Molecular characterization and biomarker identification in colorectal cancer: Toward realization of the precision medicine dream

Gaurav Goel
Division of Hematology-Oncology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Abstract: Colorectal cancer (CRC) is a major public health problem, both in the USA and globally. Over the past 20 years, significant advances have been made in the treatment of patients with metastatic CRC (mCRC). Recent efforts in the field of biomarkers have focused on the development of molecular diagnostics to define the subset of patients with mCRC that is likely to derive most benefit from anti-EGFR therapy. Herein, we review the recent advancements in molecular stratification of CRC and the role of current as well as emerging biomarkers in this disease. It is now clear that the presence of activating mutations in the KRAS and NRAS genes serves as reliable predictive markers for resistance to anti-EGFR therapy in mCRC. It is also clear that further improvements in the survival of mCRC patients will probably be made possible only with identification of new predictive molecular biomarkers and their evaluation using rational and innovative clinical trials. The recent advances in DNA sequencing technology and “omics”-based approaches have provided promising new strategies for the development of novel molecular biomarkers in this disease.

Keywords: colorectal cancer, biomarkers, circulating tumor DNA, next-generation sequencing, personalized medicine, omics

Introduction
Colorectal cancer (CRC) is a major public health problem, both in the USA and globally. In the USA, it is the second leading cause of cancer-related mortality, and in 2018, it is estimated that nearly 50,630 deaths will be attributed to this disease.1 When metastatic disease is diagnosed, CRC is usually associated with poor prognosis with 5-year survival rates less than 15%.1 With the advances in the treatment of metastatic CRC (mCRC) over the past 20 years, the median overall survival (OS) has been steadily increasing, and it has now reached beyond 40 months in a select group of patients.2–4 This improvement in OS has been made possible with the identification of new agents and incorporation of targeted biologic agents into the cytotoxic chemotherapy regimens used for treatment of mCRC patients.5,6 Although the progress made thus far is encouraging, the existing treatment paradigm usually employs a “one-size-fits-all” approach, which translates into demonstrable clinical benefit from any given chemotherapeutic regimen in only a small subset of treated patients.

It is now being increasingly realized that CRC is not a single disease entity, but a heterogeneous group of tumors, both at the intertumoral and intratumoral levels. Given the complex molecular makeup of these tumors, significant efforts have been focused in recent years on identifying potential molecular markers that would...
provide further information regarding tumor response to anticancer therapies, which in turn would help personalize treatment for mCRC. “Predictive” biomarkers identify patients who are most likely to benefit from a specific treatment, and they can, therefore, play a critical role in guiding treatment decisions. Research in the identification of predictive biomarkers for mCRC has been directed toward the targets that are downstream in the EGFR signal transduction cascade, including mutations in the KRAS, NRAS, BRAF, and PIK3CA genes, and PTEN protein expression. “Prognostic” biomarkers, on the other hand, provide information about the tumor’s aggressiveness, and they help distinguish patients based on their expected disease outcome, independent of the treatment received. It is believed that molecular stratification of CRC can identify biomarkers to provide such additional prognostic and predictive information, over and beyond the conventional TNM staging system. Such validated biomarkers will also help optimize efficacy, minimize toxicity, and reduce treatment costs.

This article will review the recent advancements in molecular stratification of CRC and the role of current as well as emerging biomarkers in this disease.

**Subtypes of colorectal tumors**

**CRC subtyping based on traditional oncogenic pathways**

CRC has been theorized as a stepwise model of genetic and epigenetic events by Fearon and Vogelstein, nearly three decades ago.7 It usually begins with the inactivation of APC tumor suppressor gene, resulting in increased Wnt pathway signaling.8 Thereafter, depending on the mechanism of underlying genomic alterations, CRC develops through one of the three distinct molecular pathways—chromosomal instability (CIN), microsatellite instability (MSI), and CpG island methylator phenotype (CIMP) pathways. Of these, CIN pathway is the most commonly occurring (85%), and it involves widespread loss of heterozygosity and gross chromosomal abnormalities leading to hypermutability. MSI and CIMP pathways result in development of colorectal tumors due to a defective DNA mismatch repair (MMR) system. MMR system is a multiprotein system responsible for proofreading and eliminating any nucleotide mismatches that escape the checking system of DNA polymerase during DNA replication process. While the MSI pathway is characterized by the presence of genetic or sporadic mutations in the MMR genes, the CIMP pathway involves epigenetic silencing of the MMR gene MLH1 by causing hypermethylation of the CpG dinucleotide sequences located in the gene’s promoter region.

**Transcriptome-based CRC subtypes**

The international Colorectal Cancer Subtyping Consortium studied six independent transcriptome-based classification systems of CRC, and the resulting collection of 27 subtypes were coalesced into four distinct Consensus Molecular Subtype (CMS) groups.9 Transcriptional molecular signatures refine disease subclassification and help to identify subtypes with distinct biologic behavior. CMS1 group (MSI immune subtype, 14%) consists of hypermutated MSI tumors with strong immune activation. These tumors have overexpression of proteins involved in DNA damage repair consistent with defective DNA mismatch repair. BRAF mutations occur frequently in these tumors, and an increased expression of genes associated with diffuse immune infiltrate is noted. CMS2 (canonical subtype, 37%) group is characterized by epithelial differentiation and chromosomally unstable tumors. The tumors in this group display high somatic copy number alterations, and upregulation of WNT and MYC downstream targets. CMS3 (metabolic subtype, 13%) tumors show epithelial differentiation and marked metabolic dysregulation. Nearly 30% of these tumors are hypermutated as well. There is a higher prevalence of CIMP and KRAS activating mutations in CMS3. CMS4 (mesenchymal subtype, 23%) consists of tumors with prominent TGF-β activation, stromal invasion, and angiogenesis.

