial type) in 2.6%. Their findings were consistent with the molecular subtype model prediction described above. A higher percentage of MMR-P tumors with $BRAF^{V600E}$ were proximal (76%), high grade (44%), N2 stage (59%), and detected in women (59%), compared to MMR-P tumors without BRAF or KRAS mutations (33%, 19%, 41%, and 42%, respectively; all $P < 0.0001$). A significantly lower 5-year disease free survival (DFS) in patients with MMR-P and $BRAF^{V600E}$ mutations was found compared to patients who were MMR-P with no mutations in either gene. DFS in patients with MMR-D sporadic or familial

subtypes was similar to that in patients with MMR-P with no *BRAF*$^{V600E}$ or KRAS mutations.

The other study is based on the Seattle Colon Cancer Family Registry. A total of 2706 patients were diagnosed with invasive CRC from 1998 through 2007 in western Washington State and followed for survival through 2012. Tumor samples were collected from 2050 participants and classified into five subtypes based on combinations of tumor markers: type 1 as MSI-H, CIMP-positive with BRAF mutation; type 2 as MSS or MSI-L, CIMP-positive with BRAF mutation; type 3 as MSS or MSI-L, non-CIMP, positive for KRAS mutation; type 4 as MSS or MSI-L, non-CIMP, with no mutations in BRAF and KRAS; and type 5 as MSI-H, non-CIMP, and negative for mutations in BRAF and KRAS. Hazard ratios (HR) and 95% confidence intervals (CI) were assessed for associations of subtypes with disease-specific and overall mortality after adjusting for age, sex, body mass, diagnosis year, and smoking history. The results showed that compared with patients with type 4 tumors (MSS or MSI-L, non-CIMP, without BRAF or KRAS mutations, the most predominant), the patients with type 2 tumors (MSS or MSI-L, CIMP-positive, with BRAF mutation) had the highest disease-specific mortality (HR = 2.20, 95% CI: 1.47–3.31). Patients with type 3 tumors (MSS or MSI-L, non-CIMP, with KRAS mutation) also had higher disease-specific mortality (HR = 1.32, 95% CI: 1.07–1.63). Patients with type 5 tumors (MSI-H, non-CIMP, and without BRAF or KRAS mutations) had the lowest disease-specific mortality (HR = 0.30, 95% CI: 0.14–0.66). Associations with overall mortality in each type were similar to those with disease-specific mortality. These two studies confirmed that CRC subtypes, defined by proposed molecular pathways, are associated with marked differences in survival.

As mentioned above, MSI-H CRCs accumulate mutations with frameshift mutations and subsequently create truncated transcripts and proteins that are neo-antigenic to the patients' immune system. A recent study confirmed that MSI status is predictive of immune checkpoint blockage in advanced CRC [32].

In summary, integration of molecular pathways and subtypes with the TNM staging system is important for us to guide treatment for patients with active disease, and for surveillance/monitoring their status post potential curative procedure and adjuvant therapy. However, CRC heterogeneity has been observed at the intratumoral, intermetastatic, and intrametastatic levels. Rare variant cell populations may have important roles in clinical outcome. New strategies such as deep sequencing of primary CRC cell populations, comprehensive single-cell analyses, and analyses of circulating tumor-derived DNA are future molecular approaches that are needed to better define prognosis and predict likely responses to existing and new targeted therapies.

## Conflicts of interest

The author indicated no potential conflicts of interest.

## References


1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin, 2016, 66: 7–30.
2. Jorgensen ML, Young JM, Solomon MJ. Optimal delivery of colorectal cancer follow-up care: improving patient outcomes. Patient Relat Outcome Meas, 2015, 6: 127–138.
3. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology, 2008, 135: 1079–1099.
4. Herzig DO, Tsikitis VL, Molecular markers for colon diagnosis, prognosis and targeted therapy. J Surg Oncol, 2015, 111: 96–102.
5. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology, 2007, 50: 113–130.
6. Guinney J, Dienstmann R, Wang X, *et al*. The consensus molecular subtypes of colorectal cancer. Nat Med, 2015, 21: 1350–1356.
7. Popat S, Hubner R, Houlston RS. Systemic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol, 2005, 23: 609–618.
8. Hoadley KA, Yau C, Wolf DM, *et al*. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell, 2014, 158: 929–944.
9. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature, 2012, 487: 330–337.
10. Roepman P, Schlicker A, Tabernero J, *et al*. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. Int J Cancer, 2014, 134: 552–562.
11. Sadanandam A, Lyssiotis CA, Homicsko K, *et al*. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med, 2013, 19: 619–625.
12. Seshagiri S, Stawiski EW, Durinck S, *et al*. Recurrent R-spondin fusion in colon cancer. Nature, 2012, 488: 660–664.
13. Vogelstein B. Papadopoulos N, Velculescu VE, *et al*. Cancer genome landscapes. Science, 2013, 339: 1546–1558.
14. Hutchison J, Cohen Z, Onyeagucha BC, *et al.* How microRNAs influence both hereditary and inflammatory-mediated colon cancers. Cancer Genet, 2013, 206: 309–316.
15. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell, 1990, 61: 759–767.
16. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology, 2010, 138: 2059–2072.
17. Drost J, van Jaarsveld RH, Ponsioen B, *et al*. Sequential cancer mutations in cultured human intestinal stem cells. Nature, 2015, 521: 43–47.
18. 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, 1997, 57: 808–811.
19. Herman JG, Umar A, Polyak K, *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci USA, 1998, 95: 6870–6875.
20. Veigl ML, Kasturi L, Olechnowicz J, *et al*. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing

human MSI cancers. Proc Natl Acad Sci USA, 1998; 95: 8698–8702.
21. Schwitalle Y, Kloor M, Eiermann S, *et al*. Immune response against frameshift-induced neopeptides in HNPCC patients and healthy HNPCC mutation carriers. Gastroenterology, 2008, 134: 988–997.
22. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology, 2008, 135: 1079–1099.
23. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology, 2010, 138: 2073–2087.
24. Carethers JM, Murali B, Yang B, *et al.* Influence of race on microsatellite instability and CD8+ T cell infiltration in colon cancer. PLoS One, 2014, 9: e100461.
25. Tahara T, Yanamoto E, Madireddi P, *et al.* Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. Gastroenterology, 2014, 146: 530–538.
26. Hinshelwood RA, Melki JR, Huschtscha LI, *et al.* Aberrant de novo methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. Hum Mol Genet, 2009, 18: 3098–3109.
27. Limsui D, Vierkant RA, Tillmans LS, *et al*. Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. J Natl Cancer Inst, 2010, 102: 1012–1022.
28. Ichimura K. Molecular pathogenesis of IDH mutations in gliomas. Brain Tumor Pathol, 2012, 29: 131–139.
29. Hinoue T, Weisenberger DJ, Lange CP, *et al*. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res, 2012, 22: 271–282.
30. Sinicrope FA, Shi Q, Smyrk TC, *et al*. Molecular markers identify subtypes of stage III colon cancer associated with patient outcomes. Gastroenterology, 2015, 148: 88–99.
31. Phipps AI, Limburg PJ, Baron JA, *et al.* Association between molecular subtypes of colorectal cancer and patient survival. Gastroenterology, 2015, 148: 77–87.
32. Le DT, Uram JN, Wang H, *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med, 2015, 372: 2509–2520.

**DOI 10.1007/s10330-016-0164-6**
**Cite this article as:** Sun WJ. Molecular subtypes of colorectal cancer: Evaluation of outcomes and treatment. Oncol Transl Med, 2016, 2: 145–149.