**Established biomarkers in current clinical practice**

**RAS (KRAS and NRAS) mutations**

The RAS gene family is composed of three well-known genes: KRAS, NRAS, and HRAS.10 KRAS mutations are estimated to be present in ~30%–45% of CRC patients and include single-point mutations in exons 2, 3, and 4.11 Another 5% of CRC tumors harbor activating mutations in NRAS, in the same exons as KRAS. These genes encode GTP/GDP binding protein, which is a key player in the MAP kinase signal transduction pathway.

The role of RAS mutations as a negative predictive marker for anti-EGFR therapy in the treatment of mCRC has been confirmed in multiple clinical trials and is now well established.12–18 Several systematic reviews and meta-analyses have also confirmed that among the anti-EGFR monoclonal antibody-treated mCRC patients, KRAS wild-type patients perform better than KRAS-mutant population; and that among the KRAS wild-type patients, treatment with anti-EGFR
monoclonal antibody-containing regimen is associated with better response than chemotherapy alone. Studies have also clearly established that in contrast to the KRAS wild-type tumors, the addition of anti-EGFR therapy to chemotherapy does not improve the outcomes in KRAS-mutant population and is even potentially harmful.

It is, however, worth noting that in these earlier studies, only mutations in exon 2 of the KRAS gene were assessed. More recent data suggest that in addition to mutations in KRAS exon 2, other RAS mutations in KRAS exons 3 and 4, and NRAS exons 2, 3, and 4 are also associated with resistance to anti-EGFR therapy in mCRC. The most convincing evidence establishing the predictive value of “expanded” RAS testing is provided by the meta-analysis of nine randomized clinical trials consisting of nearly 6,000 patients. In this meta-analysis, the tumors with new RAS mutations (KRAS exons 3 and 4, NRAS exons 2, 3, and 4) were compared with tumors without any RAS mutation, with respect to anti-EGFR treatment-related progression-free survival (PFS) and OS benefit. Patients with tumors that were KRAS exon 2 wild-type, but harbored RAS mutations in KRAS exons 3 and 4, or NRAS exons 2, 3, and 4 had significantly inferior survival outcomes compared to those without any RAS mutation.

Taken together, it is now well established that presence of any activating mutation in KRAS or NRAS predicts for resistance to anti-EGFR therapy in mCRC. A recent guideline on molecular testing for CRC issued jointly by the American Society for Clinical Pathology (ASCP), College of American Pathologists (CAP), Association of Molecular Pathology (AMP), and American Society of Clinical Oncology (ASCO), recommends expanded RAS testing, which includes mutational analysis of KRAS and NRAS codons 12 and 13 of exon 2, codons 59 and 61 of exon 3, and codons 117 and 146 of exon 4 for mCRC patients being considered for anti-EGFR therapy. The mCRC patients whose tumors harbor these mutations should not receive anti-EGFR therapy as they will not derive clinical benefit from such treatment. However, a greater concern is that such patients are likely to experience an even worse clinical outcome. It is also worth noting that the absence of RAS mutation is not the most accurate predictor of tumor response to anti-EGFR monoclonal antibody, as only 40%–50% of patients with KRAS/NRAS wild-type disease will ultimately respond to such therapy.

In contrast to the predictive value, the role of KRAS mutation in CRC as an “independent” prognostic marker has yielded inconsistent results. In one of the early efforts, several researchers with information about the KRAS tumor genotype and outcome of CRC patients came together to form a collaborative database. The data on 3,439 patients recruited from 42 centers in 21 countries were entered into a multivariate analysis, and showed that KRAS mutation in codon 12 was associated with an increased risk of tumor recurrence and worse survival. In another retrospective analysis of tumor samples obtained from CRC patients treated with chemotherapy (without anti-EGFR monoclonal antibody) in the MRC FOCUS trial, the presence of mutation in either KRAS (codons 12, 13, and 61) or BRAF was associated with significantly decreased survival when compared to the wild-type tumors. In contrast, the CO.17 study and the study by Kim et al did not demonstrate any significant survival difference between the wild-type KRAS and mutant-KRAS subgroups. In a systematic review with meta-analysis also, KRAS mutation status did not correlate with prognosis in CRC patients. Due to the inconsistencies in data, KRAS mutational status cannot be used reliably as an independent prognostic marker for mCRC patients at this time.

Table 1 provides a summary of established and emerging biomarkers in CRC and their potential role in clinical practice.

**BRAF mutations**

BRAF is a member of the RAF proto-oncogene family, and a key player of the RAS/RAF/MAP kinase pathway. Activating point mutations in BRAF are present in ~5%–10% of CRC. This mutation results in constitutive activation of BRAF kinase, which then leads to activation of the downstream RAS/RAF/MAP kinase pathway.

Several independent studies have consistently demonstrated that the presence of BRAF mutations is associated with poor prognosis in patients with mCRC. As the population of BRAF-mutant mCRC patients is relatively small due to the low mutation prevalence, several meta-analyses and systematic reviews have been performed to provide more conclusive evidence on the prognostic and predictive value of BRAF mutation as it relates to anti-EGFR therapy. In one of the largest meta-analyses of 21 clinical trials that included 5,229 mCRC patients treated with anti-EGFR monoclonal antibody, 343 patients displayed BRAF mutations out of 4,616 (7.4%) patients with known BRAF status. Patients with BRAF wild-type showed an improved PFS (HR, 0.38; 95% CI 0.29–0.51) and an improved OS (HR, 0.35; 95% CI 0.29–0.42), compared to the BRAF-mutant group. These results indicate that the presence of BRAF mutation is associated with poor prognosis in mCRC patients.

In mCRC, the presence of BRAF mutation has been suggested as a negative predictor of response toward anti-EGFR
Table 1 Summary of established and emerging biomarkers in the management of colorectal cancer

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Prognostic value</th>
<th>Predictive value</th>
<th>Comments</th>
<th>ASCP/CAP/AMP/ASCO CRC biomarker guidelinea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Somatic mutations and alterations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS – codons 12 and 13 of exon 2; codons 59 and 61 of exon 3; codons 117 and 146 of exon 4</td>
<td>Inconsistent data</td>
<td>Yes</td>
<td>CRC patients with these mutations should not receive anti-EGFR therapy</td>
<td>Perform testing in CRC patients being considered for anti-EGFR therapy</td>
</tr>
<tr>
<td>NRAS – codons 12 and 13 of exon 2; codons 59 and 61 of exon 3; codons 117 and 146 of exon 4</td>
<td>Inconsistent data</td>
<td>Yes</td>
<td>CRC patients with these mutations should not receive anti-EGFR therapy</td>
<td>Perform testing in CRC patients being considered for anti-EGFR therapy</td>
</tr>
<tr>
<td>BRAF p. V600</td>
<td>Yes</td>
<td>Inconsistent data</td>
<td>Tumors carrying BRAF mutation have poorer outcome</td>
<td>Perform testing in CRC patients for prognostic stratification</td>
</tr>
<tr>
<td>BRAF p. V600</td>
<td>–</td>
<td>–</td>
<td>Presence of BRAF mutation favors sporadic pathogenesis. Absence of BRAF mutation does not exclude Lynch syndrome</td>
<td>Perform testing in MMR-deficient tumors with loss of MLH1 to evaluate for Lynch syndrome risk</td>
</tr>
</tbody>
</table>
| PIK3CA – exons 9 and 20                        | Inconsistent data | Yes              | - Exon 20 mutations are associated with poor clinical response to cetuximab
- Improved survival with postoperative aspirin use in tumors with PIK3CA mutation | Insufficient evidence to recommend testing for therapy selection outside of clinical trial |
| Loss of PTEN expression                        | Inconsistent data | Inconsistent data | There is significant discordance (66%) in PTEN expression between primary tumor and matched metastases | Insufficient evidence to recommend testing (IHC or FISH) for therapy selection outside of clinical trial |
| HER2 amplification                             | –                | Yes              | HER2 amplification predicts response toward HER2-directed therapy         | –                                                                          |
| **Genetic biomarkers**                         |                  |                  |                                                                          |                                             |
| Microsatellite instability                     | Yes              | Yes              | - MSI-Hi is associated with favorable prognosis in patients with early stage colon cancer
- MSI-Hi is predictive for lack of efficacy toward 5-FU adjuvant chemotherapy in stage II colon cancer
- MSI-Hi is a positive predictor of durable tumor response to anti-PD-1 immune-checkpoint therapy in mCRC | Perform testing in CRC to identify patients at high risk for Lynch syndrome and/or prognostic stratification |
| Gene expression recurrence score assays        | Yes              | No               | Low recurrence score is associated with longer survival                  | –                                                                          |
| **Immunologic characteristics**                |                  |                  |                                                                          |                                             |
| CD3+ T-cells (tumor-infiltrating lymphocytes) in tumor core and at invasive margin | Yes              | –                | Tumors with elevated levels of CD3+ T-cells in the core as well as at the invasive margin are associated with the best clinical outcome, independent of TNM stage | –                                                                          |
| Immunoscore (density of CD8+ cytotoxic T-cells and CD45RO+ memory cells at invasive margin and in the center of tumor) | Yes              | –                | Time-to-tumor recurrence is shorter among patients with “low” immunoscore and longer in those with “high” immunoscore | –                                                                          |
| **Tumor location**                             |                  |                  |                                                                          |                                             |
| Left- vs right-sided colon cancer              | Yes              | –                | - Left-sided primary tumors have superior survival than right-sided tumors
- Right-sided tumors have longer survival with bevacizumab, and left-sided tumors have longer survival with cetuximab | –                                                                          |

(Continued)
therapy, based on the results from the retrospective study by Di Nicolantonio et al\textsuperscript{32} and the PICCOLO trial.\textsuperscript{42} On the other hand, in a meta-analysis of eight randomized clinical trials that included RAS wild-type mCRC patients treated with anti-EGFR therapy, there was lack of statistically significant difference in OS ($P=0.43$) between the \textit{BRAF}-mutant and \textit{BRAF} wild-type subgroups.\textsuperscript{40} In another meta-analysis, the addition of anti-EGFR therapy in the \textit{BRAF}-mutant advanced CRC patients did not significantly improve the survival compared to the control regimen.\textsuperscript{41}

The available clinical evidence suggests that the presence of \textit{BRAF} mutations has a clear prognostic value in mCRC as it confers a significantly worse clinical outcome. Recent guidelines recommend that \textit{BRAF}\textsuperscript{V600E} mutational analysis should be performed in all mCRC patients for prognostic stratification.\textsuperscript{26} However, the available data on \textit{BRAF} mutations as a predictive marker for anti-EGFR therapy are limited due to being largely retrospective in nature. The relatively low prevalence of \textit{BRAF} mutation in colon cancer poses the greatest challenge in addressing this question prospectively. Therefore, at the present time, there is insufficient evidence to recommend \textit{BRAF} mutation testing on colorectal tumors to predict response to anti-EGFR therapy.\textsuperscript{26}

\section*{MSI and tumor mutational burden}

The loss of MMR proteins leads to MMR deficiency, and these tumors are also referred to as microsatellite instability-high (MSI-Hi) tumors. In fact, MSI is considered as the molecular fingerprint of a deficient MMR system. The Cancer Genome Atlas (TCGA) project reported that $\sim$16% of all CRCs are hypermutated or MSI-Hi.\textsuperscript{43} The majority of MSI-Hi CRCs (11%–13%) are caused by somatic changes, either resulting from epigenetic silencing of the \textit{MLH1} gene in the majority of cases or due to somatic mutation in any of the four most common MMR genes (\textit{MLH1}, \textit{MSH2}, \textit{MSH6}, and \textit{PMS2}) or \textit{POLE} gene.\textsuperscript{43-45} The remaining 3%–5% MSI-Hi CRCs are caused by germline mutation in the MMR genes, and these patients are referred to as having the hereditary nonpolyposis CRC (Lynch syndrome).\textsuperscript{44-47}

MSI status has been evaluated as a prognostic and predictive biomarker in CRC. MSI-Hi CRCs are more often located in the proximal colon, are poorly differentiated, and have mucinous or medullary histology. These tumors are also considered highly immunogenic.\textsuperscript{48} It is now understood that the length variations in the microsatellites of coding sequence in MSI-Hi tumors lead to frame-shift mutations. The high mutational burden associated with these tumors results in formation of tumor-specific neoantigens and subsequent T-cell infiltration, which suggests presence of a robust antitumoral immune response.\textsuperscript{48} This explains the observation that MSI-Hi is associated with favorable prognosis in patients with early stage colon cancer.\textsuperscript{49} It is now recommended that MMR status should be checked in patients with CRC for prognostic stratification.\textsuperscript{26}
MSI status also has predictive value, both in early stage and advanced CRC.\textsuperscript{50} The presence of MSI-Hi has been shown to be predictive for lack of efficacy toward 5-FU adjuvant chemotherapy in early-stage colon cancer.\textsuperscript{51} Therefore, adjuvant 5-FU should be avoided for stage II MSI-Hi colon cancer. In patients with advanced CRC, the data from early clinical trials suggest that the presence of MSI-Hi status is a positive predictor of durable tumor response to anti-programmed death-1 (PD-1) immune-checkpoint therapy.\textsuperscript{52–55} A Phase II study evaluated the clinical efficacy of anti-PD-1 agent pembrolizumab in patients with chemotherapy refractory mCRC, with and without MMR deficiency.\textsuperscript{54} In this study, the MSI-Hi CRCs were found to have encouraging responses to pembrolizumab. The overall response rate (ORR) was 50% for MSI-Hi CRC and 0% for microsatellite stable (MSS) CRC.\textsuperscript{55} The median OS was not reached in the MSI-Hi CRC cohort, as opposed to 6 months in the MSS CRC cohort (HR, 0.247; \textit{P}=0.001). Based on these encouraging results, a phase III study (KEYNOTE-177, NCT02563002) of pembrolizumab in MSI-Hi advanced CRC is ongoing to confirm these early observations.

Similarly, in another phase II study (CheckMate-142), nivolumab monotherapy was shown to provide clinical benefit (ORR, 31%; 12-month PFS, 50%; 12-month OS, 73%) in previously treated patients with MSI-Hi mCRC patients.\textsuperscript{56} Nivolumab was also combined with the CTLA-4 inhibitor ipilimumab to synergistically promote T-cell antitumor activity and further improve the outcomes among MSI-Hi mCRC patients.\textsuperscript{57} In the nivolumab plus ipilimumab cohort of the CheckMate-142 study, patients received nivolumab plus ipilimumab every 3 weeks for four doses, followed by nivolumab once every 2 weeks. Among the 119 patients treated with this combination, ORR was 55% at a median follow-up of 13.4 months. The 12-month PFS and OS rates were 71% and 85%, respectively. Based on these encouraging results, the Food and Drug Administration granted accelerated approval to the combination of nivolumab and ipilimumab in MSI-Hi mCRC patients that have progressed beyond first-line chemotherapy.

Recent evidence suggests that the presence of a high tumor mutational load might be a more important and accurate predictive marker of response to immunotherapy than the tumor’s MSI status itself. This is illustrated by the fact that PD-1 immunotherapy has shown efficacy in colorectal tumors carrying the \textit{POLE} mutation.\textsuperscript{58} This type of tumor is classified as an MSS tumor but is characterized by the presence of very high mutational burden and tumor-infiltrating lymphocytes in its microenvironment.

### Gene expression profiling-based recurrence score assays

Multigene expression signatures have been developed for use as biomarkers in early stage colon cancer, such as stage II disease where the role of adjuvant chemotherapy is still unclear. Examples of gene expression signatures that can guide CRC prognosis include Oncotype DX and ColoPrint assays. The Oncotype DX is a 12-gene RT-PCR-based recurrence score assay that was shown to have prognostic value in patients with stage II colon cancer, independent of the traditional clinicopathologic factors.\textsuperscript{59,60} ColoPrint is an 18-gene microarray-based assay that differentiates stage II tumors into low and high risk of recurrence.\textsuperscript{61} Stage II colon cancer patients who were identified as having low-risk disease based on the ColoPrint assay had significantly improved disease-free survival (DFS) than those in the high-risk group (87.6% vs 67.2%; HR, 2.5; \textit{P}=0.005). Although these recurrence score assays have a definite prognostic value in early stage colon cancer, their ability to predict benefit from adjuvant chemotherapy has not been established yet.\textsuperscript{62}

### Emerging biomarkers

#### \textit{PIK3CA} mutations

\textit{PIK3CA} gene encodes for the p110\alpha catalytic subunit of the PI3K protein, which is a component of the PI3K/PTEN/AKT signaling pathway. \textit{PIK3CA} mutations occur in 10%–18% of CRCs, mainly in exons 9 and 20.\textsuperscript{31,43,63,64} When mutated, \textit{PIK3CA} induces phosphorylation of AKT, which promotes cell growth and suppresses apoptosis in CRCs.

The impact of \textit{PIK3CA} mutations on the outcome of mCRC has been evaluated in retrospective studies. In the study by De Roock et al, it was suggested that not all \textit{PIK3CA} mutations behave similarly in terms of response to anti-EGFR therapy.\textsuperscript{31} \textit{PIK3CA} exon 9 mutations had no effect, whereas exon 20 mutations were associated with poor clinical response to cetuximab compared to wild-type \textit{PIK3CA} tumors (OS, 34 weeks vs 51 weeks [HR, 3.29; \textit{P}=0.0057]). Several meta-analyses have also been performed to study the association between \textit{PIK3CA} mutational status and clinical response to anti-EGFR therapy in mCRC patients.\textsuperscript{39,65,66} In one such systematic review with meta-analysis, Yang et al evaluated the predictive value of specific exon mutations of \textit{PIK3CA} (exons 9 and 20) in response to anti-EGFR therapy in \textit{KRAS} wild-type mCRC patients.\textsuperscript{39} \textit{PIK3CA} exon 20 mutations (but not \textit{PIK3CA} exon 9 mutations) were associated with shorter PFS and OS. In another meta-analysis by Wu et al, the presence of a \textit{PIK3CA} mutation was associated with poorer PFS (HR, 1.53; \textit{P}<0.001) and OS (HR, 1.28; \textit{P}=0.015).
in the unselected mCRC patients treated with anti-EGFR therapy. The negative predictive value of PIK3CA mutation toward anti-EGFR therapy was even stronger when the analysis was restricted to patients with only KRAS wild-type disease (HR for PFS was 2.44; P=0.004).

Based on these results, the presence of PIK3CA mutations may be considered a negative predictive biomarker for anti-EGFR therapy in patients with wild-type KRAS mCRC, and this is largely driven by PIK3CA mutations in exon 20. However, as the available clinical data are largely retrospective in nature, the recent CRC biomarker guideline recommends against routine PIK3CA mutational analysis for selection of therapy outside of a clinical trial.

Loss of PTEN expression
PTEN acts as a tumor suppressor gene, where its normal function is to inhibit PI3K-initiated signaling. Accordingly, loss of PTEN expression results in activation of the PI3K/AKT pathway. PTEN loss is present in 20%–40% of the unselected CRC cases when evaluated by immunohistochemistry (IHC). The available data surrounding the role of PTEN expression as a biomarker in CRC are inconsistent and discordant.

Data from a few retrospective studies suggest that PTEN expression in CRC might have a prognostic value, such that low PTEN expression is associated with aggressive tumor behavior, liver metastases, lymphatic invasion, venous invasion, and poor survival. On the other hand, there are several studies that refute the prognostic influence of PTEN in CRC.

Similarly, PTEN expression has also been evaluated for its predictive value toward anti-EGFR therapy in mCRC but has shown inconsistent results. In a few studies, the loss of PTEN expression was associated with resistance to anti-EGFR therapy. The largest of these is the retrospective study by Loupakis et al, which evaluated the effect of PTEN expression loss on the clinical outcome of mCRC patients treated with cetuximab-based therapy. Patients with PTEN loss had significantly shorter PFS than those with PTEN-positive tumors (3.3 months vs 4.7 months; HR, 0.49; P<0.005). In contrast, several other published reports suggest that PTEN loss is not predictive of response to anti-EGFR therapy.

Due to these conflicting results, the precise role of PTEN status as a prognostic and predictive biomarker of CRC remains unclear. In addition to the inconsistency in data, there are other potential issues that limit the use of PTEN loss as a biomarker. PTEN loss is usually measured by IHC or fluorescence in situ hybridization (deletion), and at present, there is no international standard for IHC assessment of PTEN loss.

As a result, there can be significant interobserver variation among the different clinical trials. Moreover, there appears to be a significant discordance (66%) in PTEN expression between the primary tumor and the matched metastases. Therefore, PTEN loss cannot be viewed as a reliable predictive marker for anti-EGFR therapy at this time. The recent CRC biomarker guideline also does not recommend routine analysis of PTEN for the purposes of therapy selection.

HER2 amplification
HER2 (also known as ERBB2) is a proto-oncogene that belongs to the human EGFR family. It is amplified in ~5% of patients with KRAS exon 2 wild-type mCRC and has a high concordance between the primary tumor and the metastases. It is also mutually exclusive with alterations to KRAS, NRAS, and BRAF in CRC.

HER2 amplification has been studied as a biomarker to predict response toward HER2-directed therapy in mCRC. The phase II HERACLES trial evaluated the activity of dual HER2 inhibition using trastuzumab plus lapatinib in chemorefractory mCRC patients with HER2-positive, KRAS exon 2 (codons 12 and 13) wild-type disease. Of the 27 patients enrolled in the study, eight patients showed a response (ORR 30%). Moreover, the responses were noted to be durable with a median duration of 38 weeks, and the treatment combination was well tolerated. These results show that HER2 amplification is probably a clinically relevant genetic alteration that has a potential role as a predictive biomarker in CRC. However, the extremely low prevalence of this genetic alteration in colorectal tumors presents a significant challenge in validating these results in a confirmatory phase III clinical trial setting.

Immunologic characteristics
Galon and his colleagues were the first to demonstrate the importance of immune signatures as a prognostic marker in colon cancer, beyond the conventional TNM staging. The distribution of total T-cells (CD3+), cytotoxic T-effector cells (CD8+), and memory T-cells (CD45RO+) was evaluated in the tumor core and at the invasive margin. The investigators demonstrated that tumors with elevated levels of CD3+ T-cells in the core as well as at the invasive margin were associated with the best clinical outcome, independent of the TNM stage. In a subsequent study, the prognostic role of tumor-infiltrating lymphocytes was also demonstrated in mCRC. In this study, high tumor-infiltrating lymphocyte density in the metastatic sites conferred a greater response to chemotherapy and was associated with a longer PFS.
Recently, an “Immunoscore” methodology has been defined with an intent to quantify the immune infiltrate inside the tumor. This scoring system is based on the density of CD8+ cytotoxic T-cells and CD45RO+ memory cells at the invasive margin and in the center of the tumor. In a worldwide consortium-based analysis of tissue samples from stage I–III colon cancer patients, this standardized immunoscore assay was recently validated as a prognostic biomarker in early stage colon cancer. In the training set consisting of 700 patients, the patients with “high” Immunoscore had the lowest risk of cancer recurrence at 5 years (HR for “high” vs “low” immunoscore, 0.20; P<0.0001). These findings were confirmed independently in the internal and external validation sets, consisting of 636 and 1,345 patients, respectively. Moreover, the prognostic value of immunoscore assay was independent of the known existing prognostic factors including T stage, N stage, and MSI status.

**CDX2 expression**

CDX2 is an intestine-specific transcription factor and is a major regulator of intestinal development and oncogenesis. In fact, it is a highly sensitive and specific marker of adenocarcinomas of intestinal origin. It is known that colon cancers that lack CDX2 expression are associated with advanced stage, poor differentiation, vascular invasion, BRAF mutation, and CIMP pathway.

CDX2 expression has been recently evaluated for its role as a prognostic and predictive biomarker in early stage colon cancer. In this study’s validation data set, CDX2-negative colon cancers were associated with a lower 5-year DFS than the CDX2-positive tumors (12.1% vs 87.9%; HR, 2.42; P=0.003). These results were independent of the known prognostic factors including tumor stage and pathologic grade. Absence of CDX2 expression was also shown to be a potential predictive marker of response to adjuvant chemotherapy in colon cancer. Among the stage II CDX2-negative colon cancers, the 5-year DFS was found to be longer in patients who received adjuvant chemotherapy than those who did not (91% vs 56%; P=0.006). Although encouraging, these results are derived from retrospective patient cohorts and pooled data sets, and, therefore, require validation using prospective randomized trials before CDX2 expression can be incorporated as a biomarker in routine clinical practice.

**Tumor location**

It is now well established that there are distinct differences in the clinical and molecular characteristics of the right- vs left-sided colon tumors. The large genome-scale analysis of CRC samples performed by TCGA revealed that significant biologic differences exist between right-sided tumors and tumors originating from the other sites, such that the right-sided cancers were more frequently hypermethylated and hypermutated.

The location of colorectal primary tumor has prognostic, and potentially predictive value in CRC. In the study by Loupakis et al, the left-sided tumors were found to have better prognosis than the right-sided tumors in three independent cohorts of mCRC patients derived from the PROVETTA, AVF2107g, and NO16966 trials. The favorable outcome of left-sided tumors was independent of the known prognostic variables. In another study, Venook et al studied the effect of primary tumor location on survival in mCRC patients from the CALGB/SWOG 80405 clinical trial. In the KRAS wild-type mCRC cohort, patients with left-sided primary tumor had superior survival than those with right-sided primary tumor. Post hoc subgroup analyses showed that among patients with right-sided tumors, treatment with bevacizumab was associated with longer survival than that seen with cetuximab. Conversely, among patients with left-sided tumors, treatment with cetuximab was associated with longer OS than with bevacizumab. These results suggest that in addition to the known prognostic role, the location of colorectal primary tumor might also have a predictive value toward anti-EGFR and anti-VEGF therapies.

**Consensus molecular subtypes of CRC**

The recently defined four transcriptome-based molecular subtypes of CRC also demonstrate clinical and prognostic associations. For example, the CMS4 tumors are associated with worse OS and DFS even after adjustment for clinicopathologic features, MSI, BRAF, and KRAS mutation status. The CMS1 lesions are frequently right sided with high histopathologic grade, and these patients have very poor survival after disease relapse. Conversely, CMS2 tumors are predominantly left sided and have superior survival after relapse. Due to the biologic differences between the CMS subtypes, response to therapies is also expected to differ for each subtype. Future studies that prospectively compare the efficacy of therapies in different colorectal CMS subtypes will help establish its role as a predictive biomarker to guide therapeutic decisions.

**MicroRNAs (miRNAs)**

miRNAs are short (18–22 nucleotides), noncoding single-stranded RNA sequences involved in the posttranscriptional regulation of gene expression. They cause either
degradation or inhibition of translation by binding to the 3′-untranslated region of targeted mRNA. miRNAs are believed to play a significant role in colorectal tumorigenesis, and several candidate miRNAs are being evaluated for having either predictive or prognostic role in the treatment of CRC. More recently, circulating exosomal miR-27a and miR-130a were demonstrated as novel diagnostic and prognostic biomarkers of CRC.

**Novel molecular approaches in CRC management**

**Next-generation sequencing**

Next-generation sequencing is non-Sanger-based, high-throughput DNA sequencing technology, where an extremely large number of DNA strands can be sequenced simultaneously in parallel, yielding substantially more throughput of several hundred billion bases in a single run than by conventional sequencing methods. Next-generation sequencing of entire CRC genome has shown that each tumor harbors around 75 mutations. Whole-genome sequencing and whole-exome sequencing are two of the more recent advances in next-generation sequencing technology. Next-generation sequencing has utility in target identification for matched therapies. Specific examples include MSI-Hi for immune checkpoint inhibitor therapy, HER2 amplification for HER2-targeted therapy.

**“Omics”-based approaches**

Large-scale “omics” techniques have the ability to identify tumor biology with an extraordinarily high definition that enables understanding of complex pathologies and discriminate tumors with distinct molecular phenotypes. This growing knowledge has the potential to identify biomarkers leading to the development of pathway-directed targeted therapies and help individualize anticancer treatments to improve clinical outcomes.

Transcriptomics is the study of global mRNA expression of a particular tissue. Recent transcriptomic and immune-based characterization of CRC has further improved our understanding of the colorectal tumor biology, and the precision medicine paradigm in this disease has now evolved and shifted from clonal perspective for targeted therapies, to a clonal—stromal—immune perspective, which involves multi-molecular, multitarget approach. Epigenomics is the study of epigenetic modifications that are now widely known to play an important role in tumorigenesis and are prevalent in CRC. These alterations include DNA promoter region methylation and histone modifications, and several biomarkers of these epigenetic changes have been described. Proteomics refers to large-scale study of proteins to comprehensively map biologic processes such as molecular mechanisms of carcinogenesis. It involves identification and quantification of proteins, as well as the study of their location, modifications, interactions, and functions. Proteomics has identified five different CRC subtypes in TCGA cohort, two of which overlap with the MSI/CIMP transcriptomic subtype. Omics strategy is now being explored to guide therapies in these distinct CRC subtypes. Finally, genomics, based on the advances in the next-generation sequencing technology, is providing a great opportunity for biomarker development and realization of precision medicine in CRC.

**Circulating tumor DNA**

Circulating cell-free DNA is referred to the DNA fragments found in cell-free components of peripheral blood, and it is largely derived from the normal tissue. A small fraction of circulating cell-free DNA in patients with malignancy is also derived from the tumor cells and is referred to as circulating tumor DNA (ctDNA). CtDNA is present in a majority of patients with mCRC, and molecular analysis of ctDNA from a patient’s peripheral blood is a considerably less invasive means of identifying specific genetic mutations in tumors.

CtDNA has high accuracy for the detection of KRAS and other mutations in CRC patients. Moreover, it does not have the limitation of spatial selection bias owing to tumor heterogeneity that is frequently associated with biopsy of the tumor tissue sample. CtDNA has also made it significantly easier to study the clonal evolution of CRC and detection of emerging KRAS-mutated clones during anti-EGFR therapy. For example, in the studies by Diaz et al and Misale et al, the emergence of KRAS mutations was detectable in ctDNA about 6–10 months prior to radiographic evidence of disease progression while on anti-EGFR therapy.

CtDNA has an immense potential to be a noninvasive molecular biomarker. Given its ease for sample collection, it can be used to estimate tumor burden, predict radiologic tumor progression, detect minimal residual disease, monitor both tumor progression and treatment response, and study acquired resistance mechanisms. However, further development of ctDNA-based biomarkers would require standardization of the assay platform and prospective studies to validate ctDNA-based biomarkers in patient care.

**Circulating tumor cells**

Tumor cells detected in the peripheral blood are referred to as circulating tumor cells (CTCs). CtDNA and CTCs are
considered distinct entities, as ctDNA is often present in patients without detectable CTCs. Several studies have suggested a potential prognostic role of CTCs in CRC. In a prospective multicenter study, Cohen et al examined CTCs from 430 patients with mCRC at baseline and after starting first-, second-, or third-line chemotherapy. Patients were stratified into unfavorable and favorable prognostic groups based on CTC levels of ≥3 or <3 CTCs/7.5 mL, respectively. Patients in the baseline unfavorable CTCs group had a shorter median PFS (4.5 months vs 7.9 months; \( P = 0.0002 \)) and OS (9.4 months vs 18.5 months; \( P < 0.0001 \)) compared to those in the favorable CTCs group. Patients whose CTC level converted from the unfavorable to the favorable group after the initial 3–5 weeks of treatment had a significantly improved PFS (6.2 months vs 1.6 months; \( P = 0.02 \)) and OS (11.0 months vs 3.7 months; \( P = 0.0002 \)).

### Precision medicine – goals and challenges

#### The evolving paradigm

Precision medicine paradigm in CRC originally started with “one gene, one drug” approach, which was a considerable advancement over nonselective patient treatment based on the results of conventional population-level clinical trials. However, it was soon realized that such strategy is unlikely to result in durable remission given the complex molecular makeup of CRC with significant cross-talk between signaling pathways. As an example, there is virtually no clinical activity of BRAF inhibitors in BRAF-mutant mCRC, which is in striking contrast to their activity in metastatic melanoma. This lack of response to BRAF inhibitors in CRC is due to the rapid activation of EGFR/P13K pathway via feedback loop, resulting in the escape from downstream inhibition of targets because of the cross-talk between signaling pathways in colon cancer cells. Consequently, a “multi-gene, multi-drug” approach was evaluated and was the next big step to optimize outcomes. The encouraging activity observed with combined inhibition of BRAF, MEK, and EGFR in early trials of BRAF-mutant CRC is a perfect example of this concept.

#### Impact of tumor heterogeneity

CRC is a very heterogeneous disease, and tumor heterogeneity exists at several different levels. First, the genetic makeup of a given tumor type can vary significantly between one patient and another. Next, within an individual patient, there exists a high degree of genetic heterogeneity between the primary tumor and the metastasis as well as between the different metastatic sites (intertumoral heterogeneity). Heterogeneity can also exist in different areas within a single tumor (intratumoral heterogeneity), in different anatomical regions of the same patient (spatial tumor heterogeneity), and at different time points within the same tumor (temporal tumor heterogeneity).

Intratumoral heterogeneity relates to genetic heterogeneity, functional heterogeneity, and nongenetic (epigenetic) heterogeneity. Genetic heterogeneity can be attributed to varying number of cancer clones depending upon the mutation rates and differences in the ability of these clones to metastasize or/and their responsiveness to therapies. Genetic heterogeneity also arises due to clonal selection from treatment pressures. For example, mutations in the driver oncogenes KRAS, NRAS, and BRAF are found to emerge in patients who were initially diagnosed with wild-type disease and were treated with anti-EGFR antibody, but subsequently developed resistance to such therapy. Even though nondominant, these dynamic clonal and subclonal alterations are responsible for sustaining CRC progression.

These different levels of tumor heterogeneity can affect the signaling of multiple key oncogenic pathways resulting in extensive phenotypic variation, such that each tumor clone even within the same tumor displays its own genetic, epigenetic, transcriptomic, and proteomic profile. Consequently, intratumoral heterogeneity poses an enormous challenge to the practice of personalized medicine because any given therapeutic agent targeted at a specific molecular alteration is likely to be successful in only a minority of patients. The four CMS groups are currently the best representation of CRC heterogeneity at the gene-expression level.

#### Future directions and conclusion

Significant advances have been made in the development of molecular biomarkers for the treatment of mCRC. Unlike BRAF, mutations in RAS genes (KRAS and NRAS) do not have a definite prognostic significance in mCRC. However, the presence of activating mutations in these RAS genes is predictive for lack of response to anti-EGFR therapy. There is now a well-established body of evidence documenting that patients with RAS-mutated mCRC should not receive anti-EGFR therapy. However, it is worthwhile to note that the mere presence of wild-type RAS in the tumor does not guarantee a response to anti-EGFR therapy. This suggests that further refinement in the biomarker selection criteria beyond our current understanding is warranted.

CRC is a highly heterogeneous disease, and, therefore, molecular characterization of these tumors is necessary to
determine molecularly homogeneous subtypes, understand tumor biology, and identify the most relevant driver events. Recent advances in DNA sequencing technology and "omics"-based approaches have provided promising new strategies for the development of molecular biomarkers in this disease. We have now identified and characterized genomic, transcriptomic, and immune-based subtypes of CRC. Our future success will depend on the ability to guide therapeutic decisions based on distinct biologic subtypes identified using such gene-expression signatures. Given the significant heterogeneity associated with colorectal tumors, the clonal perspective for targeted therapies is met with only limited success. Instead, a "multimolecular, multidrug" approach is now advocated that integrates knowledge from transcriptomic and immune subtyping of CRC to guide novel therapeutic strategies and precision medicine.

For example, combination of PD-L1 blockade plus immune stimulation with MEK inhibitor in MSS colon cancer has shown early signs of clinical efficacy. For precision medicine and molecularly homogeneous subtypes of CRC, we must consider molecularly defined subgroups of CRC to optimize efficacy, minimize toxicity, and reduce cost. Of course, such initiatives will require intimate collaborations between institutions in multiple countries across continents to ensure rapid clinical trial accrual and efficacy evaluation.

Disclosure
The author reports no conflicts of interest in this work.

References

2. Lenz HJ, Niedzwiecki D, Innocenti F. CALGB/SWOG 80405: phase III trial of irinotecan/5-FU/leucovorin (FOLFIRI) or oxaliplatin/5-FU/leucovorin (MFOLFOX6) with bevacizumab (BV) or cetuximab (CET) for patients with expanded RAS analyses untreated metastatic adenocarcinoma of the colon or rectum (mCRC). Paper presented at: ESMO Congress, Madrid, Spain; 2014.


101. Dong Y, Zhao J, Wu CW, Cw W, et al. Tumor suppressor functions of 
102. Karayian Z, Maiti H, Ju J. mir-129 promotes apoptosis and enhances 
chemosensitivity to 5-fluorouracil in colorectal cancer. Cell Death Dis. 
2013;4:e659.
103. Liu X, Pan B, Sun L, et al. Circulating exosomal miR-27a and miR- 
130a act as novel diagnostic and prognostic biomarkers of colorectal 
104. Wang L, Wheeler DA. Genomic sequencing for cancer diagnosis and 
105. Wang X, Kuan YY, Xu HT, Xu H. Advances in epigenetic biomarker 
research in colorectal cancer. World J Gastroenterol. 2014;20(15): 
4276–4287.
106. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. 
Integrated genome and transcriptome sequencing of the same cell. 
J. Consensus molecular subtypes and the evolution of precision 
108. van Engeland M, Derks S, Smits KM, Meijer GA, Herman JG. 
tumor DNA in early- and late-stage human malignancies. Sci Transl 
112. Thierry AR, Moulere F, El Messaoudi S, et al. Clinical validation of the 
detection of KRAS mutations in colorectal cancer. J Gastrointestin Liver 
provides survival in colorectal cancer patients. J Gastrointestin Liver 
cells and tumor response in colorectal cancer patients treated with 
115. Kopetz S, Desai J, Chan E, et al. PLX4032 in metastatic colorectal 
2010;28(15_Suppl):3534.
116. Karaayvaz M, Zhai H, Ju J. miR-129 promotes apoptosis and enhances 
chemosensitivity to 5-fluorouracil in colorectal cancer. J Clin Oncol. 
117. Sausen M, Parpart S, Diaz LA. Circulating tumor DNA moves further 
cells to tumor response, progression-free survival, and overall 
predicts survival in colorectal cancer patients. J Gastrointestin Liver 
120. Wang X, Gao P, Song Y, et al. Relationship between circulating tumor 
cells and tumor response in colorectal cancer patients treated with 
J. Consensus molecular subtypes and the evolution of precision 
122. van Engeland M, Derks S, Smits KM, Meijer GA, Herman JG. 
tumor DNA in early- and late-stage human malignancies. Sci Transl 
126. Thierry AR, Moulere F, El Messaoudi S, et al. Clinical validation of the 
detection of KRAS mutations in colorectal cancer. J Gastrointestin Liver 